

6. ESTIMACIÓN Y PREDICCIÓN DE MODELOS NO LINEALES CAÓTICOS.

6.1. INTRODUCCIÓN.

6.2. CRITERIO DE SELECCIÓN DE MODELOS.

6.3. MÉTODOS GLOBALES.

6.4. MÉTODOS LOCALES.

6.5. CONCLUSIONES.

6.1. INTRODUCCIÓN.

En el capítulo quinto, se utilizan distintos métodos para identificar la presencia de comportamientos no lineales en series temporales. Los resultados obtenidos han permitido encontrar indicios de comportamiento caótico, pero en cualquier caso no resultan concluyentes.

La presencia de ruido en las observaciones que forman una serie temporal hace que los métodos utilizados en el capítulo anterior no lleven a resultados concluyentes, tal y como señalan Schreiber y Kantz¹, aunque dicho ruido no va a afectar a la estimación de modelos del comportamiento del sistema.

¹ Schreiber y Kantz, 1995.

Una forma alternativa de comprobar la presencia de comportamientos no lineales, tal y como propone Granger², sería encontrar la ecuación generadora de este posible comportamiento caótico. A lo largo del presente capítulo vamos a proponer distintos modelos. Nuestro objetivo va a ser predecir el rendimiento del contrato de Futuro sobre el Bono Nocional a 10 años..

Vamos a partir de las series temporales de rendimientos de Futuro sobre el Bono Nocional a 10 años que habíamos obtenido en el capítulo cuarto, tanto en su forma operación a operación como en su forma de cinco minutos, y que podemos expresar como [6.1.1].

$$\{x_{ti}\}_{i=1,\dots,T} \quad [6.1.1]$$

Los rendimientos del Futuro sobre el Bono Nocional a 10 años van a formar una serie temporal, que viene generada por un sistema dinámico que expresamos con las ecuaciones [6.1.2] y [6.1.3]. La ecuación [6.1.2] define el sistema dinámico que conforma el mercado de Futuro. Este mercado viene definido por una serie de características (precio, demandantes, oferentes, precio del activo subyacente) que expresamos de forma genérica por una serie de variables que conforman el vector $s(t)$. Como se puede comprobar, no planteamos ningún tipo de hipótesis sobre la forma del sistema dinámico que representa al mercado. En la ecuación [6.1.2] incluimos tanto una parte determinista, $F(s(t))$, como una parte aleatoria, que expresamos a través de $h(t)$.

² Granger, 1991.

$$\frac{ds}{dt} = F(s(t)) + \eta(t) \quad [6.1.2]$$

$$x_t = h(s(t)) + \mathbf{x}_t \quad [6.1.3]$$

No es necesario definir las variables que forman el vector $s(t)$ en el mercado de Futuro. Basta con analizar una serie de escalares x_t que son observables y que sabemos que vienen determinados por la dinámica del sistema, esto es, está en función de las variables que forman $s(t)$ o por lo menos de algunas de ellas. Este escalar son los rendimientos obtenidos en el mercado de Futuro sobre el Bono Nocional a 10 años. Los rendimientos del activo están proporcionando información sobre el estado del sistema. En la ecuación [6.1.3] se expresa esta conexión entre la observación y las variables del sistema a través de la función $h(s(t))$. La observación del mercado no es ajena a la existencia de ruido³. Para recoger estos efectos se incluye la variable aleatoria \mathbf{x}_t , que indica el ruido presente en la observación.

En el caso de los Futuros sobre bonos, el sistema [6.1.2] es el que lleva al establecimiento del precio de mercado del tipo de interés a largo plazo. Este va a depender de una serie de factores que en algunos casos son conocidos y en otros no. Incluso entre los conocidos no todos son observables. De lo que sí tenemos observaciones son de los rendimientos del Futuro sobre el Bono Nocional a 10 años. También sabemos que este precio está conectado con el sistema dinámico en el que se determinan los tipos de interés a largo plazo.

³Producido entre otras cosas por la existencia del llamado “tick size”. No todos los precios de compra-venta son posibles, sólo son posibles, en MEFF Renta Fija, los valores enteros.

El método de Takens que ya introducimos en el capítulo anterior nos permite reconstruir el sistema [6.1.2] a partir de la serie de rendimientos de [6.1.3]. Mediante la transformación de Takens ([6.1.4]) podemos obtener una serie de vectores que son topológicamente equivalentes a la serie de vectores $s(t)$ que describían la situación el mercado.

$$X_t = (x_t, \dots, x_{t-(m-1)}) \quad [6.1.4]$$

Como ya se indicó en el capítulo anterior, la reconstrucción del sistema por el método propuesto por Takens requiere la determinación del valor de dos parámetros: m y t . Donde m es la dimensión de inmersión y t es el intervalo de tiempo entre cada uno de los elementos del vector.

Una vez que hemos aproximado la serie de vectores $s(t)$ a partir de la serie de vectores X_t , será necesario aproximar la función $F(s(t))$ de [6.1.2] para conocer la dinámica del sistema. Esto es, vamos a buscar una función ([6.1.5]) que nos permita predecir cual será el siguiente valor de la serie.

$$x_{t+\tau} = g(X_t) \quad [6.1.5]$$

En el caso de que aceptemos que el mercado presenta un comportamiento de tipo caótico, la función g va a ser no-lineal. Debido a la sensibilidad a las condiciones iniciales que tendrá el sistema si es caótico, la predicción que podemos hacer sólo tendrá sentido a corto plazo.

La existencia de una función g no lineal que permita predecir el comportamiento del rendimiento del Futuro, la variable x_t , va a ser el mejor indicio de que efectivamente estamos ante un sistema caótico.

Vamos a empezar estableciendo el criterio de decisión (6.2) que vamos a utilizar para elegir entre los distintos modelos que iremos proponiendo. A partir de esto, vamos a comparar los resultados que se derivarían de modelos que no pueden presentar comportamientos caóticos (6.3), frente a aquellos que se propondrían para sistemas con comportamientos de tipo caótico. Entre estos últimos podemos clasificarlos entre aquellos que intentan darnos una visión global del sistema (6.4) y aquellos que se conforman con intentar conocer el comportamiento aproximado en una región concreta del sistema (6.5). Por último resumiremos los resultados obtenidos a lo largo del capítulo (6.6).

6.2. CRITERIO DE SELECCIÓN DE MODELOS.

Existen distintos modelos que pueden servirnos para aproximar la evolución de los datos. Buscamos un criterio que nos permita decidir cuál de los múltiples modelos que vamos a estimar es el más adecuado.

Un primer acercamiento nos llevaría a elegir aquel modelo que mejor ajusta los datos que tenemos. Esto es, escoger aquel modelo que reduce la varianza del error cometido en la estimación. Sin embargo, hay que tener en cuenta que siempre es posible

mejorar un ajuste añadiendo al modelo nuevas variables, aunque estas variables no sean significativas. Pero ese mejor ajuste no implica un mejor conocimiento del sistema dinámico que definíamos en [6.1.2], ni una mayor capacidad predictiva sobre unos datos distintos a los usados para estimarlo.

Se han definido multitud de criterios de selección para modelos lineales en los que se penaliza el número de parámetros que incluye el modelo. Esto seguiría el esquema del R^2 corregido, el criterio de información de Akaike⁴ o el de Schwarz⁵ que ya habíamos usado en el capítulo cuarto para la selección del modelo ARIMA adecuado para las series de rendimientos de cinco minutos. Otro criterio es el de Rissanen⁶ que en determinados casos es equivalente al criterio de Schwarz, y que se basa en el análisis bayesiano⁷.

Todos estos criterios son válidos para los modelos lineales, pero no para los no lineales. En el caso de modelos no lineales, la estructura no es la misma, y en ciertos casos incluso el término parámetro no queda claro⁸. En estos casos los supuestos que permiten definir estos criterios ya no son válidos.

Para encontrar un criterio adecuado vamos a basarnos en su capacidad de predecir, que es, al fin y al cabo lo que andamos buscando. Por eso el criterio que hemos considerado más adecuado es, sencillamente elegir aquel modelo que mejor predice. A

⁴ Akaike, 1977.

⁵ Schwarz, 1978.

⁶ Rissanen, 1981; Rissanen, 1986.

⁷ Rissanen, 1986; Schwartz, 1978.

⁸ Los métodos locales.

este criterio se le denomina *validación cruzada*⁹ y consiste en separar una parte de los datos antes de estimar los modelos, de tal forma que escogeremos aquel modelo que reduzca el error en la predicción de ese conjunto que hemos separado y que no hemos utilizado para estimar el modelo. Este método resulta homogéneo para seleccionar entre modelos con estructura diferente.

Para las series de rendimientos de cinco minutos, hemos separado en el final de cada una de las series de tres meses, el último 10% de la serie, lo que supone entre 510 y 660 datos. En las series de rendimientos operación a operación, dado que un 10% supondría una muestra excesivamente grande (hasta 20.000 datos), nos hemos limitado a separar los últimos 2.000 datos. Es sobre estas series sobre las que vamos a comprobar la bondad de los distintos ajustes propuestos.

En cada modelo vamos a comparar el error cuadrático medio cometido en la predicción con la varianza de los datos. El modelo es adecuado, si logra reducir la varianza, esto es, si consigue explicar parte de la variabilidad. No será adecuado si dicho error cuadrático medio es superior a la varianza del proceso. Para poder comparar con posterioridad, incluimos en la tabla 6.2.1 las varianzas de las series de rendimientos en el segmento no utilizado en la estimación de los modelos. Lo primero que nos llama la atención es el hecho de que la varianza de las series operación a operación es mucho más baja que las de cinco minutos. El hecho de haber aumentado el intervalo temporal de las series está provocando un aumento en el tamaño de la varianza.

⁹ Efron y Tibshirani, 1986; Stone, 1974; Stone, 1977a; Stone, 1977b.

Tabla 6.2.1: Varianza de las series de rendimientos.

	Serie operación a operación	Serie cinco minutos
Bono marzo 98	3,0971E-09	5,5011E-08
Bono junio 98	1,9014E-09	3,3625E-08
Bono septiembre 98	4,3346E-09	10,0181E-08
Bono diciembre 98	4,2676E-09	9,6631E-08

Con la predicción no sólo buscamos demostrar que hay modelos no lineales (en los que se presentan los comportamientos caóticos) que son capaces de explicar parte del comportamiento de los rendimientos del Futuro sobre el Bono Nocional a 10 años, sino que ese poder explicativo es superior al de modelos lineales, donde no resulta posible el comportamiento de tipo caótico.

El modelo lineal al que hacemos referencia en este apartado no es más que el modelo ARIMA con la estructura identificada en el capítulo cuarto. Eso sí, en este caso estimamos los modelos partiendo de la primera parte de la serie y comprobando su capacidad predictiva sobre la segunda.

La primera incógnita que se nos plantea es decidir qué es lo que queremos predecir. Para poder comparar los modelos lineales en igualdad de condiciones con los no lineales tenemos que utilizar mismo tipo de predicciones. En el caso de los modelos no lineales partimos de la reconstrucción del sistema que hacíamos en el capítulo quinto, de tal forma que lo que hacemos es obtener una estimación de $x_{t+\tau}$ en función de la m -historia X_t . Vamos a comprobar el error que cometeríamos con nuestro modelo ARIMA si intentáramos predecir el valor de $x_{t+\tau}$ conociendo los valores de la serie hasta el momento t .

Los resultados obtenidos se pueden observar en la Tabla 6.2.2. En ella se comprueba como el hecho de conocer el pasado de la serie no nos garantiza que vayamos a conocer su comportamiento futuro dado su carácter estocástico.

Tabla 6.2.2: Error cuadrático medio de la predicción con modelos ARIMA.

	Serie operación a operación		Serie cinco minutos	
	ECM(*)	ECM/ var(x)	ECM	ECM/ var(x)
Bono marzo 98	3,1103E-09	1,0043	5,5119E-08	1,0020
Bono junio 98	1,9098E-09	1,0044	3,3957E-08	1,0099
Bono septiembre 98	4,3410E-09	1,0015	10,0699E-08	1,0052
Bono diciembre 98	4,2822E-09	1,0034	9,6406E-08	0,9977

(*)ECM: Error cuadrático medio; ECM: Error cuadrático medio partido de la varianza.

El hecho de que un proceso pueda modelizarse de forma lineal, no quiere decir que esta sea la única explicación posible. Los modelos lineales se basan en el concepto de incorrelación. Sin embargo la incorrelación no implica independencia. Así, por ejemplo, una serie obtenida a partir de la función logística puede estar incorrelacionada, pero resulta posible expresarlo de forma puramente determinista, si utilizamos un modelo no lineal adecuado.

6.3. MÉTODOS GLOBALES.

Los métodos globales se basan en aproximar la función no lineal g de la expresión [6.1.5]. La forma más simple de lograr la aproximación es una combinación lineal de funciones no lineales f_i [6.3.1].

$$\hat{x}(t+1) = \sum_{i=1}^k a_i f_i(X(t)) \quad [6.3.1]$$

Dependiendo de la forma que tengan estas funciones, la estimación de los parámetros es más o menos complicada. Si las funciones f_i no dependen de ningún parámetro, la estimación se limitará al valor de los coeficientes a_i . Estos coeficientes podremos estimarlos entonces por mínimos cuadrados ordinarios. En el caso de que las funciones f_i sí dependan de otros parámetros, la estimación ya no podrá ser por mínimos cuadrados ordinarios y tendremos que buscar otras formas de estimarlos, esto es lo que haremos con las redes neuronales artificiales.

La forma más simple que pueden presentar las funciones f_i van a ser los *polinomios*¹⁰. Esto nos permite tratar nuestra aproximación a la función objetivo g , como si fuera un desarrollo de Taylor. La función [6.3.1] pasará a tener la forma:

$$\hat{x}(t+1) = \sum_{i=1}^k a_i \cdot x_1^{l_{i1}} \cdots x_m^{l_{im}} \quad [6.3.2]$$

La ecuación [6.3.2] es fácilmente estimable por mínimos cuadrados ordinarios. Consideramos los productos de los distintos x_r elevados a las potencias λ_r como una única variable, con lo que podemos estimar los valores de los parámetros a_i .

Cada combinación de los exponentes λ_r suponen una variable distinta. Dependiendo de la dimensión de inmersión m que estemos considerando y del grado del polinomio, el número de variables que tendremos en nuestra regresión será muy distinto. Este número crece de forma exponencial, a medida que aumenta el grado del

¹⁰ Casdagli, Des Jardins, Eubank, Farmer, Gibson, Theiler y Hunter, 1992; Kantz y Schreiber, 1997.

polinomio. En la Tabla 6.3.1 se representan las variables que se obtendrían incluyendo todas las combinaciones posibles de los exponentes.

Tabla 6.3.1: Numero de parámetros \mathbf{a} a estimar en la ecuación [6.3.2] según cual sea la dimensión de inmersión y el grado máximo.

		Dimensión de inmersión								
		2	3	4	5	6	7	8	9	10
Grado máximo	1	2	3	4	5	6	7	8	9	10
	2	5	10	16	23	31	40	50	61	73
	3	8	19	35	57	83	120	166	222	289
	4	12	33	69	126	199	320	486	707	994
	5	17	53	124	251	432	763	1259	1975	2977
	6	23	80	207	460	863	1656	2944	4947	7951
	7	30	115	326	789	1609	3326	6330	11336	19345
	8	38	159	490	1283	2832	6265	12701	24142	43591
	9	47	213	709	1997	4749	11185	24056	48367	92126

En vista del tamaño que podría llegar a tener la estimación, nos hemos quedado en un grado máximo igual a cuatro. Los resultados obtenidos por las distintas regresiones se pueden ver en la Tabla 6.3.2 en la que se presentan los mejores resultados para cada serie.

Tabla 6.3.2: Error cuadrático medio de la predicción con polinomios.

	Serie operación a operación		Serie cinco minutos	
	ECM	ECM/var(x)	ECM	ECM/var(x)
Bono marzo 98	3,0673E-09	0,9904	4,6715E-08	0,8492
Bono junio 98	1,8785E-09	0,9880	3,2931E-08	0,9794
Bono septiembre 98	4,3414E-09	1,0016	10,0670E-08	1,0048
Bono diciembre 98	4,2824E-09	1,0035	6,1962E-08	0,6412

Los resultados obtenidos por los polinomios no permiten explicar nada, en el caso de las series de operación a operación y en el caso de las series de cinco minutos, no se producen mejoras ni en la serie de junio ni en la serie de septiembre. Sin embargo,

sí logramos predecir tanto para la serie con vencimiento en marzo como para la de diciembre.

Un tipo de funciones \mathbf{f}_i alternativas a las funciones polinómicas son las *funciones radiales* propuestas por Smith para la estimación de sistemas con atractores extraños¹¹. Las funciones radiales no toman como argumento el valor de los vectores de las m -historias, sino la distancias de éstas sobre otras m -historias. Esto nos da funciones de la forma de la ecuación [6.3.3].

$$\hat{x}(t+1) = \sum_{i=1}^k \mathbf{a}_i \mathbf{f}_i(\|X(t) - X(i)\|) \quad [6.3.3]$$

En [6.3.3], las \mathbf{f}_i son funciones con forma acampanada con un máximo en cero y que decrecen hacia cero a medida que aumenta el valor absoluto del argumento de la función. Dicho argumento, $\|X(t)-X(i)\|$, es la distancia desde la m -historia que representamos por $X(t)$ respecto a la i -ésima m -historia de la serie.

La no linealidad del sistema se intenta recoger mediante la selección de una serie de puntos de referencia representativos de todo el sistema. En las cercanías de uno de esos puntos de referencia habrá un determinado comportamiento, mientras que ese comportamiento será distinto en la cercanía de otro de esos puntos de referencia.

¹¹ Smith, 1992.

Para determinar completamente la ecuación [6.3.3] resulta necesario fijar los que van a ser los puntos de referencia y la función f_i . Para la selección de la función hemos seguido a Kantz y Schreiber¹², que proponen una función con la forma de la figura 6.3.1 y que representamos con la ecuación [6.3.4].

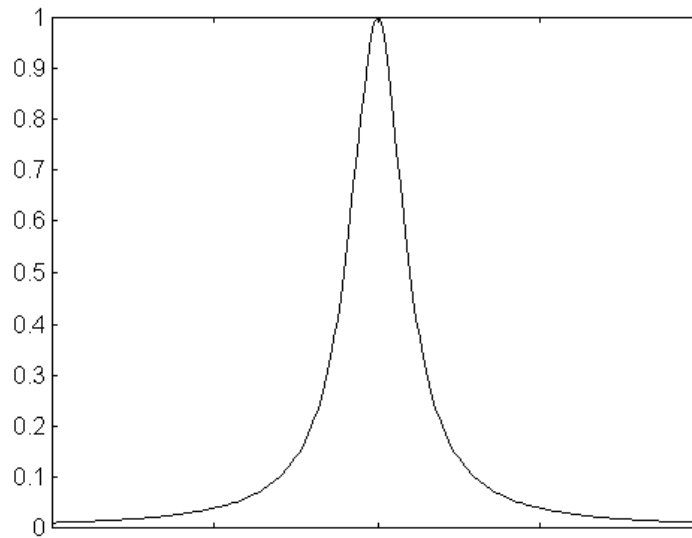


Figura 6.3.1: Representación de una función de base radial como la de la ecuación [6.3.5]

$$f(r) = \frac{1}{1 + \frac{r^2}{a^2}} \quad [6.3.4]$$

La función [6.3.5] es una función en la que únicamente aparece un parámetro, a . La presencia de un parámetro dentro de la función f_i hace que no sea posible la estimación por mínimos cuadrados ordinarios a la vez de los distintos a y \mathbf{a}_i .

La forma de proceder para estimar el modelo es fijar primero el valor que damos al parámetro a y a continuación estimamos \mathbf{a}_i por mínimos cuadrados ordinarios.

¹² Kantz y Schreiber, 1997.

Hemos propuesto 11 valores distintos para el parámetro α , que van desde dos veces la desviación típica de la serie a un medio de esta misma desviación típica.

El método utilizado para seleccionar los puntos de referencia para calcular las distancias que son los argumentos de las funciones radiales. El primer punto que se escoge es la primera m -historia de la serie. Se estima el modelo con esa sola función radial, y se calcula el error que cometemos en la estimación. A continuación, proponemos un nuevo modelo con dos funciones radiales, tomando como puntos de referencia los del modelo anterior y otro nuevo que será aquél que en la estimación anterior nos llevaba a cometer un mayor error. Este proceso continua añadiendo nuevas funciones para así ir incorporando las distintas formas de comportamiento que genera el sistema. En nuestro caso, hemos estimado modelos hasta llegar a una combinación de 50 funciones de base radial.

Tabla 6.3.3: Mejores resultados de predicción con funciones de base radial. Errores cuadráticos medios y porcentaje sobre la varianza

	Serie operación a operación		Serie cinco minutos	
	ECM	ECM/var(x)	ECM	ECM/var(x)
Bono marzo 98	3,0736E-09	0,9924	4,6364E-08	0,8428
Bono junio 98	1,8757E-09	0,9865	3,2855E-08	0,9771
Bono septiembre 98	4,3416E-09	1,0016	9,8485E-08	0,9831
Bono diciembre 98	4,2774E-09	1,0023	6,1497E-08	0,6364

Los mejores resultados obtenidos por los modelos de funciones de base radial se presentan en la Tabla 6.3.3. En este caso hemos estimado 50 modelos para cada uno de los 11 distintos valores del parámetro α y para dimensiones de inmersión entre dos y diez, lo que hace un total de 4950 modelos para cada una de las series. Los resultados obtenidos son similares a los que teníamos con los modelos de polinomios y no

presentan una mejora significativa respecto a ellos. De nuevo los únicos casos en los que conseguimos predecir son los de los rendimientos cada cinco minutos de los contratos de Futuro sobre el Bono Nocional a 10 años con vencimiento en marzo y diciembre de 1998.

El último tipo de modelo global que vamos a probar son los modelos de *redes neuronales artificiales*. Tal y como señala White¹³, las redes neuronales permiten hacer una aproximación a las funciones de tipo caótico. La ventaja que presentan es que son modelos globales no lineales que no requieren establecer ningún tipo de hipótesis sobre la forma del sistema que intentamos aproximar.

El nombre de redes neuronales procede de la supuesta similitud entre la estructura de funcionamiento del método de aproximación de funciones y las neuronas de los seres vivos. Las neuronas se estructuran de tal forma que se conectan entre sí. De igual forma, los valores que toman unas funciones sirven de argumentos para otras funciones.

El tipo de red neuronal artificial que vamos a utilizar es el propuesto por White¹⁴, consistente en una capa de neuronas de entrada que van a ser los elementos del vector de m -historias, otra capa de salida, que será el siguiente elemento al de la m -historia, y que es el que tratamos de predecir, y una serie de capas internas formadas por un conjunto de funciones que en un principio pueden ser de cualquier tipo. Esa estructura es la que aparece representada en la figura 6.3.2.

¹³ White, 1989.

¹⁴ White, 1989. También Jaeger y Kantz, 1996, utilizan este tipo de red neuronal.

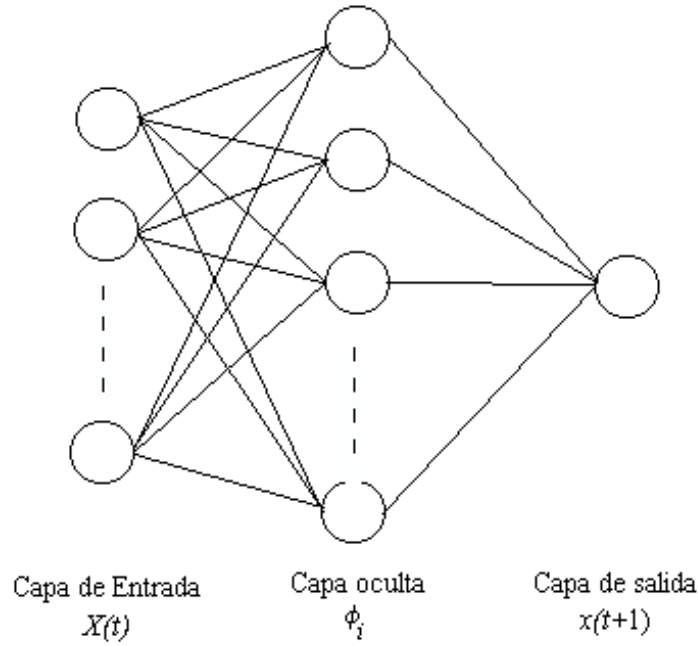


Figura 6.3.2: Estructura de la red neuronal que vamos a usar para la estimación de modelos no lineales.

La estructura de esta red se denomina de alimentación hacia delante¹⁵. Esto es, las neuronas de la capa de entrada están conectadas con la capa interna, y esta con la de salida. La entrada a la neurona intermedia es una combinación lineal de las variables de la capa de entrada más un sesgo. La salida de la capa externa es una combinación lineal de las salidas de las capas intermedias, más un nuevo sesgo. Al final habremos obtenido una función que tendrá la forma de la ecuación [6.3.5].

$$\hat{x}_{t+m \cdot t} = a + \sum_{i=1}^k \mathbf{a}_i \cdot \mathbf{f}_i(b_i + w_{i1}x_t + w_{i2}x_{t+1} + \dots + w_{im}x_{t+(m-1) \cdot t}) \quad [6.3.5]$$

¹⁵ Traducción del término inglés: *feed-forward*.

En la ecuación [6.3.5] tenemos las funciones f_i que ya teníamos en la ecuación [6.3.1] y que ahora forman la capa interna. Los coeficientes w_{ij} son los pesos de las conexiones entre la capa de entrada y la capa interna y a_i son los pesos entre la capa interna y la capa de salida. Por último, a los coeficientes a y b_i se les denomina sesgos. En nuestro caso, para las funciones f_i hemos escogido funciones sigmoidales pues son las más habituales. Las funciones sigmoidales tienen la forma de la ecuación [6.3.6] y de la figura 6.3.3.

$$f(x) = \frac{1}{1 + \frac{1}{e^x}} \quad [6.3.6]$$

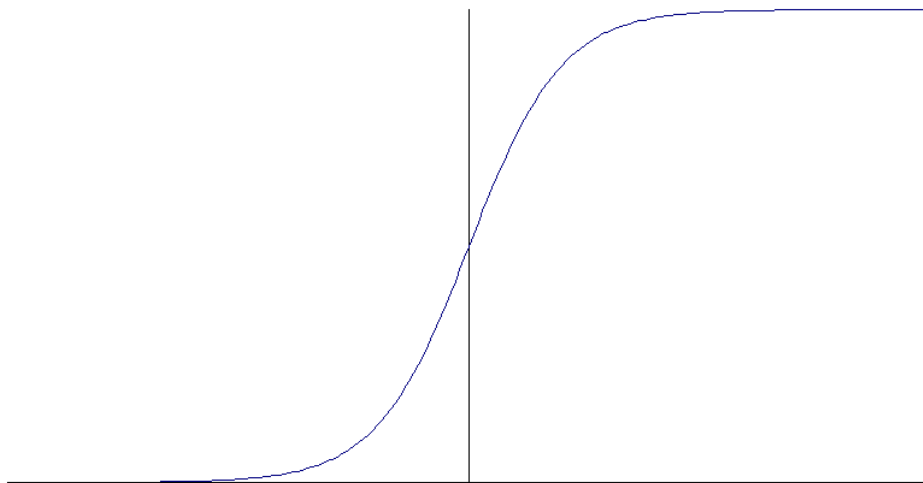


Figura 6.3.3: Forma de la función sigmoide de la ecuación [6.3.7]

La ecuación [6.3.5] es no lineal y no vamos a poder estimar los parámetros por mínimos cuadrados ordinarios. La forma de estimarlos se denomina entrenamiento de la red neuronal.

El método de entrenamiento¹⁶ que vamos a utilizar es una versión iterativa de los mínimos cuadrados denominada técnica del descenso más pronunciado. Partimos de la función de error cuadrático de la red neuronal que expresamos en la ecuación [6.3.7].

$$E = \sum_t (x_{t+m} - \hat{x}_{t+m})^2 \quad [6.3.7]$$

Sobre la función de error cuadrático [6.3.7] se calcula la derivada respecto a los parámetros, resultado que aparece en la ecuación [6.3.8], para saber en qué sentido deberíamos variar los parámetros para lograr que el error cuadrático descendiera más deprisa.

$$\frac{\partial E}{\partial a} = -2 \sum_t (x_{t+m} - \hat{x}_{t+m}) \quad [6.3.8]$$

$$\frac{\partial E}{\partial \mathbf{a}_i} = -2 \sum_t (x_{t+m} - \hat{x}_{t+m}) \cdot \mathbf{f}_i \quad [6.3.9]$$

Una vez que tenemos las derivadas [6.3.8] y [6.3.9], sabemos en qué dirección deben cambiar los parámetros para lograr un mejor ajuste. Con esta información modificamos el valor de los parámetros para reducir el error, siguiendo las ecuaciones [6.3.10] y [6.3.11].

$$a = a + \mathbf{h} \sum_t (x_{t+m} - \hat{x}_{t+m}) \quad [6.3.10]$$

¹⁶ Seguimos aquí a Freeman y Skapura, 1991.

$$\mathbf{a}_i = \mathbf{a}_i + \mathbf{h} \sum (x_{t+m} - \hat{x}_{t+m}) \mathbf{f}_i \quad [6.3.11]$$

En las ecuaciones [6.3.10] y [6.3.11] se realiza el ajuste de los parámetros, incluyendo un nuevo parámetro, \mathbf{h} , que se denomina tasa de aprendizaje¹⁷. La tasa de aprendizaje nos da la proporción en que utilizamos las derivadas [6.3.8] y [6.3.9] para modificar el valor de \mathbf{a} y de los α_i . La tasa de aprendizaje es un parámetro subjetivo que no vamos a poder optimizar de otra forma que probando distintos valores. Si escogemos una tasa de aprendizaje demasiado grande, podemos provocar un cambio demasiado grande y saltarnos el mínimo de la función de error que buscamos. Si el valor es demasiado pequeño, necesitaremos más iteraciones para alcanzar el mínimo, y más fácil será que nos quedemos en un mínimo local. En nuestro caso, tras exploraciones iniciales, hemos utilizado once valores distintos para la tasa de aprendizaje, desde $\mathbf{h} = 0.0001$ hasta $\mathbf{h} = 0.000001$.

Para los coeficientes de la capa intermedia usamos el mismo método de ajuste. En este caso tenemos un problema añadido, pues no sabemos cuál es el valor real que debe tomar la salida de cada una de estas neuronas. La forma de intervención de cada una de las neuronas en el error total de la función se denomina propagación hacia atrás¹⁸. Lo que nos va a llevar a obtener el valor de las derivadas parciales del error, en las ecuaciones [6.3.12] y [6.3.13].

¹⁷ Traducción del término inglés: *Learning-rate*.

¹⁸ Traducción del inglés: *backpropagation*.

$$\frac{\partial E}{\partial b_i} = -2 \sum (x_{t+m,t} - \hat{x}_{t+m,t}) \mathbf{a}_i \cdot \mathbf{f}_i' \quad [6.3.12]$$

$$\frac{\partial E}{\partial w_{ij}} = -2 \sum (x_{t+m,t} - \hat{x}_{t+m,t}) \mathbf{a}_i \cdot \mathbf{f}_i' \cdot x_{t+\cdot,j} \quad [6.3.13]$$

Los resultados de las derivadas [6.3.12] y [6.3.13] nos permiten definir el cambio que tenemos que hacer en los coeficientes b_i y w_{ij} , que se corresponden con las entradas de las neuronas de la capa interna. Los nuevos coeficientes serán los que aparecen en las ecuaciones [6.3.14] y [6.3.15].

$$b_i = b_i + \mathbf{h} \sum (x_{t+m,t} - \hat{x}_{t+m,t}) \mathbf{a}_i \cdot \mathbf{f}_i' \quad [6.3.14]$$

$$w_{ij} = w_{ij} + \mathbf{h} \sum (x_{t+m,t} - \hat{x}_{t+m,t}) \mathbf{a}_i \cdot \mathbf{f}_i' \cdot x_{t+\cdot,j} \quad [6.3.15]$$

Con los nuevos coeficientes obtenidos en las ecuaciones [6.3.10], [6.3.11], [6.3.14] y [6.3.15], podemos volver a calcular los errores cometidos en las estimaciones e ir iterando el proceso hasta lograr el ajuste que buscamos.

En nuestro caso, hemos probado entre dos y ocho neuronas en la capa interna, mientras que el número de neuronas de la capa de entrada viene determinado por la dimensión de inmersión, m . Hemos probado entre dos y diez dimensiones de inmersión. Esto supone 63 configuraciones distintas de la red neuronal. Para cada una de esas 63

configuraciones hemos utilizado las once tasas de aprendizaje que habíamos establecido con anterioridad para entrenar la red. Esto supone un total de 693 redes.

El entrenamiento de las redes neuronales requiere dar unos valores iniciales a los coeficientes que estamos estimando. Para evitar que los resultados obtenidos se deban a un mínimo local de la función de error cuadrático, debemos probar con distintos valores iniciales de estos coeficientes, y que hemos elegido aleatoriamente. Para cada una de las 693 redes hemos probado con 10 valores iniciales distintos lo que supone que hemos entrenado 6930 redes para cada una de las series de rendimientos del contrato de Futuro sobre el Bono Nocional a 10 años que estamos estudiando. Los mejores resultados de estas redes aparecen en la Tabla 6.3.4.

Tabla 6.3.4: Mejores resultados de predicción con redes neuronales

	Serie operación a operación		Serie cinco minutos	
	ECM	ECM/var(x)	ECM	ECM/var(x)
Bono marzo 98	2,0206E-09	0,6524	4,6481E-08	0,8449
Bono junio 98	1,8411E-09	0,9683	3,3561E-08	0,9981
Bono septiembre 98	4,2802E-09	0,9874	9,9913E-08	0,9973
Bono diciembre 98	2,2891E-09	0,5364	6,1731E-08	0,6388

Errores cuadráticos medios y porcentaje sobre la varianza

De los resultados de la Tabla 6.3.4 podemos ver que para las series de cinco minutos, aunque se produce una mejora con respecto a los sistemas anteriores, estos no son muy distintos. Sin embargo, en las series de rendimientos operación a operación, la mejora sí es significativa, sobre todo en las series con vencimiento en marzo y en diciembre.

6.4. MÉTODOS LOCALES.

Como hemos comprobado en el apartado anterior, resulta mucho más complicado intentar estimar un modelo no lineal del que desconocemos hasta su forma que estimar un modelo lineal. Dada la complejidad que presentan los sistemas dinámicos que dan lugar a los comportamientos caóticos, puede resultar mucho más práctico renunciar a encontrar la forma de la función g de la ecuación [6.1.5]. En lugar de esto podemos tratar el problema de forma local. Si nos aproximamos al sistema de forma local podremos considerar a g como una función que es lineal alrededor de esa zona y podríamos utilizar métodos lineales para tratar de predecir el comportamiento del sistema.

Este tipo de actuación supone renunciar a conocer la estructura y el comportamiento del sistema, concentrando nuestro objetivo únicamente en tratar de predecir los valores futuros de este.

La forma más simple de predecir es buscar comportamientos similares a la m -historia X_t en el pasado (m -historia X_r), pues suponemos que lo que suceda a partir de ahora se asemejará a lo que pasó en esa situación similar del pasado. Así usaríamos como predicción de $x_{t+\tau}$ al valor $x_{r+\tau}$.

En el capítulo quinto, para calcular los exponentes de Lyapunov considerábamos que dos trayectorias próximas inicialmente se irían alejando de forma exponencial, pero a corto plazo, una nos va a servir para predecir la otra. A este método se le denomina de

los *vecinos próximos*. Los primeros en proponerlos en este contexto son Eckmann y Ruelle¹⁹.

Así vamos a proponer como primer método predictivo de tipo local, vamos a buscar las m -historias más próximas a la m -historia que tratamos de explicar. La predicción de $x_{t+\tau}$ va a ser el valor medio de los $x_{r+\tau}$ [6.4.1].

$$\hat{x}_{t+\tau} = E[x_{r+\tau}] \quad \forall r / \|X_r - X_t\| < l \quad [6.4.1]$$

En [6.4.1] hay un único parámetro que tenemos que fijar, l , que determina la región alrededor de la m -historia X_t de la serie que queremos predecir, que vamos a considerar como región local. Como alternativa a determinar l tenemos la posibilidad de fijar el número de vecinos, X_r , que vamos a considerar como próximos a X_t y que vamos a usar para predecir. El número de vecinos resulta una decisión complicada, pues si bien un número reducido puede llevarnos a una predicción no demasiado ajustada, si aumentamos el tamaño demasiado, estaremos recogiendo no linealidades que es precisamente lo que queremos evitar.

Para la predicción de los rendimientos cada cinco minutos del contrato de Futuro sobre el Bono Nocional a 10 años mediante vecinos próximos, hemos utilizado para cada dimensión de inmersión cien tipos de regiones, desde los diez vecinos más próximos, hasta los mil vecinos más cercanos. En el caso de los rendimientos operación a operación, dado el mayor tamaño de la serie resulta posible considerar un mayor

¹⁹ Eckmann y Ruelle, 1985.

número de vecinos próximos por lo que hemos probado 200 tamaños de la región local, entre 25 y 5000 vecinos próximos. En total hemos considerado 900 tipos de predicciones por vecinos próximos para las series de cinco minutos y 1800 tipos para las series de operación a operación. Los mejores resultados obtenidos con cada uno de ellos se presentan en la tabla 6.4.1.

Tabla 6.4.1: Mejores resultados de predicción con vecinos próximos.²⁰

	Serie operación a operación		Serie cinco minutos	
	ECM	ECM/var(x)	ECM	ECM/var(x)
Bono marzo 98	3,0748E-09	0,9928	4,5448E-08	0,8262
Bono junio 98	1,8754E-09	0,9863	3,2805E-08	0,9756
Bono septiembre 98	4,3378E-09	1,0007	9,8814E-08	0,9864
Bono diciembre 98	4,2844E-09	1,0039	6,1386E-08	0,6353

La primera columna representa los errores cuadráticos medios y la segunda el porcentaje sobre la varianza

En la tabla 6.4.1 se ve que los resultados obtenidos son similares a los del apartado anterior, en la que destacan los resultados de las series de cinco minutos con vencimientos en marzo y diciembre. Estos casos muestran que es posible explicar en parte los rendimientos con este tipo de métodos.

Una versión refinada del método de los vecinos próximos es el método de los *vecinos próximos ponderados*²¹, en el que a la hora de calcular el valor medio de los $x_{t+\tau}$ damos más peso a los vecinos más próximos y menos a los más alejados [6.4.2]. Esto lo podemos lograr multiplicando cada vecino por la inversa de la distancia respecto al punto que queremos predecir.

²⁰ Resultados obtenidos a partir de programas de elaboración propia en MATLAB 4.0.

²¹ Casdagli, Des Jardins, Eubank, Farmer, Gibson, Theiler y Hunter, 1992.

$$\hat{x}_{t+t} = \frac{\sum x_{r+t} \frac{1}{\|X_r - X_t\|}}{\sum \frac{1}{\|X_r - X_t\|}} \quad \forall r / \|X_r - X_t\| < l \quad [6.4.2]$$

Para predecir el valor de los rendimientos usando los vecinos próximos ponderados utilizamos los mismos números de vecinos que usábamos con los vecinos próximos. En la tabla 6.4.2 se recogen los mejores resultados que hemos conseguido para cada una de las series.

Tabla 6.4.2: Mejores resultados de predicción con vecinos próximos ponderados²²

	Serie operación a operación		Serie cinco minutos	
	ECM	ECM/var(x)	ECM	ECM/var(x)
Bono marzo 98	3,1002E-09	1,0010	4,6003E-08	0,8362
Bono junio 98	1,8752E-09	0,9862	3,3923E-08	1,0089
Bono septiembre 98	4,3502E-09	1,0036	9,8880E-08	0,9870
Bono diciembre 98	4,3136E-09	1,0108	6,2392E-08	0,6457

La primera columna representa los errores cuadráticos medios y la segunda el porcentaje sobre la varianza

Los resultados de los vecinos próximos ponderados son muy similares a los de los vecinos próximos aunque en todos los casos ligeramente peores. De nuevo resaltan las mejoras logradas en las series de cinco minutos con vencimientos en marzo y en diciembre.

El siguiente paso lógico es considerar que, si tenemos una serie de puntos localmente próximos, y si consideramos que a corto plazo el comportamiento se aproximará a ser lineal²³, entonces nos bastará con hacer una regresión lineal con estos

²² Resultados obtenidos a partir de programas de elaboración propia en MATLAB 4.0.

²³ Kostelich y Yorke, 1990; Farmer, y Sidorowich, 1991.

vecinos. La idea del regresor local viene tomada del método para calcular el espectro de exponentes de Lyapunov propuesto por Eckmann y Ruelle²⁴ y que ya hemos desarrollado en el apartado correspondiente del capítulo quinto. A este método de predicción lo denominamos *vecinos próximos lineales*.

Lo que hacemos con este último método local es aproximar la dinámica del sistema mediante el sistema lineal [6.4.3]. Donde A es una matriz de $m \times m$ y b un vector de tamaño m . A es el jacobiano de la función $g(x)$ de la ecuación [6.1.5] en el punto t que tomamos como referencia para predecir. Los coeficientes de las matrices A y b son obtenidos por regresiones lineales ordinarias, en las que sólo consideramos el conjunto de vecinos próximos al vector $X(t)$ de referencia.

$$\hat{x}_{t+t} = A \cdot X_t + b \quad [6.4.3]$$

De nuevo, la aplicación del método de los vecinos próximos lineales ha sido idéntica a la que hacíamos con los vecinos próximos y los vecinos próximos ponderados. Sin embargo, en este método tendremos problemas en ciertos casos en los que un número reducido de vecinos puede llevarnos a la presencia de multicolinealidad perfecta. Seleccionar el conjunto de vecinos para hacer la predicción resulta un problema²⁵ que tenemos que resolver. Si el número es demasiado pequeño corremos el riesgo de aproximarnos a una matriz singular lo que haría que la varianza en los

²⁴ Eckmann y Ruelle, 1985.

²⁵ Kostelich, 1992.

parámetros estimados fuera muy alta. En esos casos hemos tenido que ir aumentando el número de vecinos hasta lograr acabar con la multicolinealidad perfecta.

Los mejores resultados de aplicar el método de los vecinos próximos lineales para cada una de las series de rendimientos de cinco minutos y de operación a operación se pueden encontrar en la tabla 6.4.3. Los resultados son de nuevo muy similares a los obtenidos con los vecinos próximos y los vecinos próximos ponderados. Si bien en este caso en algunas series sí se mejoran los vecinos próximos.

Tabla 6.4.3: Mejores resultados de predicción con vecinos próximos lineales ²⁶

	Serie operación a operación		Serie cinco minutos	
	ECM	ECM/var(x)	ECM	ECM/var(x)
Bono marzo 98	3,0816E-09	0,9950	4,5543E-08	0,8279
Bono junio 98	1,8820E-09	0,9898	3,2756E-08	0,9742
Bono septiembre 98	4,3172E-09	0,9960	9,9356E-08	0,9918
Bono diciembre 98	4,2714E-09	1,0009	6,1305E-08	0,6344

La primera columna representa los errores cuadráticos medios y la segunda el porcentaje sobre la varianza

Para poder comparar los resultados que se obtienen con cada uno de los métodos de predicción local, se pueden ver en las gráficas de la 6.4.1 a la 6.4.8 los resultados de todas nuestras observaciones.

²⁶ Resultados obtenidos a partir de programas de elaboración propia en MATLAB 4.0.

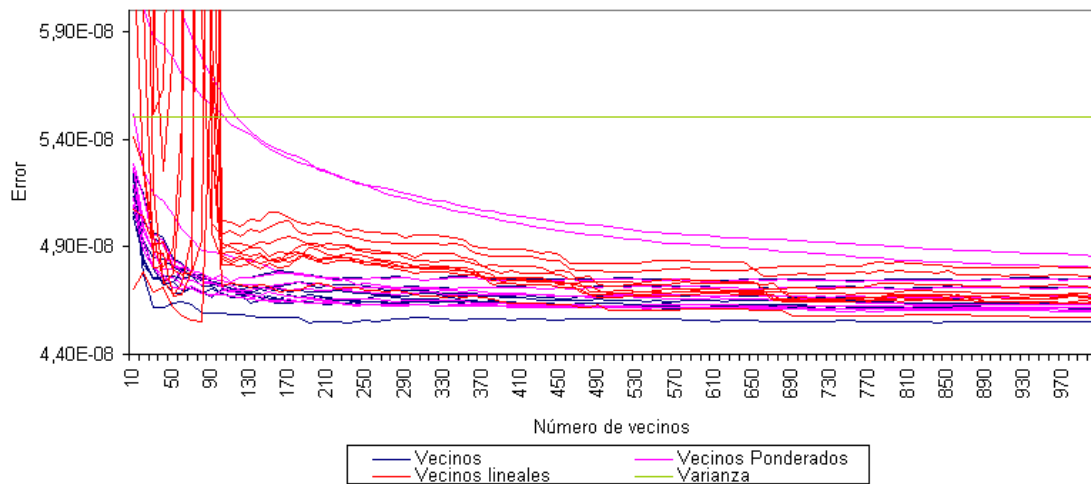


Figura 6.4.1: Errores cuadráticos medios de los vecinos próximos comparados con la varianza, para el rendimiento de los Futuro sobre el Bono con vencimiento en marzo 98 de cinco en cinco minutos

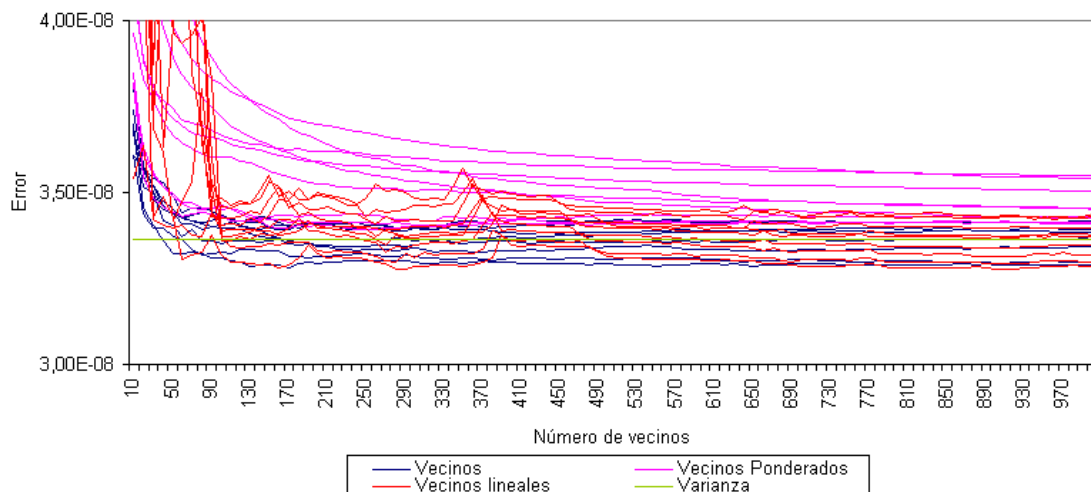


Figura 6.4.2: Errores cuadráticos medios de los vecinos próximos comparados con la varianza, para el rendimiento del Futuro sobre el Bono con vencimiento en junio 98 de cinco en cinco minutos

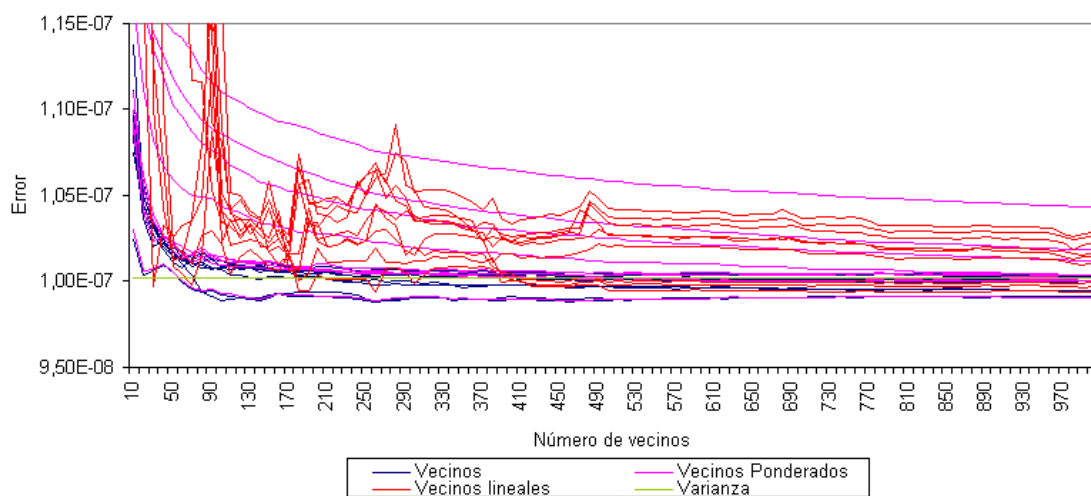


Figura 6.4.3: Errores cuadráticos medios de los vecinos próximos comparados con la varianza, para el rendimiento Futuro sobre el Bono con vencimiento en septiembre 98 de cinco en cinco minutos

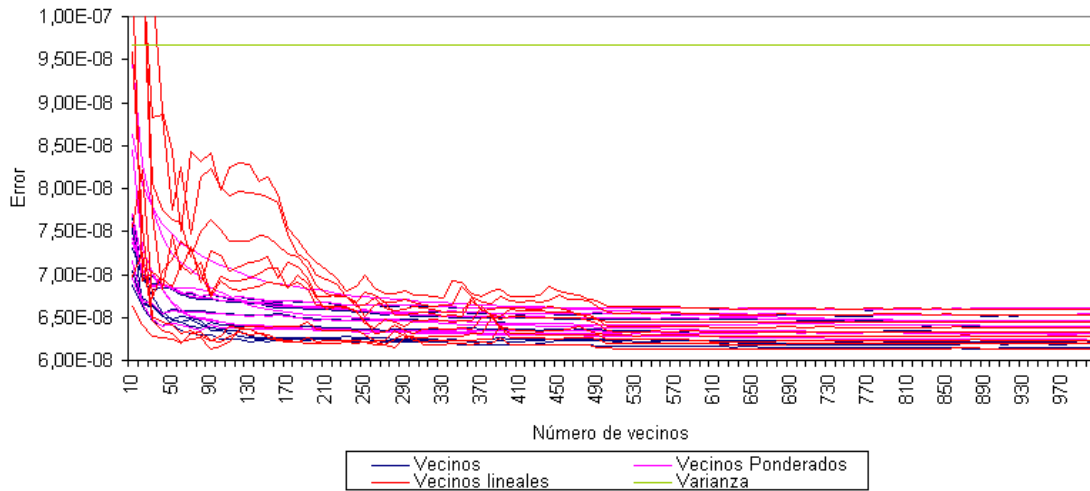


Figura 6.4.4: Errores cuadráticos medios de los vecinos próximos comparados con la varianza, para el rendimiento del Futuro sobre el Bono con vencimiento en diciembre 98 de cinco en cinco minutos

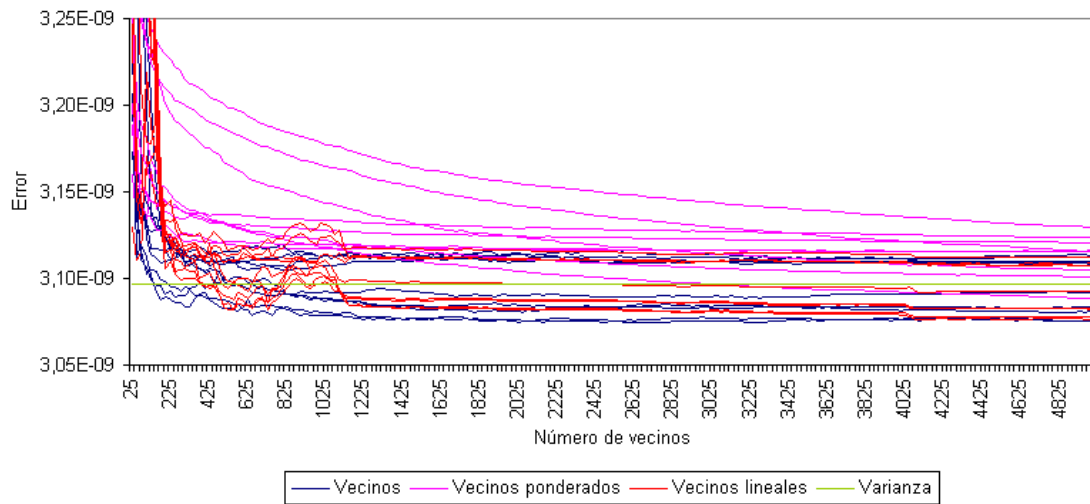


Figura 6.4.5: Errores cuadráticos medios de los vecinos próximos comparados con la varianza, para el rendimiento del Futuro sobre el Bono con vencimiento en marzo 98 operación a operación

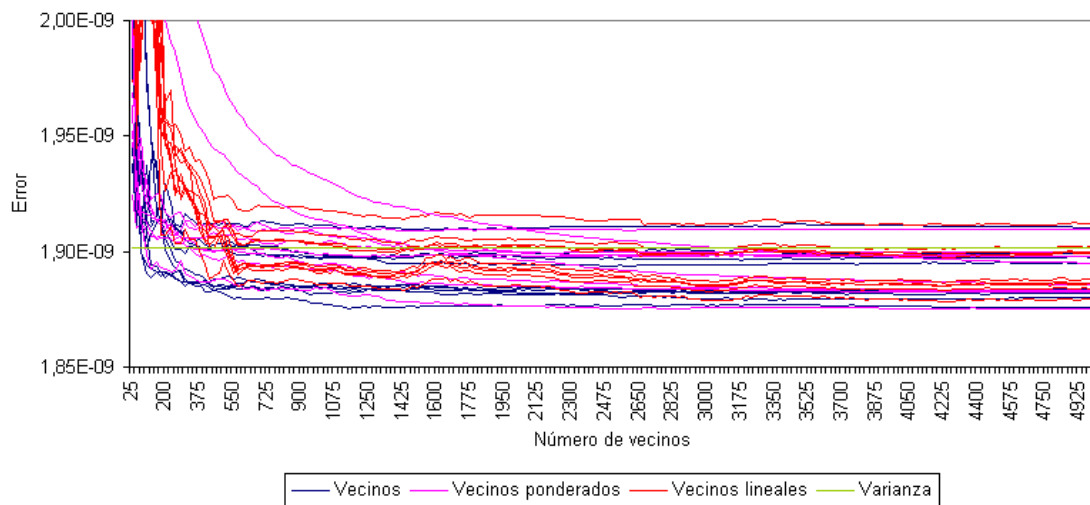


Figura 6.4.6: Errores cuadráticos medios de los vecinos próximos comparados con la varianza, para el rendimiento del Futuro sobre el Bono con vencimiento en junio 98 operación a operación

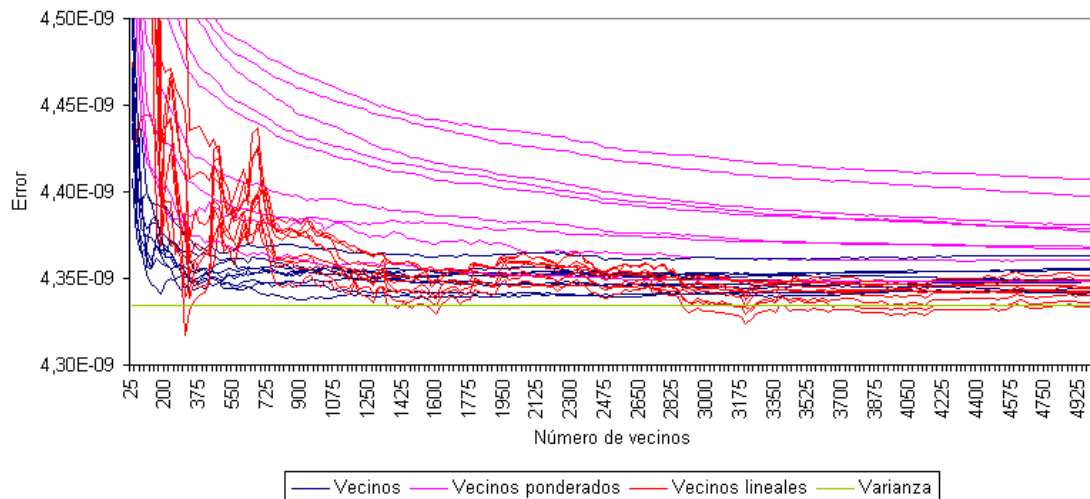


Figura 6.4.7: Errores cuadráticos medios de los vecinos próximos comparados con la varianza, para el rendimiento del Futuro sobre el Bono con vencimiento en septiembre 98 operación a operación

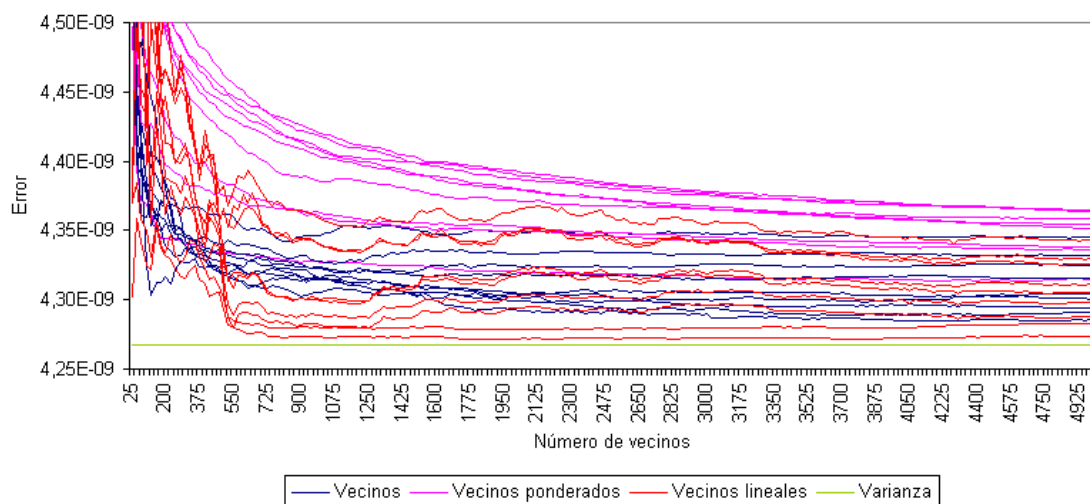


Figura 6.4.8: Errores cuadráticos medios de los vecinos próximos comparados con la varianza, para el rendimiento del Futuro sobre el Bono con vencimiento en diciembre 98 operación a operación

6.5. CONCLUSIONES

Podemos sacar las siguientes conclusiones del intento de predecir el comportamiento de los rendimientos del contrato de Futuro sobre el Bono Nacional a 10 años:

1. Los modelos lineales de tipo ARIMA tienen un nulo poder predictor del futuro de los rendimientos. Estos resultados están conformes con la teoría existente sobre mercados financieros, la hipótesis de mercado eficiente de

Fama, y en general con los resultados obtenidos en los análisis hechos hasta la fecha. La predicción lineal no funciona.

2. De los métodos no lineales globales el mejor resultado se obtiene con las redes neuronales artificiales. Los otros dos métodos empleados, los polinomios y las funciones radiales, demuestran escaso poder de mejora respecto a los modelos lineales.
3. Respecto a los métodos locales, no observamos grandes diferencias entre ninguno de los tres métodos. Sí parece ligeramente peor el método de los vecinos próximos ponderados respecto a los otros dos, aunque no está claro que esa diferencia sea significativa.
4. En total hemos ensayado para cada una de las series 36 modelos de polinomios, 4950 modelos de funciones radiales y 6930 modelos de redes neuronales, en lo que se refiere a modelos globales. Con respecto a los métodos locales, hemos estimado 2700 modelos de vecinos para las series de cinco minutos y 5400 modelos en las series operación a operación. Esto suma 14616 modelos en cada una de las series de cinco minutos y 17316 modelos en cada una de las series operación a operación, lo que hace un total de 127728 modelos no lineales estimados.
5. En las series de cinco minutos, tanto los modelos globales como los locales logran explicar una parte de la varianza de las series de rendimientos con

vencimiento en marzo y diciembre de 1998. Ese éxito no es tan claro en las series con vencimiento en junio y septiembre de 1998. En cualquier caso sí parecen mejores los resultados obtenidos con los métodos locales.

6. En las series de operaciones, las redes neuronales artificiales obtienen resultados mucho mejores a los de los demás métodos. De nuevo logramos explicar una parte importante de la varianza en los vencimientos de marzo y diciembre de 1998. Los resultados, de nuevo no resultan tan claros en las series con vencimientos en junio y septiembre.
7. Las reducciones de la incertidumbre de la predicción logradas en algunas de las series muestra que es posible encontrar dependencias no lineales en los rendimientos de los Futuros. Esto no debería ser posible si siguiéramos la hipótesis de mercado eficiente de Fama, pero sí lo sería si en el funcionamiento del mercado de Futuro sobre el Bono Nocional a 10 años existiera un comportamiento de tipo caótico.