

Utilización de Redes Neuronales en la Caracterización, Modelización y Predicción de Series Temporales Económicas en un Entorno Complejo

Elena Olmedo, Juan Manuel Valderas

Dep. Economía Aplicada I, Fac. CC. Económicas y Empresariales
Avda. Ramón y Cajal, 1
Sevilla, 41018
{olmedo,valderas}@us.es

Ruth Mateos

Dep. Empresa, Universidad San Pablo-CEU
C/ Julián Romea, 23
Madrid, 28003
matcab@ceu.es

Ricardo Gimeno

Dep. Métodos Cuantitativos, Universidad Pontificia Comillas ICADE
C/ Alberto Aguilera 23
Madrid, 28015
rgimeno@upco.es

Resumen

En el presente trabajo se realiza una revisión de las principales implicaciones del nuevo enfoque de la complejidad, y de sus relaciones con la no linealidad y la teoría del caos para el estudio de los problemas económicos. En particular, se analiza la incidencia del cambio de enfoque en el análisis de series temporales económicas, haciendo especial hincapié en la aplicación de la metodología de redes neuronales para la predicción, caracterización y modelización de series temporales económicas.

Palabras clave: Paradigma, Econometría, Análisis de Series Temporales, Redes Neuronales, Aprendizaje, Caos.

1. Introducción

En la evolución histórica de la ciencia se puede hablar de la existencia de tres paradigmas distintos (ver, por ejemplo, [Nieto de Alba, 1998], o [Morin, 1995]): el Paradigma Determinista, el Paradigma Estadístico (ambos englobados en un Paradigma de la Simplificación) y el Paradigma de la

Complejidad. Cada uno de ellos está caracterizado por un modo de pensar y de enfocar el análisis de la realidad. Estos tres paradigmas, sin embargo, no son sustitutivos entre sí, sino que se completan, conviviendo en el tiempo.

Esta evolución se ha dejado sentir tanto en Economía como en el resto de disciplinas científicas.

En este trabajo se realiza una revisión de los cada uno de los paradigmas mencionados, así como los motivos de su surgimiento y sus implicaciones para el Análisis de Series Temporales Económicas.

En concreto, como posteriormente se desarrollará, en la actualidad puede hablarse de la potenciación de una doble vía en el Análisis de Series Temporales en Economía: la basada en la reconstrucción del espacio de estados que garantiza la preservación de las propiedades dinámicas del sistema original y el sistema reconstruido, y la basada en la utilización de redes neuronales con comportamientos que logran emular, mediante procedimientos de aprendizaje, el comportamiento observado en las series temporales reales. El primer enfoque es adecuado para trabajar con series procedentes de sistemas simples y con

principios fundamentales que subyacen en poco nivel de ruido, y el segundo para series cortas y con elevado nivel de ruido.

Debido a este motivo, el enfoque de la reconstrucción tiene más aplicación teórica que empírica cuando se trabaja con la realidad económica, ya que las series temporales suelen ser bastante cortas y con ruido. Para intentar caracterizar, modelizar y predecir estas series temporales las redes neuronales están siendo, por el contrario, cada vez más utilizadas. En concreto, y a modo de resumen, según este nuevo enfoque, los pasos a seguir para el análisis de una serie temporal en Economía serían los esbozados en el siguiente cuadro:

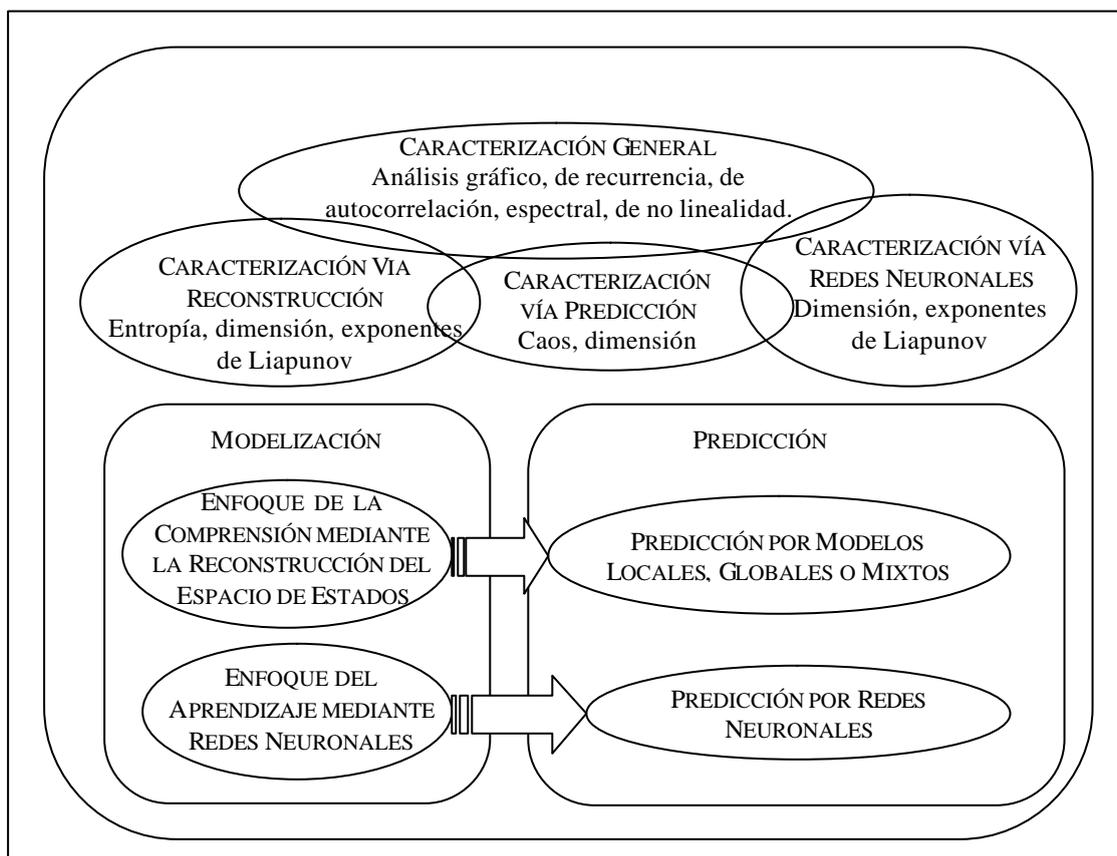


Figura 1. Caracterización, Modelización y Predicción de Series Temporales en la Econometría Compleja

2. Del Paradigma de la Simplificación al Paradigma de la Complejidad

Newton proporcionó las tres leyes dinámicas para describir el movimiento de los planetas, y estas leyes se aplicaron con éxito de un modo general a multitud de problemas físicos, pasando a ser consideradas la base del método científico moderno, también denominado, por este motivo, paradigma newtoniano. Este método marcó el desarrollo de la ciencia clásica, sustentado por el denominado reduccionismo cartesiano, que permitía estudiar el todo mediante el estudio de las partes (parábola de la máquina).

Este reduccionismo cartesiano conduce al Principio de Causalidad Fuerte que rige esta mentalidad determinista: el conocimiento completo de las causas implica el conocimiento completo de las consecuencias (el conocido como 'demonio de Laplace'). Es decir, que para la ciencia clásica, que tuvo su referente en la Física, la descripción precisa, el conocimiento y la comprensión del objeto de investigación implica la capacidad de predecir con certeza la situación de este objeto, tanto en el pasado como en el futuro. Para esta predicción se necesitaba conocer la Ley que rige su evolución, y uno de los estados del objeto considerado. En definitiva, para el conocimiento del sistema estudiado se necesitaba experimentación, que implicaba reducir el sistema a sus partes y estudiarlas en un contexto adecuado.

Pero el mismo Laplace puso de manifiesto que este determinismo no era tanto real como asintótico ya que, en la práctica, no se conocían completamente todas las causas. Debido a la imposibilidad de conocer completamente las causas implicadas, motivada por presencia de numerosos agentes, se relajó el Principio de Causalidad Fuerte, sustituyéndose por el Principio de Causalidad Débil: el conocimiento aproximado de las causas implica el conocimiento aproximado de las consecuencias, llegando así a una predicción aproximada, en términos estadísticos. La búsqueda de leyes universales deterministas se sustituye por la búsqueda de leyes estadísticas, que trabajan en promedio y que utilizan las herramientas necesarias para el manejo de la incertidumbre o azar aparejada al desconocimiento de todas las relaciones implicadas en el fenómeno estudiado. Este azar era concebido como ausencia de información, asociada con la presencia de un número demasiado elevado de relaciones y datos implicados.

Así, el mundo determinista y el estadístico convivían, cada uno dedicado al tratamiento de un

tipo distinto de problemas:

- El primero asociado a los fenómenos simples, entendiendo por simples aquellos con pocos grados de libertad, en los cuales es posible el conocimiento de las causas.
- El segundo asociado a los fenómenos complejos, entendiendo por complejos aquellos con un número elevado de grados de libertad, en los cuales no es posible el conocimiento de las causas

En ambos casos, sin embargo, existían dos principios subyacentes fundamentales, hasta ahora no mencionados:

- La independencia del objeto respecto del observador.
- La existencia de relaciones lineales entre las causas y los efectos, que permitían que el conocimiento aproximado de las causas implicara el conocimiento aproximado de las consecuencias.

Dichos principios conducían a que bajo el Paradigma de la Simplificación se trabajaba en base a conceptos como localidad, estabilidad y atemporalidad [Nieto de Alba, 1998]:

- Localidad (ciencia de objetos aislados). El todo es "aproximadamente" la suma de las partes
- Estabilidad. Es la ciencia de lo lineal, de los sistemas en equilibrio, de los sistemas cerrados.
- Atemporalidad. El tiempo se considera una variable exógena, externa al sistema; es un tiempo reversible, que no da lugar a cambios bruscos en el sistema

En este punto conviene recordar que la Economía nace como ciencia en el siglo XVIII (generalmente la fecha considerada es 1776, fecha de publicación de *An Inquiry into the Nature and Causes of the Wealth of Nations*, de Adam Smith). En esta época, esta nueva ciencia, la Economía, adopta el paradigma científico clásico, buscando el enunciado de leyes económicas universales. Aunque sigue utilizando herramientas estadísticas para la descripción de la realidad económica, no introduce explícitamente la aleatoriedad en los modelos hasta el siglo XX.

De hecho, las características particulares de la realidad económica, al igual que en el resto de las ciencias denominadas 'sociales', han cuestionado la científicidad de su aproximación al estudio de la realidad económica. Y esto es debido a que en las

ciencias sociales es difícil sostener que el observador no influye en lo observado, y que las relaciones entre causas y consecuencias son lineales. En definitiva, la excesiva complejidad de la realidad económica hace imposible mantener estas hipótesis, lo cual conduce a su vez al fallo del enfoque científico clásico, sobre todo a nivel de predicción.

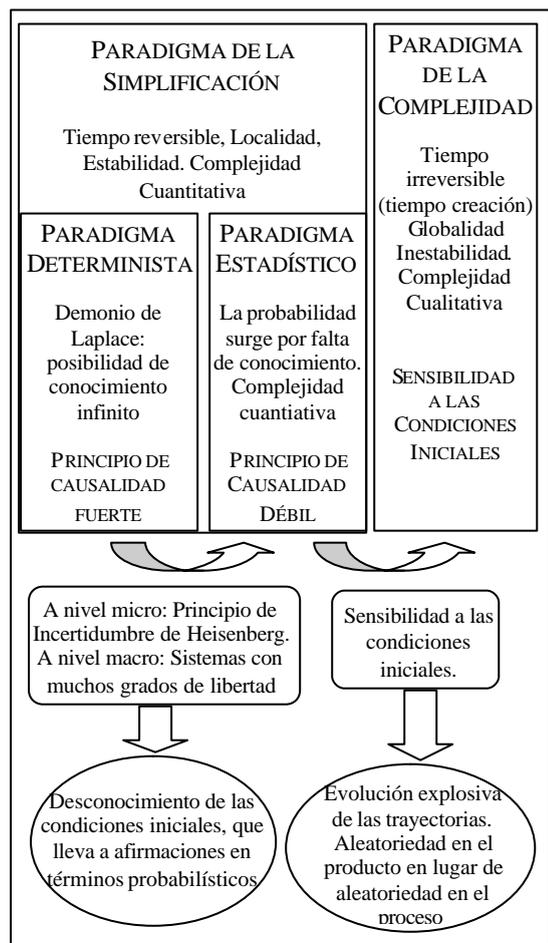


Figura 2. Evolución del Paradigma

Ahora bien, ambos supuestos han sido cuestionados en los últimos tiempos en el contexto de las ciencias usualmente calificadas como ‘duras’. El Principio de Incertidumbre de Heisenberg puso de manifiesto la dependencia entre observación y observador, mientras que la Teoría del Caos cuestionó el supuesto de relación lineal entre causas y consecuencias (ya que la propiedad de dependencia sensible a las condiciones iniciales produce que la más mínima diferencia en la descripción del estado inicial del sistema provoque cambios que hace distintos a sistemas complejos originariamente tan parecidos como se quiera suponer). En definitiva,

ambos principios han incidido en que descripciones precisas no garantizan la certeza en la predicción. Por este motivo, se puede hablar de un acercamiento entre las ‘ciencias duras’ y las ‘ciencias blandas’, trabajando las primeras con conceptos como complejidad y dependencia del pasado, usualmente vinculados a las segundas. En definitiva, la ciencia se engloba dentro del denominado Paradigma de la Complejidad (ver Figura 2).

Este Paradigma de la Complejidad no se contrapone a los ya existentes, sino que los completa, reforzando la importancia de conceptos (ver [Prigogine, 1993 y 1997]) como globalidad, inestabilidad, tiempo creación y no linealidad:

- Globalidad (ciencia de los sistemas en su conjunto, de los sistemas abiertos al entorno). El todo no es una mera suma de las partes.
- Inestabilidad. Es la ciencia del no equilibrio, de la no linealidad, del “efecto mariposa”, de los sistemas abiertos.
- Tiempo creación. El tiempo es una variable interna al sistema, es irreversible (flecha del tiempo) y es necesario considerar la inestabilidad.

3. Análisis de Series Temporales en un Entorno Complejo

3.1. Introducción

En numerosos campos de la ciencia se trabaja con modelos en los que se supone que las observaciones varían independientemente, de manera que la dependencia entre las mismas se considera aleatoria. Sin embargo en otros campos, como por ejemplo la economía, las observaciones son dependientes unas de otras y además se realizan a lo largo de cierto período temporal por lo que, en cierta medida, podemos hablar también de dependencia temporal. Estas colecciones de datos numéricos recogidos a lo largo del tiempo reciben el nombre de series temporales. El conjunto de técnicas dirigido al análisis de estas series de observaciones dependientes del tiempo recibe el nombre de análisis de series temporales.

Desde el principio de los tiempos el hombre ha buscado la capacidad de poder predecir el futuro. Este objetivo subyace en el desarrollo de la ciencia. De hecho, el objetivo último de la ciencia es la predicción mediante la utilización del método científico. Y es también esta capacidad de predicción la que, en última instancia, subyace en el análisis de series temporales.

Ahora bien, dada la evolución de una variable o conjunto de variables recogida en una serie temporal, para poder predecir los valores futuros de esta variable en principio se deben buscar las leyes dinámicas que gobiernan su evolución a lo largo del tiempo. De esta manera podemos afirmar que si existieran unas ecuaciones deterministas conocidas, en principio podrían resolverse para predecir el resultado de un experimento basándose en el conocimiento de las condiciones iniciales. Sin embargo, el caso general en economía será el desconocimiento de estas ecuaciones deterministas de manera que, para hacer predicciones primero se tienen que encontrar las reglas (modelo) que gobiernan la evolución del estado real del sistema. Estas reglas dependerán de la evolución pasada de la serie temporal, y también de la teoría subyacente. Este paso previo a la predicción es la modelización. En definitiva, el análisis de series temporales trata de extraer conclusiones acerca de un sistema complejo utilizando datos de un único sensor.

Hasta cierto punto la predicción y la modelización se solapan, pero en ningún modo son sinónimos ya que la modelización pretende encontrar una descripción que capture de forma adecuada el comportamiento aunque sólo sea a grandes rasgos del sistema. De alguna manera con el modelo pretendemos un resumen formalizado con ayuda de las matemáticas de las relaciones existentes entre las variables que interactúan en el sistema objeto de estudio. Podemos encontrar las ecuaciones que gobiernan a grandes rasgos el comportamiento del sistema pero sin embargo no lograr predicciones adecuadas. Y superponiéndose tanto a la modelización como a la predicción tenemos la mera caracterización del sistema que intenta, con poco conocimiento (o ninguno) del sistema, llegar a determinar sus propiedades fundamentales. De manera que podemos señalar tres propósitos en el análisis de series temporales (ver Figura 3):

1. Predicción de la evolución futura de la serie temporal.
2. Modelización, con la que se pretende describir el comportamiento del sistema, las reglas que gobiernan su dinámica, las relaciones entre las distintas variables que interactúan.
3. Caracterización, con la que se pretende conocer las propiedades fundamentales del sistema.

Para el conocimiento de las leyes que dirigen la dinámica del sistema lógicamente el primer paso será partir de un amplio conocimiento de la materia objeto de estudio, es decir, un dominio de la teoría subyacente. Si este requisito no se cumple, tendremos que aprovechar otro tipo de herramientas distintas. Según Weigend y Gershenfeld [Weigend y

Gershenfeld, 1994], podemos distinguir entre modelos fuertes y débiles que podemos situar en un espacio bidimensional con ejes “pobreza de datos-riqueza de datos” y “pobreza de teoría existente-riqueza de teoría existente”.

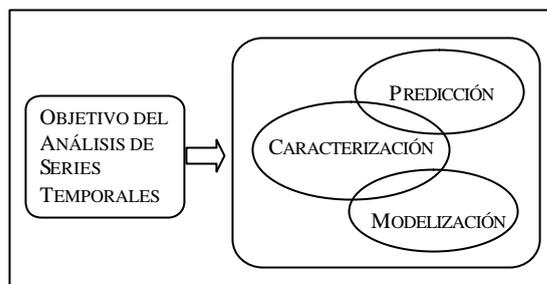


Figura 3. Objetivos del Análisis de Series Temporales

Los modelos débiles son los que manejan poca teoría subyacente y gran cantidad de datos observados y generalmente tienen pocas hipótesis generales y gran cantidad de parámetros. Por el contrario, los modelos fuertes generalmente se expresan en pocas ecuaciones con pocos parámetros, y explican gran cantidad de fenómenos. Usualmente se basan en una fundamentada teoría y no necesitan una gran cantidad de datos. Es lógico pensar que los modelos fuertes son preferibles debido a que se basan en una teoría desarrollada, algo que no puede ser sustituido en los modelos débiles. Sin embargo, ambos tipos de modelización están muy relacionados ya que los modelos débiles en gran medida son causantes del origen o generalización de los modelos fuertes puesto que pueden ayudar a fundamentar o, por lo menos, entender la teoría necesaria para explicar el funcionamiento de la serie temporal.

3.2. Implicaciones para la Modelización

Cuando se trabaja con series temporales, Casdagli [Casdagli, 1989] trata el problema de la modelización como un “problema inverso” en la teoría de sistemas dinámicos. El problema usual en este campo es, conociendo el modelo, describir su evolución conforme pasa el tiempo; el problema inverso será el contrario, es decir, dada la evolución del sistema (que es lo que se manifiesta en la serie de datos recogida a lo largo del tiempo) construir un modelo que pueda haberla originado (modelo generador de datos). Si se localiza la función (necesariamente no lineal si estamos ante un sistema caótico) que puede dar origen a esta evolución,

dispondremos de un posible modelo predictivo.

Para conocer esta función generalmente lo que se hace es separar los datos en dos grupos; el primero se utilizará para llevar a cabo una interpolación o aproximación de la función desconocida (conjunto de entrenamiento) y el segundo para evaluar la bondad del ajuste mediante los errores de predicción.

Por su propia naturaleza, en el caso de la economía y tratando con datos económicos nos centraremos en la modelización débil. Haciendo un poco de historia, hemos de decir que antes de la década de 1920 la predicción se hacía únicamente mediante la extrapolación de la serie utilizando un ajuste global en el dominio temporal. El comienzo de la predicción moderna de series temporales comienza en 1927 cuando Yule desarrolló la técnica autorregresiva para predecir el número anual de manchas solares; en esta técnica se predecía el valor siguiente como una suma ponderada de las observaciones previas de la serie. El problema era que para conseguir un comportamiento interesante de este sistema lineal era necesario introducir perturbaciones aleatorias exógenas.

El desarrollo culminante de esta modelización tuvo lugar a partir de la década de 1970, con el trabajo de Box y Jenkins que propusieron la metodología necesaria para trabajar con este tipo de modelización. Su pretensión era la obtención de modelos que poseyeran un máximo de simplicidad y un mínimo número de parámetros, pretendiendo con esta metodología conocer algo acerca del sistema que generaba la serie temporal (percepción o caracterización), hacer predicciones de valores futuros de la serie (predicción), extensión para representar las relaciones dinámicas entre distintas series e implementar políticas de control. Estos autores partían de una serie estacionaria que modelizaban mediante un proceso media móvil (basándose en la idea de Yule que mantiene que una serie temporal en la que los valores sucesivos son altamente dependientes se puede considerar generada por una transformación o filtro lineal de una serie de perturbaciones aleatorias) y posteriormente generalizaban este modelo llegando a la modelización ARIMA. Esta clase más amplia de procesos proporcionaba, según los autores, un rango de modelos, tanto estacionarios como no estacionarios, que representaban de manera adecuada la mayoría de las series temporales que nos encontramos en la práctica. Por otro lado consideraban, además de los procesos de media móvil, también el proceso autorregresivo en el que el valor actual de una serie depende linealmente de los valores pasados de la serie; la conjunción de

ambos es una generalización que proporciona mayor flexibilidad en el ajuste de las series reales: la modelización ARMA. A continuación consideraban la clase más general de series no estacionarias y, llegaban a partir del mismo a un proceso estacionario aplicando diferenciación. Así, a su modelización ARIMA se llega utilizando tres tipos de filtros (ver Figura 4): el primer filtro (lineal) es el filtro media móvil, el segundo filtro (también lineal) es el filtro estacionario autorregresivo y el tercer filtro es el filtro de integración no estacionaria (que es también lineal).

Esta modelización ha estado vigente durante más de veinticinco años, aunque presenta una serie de inconvenientes, como señala Tong (op.cit.) que se derivan básicamente de su incapacidad para ajustar comportamientos irreversibles en el tiempo, asimétricos, o irregulares, además de trabajar únicamente con la información derivada de la función de autocorrelación o espectro de potencia de la serie temporal.

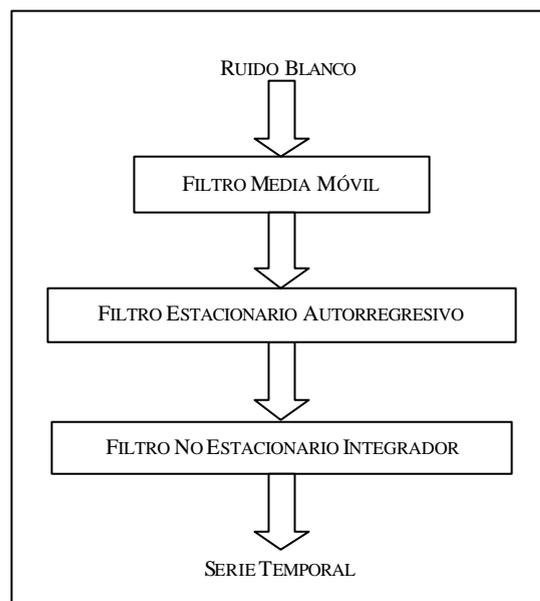


Figura 4. Esquema de la Modelización ARIMA

En este último punto radica la gran crítica a la modelización ARIMA, que ha provocado lo que algunos entienden que es un cambio de paradigma: siempre que el espectro de potencia no proporcione información o, más correctamente, proporcione información confusa, la modelización ARIMA no presentará utilidad. Y hoy día está plenamente comprobada la existencia de sistemas simples, como por ejemplo la ecuación logística, que genera un espectro de potencia ancho que no puede ser

obtenido por una aproximación lineal. Este tipo de sistemas, simples pero capaces de generar series temporales aparentemente muy complejas, llevó a la necesidad de un marco teórico más general para el análisis de series temporales y la predicción.

El paso siguiente lógico es generalizar los modelos lineales, pasando a la modelización no lineal. Este fue el camino que retomaron Tong y Lim en 1980, con su modelo autorregresivo de umbral. Aunque las raíces históricas de la modelización no lineal se remontan al siglo XIX, con los trabajos de Poincaré. A él se deben conceptos tan importantes como la sensibilidad a las condiciones iniciales y el ciclo límite (ideas ambas ausentes en el análisis lineal). Posteriormente, en los primeros sesenta años del siglo XX hubo un resurgimiento del estudio de sistemas no lineales, aparejado con el desarrollo mecánico, eléctrico y electrónico. Duffing y van der Pol estudiaron el comportamiento dinámico de osciladores forzados que generaban multitud de fenómenos que iban más allá de la capacidad de los modelos lineales. Simultáneamente, Andronov escribía su tratado Teoría de las oscilaciones (1937) en el que unía los desarrollos abstractos de Poincaré con estudios prácticos. Por último, a partir de 1970, con el descubrimiento de sistemas deterministas sencillos que podían generar trayectorias indistinguibles de las realizaciones de procesos estocásticos los estudios se han centrado en la idea del caos y la relación entre orden, desorden, determinismo y aleatoriedad.

El siguiente paso, por tanto, sería la utilización de modelos no lineales estocásticos. De hecho, se han desarrollado una enorme batería de posibilidades diferentes, todas ellas con el mismo inconveniente: se incrementa la capacidad de caracterización y predicción de la serie, pero a costa de la simplicidad con la que los modelos lineales parecen comprender y ajustar los datos. Y es que en modelización no lineal carecemos del análisis general del que sí disponemos en la modelización ARIMA. No existe un procedimiento similar a la metodología Box-Jenkins que guíe la construcción del modelo. Y esta carencia ha limitado durante mucho tiempo el uso de los modelos no lineales de series temporales, aunque actualmente están experimentando un resurgimiento. Pero esta vía de estudio de las series temporales económicas es un paso adelante en la modelización tradicional.

El gran cambio se ha suscitado en la década de los ochenta, de la mano de la generalización del uso de ordenadores, que propiciaron el cambio de paradigma en el análisis de series temporales. Puede hablarse en este sentido de una nueva doble vía para el análisis de series temporales. La primera vía está

basada en la reconstrucción del espacio de estados mediante la absorción de retardos temporales, sustentada sobre las ideas de la topología diferencial y los sistemas dinámicos [Takens, 1985]. Con esta reconstrucción se proporciona una técnica para determinar si la serie temporal ha sido generada por un sistema determinista no lineal y, de ser así, para conocer la estructura geométrica subyacente. La segunda vía fue la emergencia del campo del aprendizaje (machine learning), más concretamente las redes neuronales, capaces de explorar de forma adaptativa una gran cantidad de modelos potenciales.

A partir de aquí se puede decir que actualmente existen dos soluciones complementarias al problema de intentar conocer las relaciones entre los valores pasados y futuros de la serie temporal:

- **Comprensión:** buscar entender el comportamiento del sistema utilizando la reconstrucción del espacio de estados.
- **Aprendizaje:** desarrollar modelos implícitos utilizando un enfoque de redes, desarrollando un algoritmo que imite el comportamiento de la serie temporal

A grandes rasgos, el enfoque de la comprensión se encarga de los sistemas de dimensión baja y con poco nivel de ruido y el enfoque del aprendizaje de los sistemas de dimensión elevada o aleatorios. Estos dos diferentes enfoques atacan el problema de la modelización desde distintos puntos de vista y conducirán a diferentes tipos de predicción.

Debido a que nos interesa centrarnos en la utilización de las redes neuronales en el análisis de series temporales, nos centraremos en el segundo enfoque.

3.2.1. Modelización en el enfoque del Aprendizaje: Redes Neuronales

El enfoque de la reconstrucción conduce a buenos resultados si en verdad existe un sistema simple que detectar y modelizar. Si, por el contrario, el sistema es complejo, bien por su elevado número de grados de libertad, o bien por ser estocástico, no existe este sistema simple subyacente y fallará el enfoque de la reconstrucción del espacio de estados. En estos casos, los usuales cuando se trabaja con series temporales económicas, se aplica el enfoque de redes neuronales. A pesar de la dificultad de su interpretación, su buen funcionamiento justifica su amplia difusión y aplicación en campos muy diversos.

Debido al éxito conseguido por la utilización de

redes neuronales en el manejo de series temporales económicas, consideramos interesante dedicar, aunque sea brevemente, unas líneas a su definición y manejo.

Las redes neuronales (ver [Nychka et al., 1992], [Jungeilges, 1996] y White, 1989]) son una clase de modelos no lineales inspirados por la arquitectura neuronal del cerebro. Nacieron en el campo de la inteligencia artificial como un intento de modelizar la capacidad de aprendizaje de los sistemas biológicos neuronales mediante la modelización de la estructura del cerebro.

Esta estructura cerebral se compone de un gran número de neuronas con un elevado número de interconexiones entre ellas, activándose una neurona si la señal total recibida de todas las neuronas que conectan con ella excede un cierto nivel, que se conoce como nivel de activación.

Las redes neuronales reflejan la idea de entender el proceso de aprendizaje como un procedimiento estadístico de tipo recursivo. Son modelos de tipo entrada-salida (input-output) puesto que, dado un vector input x , la red produce un vector output $y = g(x)$, estando la función g determinada por una estructura neuronal. Las redes neuronales nacieron para el reconocimiento de patrones, donde una colección de figuras se presenta a la red y la tarea de la misma es asignar la figura que entra a una o más clases. También se utilizan en la regresión no lineal, pretendiendo encontrar una interpolación suave entre dos puntos. En estos dos casos toda la información relevante se presenta a la red simultáneamente mientras que, en la predicción de series temporales se procesa la información de patrones que evolucionan en el tiempo.

Estos modelos son capaces de cierto tipo de "aprendizaje" a través de una interacción con su entorno por medio de un proceso que, como hemos dicho antes, puede considerarse como un procedimiento estadístico de estimación recursiva. Están siendo aplicados a numerosas áreas tan diferentes como la psicología, la computación, la genética, la lingüística y la ingeniería debido a su capacidad para resolver una gran diversidad de problemas como la descodificación de lenguajes, reconocimiento de caracteres escritos a mano, descodificación de sucesiones genéticas proteicas... en definitiva, aplicando este tipo de técnicas se resuelven problemas relacionados con el reconocimiento de patrones, clasificación, detección de no linealidad y predicción no lineal. Pero algo que dejan claro todos estos autores es que a pesar de que las redes neuronales son un concepto

relativamente novedoso, los métodos que aplican son métodos estadísticos conocidos desde principios de la década de los cincuenta.

Al igual que los métodos analizados en la sección anterior dedicada al enfoque de la comprensión mediante la reconstrucción, las redes neuronales intentan solucionar el llamado problema inverso de ajustar una función no lineal utilizando para ello los valores sucesivos de una serie temporal observada. Las ventajas que presentan sobre los métodos anteriores son principalmente dos:

- Los desarrollos recientes en la literatura de redes neuronales artificiales proporcionan los fundamentos teóricos para la universalidad de redes feedforward (que posteriormente definiremos) como aproximaciones a funciones ya que redes feedforward con un número arbitrario de neuronas pueden aproximar cualquier función uniformemente continua, así como también la derivada de una función arbitraria.
- Para el ajuste utilizan un número de parámetros que aumenta linealmente con el orden de aproximación, mientras que la mayoría de los modelos utilizan el crecimiento con respecto al orden de aproximación suele ser exponencial.

Como ya se mencionó anteriormente, una red neuronal es una estructura de interconexiones, compuesta por una serie de capas, cada una de ellas formada por neuronas, de manera que cada una de ellas alimenta a todas las neuronas de la capa siguiente (la misma estructura que veíamos en el cerebro). Cada una de las neuronas proporciona una salida o output que es el resultado que proporciona una función, llamada función de activación, a la suma de sus inputs multiplicados por unas ponderaciones o pesos. Estas ponderaciones se ajustan conforme a un algoritmo de aprendizaje especificado con el que se pretende minimizar la función de coste calculada a partir del error cometido si se compara la salida de la red con el valor real.

Para especificar una red neuronal se necesitan los siguientes elementos:

- su arquitectura de interconexiones dependientes de unos pesos o ponderaciones, con neuronas organizadas en distintas capas
- las funciones de activación que relacionan el output de una neurona con su input
- la función de coste que evalúa la salida de la red

- un algoritmo de entrenamiento que cambia los parámetros de interconexión (llamados pesos) para minimizar la función de coste

Trabajando con una red lineal, si los valores de entrada (inputs) son los valores retardados de una serie temporal y la salida es la predicción para el siguiente valor, la red será equivalente al uso de un modelo autorregresivo, de manera que los pesos al final del proceso igualarán los coeficientes del modelo especificado. Por este motivo las redes lineales son igual de limitadas que los modelos autorregresivos y las redes neuronales con capacidad interesante de adaptación o imitación son precisamente las no lineales. Así, la idea clave responsable de la potencia y popularidad de las redes es la inserción de una o más capas ocultas no lineales entre el input y el output. Estas no linealidades permiten la interacción entre los inputs y por tanto permiten ajustar funciones más complicadas y así mejorar la capacidad de predicción.

En la estructura de la red es fundamental la elección del número de capas ocultas, del número de nodos o neuronas de dichas capas y del número de unidades de entrada. En cuanto al número de capas ocultas, en general es suficiente trabajar con una [Zhang y otros, 1998], aunque en determinados problemas es más útil trabajar con dos. Para determinar el número de nodos de la capa oculta se utilizan argumentos heurísticos que lo relacionan con el número de inputs, nunca superando el doble de ellos. De todos modos, como muestran algunos estudios (ver, por ejemplo [Lachtermacher y Fuller, 1995] y [Tang y Fishwith, 1993], el número de nodos de la capa oculta tiene un efecto sobre la predicción, aunque no verdaderamente significativo, siendo ésta mucho más sensible a la elección del número de nodos de entrada. ya que dicho número contiene información importante acerca de la complejidad de la estructura de la red. Este número suele ser escogido en función de los resultados de diseños experimentales, escogiéndose normalmente el mejor diseño en función del Criterio de Información de Akaike o de Schwartz.

Existe una amplia aplicación en el campo de la predicción de series temporales económicas. En concreto, se pueden citar ejemplos en mercados financieros ([Trippi y Turban, 1993], [Azoff, 1994], [Refenes, 1995], y [Gately, 1996]), predicción en mercados bursátiles ([White, 1988], [Kimono et al., 1990], [Schoneburg, 1990], [Yoon y Swales, 1991], [Grudnitsky y Osburn, 1993] y [Chen et al., 2003]), predicción de bancarrotas y fallos de mercados ([Odom y Sharda, 1990], [Coleman et al., 1991], [Salchenkerger et al., 1992] [Tam y Kian, 1992], [Fletcher y Goss, 1993] y [Wilson y Sharda, 1994]),

predicción de tipos de cambio ([Weigend et al, 1992, Refenes, 1993], [Kuan y Liu, 1995], [Wu, 1995], [Hann y Steurer, 1996], [Zhang y Hu, 1998] [Franses y Griensven, 1998] y [Giles et al., 2001]. En los últimos tiempos también se están desarrollando aplicaciones macroeconómicas, como la predicción de recesiones [Qi, 2001] o del comportamiento de magnitudes macroeconómicas (GDP en [Tkacz, 2001]).

En la práctica las redes neuronales no lineales más utilizadas en economía son las redes neuronales hacia delante de una capa oculta (abreviadamente, SLFFN) y las redes neuronales recurrentes, cuya estructura exponemos a continuación.

3.2.2. Redes Neuronales SLFFN

La estructura de la red vendría dada por la Figura 5 [Gencay y Tung, 1997], siendo dos los valores de entrada, x_1 y x_2 .

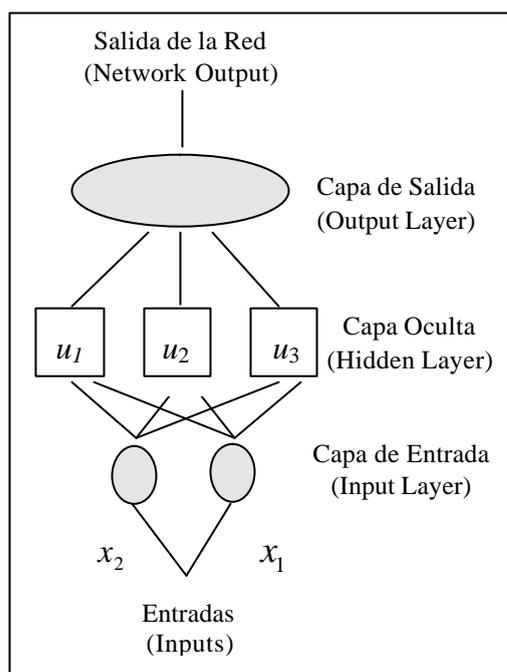


Figura 5. Esquema de una red SLFFN con una capa oculta

Los valores entrada (inputs), x_1 y x_2 , son recibidos por las dos unidades de entrada que simplemente pasan el input hacia las unidades ocultas u_i . Cada conexión (indicada por una flecha en el diagrama) produce una transformación lineal de los inputs determinada por el denominado factor de intensidad, peso o ponderación γ_{ij} , de manera

que el input total que entra en la unidad oculta u_i es $\gamma_{i1}x_1 + \gamma_{i2}x_2$ (si no se introduce el sesgo). Cada una de estas unidades ocultas lleva a cabo una transformación no lineal sobre el input total que le entra, aplicándole una función de activación y produciendo, si se introduce un sesgo, el output $h_i = \Psi(\gamma_{i1}x_1 + \gamma_{i2}x_2 + \gamma_{i0})$.

Esta función de activación es la misma para todas las unidades, pero cada una de ellas tiene su propio sesgo γ_{i0} que representa un input externo o el nivel de actividad neuronal intrínseco. Esta función de activación $\Psi(u)$ tiene en general forma de *ese*¹, con dos asíntotas, en 0 cuando $u \rightarrow -\infty$ y en 1 cuando $u \rightarrow \infty$. Los autores utilizan la función

$$\text{logística } \Psi(u) = \frac{e^u}{1 + e^u}.$$

Los outputs de la capa oculta pasan a través de las unidades de salida que llevan a cabo una transformación lineal sobre su input total. La salida de la red para tres unidades en la capa oculta y dos inputs será, por tanto,

$$o = \beta_0 + \sum_{i=1}^3 \beta_i \Psi(\gamma_{i1}x_1 + \gamma_{i2}x_2 + \gamma_{i0}).$$

A continuación vamos a generalizar la notación para una red neuronal feedforward en el instante t , con n inputs o valores de entrada y q unidades en la capa oculta. Dados los inputs $x_t = (x_{1,t}, \dots, x_{n,t})$, el output de esta red se escribe como

$$h_{i,t} = \Psi\left(\gamma_{i0} + \sum_{j=1}^n \gamma_{ij}x_{j,t}\right) \quad i = 1, \dots, q$$

$$o_t = \Phi\left(\beta_0 + \sum_{i=1}^q \beta_i h_{i,t}\right)$$

o bien

$$o_t = \Phi\left(\beta_0 + \sum_{i=1}^q \beta_i \Psi\left(\gamma_{i0} + \sum_{j=1}^n \gamma_{ij}x_{j,t}\right)\right) \equiv f(x_t, \theta)$$

¹ se toman funciones de este tipo para modelizar a grandes rasgos las propiedades "umbral" de las neuronas reales. Además al ser funciones acotadas, también los errores de predicción estarán acotados.

donde Ψ y Φ (en el ejemplo anterior simplificado, Φ es la función identidad), son funciones de activación conocidas, $\theta = (\beta_0, \dots, \beta_q, \gamma_1, \dots, \gamma_q)$ y $\gamma_j = (\gamma_{j0}, \dots, \gamma_{jn})$ son los parámetros a estimar, o_t se puede considerar como un estimador para la variable target o deseada y_t y x_t es el vector de inputs. Las unidades ocultas de las redes feedforward no son dinámicas pues no dependen de los valores pasados generados por la red. Por eso se llama red feedforward.

Dada la estructura de red que acabamos de ver y elegidas las funciones de activación, la principal cuestión empírica dentro de la predicción por redes neuronales será, por consiguiente, la estimación del vector de parámetros desconocidos a partir de una muestra de valores de targets y de inputs que compararemos y utilizaremos para ir estimando recursivamente los parámetros. El algoritmo de aprendizaje más utilizado es el de retropropagación que básicamente actualiza las ponderaciones del modelo en la dirección en la que el gradiente desciende más rápidamente:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \eta \nabla f(x_t, \hat{\theta}_t) \cdot [y_t - f(x_t, \hat{\theta}_t)]$$

donde $\nabla f(x, \theta)$ es el vector gradiente de la función f con respecto a θ y η es la tasa de aprendizaje. Aquí $\nabla f(x_t, \hat{\theta}_t) \cdot [y_t - f(x_t, \hat{\theta}_t)]$ es el vector de las primeras derivadas del error cuadrático medio $[y_t - f(x_t, \hat{\theta}_t)]^2$. Este algoritmo fue ideado a principios de los 70 por Verboos, y completado con las contribuciones de Parker [Parker, 1985] y Le Cun [Le Cun, 1985], siendo popularizado a partir del trabajo de Rumelhart y McClelland [Rumelhart y McClelland, 1986]. Este procedimiento se caracteriza por la adaptación recursiva o aprendizaje de las estimaciones de los parámetros, y recibe el nombre de método de retropropagación o del gradiente descendente. La retropropagación recibe su nombre en que básicamente consiste en propagar el error hacia atrás, es decir, de la capa de salida hacia la capa de entrada, pasando por las capas ocultas intermedias y ajustando los pesos de las conexiones con el fin de reducir el error cometido.

Imponiendo condiciones apropiadas en la tasa de aprendizaje y las funciones de activación, White [White, 1989] obtiene las propiedades estadísticas de este estimador, y muestra que converge asintóticamente al estimador que minimiza localmente el error cuadrático medio, es decir, aquel que minimiza

$$\langle (y - f(x, \theta))^2 \rangle$$

Asimismo, White muestra que este método mínimo cuadrático no lineal es más eficiente que el método recursivo.

Dentro de los algoritmos de retropropagación hay dos grandes bloques: el de actualización incremental o por dato y el de actualización por bloque. El primero actualiza los pesos en cada iteración, y el segundo después de realizar un barrido completo. Los primeros son más sensibles al ruido, más agresivos aunque muy útiles para la actualización de sistemas en tiempo real. En cuanto a los segundos son más robustos y prudentes, ya que el paso de entrenamiento se realiza sobre la media de todos los patrones de entrenamiento, aunque son mucho más costosos en tiempo de cálculo.

Este algoritmo es muy sensible a la elección de la tasa de aprendizaje, al determinar esta la magnitud de las actualizaciones en las ponderaciones. Si esta tasa es demasiado baja, el proceso es excesivamente lento mientras que, si es demasiado alta, puede ocasionar oscilaciones en la convergencia. Una manera de evitar este problema es hacer que la red responda, no sólo al gradiente local, sino también a tendencias recientes en la superficie del error. Esto se logra mediante la introducción de un parámetro momento adicional para que el nuevo peso cambie en aproximadamente la misma dirección que el anterior, reduciendo así el efecto oscilatorio de incrementar la tasa de aprendizaje.

Dentro de esta categoría, uno de los algoritmos más utilizados es el denominado algoritmo de Newton [Kuan y White, 1994a], que se basa en la inclusión de la dirección de Newton en la adaptación recursiva de las estimaciones de θ . La forma del algoritmo recursivo de Newton es

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \eta_t \hat{G}_t^{-1} \nabla f(x_t, \hat{\theta}) [y_t - f(x_t, \hat{\theta})]$$

$$\hat{G}_{t+1} = \hat{G}_t + \eta_t [\nabla^2 f(x_t, \hat{\theta}_t) - \hat{G}_t]$$

donde \hat{G}_t es una matriz de dirección de Newton estimada y $\{\eta_t\}$ es una sucesión de tasas de aprendizaje de orden $1/t$. Este algoritmo [Kuan y White, 1994a] es más eficiente que el de retropropagación tradicional y asintóticamente equivalente al método mínimo-cuadrático no lineal.

Estos métodos de retropropagación con momento

son los adoptados por la mayoría de los investigadores, aunque en los últimos tiempos se han desarrollado diversas modificaciones del algoritmo para mejorar sus resultados [Zhang y otros, 1998].

3.2.3. Redes Neuronales Recurrentes

Las aplicaciones de las redes neuronales para la predicción de series temporales se basan en un ajuste de la dinámica. Como hemos visto, las redes feedforward acomodan esta dinámica incluyendo en el algoritmo una función de coste que relaciona los valores de salida de la red a partir de los valores de entrada y comparándolos con los valores deseados o target. Sin embargo, en este tipo de representación dinámica no se incluye la figura de retroalimentación o feedback interno, algo sumamente conocido en las redes biológicas.

Volviendo al diagrama de estructura de la red feedforward, en la red recurrente de una capa oculta esta retroalimentación puede representarse como una retroalimentación del resultado o output de la capa oculta que regresa de nuevo, retardados una unidad en el tiempo, como input en la capa oculta (ver Figura 6).

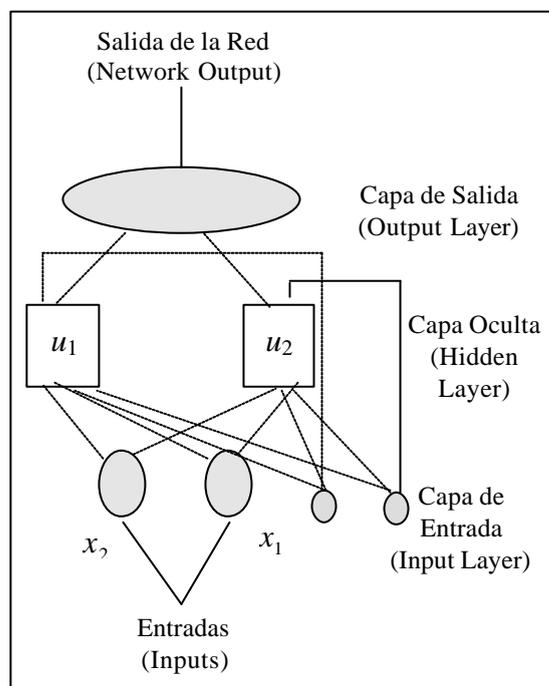


Figura 6. Esquema de una red recurrente con una capa oculta

Debido a esta dinámica interna, empíricamente se comprueba que parecen tener más memoria, y son capaces de realizar mejores predicciones, incluso en presencia de ruido. Las capas ocultas que vuelven a retroalimentar la red proporcionan una entrada adicional que da a la red propiedades dinámicas que hacen que posea memoria interna, y si bien todavía no está determinado cómo actúa esta memoria [Gencay y Tung, 1997], se aprecia que permite una mejor adaptación y un mejor tratamiento del ruido.

El output de esta red puede representarse como

$$h_{i,t} = \Psi \left(\gamma_{i0} + \sum_{j=1}^n \gamma_{ij} x_{j,t} + \sum_{l=1}^q \delta_{il} h_{l,t-1} \right) \equiv \psi_i(x_t, h_{t-1}, \theta)$$

$$i = 1, \dots, q$$

$$o_t = \Phi \left(\beta_0 + \sum_{i=1}^q \beta_i \psi_i(x_t, h_{t-1}, \theta) \right) \equiv \phi_q(x_t, h_{t-1}, \theta)$$

o bien por sustitución recursiva

$$h_{i,t} = \psi_i(x_t, \psi(x_{t-1}, h_{t-2}, \theta), \theta) = \dots \quad i = 1, \dots, q$$

$$o_t = \phi_q(x_t, h_{t-1}, \theta) \equiv g(x^t, \theta)$$

donde $x^t = (x_t, x_{t-1}, \dots, x_1)$. Como consecuencia de esta retroalimentación el output de la red depende de los valores iniciales de $h_{i,0}$ $i = 1, \dots, q$ y de la historia entera de los inputs del sistema $x^t = (x_t, x_{t-1}, \dots, x_1)$. Estas redes son capaces de generar un comportamiento dinámico muy rico, exhibiendo memoria y sensibilidad hacia el contexto. Debido a la presencia de retroalimentaciones internas, estas redes se llaman redes recurrentes.

Los parámetros de interés θ se estiman mediante la minimización de $\left\langle \left(y_t - g(x^t, \theta) \right)^2 \right\rangle$. El output de la red o depende de θ directamente y también indirectamente a través de la presencia de unidades escondidas de activación retardadas. Debido a esta estructura dependiente del estado, el método de mínimos cuadrados no lineales no es factible, y los parámetros puede estimarse mediante el algoritmo de retropropagación de Kuan y otros [Kuan y White, 1994b], y por el algoritmo recurrente de Newton, más eficiente que el anterior. Concretamente, el algoritmo recurrente de Newton [Kuan, 1994] tiene la forma

$$\hat{e}_t = y_t - \phi(x_t, \hat{h}_{t-1}, \hat{\theta}_t)$$

$$\nabla \hat{e}_t = -\phi_\theta(x_t, \hat{h}_{t-1}, \hat{\theta}_t) - \hat{\Delta}_t \phi_h(x_t, \hat{h}_{t-1}, \hat{\theta}_t)$$

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \eta_t \hat{G}_t^{-1} \nabla \hat{e}_t \quad \hat{G}_{t+1} = \hat{G}_t + \eta_t (\nabla \hat{e}_t \nabla \hat{e}_t' - \hat{G}_t)$$

donde ϕ_θ y ϕ_h son los vectores columna de las primeras derivadas de ϕ con respecto a θ y h respectivamente. La i -ésima unidad oculta de activación se ajusta de acuerdo con

$$\hat{h}_{i,t} = \Psi \left(\hat{\gamma}_{i0,t} + \sum_{j=1}^n \gamma_{ij,t} x_{j,t} + \sum_{l=1}^q \hat{\delta}_{il,t} \hat{h}_{l,t-1} \right) = \psi_i(x_t, \hat{h}_{t-1}, \hat{\theta}_t)$$

y la j -ésima columna $j=1, \dots, q$ columna de $\hat{\Delta}_{t+1}$ se ajusta de acuerdo con

$$\hat{\Delta}_{j,t+1} = \Psi_{j,\theta}(x_t, \hat{h}_{t-1}, \hat{\theta}_t) + \hat{\Delta}_t \Psi_{j,h}(x_t, \hat{h}_{t-1}, \hat{\theta}_t)$$

3.3. Implicaciones para la Predicción

Como ya señalamos anteriormente, la modelización dependerá del sistema generador de la serie temporal observada: la proporcionada por la reconstrucción en el caso de sistemas deterministas de poca dimensión y la proporcionada por redes neuronales en caso contrario. En el primer caso hay una estructura comprensible en el espacio de absorción, que puede detectarse y por tanto modelizarse. El problema se produce cuando, aunque procedamos a la reconstrucción, no se identifica ninguna figura simple que podamos modelizar. Las redes neuronales, al no intentar proporcionar una descripción explícita de un objeto, son idóneas para este tipo de modelización aunque al mismo tiempo, tienen una interpretación más difícil.

En la breve introducción que se ha hecho de las redes neuronales se ha especificado la existencia de unas interconexiones entre las neuronas de las distintas capas que dependen de unos parámetros, unas funciones de activación, una función de coste y un algoritmo de entrenamiento que busca ajustar los parámetros para minimizar la función de coste. Así, durante el entrenamiento la red neuronal compara la salida predicha con la real, calcula la función de coste y ajusta los parámetros para minimizar esta función. Se trata de una adaptación que en la terminología de redes neuronales se denomina open-loop (rizo abierto) pues tanto la entrada o input como la respuesta deseada se proporcionan por el mismo conjunto de entrenamiento, de manera que la salida o output de la red no se retroalimenta como

input durante el entrenamiento. Este esquema también se conoce como adaptación ecuación-error o bien aprendizaje forzado.

La misma arquitectura puede usarse para seguir la trayectoria en el largo plazo, a costa de posiblemente peores predicciones a corto plazo. Para ello se usará el valor predicho como input en el entrenamiento, en lugar del valor real. Es decir, que una vez que la red está entrenada, las predicciones iteradas a largo plazo se consiguen tomando la estimación de la red y retroalimentándola como input en la red. Este tipo de sistema, en contraposición al anterior, se conoce con la denominación de sistema closed-loop (o de rizo cerrado). La red se puede iterar en el tiempo para hacer predicciones tan lejos en el futuro como se desee. El inconveniente de esta metodología es que, debido a que el entrenamiento se basa únicamente en predicciones de un paso hacia delante, no se puede garantizar que las predicciones a largo plazo sean buenas. Entre estos dos extremos hay multitud de combinaciones que surgen de la combinación de los dos tipos de errores de una manera ponderada.

3.4. Implicaciones para la Caracterización

Tradicionalmente se ha pensado que el comportamiento complejo de las series temporales estaba ocasionado por la complejidad del sistema subyacente, bien sea debido al gran número de variables que interactúan en el mismo, o bien debido a algún tipo de desconocimiento o falta de información que hace necesaria la modelización mediante variables aleatorias. Sin embargo, la dinámica caótica ha puesto de manifiesto que sistemas simples pueden también producir series temporales que parecen complicadas. Estos dos extremos requieren, como posteriormente se comprobará, diferentes técnicas y, por tanto, el éxito de la predicción depende de a qué grupo pertenece nuestro sistema. Por tanto será fundamental caracterizar a qué tipo pertenece el sistema generador de la serie objeto de estudio y será primordial la utilización de métodos que caractericen las propiedades esenciales que subyacen en la serie temporal.

Existen ciertos métodos ampliamente conocidos, que no pierden en absoluto su vigencia, sino que se ven complementados por el surgimiento de otros nuevos, derivados de los nuevos desarrollos recientes. De este modo, el conjunto de herramientas a utilizar para analizar la serie y caracterizarla ha aumentado. Por hacer una clasificación, podemos distinguir entre herramientas generales, el análisis explícito de la estructura en el espacio de absorción y el enfoque implícito basado

en el análisis de propiedades de una imitación del sistema a la que se llega mediante aprendizaje.

3.4.1. Nuevas Herramientas Generales

Dentro de las herramientas generales, además de los métodos tradicionales han cobrado renovada importancia los contrastes de no linealidad. Entre ellos destacan como contrastes más potentes el BDS [Brock, Dechert y Scheinkman, 1996] y el de Kaplan [Kaplan, 1994 y 1995]. Si el sistema es no lineal, hay autores que proponen trabajar otros contrastes para distinguir el tipo de no linealidad. También existen contrastes para tipos específicos de no linealidad (por ejemplo de tercer orden el contraste de Hinich [Hinich, 1982], de varianzas los de McLeod-Li [McLeod y Li, 1983] y Engle [Engle, 1982]). Otros contrastes más intuitivos y generales son el de "barajado" de Scheinkman y LeBaron [Scheinkman y LeBaron, 1989], el contraste de residuos de Broca [Brock, 1986] y el de datos subrogados [Theiler et al., 1992a y 1992b].

Asimismo cabe destacar la utilización de y una nueva herramienta gráfica que permite analizar de manera bastante sencilla la posible dinámica subyacente en la serie temporal: el gráfico de recurrencia. A partir de esta herramienta se han desarrollado medidas cuantitativas (Análisis Cuantitativo de Recurrencia), entre las que destacan el determinismo y el tiempo de predicción [Eckman, Kamphorst y Ruelle, 1987], [Iwanski y Bradley, 1998] y [Gilmore, 1996].

En cuanto a las redes neuronales, están siendo muy aplicadas para el cálculo de los exponentes de Liapunov en series cortas y con ruido. A continuación comentamos la definición y significado de dichos exponentes, y los motivos de la utilización de las redes neuronales en su cálculo.

3.4.2. Caracterización vía Redes Neuronales. Cálculo de los Exponentes de Liapunov

Los exponentes de Liapunov se definen como los autovalores de la linealización local de la dinámica y su importancia radica en que miden la tasa de divergencia de los ejes principales de las trayectorias cercanas. De ahí que un exponente de Liapunov positivo sea indicativo (para algunos autores, incluso definitorio) de la propiedad de dependencia sensible a las condiciones iniciales o, lo que es lo mismo, de la existencia de una dinámica caótica.

El máximo exponente de Liapunov está también relacionado con el límite de la predecibilidad o tiempo de predicción ya que, a mayor valor del

exponente, menor será la capacidad de predicción. Además, la suma de los exponentes de Liapunov positivos proporciona una estimación de la entropía. Si el sistema es sencillo (sin ruido), como demuestra Brock [Brock, 1986] funciona bien el método directo de estimación de los exponentes de Liapunov, propuesto por Guckenheimer [Guckenheimer, 1982] que consiste en estimarlos a partir de la separación empírica de órbitas generadas a partir de puntos inicialmente próximos. Dentro de los métodos directos, el más utilizado es el de Wolf [Wolf et al., 1985].

Sin embargo este método no funciona bien cuando las series son cortas o están contaminadas por ruido, lo cual es usual en economía. El problema se complica cuando el modelo subyacente es además no lineal. En ambos casos el método directo no proporcionará buenas estimaciones del exponente de Liapunov.

Existe otra forma de estimar los exponentes de Liapunov, que no se basan en la estimación directa de la divergencia de la evolución seguida por dos condiciones iniciales arbitrariamente próximas, sino en estimar esa tasa de divergencia indirectamente, aplicando que el exponente de Liapunov para una condición inicial x_0 se define como

$$\lambda(x_0) = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \left[\left| Df^T(x_0) \right| \right]$$

En definitiva, estos métodos buscan estimar la matriz jacobiana (la matriz de derivadas parciales) del sistema subyacente desconocido, generador de la serie temporal, y aplicar la definición anterior. Por este motivo también se los conoce como métodos de estimación del jacobiano. Difieren unos de otros precisamente en la manera de estimar esta matriz.

Dentro de estos métodos indirectos, los primeros desarrollados se basaron en estimaciones locales lineales (ver [Eckman y Ruelle, 1985 y 1986] y [Sano y Sawada, 1985]. Debido a su excesiva simplicidad, no funcionaron bien en series cortas y con ruido (las propias de la Economía), además de presentar valores espúreos cuando la dimensión del sistema era alta. Para solucionar estos problemas se sustituyó la estimación lineal por la aplicación de ajustes no lineales.

Dentro de ellos, los métodos indirectos de estimación del jacobiano basados en redes neuronales, inicialmente propuestos por Nychka y otros [Nychka et al., 1992] y Gencay y Dechert [Gencay y Dechert, 1992], han funcionado muy bien en la práctica con series cortas y con ruido, además

de mostrar buenas propiedades teóricas (ver [Nychka et al., 1992], [McCaffrey, 1991] y [McCaffrey et al., 1992]), fundamentalmente su buen nivel de aproximación y la escasa sensibilidad al aumento de la dimensión del sistema. La propiedad de aproximación universal también se aplica a sus derivadas [Gallart y White, 1992], propiedad interesante en el contexto de los exponentes de Liapunov. Los más utilizados han sido los programados por Ellner et al. [Ellner et al., 1992] (LENNS) y Kuan, Liu y Gencay [Kuan et al., 1997] (NETLE 3.01). En ambos casos se utiliza para la estimación del jacobiano el método de mínimos cuadrados no lineales, y se trabaja con una red neuronal SLFFN de una sola capa, trabajando con un número amplio de número de nodos y de inputs, eligiendo el mejor ajuste conseguido en función de la minimización del criterio de información de Schwarz (CIS) que penaliza el error cuadrático medio por el número de nodos utilizados

Los resultados de simulación también han sido favorables a la utilización del método de redes neuronales para el cálculo de los exponentes de Liapunov (ver, por ejemplo, [Gallart y White, 1992], [Kuan y White, 1994a] y [Barnett et al., 1997]).

Adicionalmente, en los últimos tiempos se ha estado trabajando en el desarrollo de la distribución asintótica de probabilidad de los exponentes de Liapunov calculados mediante redes neuronales. En concreto, Shintani y Linton [Shintani y Linton, 2003 y 2004] han probado que convergen asintóticamente a la distribución normal. La obtención de esta distribución es fundamental para extraer conclusiones fiables acerca de la estabilidad o inestabilidad del sistema generador de la serie temporal.

Por los motivos enunciados actualmente existen numerosos ejemplos de la aplicación de redes neuronales en el cálculo de exponentes de Lyapunov en series temporales porcentes del mundo económico y financiero. Como ejemplos pueden citarse trabajos para datos de tipos de cambio [Dechert y Gencay, 1992] y [Bask y de Luna, 2001 y 2002], agregados monetarios [Serletis, 1995] y [Barnett et al., 1995], rendimientos financieros [Abhyankar et al., 1997] y [Shintani y Linton, 2004], datos reales de producción [Shintani y Linton, 2003].

4. Conclusiones

El redescubrimiento del concepto de complejidad de la mano de la propiedad de dependencia sensible a

las condiciones iniciales ha provocado un desarrollo de la modelización no lineal y, lo que es más importante, un cambio de enfoque (algunos incluso defienden un cambio de paradigma) también en la econometría y en el análisis de series temporales. Este nuevo enfoque engloba al anterior, pero además permite trabajar con nuevos conceptos unidos a esta nueva complejidad cualitativa.

Este cambio de mentalidad se traspasa también al campo del análisis de series temporales en el que, sin perder de vista el objetivo último de la predicción de los valores futuros de la serie, nace un doble enfoque distinto al empleado hasta el momento.

El primero es la predicción mediante la reconstrucción del espacio de estados y la subsiguiente identificación de una estructura clara que modelizar. El segundo es la predicción mediante el aprendizaje utilizando redes neuronales, más apropiado cuando el sistema no da lugar a una estructura geométrica suficientemente simple en la reconstrucción. Asimismo las redes neuronales han mostrado un gran potencial en el cálculo de los exponentes de Liapunov en series temporales cortas y con ruido.

Referencias

- Abhayandar, A., Copeland, L.S. y Wong, W. (1997): *Uncovering nonlinear structure in real-time stock-market indexes: the S&P 500, the DAX, the Nikkei 225 and the FTSE-100*. Journal of Business and Economic Statistics 15: 1-14
- Azoff, E.M. (1994). *Neural network time series forecasting of financial markets*. Chichester, John Wiley and Sons
- Barnett, W.A., Gallant, A.R., Hinich, M.J., Jungeilges, J.A., Kaplan, D.T. y Jensen, M.J. (1995): *Robustness of nonlinearity and chaos tests to measurement error, inference method, and sample size*. Journal of Economic Behavior and Organization 27: 301-320
- Barnett, W.A., Gallant, A.R., Hinich, M.J., Jungeilges, J.A., Kaplan, D.T. y Jensen, M.J. (1997): *A single-blind controlled competition among tests for nonlinearity and chaos*. Journal of Econometrics 82: 157-192
- Bask, M. and de Luna, X. (2001). *EMU and the stability and volatility of foreign exchange: Some empirical evidence..* Umea Economics Studies, 565, Umea University, Sweden.
- Bask, M. y de Luna, X. (2002): *Characterizing the degree of stability of nonlinear dynamic models*. Studies in Nonlinear Dynamics & Econometrics 6, 1: art. 3
- Brock, W.A. (1986). *Distinguishing random and deterministic systems. abridged version*. Journal of Economic Theory 40: 168-195
- Brock, W.A., Dechert, W.D., Scheinkman, J.A. y LeBaron, B.D. (1996). *A test for independence based on the correlation dimension*. Econometric Reviews 15: 197-235
- Casdagli, M. (1989). *Nonlinear prediction of chaotic time series*. Physica D 35: 35-56
- Casdagli, M. y Eubank, S. (1992). *Nonlinear Modeling and Forecasting* Redwood City, CA. Addison-Wesley
- Chen, A.-S., Leung, M.T. y Daouk, H. (2003). *Application of neural networks to an emerging financial market. Forecasting and trading the Taiwan Stock Index*. Computers and Operations Research 30, 6: 901-923
- Coleman, K.G., Graettinger, T.J., Lawrence, W.F. (1991). *Neural network for bankruptcy prediction. the power to solve financial problems*. AI Review 5, julio/agosto: 48-50
- Day, R.H. (1993). *Complex economic dynamics. obvious in history, generic in theory, elusive in data*. En *Nonlinear dynamics, chaos and econometrics* (Ed.. M.H. Pesaran y S.M. Potter). Nueva York. John Wiley & Sons: 1-15
- Dechert, W.D. y Gencay, R. (1992): *Lyapunov exponents as a nonparametric diagnostic for stability analysis*. Journal of Applied Econometrics 7: S41-S60
- Dopfer, K. (1991). *The complexity of economic phenomena. reply to Tinbergen and beyond*. Journal of Economic Issues 25, 1: 39-76
- Durlauf, S.N. (1997). *What should policymakers know about economic complexity?*. working paper n° 97-10-080. Instituto de Santa Fe, Santa Fe, Nuevo Méjico
- Eckman, J.-P y Ruelle, D. (1985). *Ergodic theory of chaos and strange attractors*. Reviews of Modern Physics 57, 3: 617-656
- Eckman, J.-P., Karmphorst, S.O. y Ruelle, D. (1987). *Recurrence plots of dynamical systems*; Europhysics Letters 1, 9: 973-977
- Ellner, S., Nychka, D.W. y Gallant, A.R. (1991). *LENNS, a program to estimate the dominant Lyapunov exponent of noisy nonlinear system from time series data*. Institute of Statistics Mimeo Series #2235, Statistics Department, North Carolina State University, Raleigh, NC 27695-8203.
- Engle, R.F. (1982). *Autoregressive conditional heteroskedascity with estimates of the variance of United Kingdom inflation*. Econometrica 50: 987-1007
- Fletcher, D., Goss, E. (1993). *Forecasting with neural networks. An application using bankruptcy data*. Information and Management 24: 159-167
- Frances, P.H. y Griensven K. (1998). *Forecasting exchange rates using neural networks for technical trading rules*. Studies in Nonlinear Dynamics and Econometrics 2, 4: 109-114
- Gallant, A.R. y White, H. (1992): *On learning the derivatives of an unknown mapping with multilayer feedforward networks*. Neural Networks 5: 12-138

- Gately, E. (1996). *Neural networks for financial forecasting*. Nueva York, John Wiley
- Gencay, R. y Dechert, W.D. (1992): *An algorithm for the n Lyapunov Exponents of an n dimensional unknown dynamical system*. Physica D 59: 142-157
- Gencay, R. y Tung, T. (1997). *Nonlinear modeling and prediction with feedforward and recurrent networks*. Physica D 108: 119-134
- Georgescu-Roegen, N. (1996). *La ley de la entropía y el proceso económico*. Madrid. Fundación Argentaria
- Giles, C.L., Lawrence, S. y Tsoi, A.C. (2001). *Noisy ediction using a recurrent neural network and grammatical inference*. Machine Learning 44, 1/2: 161-183
- Gilmore, C. (1996). *A new test for chaos*. En *Chaos & nonlinear dynamics in the financial market* (Ed. R.R. Trippi). Chicqgo. Irwin Proffesional Publising: 383-415
- Grandmont, J.-M. y Malgrange, P. (1986). *Nonlinear Economic Dynamics. Introduction*. Journal of Economic Theory 40: 3-12
- Granger, C.W.J. (1991). *Developments in the nonlinear analysis of economic series*. Scandinavian Journal of Economics 93, 2: 263-276
- Granger, C.W.J. (1994). *Forecasting in economics*. En *Time series prediction. forecasting the future and understanding the past* (Ed.. A.S. Weigend y N.A. Gershenfeld). Reading, Massachusetts. Addison-Wesley: 529-538
- Grassberger, P. y Procaccia, I. (1983a). *Measuring the strangeness of strange attractor*. Physica D 9 189-208
- Grassberger, P. y Procaccia, I. (1983b). *Characterization of strange attractors*. Physical Review Letters 50, 5: 346-349
- Grassberger, P. y Procaccia, I. (1983c). *Estimation of the Kolmogorov entropy from a chaotic signal*. Physical Review A 28, 4: 2591-2593
- Grudnitski, G., Osburn, L. (1993). *Forecasting S and P and gold futures prices. an application of neural networks*. The Journal of Future Markets 13, 6: 631-643
- Guckenheimer, J. (1982). *Noise in chaotic systems*. Nature 298: 358-361
- Hann, T.H., Steurer, E. (1996). *Much ado about nothing? Exchange rate forecasting. neural networks vs. linear models using monthly and weekly data*. neurocomputing 10: 323-339
- Hinich, M. J. (1982). *Testing for Gaussianity and linearity of a stationary time series*. Journal of Time Series Analysis 3: 169-176
- Iwanski, J.S. y Bradley, E. (1998). *Recurrence plots of experimental data. to embed or not to embed*. Chaos 8, 4: 861-871
- Jaditz, T. y Sayers, C.L. (1996). *Searching for nonlinearity in mean and variance*. En *Nonlinear dynamics and economics* (Ed.. W.A. Barnett, A.P. Kirman y M. Salmon). Cambridge. Cambridge University Press: 201-218
- Jungeilges, J.A. (1996). *Operational characteristics of White's test*. En *Nonlinear dynamics and economics* (Eds. W.A. Barnett, A.P. Kirman y M. Salmon). Cambridge.. Cambridge University Press: 219-266
- Kaplan, D.T. (1994). *Exceptional Events as evidence for Determinism*. Physica D 73: 38-48.
- Kaplan, D.T. (1995). *Nonlinearity and nonstationarity. the use of surrogate Data in interpreting fluctuations*. En *Interpreting Fluctuations*. Center for Nonlinear Dynamics and Dept. of Physiology. McGill University, Montreal, Canada.
- Kimoto, T., Asakawa, K., Yoda, M., Takeoka, M. (1990). *Stock market prediction system with modular neural networks*. En *Proceedings of the IEEE International Joint Conference on Neural Networks*. San Diego, California, 2: 11-16
- Kuan, C.-M. y Liu, T. (1995). *Forecasting exchange rates using feedforward and recurrent neural networks*. Journal of Applied Econometrics 10: 347-364
- Kuan, C.M., Liu, T. y Gencay, R. (1997). *Netle 3.01: feedforward neural networks and Lyapunov exponents estimation*. Ball State University
- Kuan, C.-M. y White, H. (1994a). *Artificial neural networks. an econometric perspective*. Econometric Reviews 13: 1-91

- Kuan, C.-M. y White, H. (1994b). *Adaptive learning with nonlinear dynamics driven by dependent processes*. *Econometrica* 62: 1087-1114
- Lachtermacher, G. y Fuller, G.D. (1995). *Backpropagation in time-series forecasting*. *Journal of Forecasting* 14: 381-393
- Le Cun, Y. (1985). *Une Procédure de Apprentissage pour Réseau à Seuil Asymétrique*, *Proceedings of Cognition*
- Lorenz, H.W. (1993). *Nonlinear dynamical economics and chaotic motion*. Berlín. Springer-Verlag
- McCaffrey, D.F. (1991): *Estimating Lyapunov exponents with nonparametric regression and convergence rates for feedforward single hidden layer networks*. Tesis Doctoral, Departamento de Estadística, Raleigh, North Carolina State University
- McCaffrey, D.F., Ellner, S., Gallant, A.R. y Nychka, D.W. (1992): *Estimating the Lyapunov exponent of a chaotic system with nonparametric regression*. *Journal of the American Statistical Association* 87: 682-695
- McLeod, A.I. y Li, W.K. (1983). *Diagnostic checking ARMA time series modelling using squared-residual autocorrelations*. *Journal of Time Series Analysis* 4: 269-273
- Monod, J. (1993). *El azar y la necesidad*. Barcelona. Tusquets
- Morin, E. (1995). *Introducción al pensamiento complejo*. Barcelona. Gedisa
- Nieto de Alba, U. (1998). *Historia del tiempo en economía*. Madrid. McGraw-Hill
- Nychka, D., Ellner, S., Gallant, A.R. y McCaffrey, D. (1992). *Finding chaos in noisy systems*; *Journal of the Royal Statistical Society B*. 54, 2: 399-426
- Parker, D.B. (1985). *Learning Logic*. MIT Center for Computational Research in Economics and Management science, T.R. 47.
- Pesaran, M.H. y Potter, S.M. (1993). *Nonlinear dynamics, chaos and econometrics*. Nueva York. John Wiley & Sons
- Prigogine, I. (1993). *El nacimiento del tiempo*. Barcelona. Tusquets
- Prigogine, I. (1997). *Las leyes del caos*. Barcelona.
- Grijalbo-Mondadori
- Qi, M. (2001). *Predicting US recessions with leading indicators via neural network models*. *International Journal of Forecasting* 17: 383-401
- Refenes, A.N. (1993). *Constructive learning and its application to currency exchange rate forecasting*. En *Neural Networks in Finance and Investing. Using Artificial Intelligence to Improve Real-World Performance* (Eds. R.R. Trippi y E. Turban). Chicago. Probus Publishing Company
- Refenes, A.N. (1995). *Neural networks in the capital markets*. Chichester, John Wiley
- Rumelhart, D. E. , Mc Clelland, J.L. y el PDP Research Group (Eds.) (1986). *Parallel Distributed Processing. Explorations on the Microstructure of Cognition, vol.I. Foundations*, MIT Press, Cambridge, MA.
- Salchenkerger, L.M., Cinar, E.M., Lash, N.A. (1992). *Neural networks. a new tool for predicting thrift failures*. *Decision Science* 23, 4: 899-916
- Serletis, A. (1995): *Random walks, breaking trend functions, and the chaotic structure of the velocity of money*. *Journal of Business and Economic Statistics* 13: 453-458
- Scheinkman, J.A. y LeBaron, B. (1989). *Nonlinear dynamics and stock returns*. *Journal of Business* 62, 3: 311-337
- Schittenkopf, C. y Deco, G. (1997). *Identification of deterministic chaos by an information-theoretic measure of the sensitive dependence on the initial conditions*. *Physica D* 110: 173-181
- Schoneburg, E. (1990). *Stock price prediction using neural networks. a project report*. *Neurocomputing* 2: 17-27
- Shintani, M. y Linton, O. (2003). *Is there chaos in the world economy? A nonparametric test using consistent standard errors*. *International Economic Review* 44, 1: 331-358
- Shintani, M. y Linton, O. (2004). *Nonparametric neural network estimation of Lyapunov exponents and a direct test of chaos*. *Journal of Econometrics* 120: 1-33
- Tam, K.Y. y Kiang, M.Y. (1992). *Managerial applications of neural networks. the case of bank failure predictions*. *Management Science* 38: 7. 926-947

- Tang, Z. y Fishwick, P.A. (1993). *Feedforward neural nets as models for time series forecasting*. ORSA Journal of Computing 5, 4: 374-385
- Takens, F. (1986). *Detecting strange attractors in turbulence*. En *Lecture Notes in Mathematics. Dynamical Systems and Turbulence* (Ed. D.A. Rand y L.-S. Young). New York. Springer-Verlag: 366-381
- Theiler, J., Eubank, S. Longtin, A., Galdrikian, B. y Farmer, J.D. (1992a). *Testing for nonlinearity in time series. the method of surrogate data*. Physica D 58: 77-94
- Theiler, J., Eubank, S. Longtin, A., Galdrikian, B. y Farmer, J.D. (1992b). *Using surrogate data to detect nonlinearity in time series*. En *Nonlinear modeling and forecasting* (Eds. M. Casdagli y S. Eubank). Reading, Massachusetts. Addison-Wesley: 163-188
- Tong, H. (1990). *Non-linear time series. A dynamical system approach*. Oxford. Clarendon Press Oxford
- Tkacz, G. (2001). *Neural network forecasting of Canadian GDP growth*. International Journal of Forecasting 17: 57-69
- Tsay, R. S. (1986). *Nonlinearity tests for time series*. Biometrika 73, 2: 461-466.
- Trippi, R.R. and Turban, E. (1993). *Neural networks in finance and investment. using artificial intelligence to improve real-world performance*. Chicago. Probus
- Wagensberg, J. (1994). *Ideas sobre la complejidad del mundo*. Barcelona. Tusquets
- Weigend, A.S., Huberman, B.A. y Rumelhart, D.E. (1992). *Predicting sunspots and exchange rates with connectionist networks*. En *Nonlinear Modeling and Forecasting* (Eds. M. Casdagli y S. Eubank). Redwood City, CA. Addison-Wesley: 395-432
- Werbos, P.J. (1974). *Beyond Regression. New tools for Prediction and Analysis in the Behavioral Sciences*. Ph.Th. Harvard University. Department of Applied Mathematics
- White, H. (1988). *Economic prediction using neural networks. the case os IBM daily stock returns*. En *Proceeding of the IEEE International Conference on Neural Networks*, San Diego, California, 2: 451-458
- White, A. (1989). *Some asymptotic results for learning in single hidden-layer feedforward network models*. Journal of the American Statistical Association 84, 408: 1003-1013
- Wilson, R. y Sharda, R. (1994). *Bankruptcy prediction using neural networks*. Decision Support Systems 11: 545-557
- Wolf, A., Swift, J.B., Swinney, H.L. y Vastano, J.A. (1985). *Determining Lyapunov exponents from a time series*. Physica D 16: 285-317
- Zhang, G. y Hu, M.Y. (1998). *Neural network forecasting of the British Pound/US Dollar exchange rate*. Omega. International Journal of Management Science 26, 4: 495-506
- Zhang, G., Patuwo, B.E. y Hu, M.Y. (1998). *Forecasting with artificial neural networks. the state of the art*. International Journal of Forecasting 14: 35-62