Universidad de Los Andes
Facultad de Ingeniería
Maestría en Modelado y Simulación

Evaluación de Métodos de Predicción para un Conjunto de Agregados Económicos Venezolanos utilizando Modelos Univariantes

Autor:
Roberto Ferrer

Tutor:
Dr. Kay Tucci

Trabajo de Grado presentado para optar al Título de Magister Scientiarium en Modelado y Simulación.

30 de mayo de 2009

Resumen

En este trabajo abordamos el tema de la predicción utilizando 178 series de tiempo venezolanas, en su mayoría económicas. Se plantea la necesidad de construir y evaluar modelos de pronósticos para determinar su desempeño y pertinencia en el caso venezolano. La clase de modelos abordados es la de modelos univariantes. Entre las distintas categorías de estimación consideradas se encuentran: I) Modelos ingenuos de cambio cero (NCH); II) Suavización exponencial (GXSM); III) Modelos estimados por mínimos cuadrados ordinarios (OLS); IV) Regresión robusta (RROB); V) Árboles de regresión (REGT); VI) Redes neuronales artificiales (NN); VII) Máquinas de soporte vectorial (SVM). Concluimos que en efecto, parecen existir diferencias significativas entre los distintos métodos de estimación. Resulta conveniente dividir a los siete métodos en tres grupos de desempeños estadísticamente similares. El primer grupo conformado por RROB, SVM y OLS representan el grupo de mejor desempeño. El grupo con el segundo mejor desempeño lo conforman REGT y NN, y por último, un tercer grupo que contiene a GXSM y NCH. Esta clasificación, sin embargo, se hace borrosa en la medida en que el horizonte de pronóstico se hace más grande; para éste escenario los resultados favorecen más la opción de dos grandes grupos: 1) RROB, SVM, REGT, NN, OLS; 2) GXSM, NCH.

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

El Problema

El hombre que pretende ver todo con claridad antes de decidir, nunca decide.

HERÁCLITO

La predicción económica es una tarea compleja, pero necesaria –puesto en palabras simples, buenos pronósticos conducen a buenas decisiones; y por extensión, mejores pronósticos a mejores decisiones. Sin embargo, la labor puede resultar sumamente difícil a causa de factores como los que siguen: I) Las economías evolucionan en el tiempo y están sujetas a *shocks* no anticipados que son a su vez intermitentes y en ocasiones, de tamaños considerables. También cambios en la legislación vigente, repentinos deslizamientos en la política económica o agitación política, pueden precipitar cambios estructurales (i.e. controles de cambio, regulaciones en el sistema bancario, privatizaciones, golpes de estado, etc.). II) Los modelos utilizados para entender y predecir complejos procesos económicos están muy lejos de ser representaciones perfectas de comportamiento. III) Puede decirse que, en términos generales, los modelos están constituidos por tres componentes: términos de-

terministas que capturan promedios y crecimiento estable (y cuyos valores futuros son conocidos); variables estocásticas observadas con valores futuros desconocidos (i.e. consumo, precios, etc.); y errores no observados para los cuales no se conoce ningún valor pasado, presente, ni futuro (aunque posiblemente estimable dependiendo del contexto del modelo). Si las relaciones entre cualquiera de estos tres componentes se formula inapropiadamente, son estimadas con poca precisión, o cambian en modos no anticipados, los resultados pueden verse afectados negativamente (Clements y Hendry 2002). IV) Pronósticos acertados incentivan a la creación de "máquinas de dinero" que finalmente, modifican la predicción original (Hendry 2004). V) La calidad de los datos.

Los estudios dedicados a la evaluación de métodos y modelos diseñados para predecir variables macroeconómicas, históricamente presentan conclusiones mixtas que favorecen una u otra aproximación. Son tan variadas las posibles formulaciones de métodos y modelos, así como también sus insumos (i.e. datos), que resulta una imposibilidad práctica determinar a ciencia cierta, cuál de todas ellas ofrece el mejor desempeño.

Las constantes y aparentes contradicciones que a menudo se observan, no parecen ser más que el reflejo de un hecho trivial: los resultados de los distintos modelos, y por ende, de las competencias de pronósticos, son caso-dependientes o contingentes al problema particular que se está abordando. Qué formulación predomina en la competencia depende entre otras cosas, del tipo de modelo y su especificación (entiéndase, la formulación per se), de los criterios de evaluación que se han aplicado, de las muestras seleccionadas y de sus horizontes.

Existe una amplia literatura que evalúa el desempeño predictivo de modelos en economía. Hendry y Clements (2003) nos hacen saber que Theil (1966), Mincer y Zarnowitz (1969), Dhrymes et al. (1972) y Cooper y Nelson (1975), comparan predicciones de modelos econométricos con aquellas

de modelos de series de tiempo "ingenuos" (naive, en inglés), como los "sin cambio" (no-change models).[1] Su principal conclusión es que estos últimos superan a los primeros. Wallis (1989) y McNees (1990) con datos del Reino Unido y EE.UU., respectivamente, concluyen lo contrario. Adicionalmente, Burns et al. (1986) ha detectado una leve mejoría en la precisión predictiva, a pesar de las mejoras sustanciales presentes en los modelos.

Esta inevitable (y posiblemente irresoluble) variabilidad de resultados es una de las razones que nos ha llevado a plantear un caso de estudio basado en variables económicas de Venezuela. Nuestra intención es abordar el tema de las predicciones utilizando exclusivamente datos económicos venezolanos, y se plantea la necesidad de construir y evaluar modelos de pronósticos para determinar su desempeño y pertinencia en el caso venezolano.

La clase de modelos abordados es la de modelos univariantes; como indica su nombre, supone modelos en los cuales la variable a predecir es calculada utilizando sólo una variable (ella misma) y una constante opcional.

Esta clase de modelos resulta apropiada para el caso en que se trabaja con un gran número de series de tiempo, y resulta inviable ajustar modelos multivariantes *ad hoc* que consideran explícitamente las teorías respectivas, en la especificación de los modelos. Resulta difícil hacer afirmaciones generales sobre el desempeño relativo de los métodos multivariantes. En general, se esperaría que los pronósticos multivariantes sean al menos tan buenos como los univariantes, pero esto no es cierto ni en teoría, ni en práctica. El cómputo de pronósticos multivariantes puede requrir el cómputo previo de pronósticos para las variables explicativas, y si los últimos no son lo suficientemente buenos, echan a perder los primeros. Así mismo, se tiene que los modelos multivariantes son menos robustos a desviaciones de los supuestos del mo-

[1] Considerando el contexto del documento fuente, se entiende por *modelos econométricos* aquellos cuyas especificaciones vienen fundamentadas en teoría económica. Suele también llamárseles modelos estructurales.

delo. Los modelos multivariantes usualmente consiguen un mejor ajuste a los datos de estimación, pero esta superioridad no se traduce necesariamente en mejores pronósticos. Una posible causa sea su sensibilidad a cambios estructurales. (Chatfield 2004)

Entre las distintas categorías de estimación consideradas se encuentran: I) Modelos ingenuos de cambio cero (NCH); II) Suavización exponencial simple y doble (XS y DXS); III) Modelos estimados por mínimos cuadrados ordinarios (OLS); IV) Regresión robusta (RROB); V) Árboles de regresión (REGT); VI) Redes neuronales artificiales (NN); y VII) Máquinas de soporte vectorial (SVM).

Los datos están conformados por series temporales de índole económica, en su mayoría provistas por el Banco Central de Venezuela (BCV).[2] Las series se pueden enmarcar en las siguientes categorías generales: (a) índices de producción, (b) producción física, (c) ventas, (d) índices de precios, (e) tasas de interés, (f) agregados monetarios, (g) ingresos y gastos públicos, (h) variables demográficas, (i) otras.

1.1. Justificación

Para algunos, la predicción en economía no es más que una gran pérdida de tiempo (Marget 1929). Sin embargo, esta actividad resulta de gran utilidad en los procesos de toma de decisiones. Los agentes desean tener en su poder la mayor cantidad de información auxiliar disponible, en los momentos previos a la puesta en marcha de una acción. Mirar hacia adelante, por más difícil que eso sea, debe ayudar a alcanzar mejores resultados. Este es particularmente el caso cuando la decisión es una que, como en política monetaria, toma largo tiempo en revelar sus consecuencias (Stevens 2004).

[2]Algunas disponibles al público por la página web del Banco Central de Venezuela (`http://www.bcv.org.ve`).

Los gobiernos hacen predicciones para cuadrar sus presupuestos. Quieren saber cuánto será el ingreso disponible (aproximado) para cumplir apropiadamente con el proceso de distribución de los recursos públicos, determinar políticas de préstamos nacionales, entre otras cosas (Stevens 1999). He allí, un ejemplo de la utilidad de las predicciones para la política fiscal de cualquier nación. De manera similar, en el campo de la política monetaria los tomadores de decisiones están en la obligación de considerar cambios preventivos de políticas (Stevens 2004).[3]

En el caso venezolano, no nos queda del todo claro qué tan avanzados estamos en relación al tema de la predicción de agregados económicos. Ciertamente, no existe mayor cantidad de documentación disponible al público, que indique dónde estamos ubicados.[4] Una búsqueda inicial[5] revela que el BCV tiene a disposición del público un documento que compara modelos de predicción para la inflación y otro que informa sobre algunos aspectos de la metodología de la programación financiera en Venezuela.[6] Además, la literatura que abarca la utilización de métodos relativamente recientes como los REGT, las NN y los modelos SVM, es escasa y se encuentra poco desarrollada.[7]

[3]Stevens aclara un punto muy importante en relación a las predicciones en Economía. A diferencia de otros pronosticadores (i.e. pronósticos del clima) hay una dimensión adicional en su dificultad, dado que los pronósticos en economía se hacen muchas veces con la intención de evocar respuestas de hacedores de política, que a su vez pueden afectar lo que originalmente ha sido pronosticado.

[4]Sospechamos que sí existen avances importantes en el área, pero constituye información de uso interno de las instituciones respectivas que no se hace disponible al público en general.

[5]En página web del BCV: `http://www.bcv.org.ve`.

[6]La programación financiera implica desarrollar y evaluar ejercicios de predicción. Los dos documentos son los siguientes: Fleitas et al. (2002) y Guerra et al. (1997).

[7]Estamos al tanto de un proyecto de Riesgo Bancario, auspiciado por el Instituto de Investigaciones Económicas y Sociales (IIES) de la Universidad de los Andes (ULA), en colaboración con BCV, en el cual se consideran NN. Vea: `http://webdelprofesor.ula.`

Resumiendo, existen razones importantes que justifican un estudio como el que acá se propone. En primer lugar, las predicciones económicas son elementos que auxilian el proceso de tomas de decisiones en el mismo ámbito, y su progresiva mejora constituye un avance para la comunidad usuaria. En segundo lugar, no parecen existir suficientes estudios (al menos públicos) en Venezuela que evalúen el desempeño de distintos métodos en el pronóstico de agregados económicos; y tercero, nos parece apropiada y novedosa la incorporación de métodos de predicción poco desarrollados en nuestro país.[8] como son los REGT, las NN y las SVM.

1.2. Objetivos

1.2.1. Objetivo General

Construir distintos modelos de predicción para un conjunto de agregados económicos venezolanos y evaluar sus desempeños.

1.2.2. Objetivos Específicos

- Seleccionar los criterios que servirán para evaluar el desempeño de los modelos estimados.
- Especificar y estimar los modelos propuestos.
- Evaluar el desempeño de los distintos modelos partiendo de los criterios establecidos para tal fin.

ve/economia/gcolmen/investigacion.html.

[8]En particular en el área de la predicción económica.

Capítulo 2

Marco Teórico

El mejor profeta del futuro, es el pasado.

LORD BYRON

En general, los modelos a ser utilizados consisten en especificaciones autorregresivas. Popularizados por Box y Jenkins (1970),[1] son catalogados como modelos a-teóricos dado que no derivan de teoría económica alguna. En principio, estos modelos pretenden extraer toda la información necesaria para predecir, de la misma serie a estudiar.[2]

Se han planteado distintos métodos para llevar a cabo las estimaciones de nuestros modelos, pero por su relativa novedad y por su mayor complejidad, estaremos haciendo énfasis en los modelos no lineales.

[1] En términos más precisos, Box y Jenkins trataron una clase más general de modelos denominados ARIMA (autorregresivos integrados de promedios móviles), de los cuales los autoregresivos (AR) consituyen un caso particular.

[2] Esto en contraposición a los modelos teóricos que generalmente son planteados en marcos multiecuacionales (aunque existen también especificaciones autorregresivas multivariantes - Vectores Auto-Regresivos, por ejemplo).

2.1. Suavización Exponencial Simple

Suavización exponencial es el nombre dado a una clase general de procedimientos de pronósticos que se basan en ecuaciones actualizadas para generar sus estimaciones.

La forma más básica, suavización exponencial simple (XS), es utilizada con series de tiempo no estacionales y sin tendencia.

Dada una serie temporal con las características señaladas $x_1, x_2, \ldots, x_n$ resulta natural calcular $x_n(S)$ (valores Suavizados) por medio de suma ponderada de los valores pasados de la serie:

$$x_n(S) = c_0 x_n + c_1 x_{(n-1)} + c_2 x_{(n-2)} + \ldots \tag{2.1}$$

donde c_i son los pesos. Parece sensato dar más importancia a los valores recientes y menos a los observados lejos en el tiempo. El juego de pesos geométricos hace precisamente ésto, decreciendo en una razón constante por cada incremento unitario del rezago.

Para que los pesos sumen uno, hacemos:

$$c_i = \alpha(1-\alpha)^i, \qquad \text{con } i = 0, 1, \ldots \tag{2.2}$$

donde α es una constante tal que $0 < \alpha < 1$. Así, (2.1) se vuelve,

$$x_n(S) = \alpha x_n + \alpha(1-\alpha)x_{(n-1)} + \alpha(1-\alpha)^2 x_{(n-2)} + \ldots \tag{2.3}$$

En términos estrictos (2.3) implica un número infinito de observaciones pasadas, pero en práctica sólo hay un número finito. Escribiendo (2.3) recursivamente obtenemos:

$$\begin{aligned} x_n(S) &= \alpha x_n + (1-\alpha)[\alpha x_{(n-1)} + \alpha(1-\alpha)x_{(n-2)} + \ldots] \\ &= \alpha x_n + (1-\alpha)x_{(n-1)}(S) \end{aligned} \tag{2.4}$$

El procedimiento definido por (2.4) es la suavización exponencial simple; y se le llama exponencial porque los pesos geométricos descansan, precisamente, sobre una curva exponencial.

El parámetro α se define como una constante de suavización y debe ser fijado para llevar a cabo el procedimiento de estimación. Valores entre 0.1 y 0.3 son comúnmente utilizados y producen estimaciones que dependen de un número grande de observaciones pasadas. Valores cercanos a 1 son utilizados con menos frecuencia y las estimaciones en este caso dependen mucho más de los valores recientes (Chatfield 2004).

Para pronosticar con el modelo XS, fijamos,

$$\hat{x}_{(n+1)} = x_n(S) \tag{2.5}$$

2.2. Cambio-Cero

Si examinamos el caso en que $\alpha = 1$, tenemos que (2.4) se hace,

$$x_n(S) = x_n \tag{2.6}$$

lo que describe un modelo de cambio-cero (NCH). Por lo tanto, el modelo de cambio-cero no es más que un caso particular del modelo de suavización exponencial simple. Este modelo sugiere exactamente lo que indica su nombre: el pronóstico del periodo siguiente $t + 1$, es el valor observado en el periodo inmediatamente anterior t; por (2.5) y (2.6):

$$\hat{x}_{(n+1)} = x_n(S) = x_n \tag{2.7}$$

2.3. Suavización Exponencial Doble

Para el caso de series de tiempo con tendencia, resulta útil utilizar un modelo de suavización de orden mayor como el modelo de suavización doble

(DXS). Esta variante resulta de aplicar el modelo XS dos veces. Reescribiendo (2.4) tenemos:

$$x_n(S') = \alpha x_n + (1-\alpha)x_{(n-1)}(S') \tag{2.8}$$

Aplicando suavización simple (con el mismo valor de α) a la serie ya suavizada resultante de (2.8), logramos la suavización doble:

$$x_n(S'') = \alpha x_n(S') + (1-\alpha)x_{(n-1)}(S'') \tag{2.9}$$

Por último, para pronosticar con DXS, fijamos,

$$\hat{x}_{(n+1)} = a_n + b_n \tag{2.10}$$

donde,

$$a_n = 2x_n(S') - x_n(S'') \tag{2.11}$$

$$b_n = (\alpha/(1-\alpha))x_n(S') - x_n(S'') \tag{2.12}$$

y a_n corresponde al nivel en el periodo n; y b_n a la tendencia en el periodo n.

2.4. Mínimos Cuadrados Ordinarios

El modelo de regresión múltiple estimado por mínimos cuadrados ordinarios (OLS) puede expresarse como sigue (Pindyck y Rubinfeld 1998):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.13}$$

donde $\mathbf{Y}$ es un vector $n \times 1$ de observaciones de la variable dependiente, $\mathbf{X}$ es una matriz $n \times p$ de variables regresoras, $\boldsymbol{\beta}$ es un vector $p \times 1$ de parámetros desconocidos y $\boldsymbol{\epsilon}$ es un vector $n \times 1$ de errores.

El objetivo es encontrar un vector de parámetros $\hat{\boldsymbol{\beta}}$ tal que minimicemos,

$$SSE = \sum_{i=1}^{n} \hat{\epsilon}_i^2 = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} \tag{2.14}$$

donde,

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \hat{\mathbf{Y}}$$
$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

Resolviendo para (2.14):

$$\frac{\partial SSE}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) \tag{2.15}$$

Garantizamos que la llamada *matriz de productos cruzados* $(\mathbf{X}'\mathbf{X})$ tiene inversa asegurando que $\mathbf{X}$ tiene rango completo.

2.5. Regresión Robusta

Los procedimientos estadísticos "robustos" proveen de métodos formales para lidiar con datos atípicos (Hogg 1979). Uno de estos métodos, denominados *regresión robusta* (RROB), emplea un criterio de ajuste que no resulta tan vulnerable a datos inusuales como el método de mínimos cuadrados.

El método general más común de regresión robusta es la *estimación-M* (Fox 2002). Introducido por Huber (1964), puede plantearse en los siguientes términos:
Considere el modelo lineal:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.16}$$

donde $\mathbf{Y}$ es un vector $n \times 1$ de observaciones, $\mathbf{X}$ es una matriz $n \times p$ de variables regresoras, $\boldsymbol{\beta}$ es un vector $p \times 1$ de parámetros desconocidos y $\boldsymbol{\epsilon}$ es un vector $n \times 1$ de errores no necesariamente normales, pero simétricos.

El *estimador-M* minimiza la función objetivo,

$$\sum_{i=1}^{n} \rho(\epsilon_i) = \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i'\boldsymbol{\beta}) \tag{2.17}$$

donde y_i es el elemento i-ésimo de $\mathbf{Y}$ y $\mathbf{x}_i$ es la fila i-ésima de $\mathbf{X}$. La función ρ da la contribución de cada residual a la función objetivo. Por ejemplo, para la estimación de mínimos cuadrados ordinarios $\rho(\epsilon_i) = \epsilon_i^2$.

Dejemos que $\psi = \rho'$ sea la derivada de ρ. Diferenciando la función objetivo con respecto a los coeficientes $\boldsymbol{\beta}$ e igualando las derivadas parciales a cero obtenemos el equivalente de la solución maximizadora asociada con un sistema de p ecuaciones de coeficientes:

$$\sum_{i=1}^{n} \psi(y_i - \mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i' = 0 \tag{2.18}$$

Definamos la función de peso $w(\epsilon) = \psi(\epsilon)/\epsilon$ y hagamos $w_i = w(\epsilon_i)$. Entonces podemos escribir las ecuaciones como,

$$\sum_{i=1}^{n} \sum w_i(y_i - \mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i' = 0 \tag{2.19}$$

Resolver las ecuaciones anteriores es un problema de mínimos cuadrados ponderados, minimizando $\sum w_i^2\epsilon_i^2$. Sin embargo, los pesos dependen de los residuos, los residuos dependen de los coeficientes estimados y los coeficientes estimados dependen de los pesos. Se requiere entonces una solución iterativa: mínimos cuadrados ponderados iterativo[3] y puede plantearse como sigue:

1. Obtenga estimados iniciales de $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(0)}$, i.e. estimadores mínimos cuadrados ordinarios.

2. En cada iteración t calcule los residuos $\epsilon_i^{(t-1)}$ y los pesos asociados $w_i^{(t-1)} = w\left[\epsilon_i^{(t-1)}\right]$ a partir de la iteración previa.

3. Resuelva para los nuevos estimados mínimos cuadrados ponderados,

$$\boldsymbol{\beta}^{(t)} = \left[\mathbf{X}'\mathbf{W}^{(t-1)}\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{W}^{(t-1)}\mathbf{Y} \tag{2.20}$$

donde $\mathbf{W}^{(t-1)} = \text{diag}\left\{w_i^{(t-1)}\right\}$ es la matriz de pesos corriente.

[3]En inglés, *Iteratively Reweighted Least Squares.*

Los pasos (2) y (3) se repiten hasta que el algoritmo converge (i.e. los residuos no cambian en dos pasos consecutivos) o cuando se cumpla con algún criterio de convergencia alterno (Heiberger y Becker 1992).

2.6. Árboles de Regresión

Los árboles de regresión (REGT) forman parte de la metodología CART[4] desarrollada para resolver problemas de clasificación (Breiman et al. 1984). En esta metodología, variables dependientes continuas dan lugar a *árboles de regresión* y variables dependientes categóricas producen *árboles de clasificación*. En ambos casos, el objetivo consiste en obtener un conjunto de clasificadores de datos, poniendo al descubierto la estructura predictiva del problema que se está considerando (Yohannes y Webb 1999).

Un árbol de regresión representa un proceso de decisión multi-etapa donde se toma una decisión binaria en cada una de ellas. El árbol está formado por *nodos* y *ramas*, y los nodos pueden ser internos o terminales. Los *internos* son aquellos que se dividen en dos hijos, mientras que los *terminales* no se dividen. El proceso puede entenderse como una forma de *división binaria recursiva* (Lewis 2000). El término "recursivo" significa que el proceso de división puede aplicarse una y otra vez: un nodo padre da lugar a dos nodos hijos, y a su vez, estos dos hijos pueden dividirse en otros dos cada uno y así sucesivamente hasta formar un árbol completo. La Figura 2.1 muestra un árbol de regresión genérico.

Se requieren tres componentes para construir un árbol de regresión (Yohannes y Webb 1999): 1) un conjunto de preguntas de la forma *es* $X \leq d$, donde X es una variable y d es una constante; las respuestas serán de la forma *sí* o *no*. Las divisiones de los nodos estarán basadas en estas preguntas y a menos que se especifique de otra forma, cada división dependerá de una sóla

[4] *Classification And Regression Trees.*

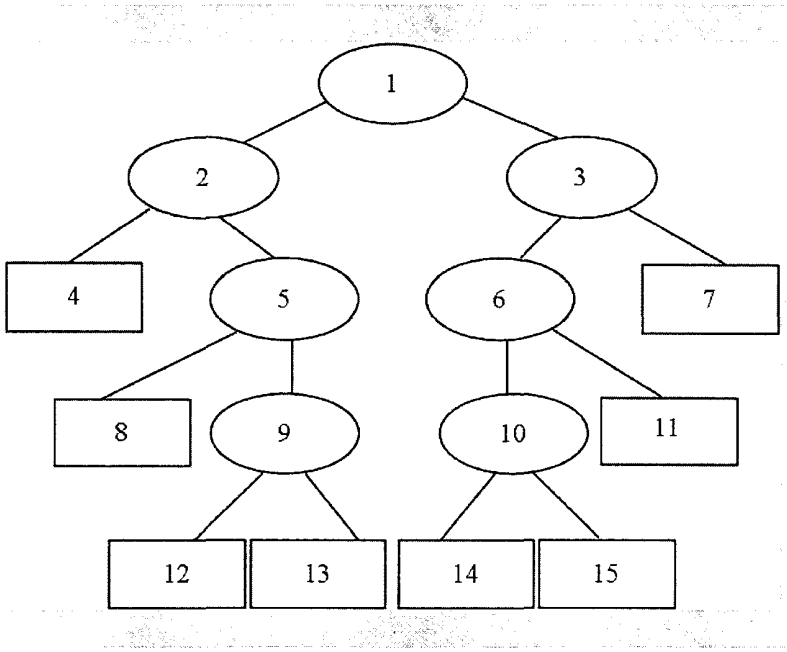


Figura 2.1: Ejemplo de un Árbol de Regresión Genérico. Los nodos rectangulares corresponden a nodos terminales. Fuente: Yohannes y Webb (1999)

de todas las variables explicativas entregadas; II) un criterio de la bondad del ajuste para elegir la mejor división en una variable; y III) la generación de estadísticas descriptivas para los nodos terminales.

Con los REGT se construyen predictores detectando la heterogeneidad que existe en los datos y luego purificándolos. La división recursiva separa a los datos en nodos terminales que internamente son relativamente homógeneos, pero entre ellos, heterogéneos. Las predicciones se consiguen *deslizando* variables regresoras por el árbol, contestando preguntas y finalmente, tomando el valor medio de la variable dependiente en el respectivo nodo terminal.

En términos formales, un árbol de regresión se construye de la siguiente manera:

1. Empezando con el nodo raíz (es decir los datos completos), se realizan

todas las posibles divisiones en cada una de las variables explicativas, se aplica una *medida de impureza* a cada división y se determina la reducción en impureza que se puede alcanzar con cada una de ellas.

Una función de impureza o regla de división que puede utilizarse es la de mínimos cuadrados (MC). Bajo este criterio, la impureza de un nodo se mide por la suma de cuadrados intra-nodo, definida por:

$$R(t) = \sum_{i=1}^{n_{(t)}} (y_{i(t)} - \bar{y}_{(t)})^2 \tag{2.21}$$

donde $y_{i(t)}$ son los valores individuales de la variable dependiente en el nodo t, $\bar{y}_{(t)}$ es la media de los valores de la variable dependiente contenidos en el nodo t y n_t es el número de observaciones en ese mismo nodo. Dada la función de impureza $R(t)$ y la división s que envía casos al nodo de la izquierda (t_L) y el resto a la derecha (t_R), la bondad de una división se mide por la función:

$$\Delta(s,t) = R(t) - R(t_R) - R(t_L) \tag{2.22}$$

donde $R(t_R)$ es la suma de cuadrados del nodo hijo derecho y $R(t_L)$ es la suma de cuadrados del nodo hijo izquierdo.

El error total del árbol resulta de la suma de los errores cuadráticos de todos los nodos terminales:

$$R(T) = \sum_{t \in \breve{T}} R(t) \tag{2.23}$$

donde $\breve{T}$ es el conjunto de todos los nodos terminales en el árbol T.

2. En un segundo paso se selecciona la "mejor" división y se dividen los datos entre dos nodos hijos. La mejor división es aquella para la cual $\Delta(s,t)$ es mayor. La regla implica elegir aquella división que resulte en la máxima reducción de impureza del nodo padre.

3. Como REGT es recursivo, los pasos (1) y (2) se repiten para cada nodo interno hasta formar el árbol más grande posible. Se suele usar una

regla de parada que detiene el crecimiento del árbol una vez se ha alcanzado un número determinado de casos en los nodos terminales.

4. Finalmente, se aplica un algoritmo de *poda* para contrarrestar el eventual sobreajuste que se ha producido. El paso de la poda pretende simplificar el árbol de decisión, sin afectar su capacidad para predecir; en efecto, un árbol completo tendrá una precisión máxima si se evalúa con los mismos datos con los que ha sido contruído, pero su capacidad para predecir fuera de muestra se verá negativamete afectada por el sobreajuste (Yohannes y Webb 1999).

2.7. Redes Neuronales Artificiales

Una red neuronal artificial (NN) es un modelo biológicamente inspirado que consiste en un conjunto de elementos procesadores llamados neuronas y las conexiones entre ellas. Aun cuando las redes neuronales presentan ciertas similitudes con el cerebro humano, no están destinadas a modelarlo (Kasabov 1996): las NN han sido diseñadas para imitar ciertos aspectos del procesamiento de información que realizan redes neuronales biológicas.

Probablemente resulte mucho más fácil entender en qué consiste una NN si discutimos la contraparte artificial de una neurona biológica. La neurona artificial constituye la pieza fundamental en la estructuración del modelo. El primer modelo matemático de una neurona se lo debemos a McCulloch y Pitts (1943) y tomaba como entradas elementos binarios, arrojando como resultados elementos también binarios. El modelo matemático de la neurona puede ser descrito por los siguientes parámetros (Kasabov 1996) (véase Figura 2.2):

- *Conexiones de entrada:* $x_1, x_2, \ldots, x_p$. Las conexiones de entradas no son más que las variables (incluyendo el sesgo) que alimentan el modelo.

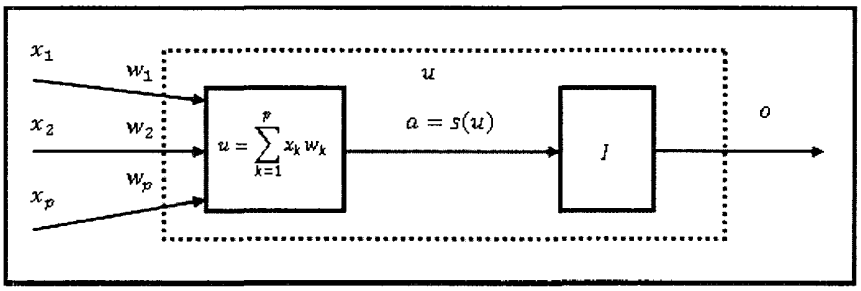


Figura 2.2: Ejemplo de una Neurona Artificial. Fuente: adaptado de Kasabov (1996)

Cada conexión de entrada tiene atada a sí misma, un coeficiente o peso: $w_1, w_2, \ldots, w_p$.

- *Función de entrada g:* calcula la señal de entrada neta agregada $u = g(\mathbf{x}_j, \mathbf{w}_j)$ con $j = \{1, 2, \ldots, n\}$, donde $\mathbf{x}_j$ corresponde al vector de entradas del caso de entrenamiento $n-$ésimo y $\mathbf{w}_j$ al vector de pesos del mismo caso. Usualmente g es la función de suma $u = \sum_{k=1}^{p} x_k w_k$.
- *Función de activación s:* calcula el nivel de activación de la neurona $a = s(u)$. Generalmente es una función sigmoidea.
- *Función de salida:* calcula el valor de la señal de salida emitida por la neurona, $y = I(a)$. La señal de salida se asume, por lo general, igual al nivel de activación de la neurona, es decir, I es la función identidad y por lo tanto, $o = a$.

Puede decirse que la tarea de la neurona es recibir entradas de alguna fuente externa y procesarlas. El *procesamiento* equivale al conjunto de transformaciones que sufren las entradas al pasar por la neurona. Cada entrada tiene un peso asociado (w_i) que puede ser modificado para simular el aprendizaje sináptico. La salida de esta unidad es una función s de la suma ponderada de los valores de entrada, y a su vez, puede constituir la entrada de otra neurona.

De acuerdo a los valores que tomen los distintos parámetros mencionados anteriormente, se pueden obtener distintos tipos de neuronas (las llamaremos también unidades o nodos). Por ejemplo, puede plantearse el caso en que las entradas que alimentan al nodo son binarias, {0,1}: la función de salida es la identidad I; y la función de activación s está definida por:

$$s(u) = \begin{cases} 1 & \text{si } u \geq \theta \\ 0 & \text{si } u < \theta \end{cases} \tag{2.24}$$

donde $\theta = 0$ es un límite definido por el modelador. En este caso, las salidas son también binarias y si quisiéramos elaborar la analogía biológica, podríamos decir que la neurona se *activa* cuando $s(u) = 1$ y permanece *inactiva* cuando $s(u) = 0$. Esta función s es comúnmente llamada *función límite o de escalón.*[5]

Además de la recién mostrada, puede definirse toda una variedad de funciones de activación. Las entradas y salidas de las neuronas no tienen por qué estar limitadas a valores booleanos como se especifica en el ejemplo anterior, que recoge lo planteado originalmente por McCulloch y Pitts (1943). Incorporar entradas de un subconjunto de $\mathbb{R}$ es una operación directa y establecer funciones de activación continuas que permitan calcular valores de salida, también continuos, es una posibilidad. Algunas funciones de activación comúnmente utilizadas son:

- *Función límite lineal.* El valor de activación (a) aumenta linealmente con incrementos en la señal de entrada neta (u); pero más allá de un límite, el nodo se satura y arroja como salidas un valor único c. Dependiendo del rango (o rangos) estipulado(s) para la saturación, pueden plantearse diferentes variantes. Un ejemplo es:

$$f(u) := \begin{cases} c & \text{si } u \geq 2 \\ a + bu & \text{si } u < 2 \end{cases} \tag{2.25}$$

[5] *Hard-limited threshold* o *step*, en inglés.

donde a y b son parámetros que definen la relación lineal.

- *Función sigmoidea.* Corresponde a cualquier función $g(u)$ de transformación no lineal con forma de 'S', que es: acotada, es decir, los valores de salida están limitados a un rango particular; monotónicamente creciente, significando que el valor de $g(u)$ nunca decrece cuando u crece; continua y suave, y por lo tanto, diferenciable en todo su dominio.

  Algunos ejemplos son:

  - La función logística:

$$h(u) = 1/(1 + e^{-cu}), \quad \text{donde } c \text{ es una constante.} \tag{2.26}$$

  - La función logística bipolar:

$$bh(u) = (1 - e^{-u})/(1 + e^{-u}) \tag{2.27}$$

  - La tangente hiperbólica:

$$tanh(u) = (e^{u} - e^{-u})/(e^{u} + e^{-u}) \tag{2.28}$$

El poder que ofrece una neurona artificial para resolver problemas es realmente limitado; pero por unión sistemática de estas estructuras básicas. se pueden armar redes que son capaces de procesar información de distintos grados de complejidad. El modelo de NN puede definirse por los siguientes cuatro parámetros (Kasabov 1996):

1. *El tipo de neuronas.* Este parámetro está directamente vinculado con lo descrito en los párrafos anteriores (los distintos tipos de funciones).

2. *La arquitectura de la red.* El tipo de conexiones existentes entre las distintas neuronas de una red determina su topología. Las neuronas o nodos pueden estar completamente conectadas unas con otras, pero en

general se encuentran parcialmente conectadas y distribuidas en capas. En la arquitectura de capas suelen ser tres: una capa de entradas, que contiene una neurona para cada variable explicativa; una capa de salida, que contiene una neurona para cada variable explicada; y una capa oculta compuesta por neuronas, definidas simplemente como aquellas que se encuentran entre la de entrada y la de salida. Esta última capa puede estar dividida en sub-capas. La Figura 2.3 muestra la forma de una NN como la descrita. Cada círculo representa una neurona. Las neuronas en la capa de entradas no necesariamente transforman los datos (así lo asumiremos a menos que se indique lo contrario) y cada una de ellas recibe a una variable. Así, para la red de la Figura 2.3, introducimos cuatro datos (una instancia de cuatro variables distintas, por ejemplo), a partir de los cuales se calculan sumas ponderadas (u), que luego pasan a ser procesadas por las neuronas de las capas ocultas. Las neuronas de las capas ocultas suelen presentar funciones de activación distintas a la *identidad*, que luego de aplicadas a sus respectivas entradas, pasan a un último proceso de transformación para finalmente salir por la capa de salida. El número total de neuronas en la capa de salida puede ser mayor que uno.

Puede observarse también que cada neurona de cada capa se conecta con cada neurona de la capa siguiente (excepto la capa de salida), pero las neuronas de una capa no están conectadas directamente entre ellas mismas. En el sentido estricto, esta red está parcialmente conectada (justamente por lo anterior), pero suele encontrarse en la literatura referencias a redes "completamente conectadas" que presentan topologías de conexión como la que se muestra en la Figura 2.3. Es decir se encuentran completamente conectadas pero capa a capa.

Existen dos grandes arquitecturas de conexiones de acuerdo al número de capas de entrada y salida. Cuando hay una sóla capa (es decir,

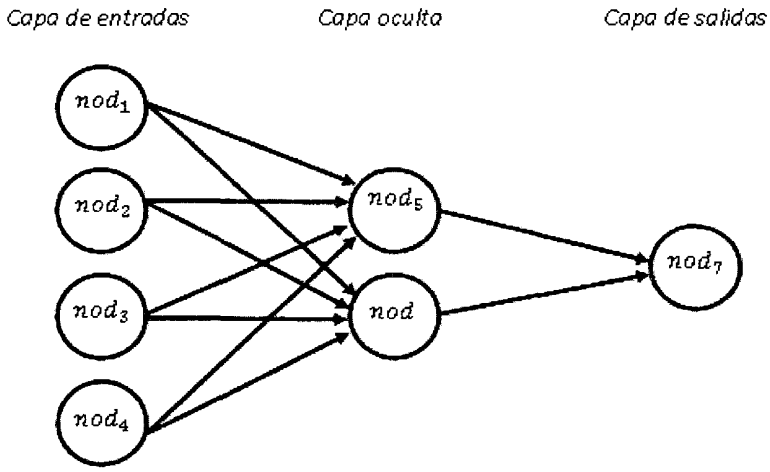


Figura 2.3: Ejemplo de una NN (sin los pesos explícitos). Cada círculo representa una neurona o nodo.

los nodos de entrada son los mismos nodos de salida), se dice que la arquitectura es *autoasociativa*; cuando existen capas separadas para las unidades de entrada y de salida, se dice que la arquitectura es *heteroasociativa*.

Pueden también las NN distinguir en su arquitectura la dirección en que fluye la información. Por un lado, tenemos las NN con *alimentación hacia delante (feedforward)*. En este caso no hay conexiones que vayan desde las neuronas de salida hasta las neuronas de entrada. La red no tiene memoria en el sentido de que no guarda los valores de salida anteriores. ni tampoco los valores de activación de las neuronas. Por otro lado, tenemos las NN con *alimentación hacia atrás*. En este caso existen conexiones que van desde las neuronas de salida hasta las neuronas de entrada. En esta modalidad el próximo estado de la red depende, no solo de las entradas, sino también de los estados previos que se hayan alcanzado.

3. *Un algoritmo de aprendizaje.* Una de las características más interesantes de las NN es su habilidad para "aprender". Una NN se entrena para que la aplicación de un conjunto de vectores de entrada $\mathbf{X}$ produzca un conjunto deseado de vectores de salida $\mathbf{Y}$; o para que la red aprenda algunas características internas y/o estructurales de los datos $\mathbf{X}$. Al conjunto $\mathbf{X}$, utilizado para entrenar, lo llamamos conjunto de entrenamiento. A los elementos $\mathbf{x}_j$ de $\mathbf{X}$, los llamamos ejemplos o casos de entrenamiento. El proceso de entrenamiento se ve reflejado en el cambio secuencial de los pesos asociados a cada conexión de la NN. Durante el entrenamiento, los pesos deben converger gradualmente a valores tales que, cada vector $\mathbf{x}_j$ del conjunto de entrenamiento provoque en la red la producción de un vector deseado $\mathbf{y}_j$.

   El aprendizaje de una NN se alcanza por medio de un *algoritmo de aprendizaje (o entrenamiento).* Estos algoritmos pueden clasificarse en tres grupos:

   Supervisados. Los ejemplos de entrenamiento comprenden vectores de entrada $\mathbf{x}_j$, junto con vectores de salida deseados $\mathbf{y}_j$. El entrenamiento se ejecuta hasta que la red *aprende* a asociar cada vector de entrada, con su correspondiente vector de salida deseado. Por ejemplo, una NN puede aprender a aproximar una función $\mathbf{Y} = f(\mathbf{X})$, representada por un conjunto de ejemplos de entrenamiento $(\mathbf{x}_j, \mathbf{y}_j)$.

   No supervisados. En esta modalidad solo vectores de entrada $\mathbf{x}_j$ son suplidos a la red. La NN aprende algunas características internas y/o estructurales del conjunto completo de datos. Por ejemplo, la red podría ser utilizada para llevar a cabo un análisis de conglomerados y determinar taxonomías aún no descubiertas.

   Aprendizaje reforzado. Esta clasificación abarca una combinación de las dos modalidades anteriores. Estos métodos se caracterizan por pre-

sentar a la NN un vector de entrada $\mathbf{x}_j$ y luego observar el vector de salida $\mathbf{y}_j$. Si el vector de salida es considerado "bueno", entonces los pesos de la red aumentan; en caso contrario, los pesos disminuyen.

4. *Algoritmo de rememorización.*[6] La última característica general de las NN es su capacidad para generalizar. El principio que funciona detrás de esto es que estímulos similares causan reacciones similares. Las NN buscan producir una salida $\mathbf{y'}_j$ que es similar a la salida $\mathbf{y}_j$ original de los ejemplos de entrenamiento, si $\mathbf{x'}_j$ es similar al vector de entrada $\mathbf{x}_j$.

Matemáticamente, una red con p variables de entrada, h unidades ocultas y una unidad (o variable) de salida, aproxima a un conjunto de datos de acuerdo a:

$$f = \sum_{i=1}^{h} o_i w_i + w_0 \tag{2.29}$$

donde $o_i = s(\sum_{k=1}^{p} x_k w_{ki} + w_{0i})$ es el valor de salida del $i-$ésimo nodo intermedio; w_i y w_{ki} son los pesos de las conexiones desde la segunda y la primera capa, respectivamente; w_0 son los pesos de las conexiones que parten desde la neurona con entradas constantes e iguales a 1 (el sesgo). s es una función sigmoidea.

De ahora en adelante estaremos haciendo énfasis en NN heteroasociativas, de alimentación hacia delante y algoritmos de aprendizaje supervisados; pues son el tipo de redes que estaremos estimando en nuestra aplicación.

Como algoritmo de entrenamiento usaremos el de la *propagación hacia atrás del error*,[7] propuesto en la década de los 80 por Rumelhart et al. (1986). La idea central detrás de este procedimiento es que los errores de las unidades ocultas son determinados propagando "hacia atrás" los errores de las unidades de la capa de salidas. El algoritmo de propagación hacia atrás puede

[6] En inglés, *recall algorithm.*

[7] *Error back-propagation algorithm*, en inglés.

ser aplicado a NN con cualquier número de capas, pero se ha demostrado que hace falta tan solo una capa oculta para aproximar cualquier función continua a cualquier nivel deseado, siempre que el número de neuronas en la capa oculta sea suficiente y contengan funciones de activación no lineales (i.e. sigmoideas) (Hornik et al. 1989, Funahashi 1989, Cybenko 1989). Por esta razón las NN son catalogadas de *aproximadores universales.*

El algoritmo de propagación hacia atrás del error usa iterativamente una regla de gradiente descendente para conseguir los pesos de conexión óptimos w_{ij} tales que minimicen el error global E (Kasabov 1996). En el ciclo $(t+1)$ el cambio de peso δw_{ij} viene dado por:

$$\delta w_{ij}(t+1) = -\eta(\partial E/\partial w_{ij}(t)) \tag{2.30}$$

donde η es la *tasa de aprendizaje.*[8] Después de un número de ciclos, la regla del gradiente garantiza que el error E alcanzará un valor mínimo o la *meseta* más baja, si el error se representa como una superficie en el espacio vectorial de los pesos. Una posible expresión para el cálculo del error global para todos los ejemplos de entrenamientos puede ser:

$$E = \sum_{(p)} \sum_{(j)} Err_j^{(p)} \tag{2.31}$$

donde el error para un ejemplo p, $Err_j^{(p)}$ puede medirse, por ejemplo, con el *error cuadrático medio*: $Err_j(p) = \left(y_j^{(p)} - o_j^{(p)}\right)^2/2$.

La regla del gradiente descendiente para cambiar el peso entre una neurona i y una neurona j puede expresarse por la *regla generalizada delta*:

$$\delta w_{ij}(t+1) = \eta \cdot Err_j \cdot g'(u_j) \cdot o_i \tag{2.32}$$

[8]La tasa de aprendizaje multiplica el negativo del gradiente para determinar los cambios de los pesos. Entre más alta, más grande es el paso. Si es demasiado alta, el algoritmo se vuelve inestable. Si es muy pequeña el algoritmo toma mucho tiempo en converger. La elección de una tasa apropiada puede revisarse en Hagan et al. (1996).

donde Err_j es el error entre el valor deseado y_j y el valor o_j producido por la neurona j. El valor $g'(u_j)$ es la derivada $\partial g/\partial u$ de la función de activación g con respecto a la entrada de la red u, para un valor particular de u_j. o_i es el valor de salida de la neurona i.

Una iteración (o época) del ciclo de aprendizaje se define como el proceso de pasar por la red uno o varios de los ejemplos de entrenamiento y calcular el error para ellos. En cada una de estas iteraciones, el algoritmo de entrenamiento hace dos cosas: I) un pase hacia adelante, cuando las entradas son entregadas a la red hasta llegar a la última neurona y generar una salida; II) un pase hacia atrás, cuando se calcula un error y se propaga hacia atrás para ajustar los pesos. Durante el pase hacia atrás, un error Err_i para un nodo intermedio i es calculado multiplicando los errores Err_j de todas la neuronas j conectadas a la neurona i, por los respectivos pesos w_{ij}. Se utiliza luego este error hacia atrás para ajustar los pesos de las neuronas k de una capa anterior conectada con la neurona i. El procedimiento se repite por muchas épocas hasta que el error global E es suficientemente pequeño.

2.7.1. Algunos Problemas

Los MLP son probablemente los modelos NN más utilizados hasta ahora. Sin embargo, hay problemas asociados a ellos. Algunos son (Kasabov 1996):

- *El problema de la estructura de la red.* Elegir el número de unidades de la capa oculta y en general la estructura de la NN, no es asunto fácil. Hay cierta heurística disponible a lo largo y ancho de la literatura para llevar a cabo esta tarea, pero no existe ninguna regla de oro. Por ejemplo, según un criterio, el número mínimo de neuronas ocultas h debería ser $h \geq (p-1)/(n+2)$, donde p es el número de ejemplos de entrenamiento y n es el número de entradas de la red. Herbrich et al. (2000) exponen un método sistemático utilizado con tal fin. Adicional-

mente ofrecen referencias a otros métodos utilizados. Elegir también el número de capas ocultas resulta difícil. Algunos problemas requieren más de una capa oculta para obtener una mejor solución.

- *El sobreajuste.* Las NN son exaltadas por su capacidad de reconocer patrones complejos. Esta capacidad está enunciada implícitamente en la propiedad de aproximación universal. Más allá de esto, nos interesa ver para qué se está utilizando la NN. Si resulta que solo se requiere para ajustar una determinada forma funcional a un conjunto de datos, con el mínimo error posible, el sobreajuste no sería un problema.[9] En cambio, si queremos utilizar la red para pronosticar valores de ese conjunto de datos, entonces el sobreajuste es un problema fundamental. El sobreajuste no es más que una especie de "abuso de poder" que ejerce la red sobre los datos de estimación (o de entrenamiento, porque recordemos que en la jerga de NN, ellas aprenden). Si bien es cierto que una red muy ajustada nos proporciona un error dentro de muestra muy pequeño, aumentan considerablemente las probabilidades de malos pronósticos fuera de muestra. Una red sobreajustada habrá capturado no solo el patrón general subyacente en los datos (si existiera alguno), sino también los movimientos aleatorios presentes en la misma, erosionando su capacidad para hacer buenas predicciones con datos fuera del rango utilizado para su estimación.

  La Figura 2.4 muestra un caso de sobreajuste. La línea dada por la red (línea continua) se ajusta casi con perfección a los datos utilizados en su estimación (círculos). Cuando hacemos pronósticos, buscamos algo más parecido a la línea punteada. Es decir, una forma general que haga abstracción del patrón subyacente. Una de las razones del sobreajuste podría ser la utilización de demasiadas unidades ocultas dentro de la

[9] No estamos seguros para qué se querría hacer algo cómo eso.

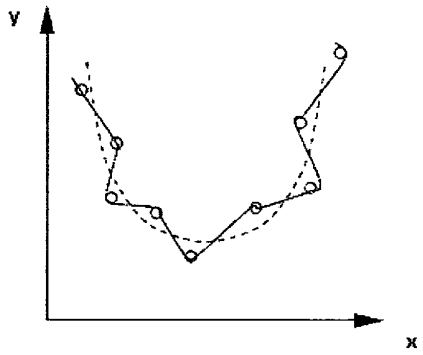


Figura 2.4: Sobreajuste de una Red Neuronal.
Fuente: `http://www.willamette.edu/~gorr/classes/cs449/linear1.html`

red.

- *El mínimo local.* Este problema está relacionado con la convergencia a errores mínimos locales[10]. La superficie de error, a diferencia de casos más sencillos (i.e. modelos lineales), no tiene por qué ser bien comportada. Una vez el algoritmo ha llegado a un mínimo, no se puede saber con certeza si se trata de uno local o global; siempre quedará la duda de si el algoritmo ha podido encontrar un mínimo aún más bajo (en cuyo caso convergió a un mínimo local).

2.8. Máquinas de Soporte Vectorial [11]

El algoritmo de máquinas de soporte vectorial (SVM) es una generalización no lineal del algoritmo **Generalized Portrait Algorithm** desarrollado por Vapnik y Lerner (1963). Como consecuencia, está fundamentado en el marco de la teoría estadística del aprendizaje (Scholkopf et al. 2000).

[10]Convergencia es el término utilizado para indicar que se ha llegado a un mínimo. En ese momento el algoritmo de aprendizaje se detiene.

[11]Tomado principalmente de Smola y Scholkopf (1998).

Las SVM son herramientas ideadas en un principio para resolver problemas de clasificación. Sin embargo, su uso ha sido extendido al dominio de problemas de regresión (Vapnik et al. 1997). Las máquinas de soporte vectorial para regresiones (SVR) explotan la idea de mapear los datos de entrada a un espacio de dimensiones muy altas (frecuentemente infinito), en donde se realiza una regresión lineal (Chu et al. 2004).

La idea básica es como sigue (Smola y Scholkopf 1998): suponga que tenemos unos datos de entrenamiento $\{(x_1, y_1), \ldots, (x_l, y_l)\} \subset \mathcal{X} \times \mathbb{R}$, donde $\mathcal{X}$ denota el espacio de los patrones de entradas (por ejemplo, $\mathbb{R}^d$). En la ϵ-SVR (Vapnik 2000) se quiere encontrar una función $f(x)$ que tenga cuando mucho una desviación ϵ de los valores objetivos y_i, para todos los datos de entrenamiento, y al mismo tiempo, sea lo menos compleja posible. En otras palabras, no nos importan los errores mientras sean menores que ϵ, pero no aceptaremos cualquier desviación mayor a esta.

2.8.1. Funciones Lineales

Empecemos describiendo el caso de funciones lineales f, que toman la forma:

$$f(x) = \langle w, x \rangle + b \quad \text{con } w \in \mathcal{X}, b \in \mathbb{R} \tag{2.33}$$

donde $\langle \cdot, \cdot \rangle$ denota el producto escalar en $\mathcal{X}$. El término w de (2.33) representa la complejidad del modelo, lo que propicia la búsqueda de un valor pequeño. Una forma de asegurar esto es minimizando la norma euclidiana, por ejemplo, $\|w\|^2$. Formalmente, este problema puede escribirse como un problema de optimización convexo requiriendo:

$$\text{mín } \frac{1}{2}\|w\|^2 \tag{2.34}$$

$$\text{Sujeto a} \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon \\ \langle w, x_i \rangle + b - y_i \leq \epsilon \end{cases}$$

El supuesto tácito en (2.34) es que tal función f que aproxima todos los pares (x_i, y_i), con precisión ϵ, realmente existe; o dicho en otras palabras, que el problema de optimización convexo sea factible. Este no siempre es el caso, o puede ser que nos interese permitir algunos errores. Pueden introducirse variables auxiliares ξ_i, ξ_i^* para lidiar con lo que de otra forma serían restricciones no factibles del problema de optimización (2.34). Así se llega a la formulación ofrecida en Vapnik (2000):

$$\text{mín } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{2.35}$$

$$\text{Sujeto a} \begin{cases} y_i - \langle w, x_i\rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

La constante $C > 0$ determina el sacrificio entre la complejidad del modelo medida por $\|w\|^2$ y el monto hasta el cual desviaciones mayores que ϵ son toleradas. La formulación de arriba corresponde a la llamada función de pérdida ϵ-insensible $|\xi|_\epsilon$ descrita por:

$$|\xi|_\epsilon := \begin{cases} 0 & \text{si } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{en otro caso} \end{cases} \tag{2.36}$$

La Figura 2.5 muestra lo anterior. Sólo los puntos por fuera de la región sombreada contribuyen al costo, y las desviaciones son penalizadas de forma lineal.

El problema de optimización (2.35) puede ser resuelto con mayor facilidad en su formulación dual. La idea clave es construir una función Lagrangiana desde la función objetivo (esta última llamada de ahora en adelante función objetivo primal) y las restricciones correspondientes, introduciendo un

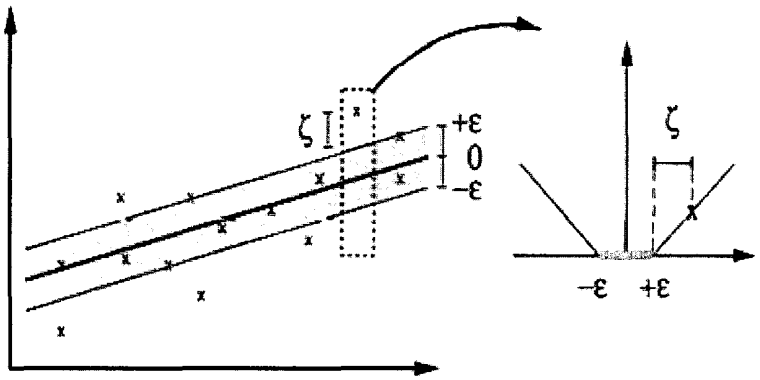


Figura 2.5: Errores en una SVM Lineal. Fuente: Smola y Scholkopf (1998)

conjunto dual de variables. Se prosigue así:

$$L := \frac{1}{2}\|w\|^2 + C\sum_{i+1}^{l}(\xi_i + \xi_i^*) - \sum_{i+1}^{l}\alpha_i(\epsilon + \xi_i - y_i + \langle w, x_i\rangle + b)$$
$$- \sum_{i=1}^{l}\alpha_i^*(\epsilon + \xi_i^* + y_i - \langle w, x_i\rangle - b) - \sum_{i+1}^{l}\eta_i\xi_i + \eta_i^*\xi_i^* \qquad (2.37)$$

donde L es el Lagrangiano y $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ son multiplicadores de Langrange.

Se entiende que las variables duales en (2.37) tienen que satisfacer restricciones positivas, i.e. $\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0$. Sigue de la condición de punto de silla que las derivadas parciales de L con respecto a las variables primales (w, b, ξ_i, ξ_i^*) desvanecen en optimalidad:

$$\partial_b L = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \qquad (2.38)$$

$$\partial_w L = w - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i = 0 \qquad (2.39)$$

$$\partial_{\xi_i^{(*)}} L = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \tag{2.40}$$

donde $\xi_i^{(*)}$ se refiere a ξ_i y también a ξ_i^*.

Sustituyendo (2.38), (2.39) y (2.40) en (2.37) resulta en el problema de optimización dual,

$$\text{máx} \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ -\epsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{cases} \tag{2.41}$$

$$\text{sujeto a} \begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases}$$

Derivando (2.41) ya hemos eliminado las variables duales η_i, η_i^* por medio de la condición (2.40), dado que estas variables no aparecieron más en la función objetivo dual y estaban presentes sólo en las condiciones de factibilidad duales. La ecuación (2.39) puede ser reescrita así,

$$w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i \tag{2.42}$$

y por lo tanto,

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b \tag{2.43}$$

La ecuación (2.43) es la llamada *expansión de soporte vectorial*. Es decir, w puede ser completamente descrita como una combinación lineal de los patrones de entrenamiento x_i. Más aun, el algoritmo completo puede ser descrito en términos de productos escalares entre los datos.

Según las condiciones *Karush-Kuhn-Tucker* (KKT) (Karush 1939, Kuhn y Tucker 1951), en la solución óptima, el producto entre variables duales y las restricciones tiene que desvanecerse. En este caso significa que,

$$\begin{aligned} \alpha_i(\epsilon + \xi_i - y_i + \langle w, x_i\rangle + b) &= 0 \\ \alpha_i^*(\epsilon + \xi_i^* + y_i - \langle w, x_i\rangle - b) &= 0 \end{aligned} \tag{2.44}$$

y

$$\begin{aligned}(C-\alpha_i)\xi_i &= 0 \\ (C-\alpha_i^*)\xi_i^* &= 0\end{aligned} \tag{2.45}$$

De esto se deduce que b puede calcularse como sigue:

$$\begin{aligned}b &= y_i - \langle w, x_i\rangle - \epsilon \quad \text{para } \alpha_i \in (0, C) \\ b &= y_i - \langle w, x_i\rangle + \epsilon \quad \text{para } \alpha_i^* \in (0, C)\end{aligned} \tag{2.46}$$

De (2.44) se deriva que sólo para $|f(x_i) - y_i| \geq \epsilon$ los multiplicadores de Lagrange pueden ser distintos de cero. En otras palabras, para todos los ejemplos dentro del tubo ϵ (la región sombreada de la Figura 2.5), el α_i, α_i^* desvanece; para $|f(x_i) - y_i| < \epsilon$, el segundo factor en (2.44) es distinto de cero, y por tanto, α_i, α_i^* tiene que ser cero para satisfacer las condiciones KKT.

La complejidad de la representación de una función, por SVM, es independiente de la dimensionalidad de $\mathcal{X}$ y depende únicamente del número de soportes vectoriales.[12] En este sentido, no necesitamos todos los x_i para describir w. Los ejemplos que vienen acompañados con coeficientes que no desvanecen se les llaman soportes vectoriales (o vectores de soporte).

2.8.2. Funciones no lineales

El paso necesario para introducir no-linealidad en las SVM consiste en pre-procesar los patrones de entrenamiento x_i con un mapeo $\Phi : \mathcal{X} \rightarrow \mathcal{F}$, donde $\mathcal{F}$ es un espacio característico de altas dimensiones (posiblemente infinito), y luego aplicar el algoritmo de SVR lineal. Esta tarea sin embargo, no es computacionalmente factible para problemas caracterizados con polinomios de alto orden y/o alta dimensionalidad (como suelen plantearse) y resulta necesario implementar una idea alternativa.

[12] Se define en breves instantes este importante concepto.

Como se ha mencionado, el algoritmo de SVM sólo depende de los productos escalares entre los distintos patrones. Por lo tanto, es suficiente conocer y usar $k(x, x') := \langle \Phi(x), \Phi(x') \rangle$, en lugar de $\Phi(\cdot)$ explícitamente. Esto permite reescribir el algoritmo de SVM como sigue:

$$\text{mín} \begin{cases} -\frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) \\ -\epsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i (\alpha_i - \alpha_i^*) \end{cases} \tag{2.47}$$

$$\text{Sujeto a} \begin{cases} \sum_{i=1}^{l} (\alpha_i \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases}$$

La análoga de (2.42) puede ser escrita como:

$$w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \Phi(x_i) \tag{2.48}$$

y por lo tanto la de (2.43),

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(x_i, x) + b \tag{2.49}$$

La diferencia con el caso lineal es que w ya no está dada explícitamente. Sin embargo por el *teorema de Fischer-Riesz* (Riesz y Szőkefalvi-Nagy 1990). se encuentra definida en un sentido débil por los productos escalares $\langle w, \Phi(x) \rangle$. Nótese también que para el caso no lineal, el problema de optimización corresponde a conseguir la función más plana en el espacio característico $\mathcal{F}$ y no en el espacio de entrada $\mathcal{X}$.

A la función k se le llama *función núcleo*;[13] y para que una función $k(x, x')$ corresponda a un producto escalar en algún espacio característico $\mathcal{F}$, debe cumplir con la *condición de Mercer* (Mercer 1909) que en términos más o menos formales dice,

$$\int_{xxx} k(x, x') f(x) f(x') dx dx' \geq 0, \forall f \in L_2(\mathcal{X}) \tag{2.50}$$

[13] *Kernel function*, en inglés.

A partir de esta condición pueden derivarse reglas simples para componer núcleos que también satisfagan la condición de Mercer (Scholkopf et al. 1998)).

2.8.3. El Proceso General

La Figura 2.6 muestra en resumidas cuentas las propiedades básicas de una SVR:

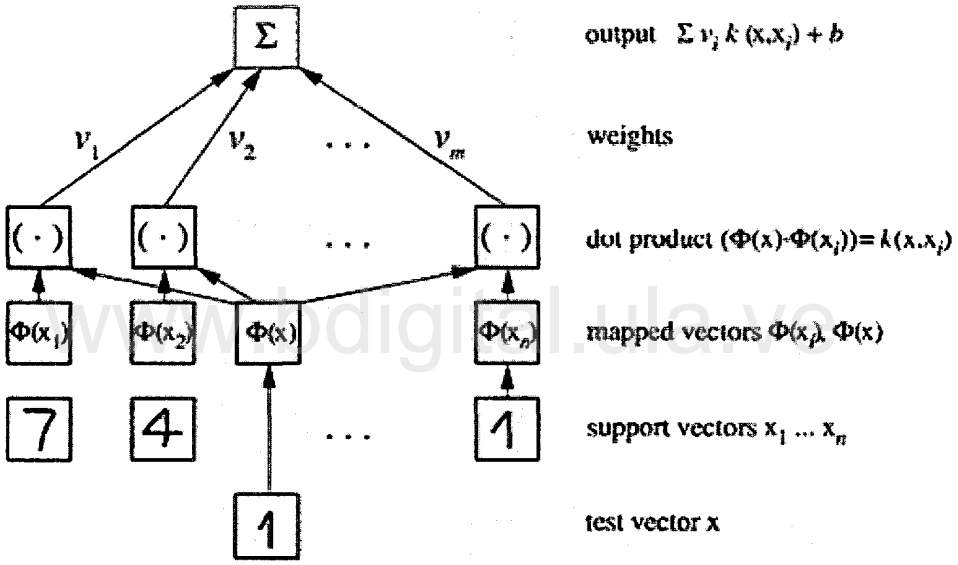


Figura 2.6: Elementos Básicos de una SVM. Fuente: Smola y Scholkopf (1998)

Los patrones de entrada (para los cuales queremos hacer predicciones) son proyectados a un espacio característico por un mapa Φ. Los productos escalares son computados con las imágenes de los patrones de entrenamiento proyectados por Φ. Esto corresponde a evaluar funciones núcleos $k(x_i, x)$. Finalmente los productos escalares son sumados utilizando los pesos $\alpha_i - \alpha_i^*$. Esto, más el término constante b, arrojan la predicción final de salida. El

proceso descrito es muy similar a una regresión en una NN, con la diferencia de que en el caso SVR los pesos en la capa de entrada constituyen un subconjunto de los patrones de entrenamiento.

Una de las grandes ventajas esgrimidas por aquellos que promulgan el uso de SVM, es que en efecto, se trata de un problema de optimización cuadrática estrictamente convexo (Smola y Scholkopf 1998), y por tanto, tiene solución única. Esto en contraste con las NN que, como se ha mencionado, pueden quedarse atascadas en mínimos locales. Adicionalmente, al igual que las NN, puede demostrarse que para muchos núcleos, las SVM son aproximadores universales (Micchelli 1986, Dyn 1987).

Capítulo 3

Marco Metodológico

Me interesa el futuro porque en él.
voy a pasar el resto de mi vida.

NICOLÁS MANCINI

3.1. Aspectos Generales del Diseño Experimental

Todos los modelos considerados tienen la forma:

$$Y_{t+h} = f(Z_t; \theta_h) + E_{t+h} \tag{3.1}$$

donde Y_t es el vector de la serie pronosticada en tiempo t, h es el horizonte de pronóstico, θ_h es un vector de parámetros desconocidos, E_t es un vector de errores y Z_t es un vector de variables predictoras. Específicamente, $Z_t = (Y_t, \ldots, Y_{t-q}, \Delta Y_t, \ldots, \Delta Y_{t-q}, \mathbf{1})$, donde q es el rezago máximo, Δ corresponde a un operador de diferencias y $\mathbf{1}$ es el vector de la constante. Cada modelo de pronóstico utiliza sólo un subconjunto $X_t \in Z_t$.

Los parámetros de estos modelos univariantes se estiman utilizando un esquema de *muestra de estimación fija* (vea por ejemplo Clements y Hendry 2005, West 2001); significa esto que se obtiene para cada modelo un único conjunto de parámetros que son utilizados para pronosticar todos los datos de una *muestra de prueba*. Otra posibilidad es configurar una ventana móvil o expansiva en donde los parámetros se estiman recursivamente actualizando la muestra para incluir (y posiblemente excluir) algún dato adicional. El costo computacional de esta última versión es excesivo y aunque más realista, debemos descartarla.

En relación a la construcción de los pronósticos, los de h pasos en adelante con $h \neq 1$ se hacen utilizando el *método de pronósticos directos*. El *método indirecto o iterado* de pronósticos multipasos se lleva a cabo utilizando un modelo de un paso hacia delante, iterando un número deseado de periodos. Los pronósticos directos, en cambio, se hacen utilizando un modelo estimado para un horizonte específico, en donde la variable dependiente es el valor que se quiere pronosticar múltiples pasos adelante (Marcellino et al. 2006, Chevillon 2007).

En relación a cuál método resulta "mejor", si el modelo de pronósticos de un paso hacia delante es correcto, entonces estimarlo para un paso e iterar hacia delante h pasos es más eficiente que estimar el modelo de h pasos directamente. Si por el contrario, los modelos están mal especificados, estimar el pronóstico h pasos hacia delante directamente permite reducir los efectos de la especificación incorrecta para el horizonte h. Desde el punto de vista práctico, los pronósticos directos requieren más tiempo de cómputo, pero simplifican considerablemente el cálculo de los pronósticos multipasos de los modelos no lineales (Stock y Watson 1999, Teräsvirta 2005, Wahlstrom 2004).

La ecuación de pronósticos directos de h pasos hacia adelante está dada

por:

$$Y_{t+h|t} = f(X_t; \hat{\theta}_{h|t_N}) \tag{3.2}$$

donde $Y_{t+h|t}$ es la serie a pronosticar h pasos hacia adelante desde el periodo t, X_t es la matriz de variables regresoras y $\hat{\theta}_{h|t_N}$ es un vector de parámetros estimados usando la muestra de entrenamiento desde el dato inicial t_0 hasta el dato final t_N.

La principal característica de estos modelos es que pretenden explicar el comportamiento de una variable, basados en valores rezagados de esa misma variable.[1]

3.2. Los Datos

Los datos consisten en 178 series de tiempo mensuales que describen distintos aspectos económicos y demográficos de Venezuela. En total, 32.637 observaciones están disponibles para el estudio, con una media de 183 observaciones por serie. Las series que menos observaciones tienen, poseen 80; la que más, 594. La fecha más antigua es enero 1958, y la más actual septiembre 2007.

Las series han sido clasificadas dentro de las siguientes categorías generales (el número corresponde al número de series en la categoría):

(a) índices de producción, 11

(b) producción física, 11

(c) ventas, 20

(d) índices de precios, 9

[1] Ormerod (2001) en su Apéndice I nos ofrece una explicación que podría convencer a algunos de la utilidad de este tipo de modelos autorregresivos.

(e) tasas de interés, 37

(f) agregados monetarios, 9

(g) ingresos y gastos públicos, 28

(h) demográficas, 47

(i) otras, 6

Para la estimación de los parámetros y el cálculo de los errores, cada serie ha sido dividida en dos juegos de datos: I) un conjunto de estimación; y II) un conjunto de prueba. Como su nombre lo sugiere, a partir del primero cada método estima sus correspondientes parámetros. Con el conjunto de prueba evaluaremos fuera de muestra la capacidad predictiva de los modelos estimados. Para los datos de estimación hemos tomado el 85 % de las observaciones de cada serie y para la prueba el restante 15 %.

Entre las manipulaciones que sufrió la base de datos, empezamos por aquellas relacionadas con los valores perdidos presentes en las distintas series (65 aproximadamente). En este caso se procedió a imputar los valores desde la fecha más antigua hasta la más actual, haciendo los siguientes ajustes:

Si el valor perdido (VP) está rodeado por valores observados (VO),

$$VP_t = (0{,}10)VO_{t-2} + (0{,}40)VO_{t-1} + (0{,}40)VO_{t+1} + (0{,}10)VO_{t+2} \qquad (3.3)$$

Si por el contrario, no hay cómo hacer el cálculo anterior pues hay un conglomerado de valores perdidos, empezando por el que está más alejado en el tiempo, estimamos:

$$VP_t = (0{,}10)VO_{t-3} + (0{,}15)VO_{t-2} + (0{,}75)VO_{t+1} \qquad (3.4)$$

Así, sucesivamente, estimamos los valores restantes del grupo de valores perdidos. Esto resulta conveniente sólo para pequeños conglomerados (nuestro caso), dado que para grupos más grandes, el cálculo converge a un valor fijo.

Las series de tiempo también sufrieron otras transformaciones. Luego de corregir los valores perdidos, se procedió a hacer los respectivos ajustes estacionales aplicando el método de Tramo/Seats[2] (Gómez y Maravall 1997, 2001). En tercer lugar, a las series que consisten en índices y cantidades se les aplicó una transformación logarítmica (natural);[3] series que representan tasas, proporciones y coeficientes no fueron transformadas.

Las series con las transformaciones anteriores representan la base de datos *original* y cualquier otra transformación adicional que se haga, será revertida a su estado original.

Otras transformaciones que se han considerado son la diferenciación de orden 1 y el reescalado de los datos entre 1 y -1. Éstas dos modificaciones serán revertidas adecuadamente para comparar las predicciones y su contraparte real. En las secciones siguientes ampliamos este punto.

En el Apéndice B, Cuadros B.7 y B.8, se muestra un resumen completo de las características de la base de datos.

3.3. Especificación de los Modelos

El primer punto que trataremos es la diferenciación de los datos. En el análisis de series de tiempo convencional resulta obligatorio verificar que los datos sean *estacionarios* o estén *cointegrados* (vea por ejemplo Gujarati 2004, Chatfield 2004). Cuando se trata de modelos relativamente nuevos como redes neuronales y máquinas de soporte vectorial, esta preocupación no surge muy a menudo en la literatura, y suele dejarse a un lado dando prioridad a los resultados de predicción y su precisión.

[2]El software utilizado fue DEMETRA Version 2.1, desarrollado por Eurostat y disponible en `http://circa.europa.eu/irc/dsis/eurosam/info/data/demetra.htm`

[3]Esto tiene el propósito de estabilizar las varianzas de las series, lo que facilita el trabajo de los modelos lineales.

Para garantizar la propiedad de estacionariedad, hemos decidido aplicar o no, diferenciación de orden uno a los datos, en función de una prueba estadística de raíz unitaria ERS-P_T (Elliott et al. 1996), con un número máximo de doce rezagos.[4] La diferenciación de orden uno es ampliamente utilizada y a menudo funciona bien (Chatfield 2004). Por ejemplo, Franses y Kleibergen (1996) muestran que con frecuencia se obtienen mejores pronósticos fuera de muestra en series de tiempo económicas, cuando los datos son diferenciados en lugar de especificando modelos con tendencias determinísticas.

En la especificación de esta prueba hacemos uso de una regresión Dickey-Fuller Aumentada estimada por mínimos cuadrados ordinarios (Said y Dickey 1984), con doce rezagos y tendencia. Si el parámetro de la tendencia resulta estadísticamente significativo al 5 %, suponemos que la serie presenta este componente y se incluye en la especificación de la prueba de raíz unitaria. Si no rechazamos la hipótesis nula de raiz unitaria al 5 %, aplicamos la primera diferencia a los datos; de lo contrario, la dejamos como está.

Las especificaciones finales de los modelos no contienen tendencia pues si los datos son diferenciados (porque presentan raíz unitaria), suponemos que se ha eliminado la tendencia, y en caso de que no sean diferenciados (porque no presenta raíz unitaria), entonces no pueden tener tendencia.

El número de rezagos que se utilizará para cada serie se estima como sigue. Consideramos modelos de orden[5] uno hasta orden doce; el conjunto $X_t \in Z_t$ de la ecuación (3.2) no considera brechas en las especificaciones de

[4] Al problema de no-estacionariedad se le llama también *problema de raíz unitaria*; la mayoría de las series de tiempo económicas presentan este fenómeno (Gujarati 2004) y pueden ser transformadas a estacionarias *diferenciando* los datos una vez. Cuando aplicamos diferenciación de *orden d* a una serie de tiempo, obtenemos una nueva serie a partir del operador: $\Delta^d X_t = (1 - L)^d X_t$, donde Δ^d es el llamado *operador de diferencias* y L el *operador de rezagos*. definido como $L^k X_t = X_{t-k}$.

[5] En este caso el orden es el término técnico utilizado para especificar qué rezago(s) se utilizará(n) como elemento(s) explicativo(s) (i.e. orden = 3, son tres rezagos: $t-1, t-2, t-3$.

los rezagos. Por ejemplo, un modelo para pronosticar $Y_{t+1|t}$ de orden cuatro (sin constante y sin diferenciación) tendría como conjunto regresor a $X_t = (Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3})$ y no (por ejemplo) $X_t = (Y_{t-1}, Y_{t-4})$.

El orden de los rezagos para cada serie se elige en atención al siguiente procedimiento:

1. Estimamos modelos desde orden uno hasta doce con mínimos cuadrados ordinarios utilizando los resultados anteriores que hacen referencia a la diferenciación.

2. Calculamos el criterio de información de Akaike (Akaike 1974).

3. Elegimos la especificación que minimza el criterio de Akaike.

El reescalado de los datos en el rango $[-1, 1]$ aplica sólo para los modelos estimados por NN y SVM. Cuando se reescala, no se hace ninguna transformación a la base de datos *original*. Estos dos métodos se estiman con y sin reescalado. Los modelos de suavización exponencial y cambio cero se estiman utilizando la base de datos original en vista de que las estimaciones no implican el cálculo de parámetros en un sentido equivalente al del resto de los métodos; son básicamente, promedios actualizados en el tiempo.

Pudiera pensarse que algunos de los procedimientos anteriores, como están basados en la estimación de modelos por medio de mínimos cuadrados ordinarios, podrían ofrecerle alguna ventaja precisamente a esos modelos. Lo cierto es que aún no hay metodologías teóricamente fundadas y de aceptación amplia y abierta que apoyen la especificación de modelos no tradicionales. Algunos intentos son de Medeiros et al. (2006), Murata et al. (1993), Moody (1992) y Sindelar y Babuska (2004).[6] Como el objetivo es evaluar el desempeño de unos modelos frente a otros, parece acertado utilizar los sólidos

[6]Medeiros et al. (2006) apunta a otras referencias en las que se discuten métodos no paramétricos utilizados para tal fin.

fundamentos teóricos que están disponibles (aún cuando están elaborados sobre la base de una clase particular de modelos) para especificar los modelos a comparar. Ésto coincide con nuestro objetivo de evaluar exactamente las mismas especificaciones (en la medida de los posible) para cada uno de los métodos de estimación. Otra aproximación podría ser la de usar, no las mismas especificaciones, sino el mismo criterio o método. Stock y Watson (1999) calculan criterios Akaike para los distintos métodos de estimación pero terminan comparando modelos con especificaciones disímiles (i.e. el número de rezagos para un método en particular es distinto al de otros métodos). Autores que proceden en forma similar a la nuestra son Moshiri y Cameron (2000) y Gonzalez (2000).

Es importante señalar al lector que no procuramos conseguir los mejores modelos que pronostiquen cada una de las series, sino especificar todo un conjunto de ellos que nos permita comparar resultados de predicción para una pequeña gama de métodos de estimación.

3.4. Procedimientos de Estimación y otros Detalles Relacionados

En las siguientes subsecciones se comentan los procedimientos de estimación para cada método.[7]

3.4.1. Suavización Exponencial y Cambio-Cero

Los modelos de suavización exponencial dobles y simples, están formulados cada uno para casos particulares; series con y sin tendencia, respectiva-

[7]El software utilizado en todos los casos es Matlab® 7.2. Se hizo uso de una librería externa (toolbox) para las máquinas de soporte vectorial (LIBSVM 2.82, disponible en `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`).

mente (Chatfield 2004). Haciendo uso de la prueba para verificar tendencias en las series. comentada en la sección 3.3, aplicamos uno u otro método. La suavización exponencial simple y doble se fusionan en un solo método de estimación.

Los distintos modelos serán estimados utilizando valores de α óptimos [vea las ecuaciones (2.4) y (2.9)]. Son óptimos en el sentido de que minimizan la suma de los errores al cuadrado (dentro de muestra) para el conjunto de valores que van desde 0.001 hasta 0.999, tomando pasos de 0.001.

Como ya se ha mencionado, al modelo de cambio-cero corresponde un $\alpha = 1$, por ser un caso especial de la suavización exponencial simple.

Las ecuaciones de pronósticos para los modelos XS, NCH y DXS, están dadas por (2.5), (2.7) y (2.10), respectivamente.

Estos modelos son estimados utilizando la base de datos *original*, sin ninguna tranformación adicional.

3.4.2. Mínimos Cuadrados Ordinarios

Mann y Wald (1943) demostraron que mucha de la teoría de la regresión clásica puede ser aplicada a modelos autorregresivos (AR).

Suponga un proceso AR de orden p, con media μ, dado por:

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \ldots + \alpha_p(X_{t-p} - \mu) + E_t \tag{3.5}$$

Dadas N observaciones $x_1, \ldots, x_N$, los parámetros $\mu, \alpha_1, \ldots, \alpha_p$ pueden ser estimados por mínimos cuadrados minimizando

$$S = \sum_{t=p+1}^{N} [x_t - \mu - \alpha_1(x_{t-1} - \mu) - \ldots - \alpha_p(x_{t-p} - \mu)]^2 \tag{3.6}$$

con respecto a $\mu, \alpha_1, \ldots, \alpha_p$. Si el proceso E_t es normal, entonces los estimadores mínimos cuadrados también son estimadores de máxima verosimilitud

(Jenkins y Watts 1968). En el caso $p = 1$, esto nos lleva a:

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N-1}(x_t - \bar{x})^2} \tag{3.7}$$

donde $\bar{x}$ es el estimador aproximado de μ, basado en la muestra.

El resultado anterior es exactamente el mismo estimador que obtendríamos al tratar la ecuación autorregresiva,

$$X_t - \bar{x} = \alpha_1(X_{t-1} - \bar{x}) + E_t \tag{3.8}$$

como una regresión ordinaria con $(X_{t-1} - \bar{x})$ como variable independiente.

Procesos autorregresivos de órdenes más altos pueden ser ajustados por mínimos cuadrados de una manera directa, y nosotros procedemos a estimarlos como se acaba de indicar.

3.4.3. Regresión Robusta

En la implementación de este método debe considerarse la función objetivo, que a su vez determina la función ψ y la función de pesos. En nuestro caso estaremos utilizando la función objetivo de *Tuckey bicuadrada* con un parámtero de *entonación* $k = 4{,}685\sigma$, donde σ es la desviación estándar de los errores. La función objetivo se define entonces como:

$$\rho(\epsilon) = \begin{cases} \frac{k^2}{6}\left\{1 - \left[1 - \left(\frac{\epsilon}{k}^2\right)\right]^3\right\} & \text{para } |\epsilon| \leq k \\ k^2/6 & \text{para } |\epsilon| > k \end{cases} \tag{3.9}$$

y la función de pesos como:

$$w(\epsilon) = \begin{cases} \left[1 - \left(\frac{\epsilon}{k}^2\right)\right]^2 & \text{para } |\epsilon| \leq k \\ 0 & \text{para } |\epsilon| > k \end{cases} \tag{3.10}$$

Para fijar las ideas, la función objetivo de *mínimos cuadrados* asigna igual peso a cada observación. mientras que los pesos de la función bicuadrada

declinan tan pronto como ϵ se aparta de cero y se hacen cero cuando $|\epsilon| > k$. Esto puede observarse en la Figura 3.1. El valor de k elegido asegura una eficiencia de 95 % cuando los errores resultan normales mientras que aún ofrece protección contra los datos atípicos (Fox 2002). Note que valores más pequeños de k ofrecen mayor protección contra los datos atípicos pero sacrificando eficiencia en el caso de errores normales.

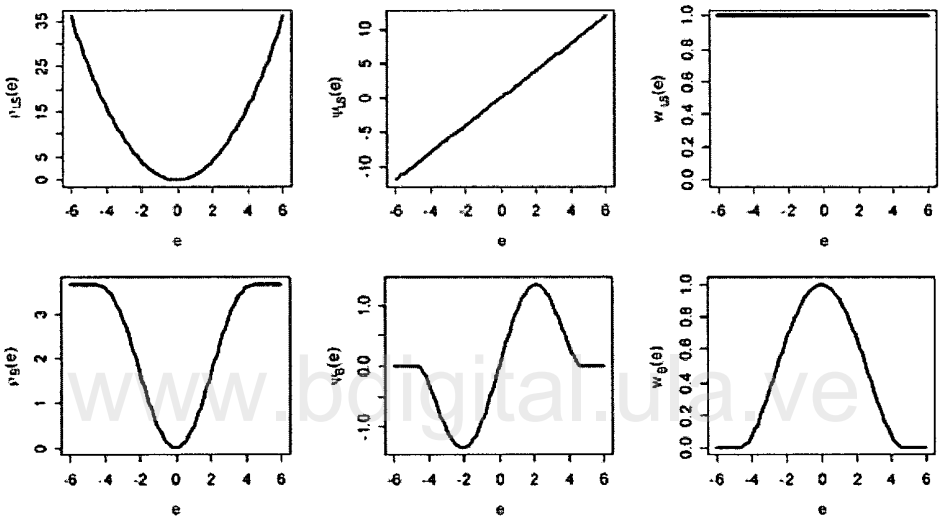


Figura 3.1: Funciones objetivos para Regresión Robusta, con sus respectivas funciones ψ y funciones de peso (de izquierda a derecha). Panel superior: estimador mínimos cuadrados. Panel inferior: estimador bicuadrado de Tuckey ($k = 4{,}685$). Ejes x = error (ϵ), ejes y primera columna = $\rho(\epsilon)$, segunda columna = $\psi(\epsilon)$, tercera columna = $w(\epsilon)$. Fuente: Fox (2002)

3.4.4. Árboles de Regresión

Los árboles de regresión pueden sufrir del problema de sobreajuste si son muy grandes y presentan un excesivo número de nodos (Yohannes y Webb 1999). Además, árboles más pequeños son más fáciles de interpretar

(Quinlan 1987). Por estas razones utilizamos el procedimiento de *poda costo/complejidad* (Breiman et al. 1984) que reduce en forma óptima (según un determinado criterio) el árbol.

La complejidad de un árbol se mide por el número de nodos terminales. La medida costo/complejidad se define como:

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}|, \quad \text{con } \alpha \geq 0 \tag{3.11}$$

donde $|\tilde{T}|$ es el número de nodos terminales en el árbol T. α representa el costo de complejidad por nodo terminal y si es pequeño, la penalidad por tener un árbol complejo es baja, resultando de la poda un árbol grande. El árbol que minimiza $R_\alpha(T)$ tenderá a tener pocos nodos y un α grande.

Antes de proseguir, resulta útil definir la *rama* T_t de un árbol T como el nodo t y todos sus descendientes. Cuando se poda esta rama se eliminan a todos sus descendientes quedando sólo la raíz t.

El procedimiento nombrado tiene dos etapas. En la primera de ellas, a partir de un árbol completo (T_{max}) se construye una secuencia de árboles más pequeños:

$$T_{max} > T_1 > T_2 > \ldots > T_K = \{t_1\} \tag{3.12}$$

donde t_1 es un árbol que está formado por un único nodo (el que contiene todos los casos).

El árbol T_1 se construye de forma distinta al resto de la secuencia. Empezando con T_{max} se revisan los errores de cada nodo terminal. Breiman et al. (1984) muestra que:

$$R(t) \geq R(t_L) + R(t_R) \tag{3.13}$$

De abajo hacia arriba, todos los pares de nodos terminales en T_{max} que satisfagan:

$$R(t) = R(t_L) + R(t_R) \tag{3.14}$$

son podados porque esas divisiones no contribuyen a disminuir el error total del árbol. Resuelto este paso, tenemos T_1.

Si un árbol $T(\alpha)$ minimiza $R_\alpha(T)$ para un α específico, entonces continuará minimizándolo hasta alcanzar un punto determinado de α. Por lo tanto, buscaremos una secuencia de valores α y los árboles que minimizan la medida costo/complejidad para cada uno de esos niveles. Una vez se tiene el árbol T_1, empezamos a podar las ramas con los vínculos más débiles. Para conseguir el vínculo más débil definimos:

$$g_k(t) = \frac{R(t) - R(T_{kt})}{|\tilde{T}_{kt}| - 1} \tag{3.15}$$

donde T_{kt} es la rama T_t que corresponde al nodo interno de t del subárbol T_k. Con la ecuación 3.15 determinamos el valor de $g_k(t)$ para cada nodo interno del árbol T_k. El vínculo más débil t_k^* en el árbol T_k es el nodo interno t que minimiza 3.15:

$$g_k(t_k^*) = \min_t\{g_k(t)\} \tag{3.16}$$

Una vez tenemos el vínculo más débil, podamos la rama definida por ese nodo. El nuevo árbol de la secuencia se obtiene por:

$$T_{k+1} = T_k - T_{t_k^*} \tag{3.17}$$

donde la resta indica el proceso de poda. El valor del parámetro de complejidad se fija a:

$$\alpha_{k+1} = g_k(t_k^*) \tag{3.18}$$

El resultado del proceso de poda será una secuencia decreciente de árboles:

$$T_{max} > T_1 > T_2 > \ldots > T_K = \{t_1\} \tag{3.19}$$

junto con una secuencia creciente de valores para el parámetro de complejidad:

$$0 = \alpha_1 < \ldots < \alpha_k < \alpha_{k+1} < \ldots < \alpha_K \tag{3.20}$$

En la segunda etapa se evalúa la secuencia de árboles y se elige a uno de ellos como modelo final. Con la secuencia de árboles podados, queremos elegir el mejor árbol tal que minimicemos la complejidad y el error de estimación $R(T)$. Recordemos además que hay un compromiso entre estos dos criterios.

Para seleccionar el árbol de tamaño óptimo debemos tener estimadores honestos del verdadero error $R^*(T)$. Esto lo hacemos calculando un error para datos que no han sido utilizados en la elaboración del árbol; la *v-validación cruzada* es una alternativa que no requiere datos adicionales (Stone 1974, Geisser 1975).

En la estimación del error de predicción para cada árbol de la secuencia, dividimos la muestra de entrenamiento L en los conjuntos $L_1, \ldots, L_V$. La muestra de estimación $v-$ésima está representada por $L^{(v)} = L - L_v$ y dejamos al conjunto L_v para estimar el error de predicción. Con cada $L^{(v)}$ se arma un árbol completo y su respectiva secuencia de subárboles:

$$T_{max}^{(v)} > T_1^{(v)} > \ldots > T_k^{(v)} > T_{k+1}^{(v)} > \ldots > T_K^{(v)} = \{t_1\} \tag{3.21}$$

Cada una de estas secuencias tiene asociado su correspondiente secuencia de parámetros de complejidad:

$$0 = \alpha_1^{(v)} < \ldots < \alpha_k^{(v)} < \alpha_{k+1}^{(v)} < \ldots < \alpha_K^{(v)} \tag{3.22}$$

En este punto tenemos $V + 1$ secuencias de árboles y parámetros de complejidad. Usaremos las secuencias $T_k^{(v)}$ para evaluar el desempeño de cada árbol en la secuencia original. Con las muestras de prueba L_v y los árboles $T_k^{(v)}$. determinamos los errores de predicción de los subárboles T_k. Para lograr eso, debemos encontrar en la secuencia $T_k^{(v)}$ árboles con complejidad equivalentes a T_k. Recuerde que un árbol T_k es el árbol de menor costo/complejidad en el rango $\alpha_k \leq \alpha < \alpha_{k+1}$.

Para usar validación cruzada en la escogencia del mejor subárbol de T_k, definimos un parámetro de complejidad representativo para ese intervalo ha-

ciendo uso de la media geométrica:

$$\alpha'_k = \sqrt{\alpha_k \alpha_{k+1}} \tag{3.23}$$

Ahora denotamos al predictor del árbol $T^{(v)}(\alpha'_k)$ como $d_k^{(v)}(x)$. El estimador del verdadero error de predicción $R^*(T)$ es:

$$\hat{R}^{VC}(T_k)) = \frac{1}{n} \sum_{v=1}^{V} \sum_{(x_i,y_i)\in L_v} \left(y_i - d_k^{(v)}(x_i)\right)^2 \tag{3.24}$$

donde n es el tamaño de la muestra completa, (x_i, y_i) es un caso de los datos con $x_i = (x_{1_i}, \ldots, x_{d_i})$, $i = 1, \ldots, n$ y d un árbol de regresión. Utilizamos cada caso de la muestra de prueba L_v y la regla $d_k^{(v)}(x)$ para obtener una predicción, para luego calcular un error cuadrático entre el valor predicho y el real. Hacemos esto para muestra de prueba y los n casos. Con la ecuación 3.24 terminamos de calcular el error de predicción para para el árbol k.

Para elegir al mejor subárbol necesitamos la expresión para el error estándar de $\hat{R}^{VC}(T_k)$ dada por:

$$\hat{SE}(\hat{R}^{VC}(T_k)) = \sqrt{\frac{s^2}{n}} \tag{3.25}$$

donde

$$s^2 = \frac{1}{n} \sum_{x_i,y_i} \left[\left(y_i - d_k^{(v)}(x_i)\right)^2 - \hat{R}^{VC}(T_k)\right]^2 \tag{3.26}$$

Finalmente, el árbol óptimo T_k^* lo identificamos con la siguiente regla: buscamos al árbol con error de predicción estimado más pequeño y lo denotamos T_0. Luego elegimos el árbol con la complejidad más pequeña tal que:

$$\hat{R}^{VC}(T_k^*) \leq \hat{R}^{VC}_{min}(T_0) + \hat{SE}(\hat{R}^{VC}_{min}(T_0)) \tag{3.27}$$

Una vez elegido el árbol óptimo, la media del nodo terminal se convierte en el valor predicho de la variable dependiente para los casos que lleguen a ese nodo.

El valor de v para la validación cruzada se ha fijado en 10 y la regla de parada en la construcción del árbol T_{max} es que un nodo deja de dividirse cuando alcanza un número de casos ≤ 10.

3.4.5. Redes Neuronales Artificiales

En esta sección describiremos dos procedimientos que estaremos ejecutando para lidiar con dos de los problemas planteados en la sección 2.7.

El sobreajuste

Existen varios procedimientos que pretenden minimizar la posibilidad de un sobreajuste. Entre ellos se encuentran *la parada temprana* y la *regularización bayesiana* (vea por ejemplo Hagiwara y Kuno 2000, Sarle 1995, y las citas en ellas presentes).

El primero consiste en entrenar la red hasta el punto en que empieza a dar señales de sobreajuste. Esto último se consigue observando el error generado en una muestra adicional llamada *de validación*; en las primeras iteraciones este error disminuye progresivamete, y una vez alcanza su mínimo y empieza a aumentar, es el momento de detener el entrenamiento (la red se especializa en la muestra de estimación y está perdiendo la capacidad de predecir correctamente datos nunca vistos: los de la muestra de validación).

Por su parte, la regularización está basada en un cambio en la función de error que minimiza la red. Típicamente estamos interesados en minimizar una función como el *error cuadrático medio*, definida como:

$$mse = \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2 \tag{3.28}$$

donde $\hat{y}_t$ es el valor pronosticado del dato real y_t y n es el tamaño de la muestra.

Es posible mejorar la generalización de la red si se modifica la función de desempeño agregando un término que consiste en la media de la suma de los pesos de la red al cuadrado (msw), de forma que:

$$msereg = \gamma mse + (1 - \gamma) msw \tag{3.29}$$

sea la nueva función a minimizar en el entrenamiento de la red, γ sea el *coeficinete de desempeño* y msw la suma de los cuadrados de los pesos de la red:

$$msw = \sum_{i=1}^{N_w} w_i^2 \tag{3.30}$$

donde w_i son los pesos y N_w es el número de pesos en la red.

Utilizando la función de desempeño modificada obliga a la red a estimar pesos más pequeños, haciendo que la red responda con más suavidad y con menos probabilidad de sobre-ajustar (Demuth et al. 2008).

Si estimamos el parámetro γ óptimo apoyándonos en el marco bayesiano de MacKay (1992), llegamos finalmente a la regularización bayesiana. Calcular el parámetro óptimo es de extrema utilidad porque con un γ muy grande podríamos tener sobreajuste y con uno muy pequeño la red no podría ajustarse adecuadamente a los datos. Esta forma de entrenamiento permite que la red neuronal utilice sólo los pesos necesarios para reproducir la función subyacente en los datos, y no el total dictado por la estructura morfológica.

Nosotros estaremos aplicando en nuestras estimaciones la regularización bayesiana.[8] Este método resulta conveniente por dos razones: en primer lugar, no hace falta reservar un tercer conjunto de datos para hacer la validación; esto resulta especialmente importante porque el 85 % de nuestras series tienen como máximo 250 observaciones. Y en segundo lugar, muestra mejor desempeño que la parada temprana (Sarle 1995).

[8]En el Apéndice A, Figura A.4, se pueden observar gráficos que corresponden a resultados específicos para este procedimiento.

El mínimo local

Para atacar este problema podemos correr la red utilizando distintos parámetros de inicialización (Sarle 1997). Como se trata de un proceso iterativo, los caminos que se transitan para llegar al mínimo son distintos, y en efecto, estaremos evaluando el comportamiento de los errores. Nos quedaremos con aquellos parámetros iniciales que minimizan el error de la red. Cada red será estimada con 50 valores de inicialización distintos. No resulta práctico en nuestro contexto estimar un múmero mayor dada la gran cantidad de series de tiempo a procesar y el tiempo de ejecución de cada red neuronal.

Arquitectura y estimación

La arquitectura de nuestras redes presentan tres capas con sesgos. Trabajaremos con una sóla capa oculta[9] de seis unidades y las funciones de activación correspondientes son tangentes hiperbólicas (tangh - ecuación 2.28) acotadas entre -1 y 1. La función de activación de la unidad de salida es una función lineal (i.e. combinación lineal) y las unidades de sesgo reciben como entradas valores constantes e iguales a uno.

Esta NN es de la clase de alimentación hacia delante porque la información sólo fluye desde la capa de entrada hacia la capa de salida, y se encuentra además, totalmente conectada (capa a capa). El número de parámetros N_w a estimarse es $N_w = (I + 1) \times H + (H + 1) \times S$, donde I es el número de variables de entrada, H es la cantidad de nodos ocultos y S es el número de variables de salida (pero recuerde que usamos regularización Bayesiana).

El algoritmo iterativo usado para estimar los pesos es propagación hacia atrás con regularización bayesiana. La regularización toma lugar dentro de una optimización Levenberg-Marquardt (pueden consultarse Levenberg 1944,

[9]Tkacz y Hu (1999) hacen referencia a Kuan y White (1994) como proponentes de una sola capa oculta para la mayoría de las aplicaciones económicas.

Marquardt 1963, Foresee y Hagan 1997). La función de error corresponde al error absoluto medio modificado [vea las ecuaciones (3.28) a la (3.30)].

Para la estimación de las redes neuronales se ha procedido a un reescalado de los datos en el rango $[-1, 1]$ como lo sugieren Demuth et al. (2008), Sarle (1997), Foresee y Hagan (1997), entre otros. Zhang et al. (1998) cita como argumentos para esto, evitar problemas computacionales, cumplir con requerimientos algorítmicos y facilitar el aprendizaje de la red. Con los datos reescalados no se verifica por raíces unitarias. Las estimaciones también se hacen sin reescalar.

3.4.6. Máquinas de Soporte Vectorial

Para esta categoría de modelos estimaremos SVR's del tipo ν-SVR. Se trata de una variante del modelo inicial (ϵ-SVR), propuesta en Smola y Scholkopf (1998). Tres parámetros deben ser fijados. En lugar de especificar a priori un nivel de certeza deseado ϵ, se fija ν, que permite controlar el número de errores.[10] También deben fijarse la constante de regularización C que penaliza los errores fuera del rango $\pm\epsilon$, el núcleo k y sus respectivos parámetros.

Con respecto a la elección del núcleo, Hsu et al. (2003) recomiendan utilizar como primera aproximación un núcleo función de base radial (RBF):

$$k(x_i, x_j) = exp(-\gamma \|x_i - x_j\|^2) \quad \text{con } \gamma > 0 \tag{3.31}$$

donde x_i, x_j corresponden a un soporte vectorial y a un punto en el espacio de datos; γ es un parámetro. El núcleo RBF, además de ser no lineal, ofrece la ventaja de poseer un único parámetro que establecer para iniciar la estimación (γ); este último mide la influencia que cada vector de soporte

[10] ν está estrechamente relacionada con ϵ, y en cierto sentido permite optimizar el valor de esta última automáticamente. Específicamente, nos permite controlar la fracción de soportes vectoriales y de puntos fuera del tubo ϵ (vea Smola y Scholkopf 1998, para los detalles).

ejercerá sobre puntos a su alrededor. Un γ pequeño permitirá a un soporte vectorial ejercer una fuerte influencia sobre un área grande, lo que se traduce en una superficie de decisión más suave y regular, junto con una reducción en el número de soportes vectoriales. Un γ más grande tiene el efecto contrario.[11] Por otra parte, este núcleo también tiene menos dificultades numéricas en comparación con otros núcleos.

Porque no se sabe de antemano que combinación de C y γ resulta mejor para un determinado problema, los mismos autores recomiendan hacer una búsqueda de malla con *v-validación cruzada*. En la validación cruzada la muestra de estimación es dividida en v submuestras de igual tamaño. Se entrena secuencialmente una SVM con $v-1$ de las submuestras y se prueba el modelo obtenido sobre la submuestra restante. Esto nos proporciona un error de generalización para cada secuencia y el promedio de estos errores nos da un error promedio para determinados parámetros C y γ. En la búsqueda de malla probamos cómo se comporta ese error promedio para distintas combinaciones de C y γ. Finalmente, usamos los parámetros que minimizan ese error para ajustar el modelo sobre la muestra completa de estimación.[12]

El procedimiento de v-validación cruzada opera en contra de un posible sobreajuste (mismo problema que en las redes neuronales y árboles de regresión) y la búsqueda de malla nos da una idea de la combinación óptima de C y γ. Hsu et al. (2003) recomiendan secuencias exponencialmente crecientes de los parámetros como una forma práctica de identificar buenos parámetros. En nuestra aplicación, dados los grandes costos computacionales, la búsqueda de malla se efectúa en dos fases. En la primera $C = 2^{-20}, \ldots, 2^{8}$ y $\gamma = 2^{-30}, \ldots, 2^{8}$ con pasos $st = 5$; en la segunda fase se hace una búsqueda

[11] Más detalles en `http://svr-www.eng.cam.ac.uk/~kkc21/thesis_main/node31.html`.

[12] En el Apéndice A, Figura A.5, se pueden observar gráficos que corresponden a resultados específicos para este procedimiento.

más fina con $st = 1$ para la región definada por los parámetros "óptimos" de la primera fase ± 2. El número de submuestras para la validación cruzada se ha fijado en $v = 10$. Duan et al. (2003) sugieren que $v = 5$ es suficiente para obtener buenos estimados del error de generalización manteniendo el costo computacional relativamente bajo. Añaden además que entre varios métodos probados para fijar los parámetros, la validación cruzada mostró el mejor desempeño global.

El parámetro $\nu = 0{,}5$ y es el valor predeterminado del programa de ejecución. En Chalimourda et al. (2004) se ofrece evidencia teórica y empírica del porqué éste (y otros valores cercanos) puede considerarse un buen valor.

Estos modelos son estimados utilizando datos reescalados en el rango $[-1, 1]$ (en cuyo caso no se verifica por raíces unitarias) y sin reescalar (verificando por raíces unitarias). El reescalado está fundamentado en los mismos argumentos que en el caso de las NN.

3.5. Las Comparaciones

Las comparaciones tienen como objetivo fundamental discernir qué categoría de modelos logra hacer predicciones puntuales con mayor precisión. Esto se hará en una base global.

Revisaremos también la precisión de los pronósticos en función del método de estimación en combinación con otro elemento característico de las series de tiempo, como es su categoría.

Podríamos además concluir sobre la efectividad de los modelos no lineales para pronosticar estas series de tiempo considerando la siguiente taxonomía:

- Modelos lineales: (1)cambio-cero, (2)suavización exponencial, (3)mínimos cuadrados ordinarios, (4)regresión robusta.

- Modelos no lineales: (5)árboles de regresión, (6)redes neuronales artificiales, (7)máquinas de soporte vectorial.

Para esta comparación partimos de la hipótesis que modelos de las categorías (5)-(7) deben pronosticar con mayor precisión que las categorías (1)-(4) porque su definición admite la posibilidad de incorporar elementos no lineales que pudiesen estar presentes en los datos. Aun más, los modelos de máquinas de soporte vectorial parecieran ofrecer una ventaja adicional, y es que tienen solución óptima única, en contraste con los modelos de redes neuronales que pudiesen estancarse en mínimos locales.

Todo lo anterior se hace considerando los distintos horizontes de predicción.

Para evaluar el desempeño haremos uso del *error porcentual medio absoluto simétrico* (SMAPE),[13] que es una medida basada en errores porcentuales y por tanto independiente de escala. Utilizado en Makridakis y Hibon (2000) y en Ahmed et al. (2007) lo definimos como:

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \tag{3.32}$$

donde y_i es el valor real a pronosticar, $\hat{y}_i$ es el pronóstico y n el número de observaciones en la muestra.

3.5.1. Pruebas Estadísticas

Siguiendo a Koning et al. (2005) implementamos algunas pruebas estadísticas formales para evaluar la significancia estadística de la precisión alcanzada por los distintos métodos de estimación. Estas pruebas son de naturaleza no paramétrica y están basadas en *ranqueos*. La metodología basada en ranqueos es de sencilla interpretación, implementación y está libre de supuestos distribucionales.

[13] *Symmetric Mean Absolute Percentage Error*, en inglés

La situación que se considera es la existencia de K métodos ($k = 1, 2, \ldots, K$) que se han aplicado a N series de tiempo ($n = 1, 2, \ldots, N$) para pronosticar H periodos ($h = 1, 2, \ldots, H$). Para cada uno de los K métodos y para cada h, tenemos un ranqueo en términos del SMAPE, promediado sobre las N series de tiempo. Nos interesa comparar los ranqueos en cada horizonte h y ver si hay diferencias significativas entre ellos.

Para motivar el uso de estos métodos, mostramos los resultados de aplicar a los errores la prueba de normalidad Lilliefors y gráficos de probabilidad normal.

Prueba General: Friedman

La prueba de Friedman (Friedman 1937, 1939) puede utilizarse para determinar si las desviaciones de los pronósticos con respecto a sus valores verdaderos (medidas por el SMAPE) son significativamente diferentes en el sentido estadístico para los distintos métodos de estimación en su conjunto.

Esta prueba evalúa la hipótesis nula:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_K \tag{3.33}$$

donde τ_k es el efecto desconocido aditivo con el que contribuye el método k (Koning et al. 2005).

Si se cumple la hipótesis anterior, significa que los ranqueos $R_{n1}, R_{n2}, \ldots, R_{nK}$ se han obtenido ranqueando variables aleatorias i.i.d. $V_{n1}+\tau_1, V_{n2}+\tau_2, \ldots, V_{nK}+\tau_K$ para cada serie de tiempo n. La hipótesis alternativa se corresponde con:

$$H_1 : \tau_1, \tau_2, \ldots, \tau_K \qquad \text{no todos iguales} \tag{3.34}$$

Si se cumple la hipótesis alternativa, las variables aleatorias $V_{n1}+\tau_1, V_{n2}+\tau_2, \ldots, V_{nK}+\tau_K$ siguen siendo independientes pero pueden diferir en locación unas con respecto a otras. El estadístico de prueba puede escribirse como:

$$S = \frac{12N}{K(K+1)} \sum_{k=1}^{K} \left(\bar{R}_k - \frac{K+1}{2} \right)^2 \tag{3.35}$$

donde $\bar{R}_k$ es el ranqueo promedio asignado al método k luego de aplicado a las N series de tiempo. Bajo la hipótesis nula (3.33), S converge en distribución a una variable aleatoria Chi-cuadrada con $K-1$ grados de libertad, mientras N tiende a infinito. Los valores críticos de S pueden encontrarse en Hollander y Wolfe (1973).

Los autores Koning et al. (2005) comentan que Stekler (1991), Stekler (2001), Kolb y Stekler (1996) y De˜Gooijer y Zerom (2000), entre otros, han utilizado esta prueba en el ámbito de la predicción.

Pruebas Multicomparativas de Igualdad de Ranqueos

Si hemos podido rechazar la hipótesis de que todos los ranqueos son iguales, todavía podemos comparar si para algún par en particular se cumple la igualdad. Éste resultado puede obtenerse utilizando una prueba *multicomparativa de igualdad de ranqueos.*

Este procedimiento de comparaciones múltiples es similar a la prueba t pero utiliza un valor crítico ajustado por la cantidad de comparaciones que se hacen. El valor crítico de Tukey (o diferencia significativa honesta de Tukey) considera que hay g grupos y $g(g-1)/2$ posibles pares de combinaciones. Tukey encontró la distribución del mayor de estos estadísticos t cuando la hipótesis nula de indiferencia es verdadera.[14] Por ejemplo, cuando hay cuatro tratamientos y seis sujetos por tratamiento, hay veinte grados de libertad para los múltiples estadísticos de prueba. En la prueba t de Student, el valor crítico es 2.09. Para que el resultado sea estadísticamente significativo de acuerdo a Tukey el estadístico t debe superar 2.80 (Dallal 2001).

A continuación se presentan dos pruebas de este tipo.

[14]Student hizo lo mismo para *dos* grupos.

Prueba Multicomparativa con el Mejor

McDonald y Thompson (1967) y McDonald y Thompson (1972) desarrollan un procedimiento multicomparativo en el cual son probadas hipótesis componentes de la forma,

$$H_0, k_1, k_2 : \tau_{k_1} = \tau_{k_2} \tag{3.36}$$

donde $k_1 = 1, 2, \ldots, k_2 - 1$ y $k_2 = 1, 2, \ldots, K$. Cada hipótesis componente es rechazada si y sólo si,

$$|\bar{R}_{k_1} - \bar{R}_{k_2}| \geq r_{\alpha.K.N} \tag{3.37}$$

donde el valor crítico $r_{\alpha,K.N}$ se elige adecuadamente para el caso multicomaprativo de forma que el error sea igual a α. Note que el valor crítico depende de α, K y N. Para N grande está demostrado que (Hollander y Wolfe 1999) :

$$r_{\alpha.K.N} \approx q_{\alpha.K}\sqrt{\frac{K(K+1)}{12N}} \tag{3.38}$$

donde $q_{\alpha.K}$ es el percentil superior α del rango de K variables normales estándar independientes. Estos valores pueden conseguirse en Harter (1960).

Para visualizar los resultados podemos construir gráficos de la siguiente manera. Para cada método k dibujamos un intervalo de longitud $r_{\alpha,K,N}$ centrado en R_k. Si los intervalos para los métodos k_1 y k_2 no se solapan, rechazamos H_{0,k_1,k_2}. En esta forma podemos comparar cada uno de los métodos con respecto al mejor. Dibujando una línea de referencia horizontal en el límite superior del método con mejor desempeño, podemos concluir cuáles de cada uno de ellos se desempeña significativamente peor que el mejor (para detalles vea Koning et al. 2005, Hsu 1996).

Prueba Multicomparativa con la Media

En esta prueba las hipótesis componentes toman la forma,

$$H_{0,k_1} : \tau k_1 = \bar{\tau} \tag{3.39}$$

donde $k_1 = 1, 2, \ldots, K$ y $\bar{\tau}$ es el efecto promedio de los métodos. Dejemos que $\bar{\bar{R}} = (K+1)/2$ denote la media del promedio de ranqueos. Cada hipótesis componente H_{0,k_1} se rechaza si y sólo si:

$$|\bar{R}_{k_1} - \bar{\bar{R}}| \geq r'_{\alpha,K.N} \tag{3.40}$$

donde el valor crítico $r'_{\alpha,K.N}$ se elige adecuadamente para el caso multicomaprativo de forma que el error sea igual a α. Para N grande tenemos,

$$r'_{\alpha.K.N} \approx H_{\alpha.K}\sqrt{\frac{K(K+1)}{12N}} \tag{3.41}$$

donde $H_{\alpha,K}$ es el valor crítico correspondiente a la *desviación absoluta máxima*. Esta última se define como $\text{máx}_k\,|Z_k - \bar{Z}_K|$ donde $Z_1, \ldots, Z_K$ son variables aleatorias normales estandarizadas e independientes y $\bar{Z}_K = K^{-1}\sum_{k=1}^{K} Z_k$. Las tablas de $H_{\alpha,K}$ pueden encontrarse en Halperin et al. (1955).

Al igual que en el caso anterior podemos visualizar estos resultados graficando el ranqueo medio de cada método $\hat{R}_k$ contra k y dibujando tres líneas horizontales. La línea central a una altura $\bar{\bar{R}}$, y la líneas de control superior e inferior a una distancia $r'_{\alpha,K,N}$ de la central. Todos los métodos por fuera del intervalo formado por las líneas de control tienen desempeño significativamente distinto al promedio.

Desempeño Relativo de cada Método

También resulta de interés comparar el desempeño relativo de cada método con cada otro método y determinar cuáles métodos se comportaron significativamente mejor que otros.

La medida de precisión utilizada es el porcentaje de veces que el método i superó al método j. La estadística de prueba es la binomial con $p = 0{,}5$.

Capítulo 4

Resultados

Resulta sumamente difícil hacer un pronóstico acertado, especialmente sobre el futuro.

NIELS BOHR

4.1. Descriptivos Globales

Los primeros resultados que presentamos pueden considerarse de naturaleza descriptiva y consisten en los errores promedios medidos por el SMAPE, para cada método de estimación y cada horizonte. Los resultados que discutimos de aquí en adelante excluyen las redes neuronales y las máquinas de soporte vectorial que han sido reescaladas. Estos resultan marcadamente erráticos comparados con su versión sin reescalar.

El Cuadro 4.1 muestra que tanto la regresión robusta (RROB) como las máquinas de soporte vectorial (SVM) presentan mejores desempeños promedio que el resto de los métodos. En tercer lugar han quedado las estimaciones por mínimos cuadrados ordinarios (OLS) y luego los árboles de regresión

(REGT). En quinto, sexto y séptimo lugar tenemos respectivamente a las redes neuronales (NN), los modelos de cambio-cero (NCH) y la suavización exponencial (GXSM). El comportamiento relativo de los errores promedios en función de los horizontes es estable para todo el grupo exceptuando los modelos estimados por OLS y RROB, que llegan a estar entre los últimos lugares para el horizonte = 12. Esto puede observarse en la Figura 4.1, donde además podemos ver cómo el error aumenta a medida que lo hace el horizonte de predicción. Este resultado es naturalmente esperado: los pronósticos alejados en el tiempo son consistentes con mayor incertidumbre y con errores más grandes (vea por ejemplo Sampson 1991).

Horizonte	**1**	**3**	**6**	**12**	**Media**
OLS	0.058	0.081	0.110	0.160	0.102
RROB	0.058	0.079	0.105	0.155	0.099
REGT	0.068	0.089	0.112	0.148	0.104
SVM	0.058	0.083	0.114	0.145	0.100
SVMR	0.137	0.164	0.181	0.207	0.172
NN	0.064	0.094	0.121	0.146	0.106
NNR	0.070	0.117	0.128	0.173	0.122
GXSM	0.139	0.146	0.159	0.176	0.155
NCH	0.122	0.133	0.140	0.154	0.137

Cuadro 4.1: Errores Medios Medidos por la Función de Pérdida SMAPE (incluye NN y SVM reescalando y sin reescalar).

La diferencia promedio entre el método con mejor y peor desempeño (RROB y GXSM) es poco menos de 6 %. Los siete métodos de estimación pueden dividirse en dos grandes grupos según su desempeño. El *grupo 1* incluye a todos los métodos excepto la suavización exponencial y los modelos de cambio cero. El *grupo 2* incluye sólo a estos dos últimos. La diferencia promedio entre el mejor y el peor del primer grupo es de aproximadamente

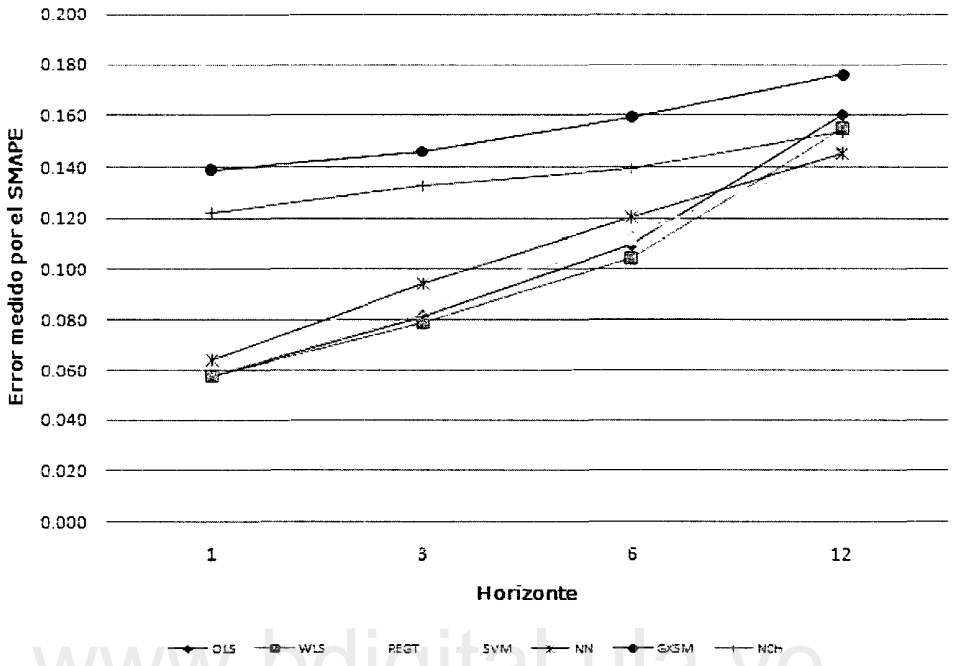


Figura 4.1: SMAPEs y Horizontes de Pronóstico.

3/4 de un punto porcentual.

En el Cuadro 4.2 se muestran los percentiles promedio asociados a los SMAPEs. Los percentiles corresponden al porcentaje de series para los que cada categoría de estimación genera errores promedio máximos del tamaño indicado. Por ejemplo, la estimación con OLS produce errores máximos de 0.036 para el 50 % de las series, de 0.167 para el 75 % de las series y así sucesivamente.

Puede observarse que las distribuciones de los errores de los integrantes del grupo 1 son bastante parecidas, e inclusive, hasta el percentil 10, los métodos del grupo 2 arrojan resultados que no son marcadamente distintos del otro grupo (aunque aún resultan mayores). Esto hace pensar que para un

	Percentiles Promedio de SMAPEs						
	5	**10**	**25**	**50**	**75**	**90**	**95**
OLS	0.001	0.002	0.008	0.036	0.167	0.307	0.365
RROB	0.001	0.002	0.007	0.034	0.160	0.301	0.368
REGT	0.001	0.003	0.008	0.040	0.171	0.301	0.399
SVM	0.001	0.002	0.007	0.036	0.156	0.313	0.360
NN	0.001	0.003	0.009	0.041	0.169	0.312	0.392
GXSM	0.001	0.004	0.017	0.062	0.208	0.460	0.583
NCH	0.002	0.005	0.019	0.064	0.174	0.413	0.532

Cuadro 4.2: Percentiles Promedio de SMAPEs. Cada fila representa el promedio de los horizontes 1, 3, 6, 12.

subconjunto de las series, los dos métodos del grupo 2 presentan resultados comparables con los métodos del grupo 1.

El resultado anterior puede conectarse también en el Cuadro 4.3. Las proporciones allí nos dicen, por ejemplo, que para el 34 % de las series, las máquinas de soperte vectorial llegan entre primer y segundo lugar; también que, para aproximadamente el 39 % de las series, la suavización exponencial obtiene ranqueos entre el primer y quinto lugar. Y así sucesivamente.

Resulta interesante notar que los métodos del grupo 2 obtienen resultados relativamente buenos para un subcojunto de los datos; para aproximadamente el 25 % de ellos. Véase, por ejemplo, que la suavización exponencial obtiene el primer lugar entre los distintos métodos, para 17 % de los datos, superado únicamente por la regresión robusta que lo hace para un 20 % de los datos. A partir de la proporción acumulada (3), se observa un estancamiento de los resultados del grupo 2. De hecho, puede decirse que logran resultados muy buenos para cierta proporción de los datos (25 %) y resultados relativamente malos para el resto. Números como estos podrían llevarnos a utilizar modelos

	Ranqueo	Prop. Acum. por Posición de Llegada					
	Promedio	**1**	**2**	**3**	**4**	**5**	**6**
OLS	3.581	0.098	0.271	0.486	0.704	0.895	0.965
RROB	3.097	0.206	0.390	0.600	0.795	0.921	0.990
REGT	3.951	0.166	0.295	0.431	0.560	0.747	0.850
SVM	3.381	0.153	0.340	0.555	0.725	0.882	0.965
NN	4.079	0.076	0.206	0.355	0.569	0.812	0.903
GXSM	4.904	0.171	0.250	0.292	0.337	0.388	0.657
NCH	5.007	0.129	0.247	0.281	0.310	0.355	0.670

Cuadro 4.3: Ranqueos y Proporciones Acumuladas Promedio por Posición de Llegada. Cada fila representa el promedio de los horizontes 1, 3, 6, 12.

del grupo 2 para pronosticar cierto tipo de series, aún cuando globalmente no han logrado un buen posicionamiento.

La sección B del Apéndice contiene cuadros detallados de los resultados presentados hasta el momento (Cuadros B.1 y B.2).

4.2. Pruebas Estadísticas Globales

Presentamos acá los resultados de las pruebas estadísticas discutidas en la sección 3.5.1. Para motivar el uso de pruebas basadas en ranqueos, hacemos mención de los resultados de la prueba de normalidad de los errores SMAPE; que en este caso, indican que los errores no siguen una distribución normal, y por lo tanto, el uso de pruebas basadas en los ranqueos es una opción apropiada en este caso. En la Figura A.1, sección A del Apéndice, se pueden hacer inspecciones gráficas de normalidad para estos datos.

Horizonte	Estadística Chi-cuadrada	P-valor
1	312.31	0
3	172.82	0
6	98.39	0
12	28.71	6.9×10^{-5}

Cuadro 4.4: Prueba Friedman de Igualdad Conjunta de Ranqueos. Valores chi-cuadrados y valores p.

Prueba General: Friedman

Los resultados de la prueba conjunta de Friedman se presentan en el Cuadro 4.4. Puede observarse que los valores chi-cuadrados son estadísticamente significativos a un nivel de significancia muy bajo (menores que 0.000001 % para los horizontes 1-6 y menores que 0.007 % para el horizonte 12), lo que nos permite rechazar la hipótesis nula de ranqueos aleatorios, a favor de la hipótesis en la que hay diferencias significativas entre los distintos métodos de predicción, en su conjunto.

Prueba Multicomparativa con el Mejor

Como se ha determinado que no todos los métodos son iguales, pasamos a discutir el resultado de la prueba multicomparativa con el mejor. El Cuadro 4.5 muestra el número de horizontes para los cuales cada método se desempeña significativamente peor que el mejor método. Los resultados indican que OLS, RROB y SVM son los mejores métodos, y que REGT, NN, GXSM y NCH son peores que el mejor método para al menos dos de los cuatro horizontes estimados.

Se observa ya un grupo de métodos (OLS, RROB, SVM) que, estadísticamente, presentan los mismos resultados en términos de la comparación planteada. Ni OLS, ni SVM presentan peor desempeño que el mejor (RROB), en

Método	Peor
OLS	0
RROB	0
REGT	2
SVM	0
NN	3
GXSM	3
NCH	4

Cuadro 4.5: Prueba Multicomparativa con el Mejor. Número de horizontes para los cuales cada método tiene peor desempeño que el mejor método.

ningún horizonte. Los mejores métodos lo son consistentemente en todos los horizontes. Sólo NCH es consistemente peor para todos los horizontes. Los detalles de estos resultados pueden verse en el Apéndice B, Cuadro B.3 y los gráficos relacionados en el Apéndice A, Figura A.2.

Prueba Multicomparativa con la Media

El Cuadro 4.6 muestra los resultados de la prueba multicomparativa con la media. Allí puede observarse que RROB tiene mejor desempeño que el medio en tres de los cuatro horizontes. Si investigamos un poco más a fondo nos damos cuenta que el horizonte para el cual éste método no es mejor que el medio es el 12, como se muestra en el Apéndice B, Cuadro B.4. Adicionalmente, en este mismo cuadro podemos ver que los ranqueos promedios para este horizonte son más parecidos entre los distintos métodos de estimación; seis de siete métodos tienen desempeño significativamente igual al medio (NCH tiene peor desempeño. Vea la Figura A.3 del Apéndice A).

Este hecho se refleja también en la prueba multicomparativa con el mejor. Para el horizonte 12 se detalla el desvanecimiento de las diferencias estadísti-

Método	Peor	Mejor	Igual
OLS	0	2	2
RROB	0	3	1
REGT	0	0	4
SVM	0	2	2
NN	0	0	4
GXSM	3	0	1
NCH	4	0	0

Cuadro 4.6: Prueba Multicomparativa con la Media. Número de horizontes para los cuales cada método tiene peor, mejor e igual desempeño que el promedio.

cas entre los mejores y peores métodos, principalmente como consecuencia del empeoramiento del desempeño del mejor método hasta ese momento (RROB). En este horizonte sólo un método es significativamente peor que el mejor: el NCH (vea Figura A.2 del Apéndice A).

Los ranqueos promedios similares para el horizonte 12 son un indicio de que cada método podría tener un buen desempeño para distintos subconjuntos de los datos, en horizontes similares (i.e. de largo plazo).

Los resultados detallados pueden observarse en el Apéndice B y los gráficos relacionados en el Apéndice A.

Desempeño Relativo de cada Método

También nos es de interés verificar el desempeño relativo de cada método con respecto a cada uno de los otros. Esto lo hacemos utilizando como estadístico de prueba el *binomial* con $p = 0{,}5$ y probabilidad de cometer error tipo 1 en la prueba de hipótesis de $\alpha = 0{,}05$. El Cuadro 4.7 muestra el número de métodos que tienen un desempeño significativamente más bajo que los respectivos métodos allí mostrados; utilizando como medida de pre-

Horizonte	1	3	6	12
OLS	4	4	3	1
RROB	5	5	6	3
REGT	2	2	2	2
SVM	4	5	3	3
NN	2	2	2	2
GXSM	0	0	0	0
NCH	0	0	0	0

Cuadro 4.7: Desempeño Relativo de cada Método: Prueba Binomial. Número de métodos que tienen un desempeño significativamente más pobre que el nombrado.

cisión relativa el número de veces que el método i tiene mejor desempeño que el método j. Los resultados indican cierta estabilidad en las posiciones relativas de cada método en función del horizonte de pronóstico. Los métodos OLS, RROB y SVM parecen tener una baja importante en su desempeño en la medida en que el horizonte se hace mayor.

El método RROB domina completamente en el mediano plazo ($h = 6$) y también en términos globales. Le siguen los métodos SVM, OLS, luego REGT, NN y en los dos últimos lugares, GXSM y NCH.

4.3. Otras Comparaciones

Vistos los resultados globales, nos ocupamos también de revisar el comportamiento de los métodos de acuerdo a las categorías generales a las que pertenecen las series de tiempo. En particular, queremos investigar si los métodos presentan resultados parecidos a los globales, considerando cada categoría de datos por separado.

El Cuadro 4.8 nos da información para tratar este asunto. Resulta revelador saber que en términos descriptivos el orden obtenido para los resultados

	Categorías de Series de Tiempo								
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
OLS	0.119	0.032	0.118	0.006	0.294	0.009	0.046	0.033	0.057
RROB	0.115	0.029	0.116	0.006	0.284	0.009	0.045	0.031	0.055
REGT	0.126	0.035	0.126	0.006	0.280	0.009	0.054	0.041	0.057
SVM	0.118	0.028	0.116	0.005	0.278	0.010	0.049	0.036	0.058
NN	0.140	0.030	0.122	0.007	0.283	0.012	0.050	0.044	0.085
GXSM	0.111	0.031	0.149	0.079	0.476	0.072	0.066	0.031	0.127
NCH	0.109	0.027	0.156	0.056	0.400	0.068	0.072	0.027	0.097

Cuadro 4.8: SMAPEs por Categoría de Datos. (a) índices de producción, (b) producción física, (c) ventas, (d) índices de precios, (e) tasas de interés, (f) agregados monetarios, (g) ingresos y gastos públicos, (h) variables demográficas, (i) otras. Cada intersección entre fila y columna recoge el promedio de los horizontes 1, 3, 6, 12. para el método y categoría de datos, respectivo.

globales no necesariamente se mantiene cuando vemos los datos por categorías. Esto sugiere que alguna(s) categoría(s) podría(n) estar influyendo marcadamente sobre los resultados globales.

Como se mencionó con anterioridad, se materializa la posibilidad de que métodos que han quedado mal parados en las comparaciones globales, presenten resultados competitivos para un subconjunto de los datos. Los métodos GXSM y NCH destacan en las categorías (a), (b) y (h), donde registran los errores más pequeños, aunque muy cercanos todos unos a otros. Errores relativamente altos en las categorías (d), (e) y (f) le han valido a estos dos métodos, con toda seguridad, los últimos lugares en las comparaciones globales.

En relación al horizonte, como se espera, en general los errores aumentan con su incremento para todas las categorías. El Cuadro B.5 del Apéndice B tiene la información detallada.

	Categorías de Series de Tiempo								
H.	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
1	0.006	0.009	0.002	0.000	0.000	0.000	0.000	0.000	0.000
3	0.464*	0.025	0.006	0.000	0.000	0.000	0.000	0.000	0.030
6	0.381*	0.349*	0.676*	0.000	0.000	0.000	0.000	0.002	0.667*
12	0.021	0.153*	0.137*	0.000	0.005	0.000	0.005	0.881*	0.182*

Cuadro 4.9: Prueba Friedman de Igualdad Conjunta de Ranqueos por Categoría de Datos. Valores *p*. (a) índices de producción, (b) producción física, (c) ventas, (d) índices de precios. (e) tasas de interés, (f) agregados monetarios, (g) ingresos y gastos públicos, (h) variables demográficas, (i) otras. H.: horizonte. Los * se refieren a valores no significativos al 5 %.

En relación a pruebas estadísticas para estas comparaciones, la prueba de Friedman no nos permite rechazar la hipótesis nula de aleatoriedad conjunta de los ranqueos (al 5 %) para el grupo (a), horizontes 3 y 6; grupo (b), horizontes 6 y 12; grupo (c), horizontes 6 y 12; grupo (h), horizonte 12 y grupo (i), horizontes 6 y 12. Los detalles pueden verse en el Cuadro 4.9. Esto nos lleva a concluir que para algunos grupos de datos, y para algunos horizontes, no hay distinción estadística entre la precisión de los resultados. Sin embargo, para el horizonte 1, siempre hay diferencias significativas.

El Cuadro 4.10 muestra el desempeño relativo de cada método utilizando la prueba binomial. Los métodos RROB y SVM tienden a dominar, en general; la SVM resulta mejor que el RROB en las categorías (b), (f), (g), y el método RROB mejor que la SVM en las categorías (a), (d), (e) y (h). Le siguen a estos dos métodos, OLS, REGT y NN, GXSM y NCH. El método REGT presenta mejores resultados que la NN en las categorías (d), (f) e (i), mientras que sucede lo contrario con las categorías (b), (e), (g) y (h). De nuevo, se observa un desempeño relativamente bueno por parte de GXSM en la categoría (a). El Cuadro B.6 del Apéndice B presenta la información

	Categorías de Series de Tiempo								
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
OLS	2	1	7	11	7	12	11	9	4
RROB	6	7	6	18	13	10	13	11	4
REGT	0	0	3	10	8	12	4	1	5
SVM	3	10	6	13	7	14	14	7	4
NN	0	1	3	8	10	8	9	2	1
GXSM	3	0	0	0	0	1	2	2	0
NCH	0	0	0	4	3	1	0	0	0

Cuadro 4.10: Resumen del Desempeño Relativo de cada Método: Prueba Binomial por Categoría de Datos. Número total de métodos que tienen un desempeño significativamente más pobre que el nombrado, considerando todos los horizontes. (a) índices de producción, (b) producción física, (c) ventas, (d) índices de precios, (e) tasas de interés, (f) agregados monetarios, (g) ingresos y gastos públicos, (h) variables demográficas, (i) otras.

detallada para todos los horizontes.

4.4. Otros Resultados

Esta sección describe brevemente lo que hemos considerado conveniente tildar de *series problemáticas*. El cálculo de los errores SMAPE para cada una de las series, utilizando cada uno de los métodos, arrojó como resultado un hecho interesante: existen grupos de series de tiempo para los cuales todos los métodos reportaron errores más altos que otros grupos. Como el SMAPE es una medida relativa que mide el error como porcentaje, se hace fácil comparar entre las distintas series de tiempo. Los resultados sugieren que hay determinados subconjuntos de los datos que son más difíciles de pronosticar que otros.

La Figura 4.2 nos da una idea de lo que ocurre con las series y sus errores.

En ella se presentan los SMAPEs de cada serie, promediados por todos los horizontes, y luego apilados (sumados) para formar el gráfico. Las series rodeadas por una elipse son aquellas que presentan sumas relativamente altas, y por tanto, han sido pronosticadas con errores más grandes.

Para formalizar la noción de serie problemática que se muestra en la Figura 4.2, debemos definirla en términos concretos. Haremos esa definición describiendo el proceso que hemos utilizado para categorizar a una serie como problemática o no.

En primer lugar, promediamos el error SMAPE de cada serie (i) y para cada uno de los métodos de estimación (j), por todos los horizontes ($SMAPEp_{i,j}$). El segundo paso es calcular la *mediana* y la *desviación absoluta mediana* que corresponden a cada método de estimación ($Med_j = Med(SMAPEp_{.,j}$ y $MAD_j = MAD(SMAPEp_{.,j})$, respectivamente). El tercer paso es promediar las Med_j y las MAD_j; estos promedios ($Medp$ y $MADp$) servirán de criterio para la definición. Por último, construimos la regla de decisión que necesitamos:

Para cada serie (i), si para por lo menos cuatro metodos (j), el $SMAPEp_{i,.} \geq Medp + 3 * MADp$, entonces la definimos como ***serie problemática.***

El resultado del procedimiento anterior puede verse gráficamente en la Figura 4.3. Las series problemáticas se corresponden con 4 (de 11) índices de producción, 4 (de 20) ventas, 36 (de 37) tasas de interés bancarias, y 3 (de 46) series demográficas. Estas series, en particular las de tasas de interés, presentan una volatilidad relativamente alta, lo cual podría explicar parte de estos resultados. Los detalles pueden verse en el Cuadro B.8 del Apéndice B.

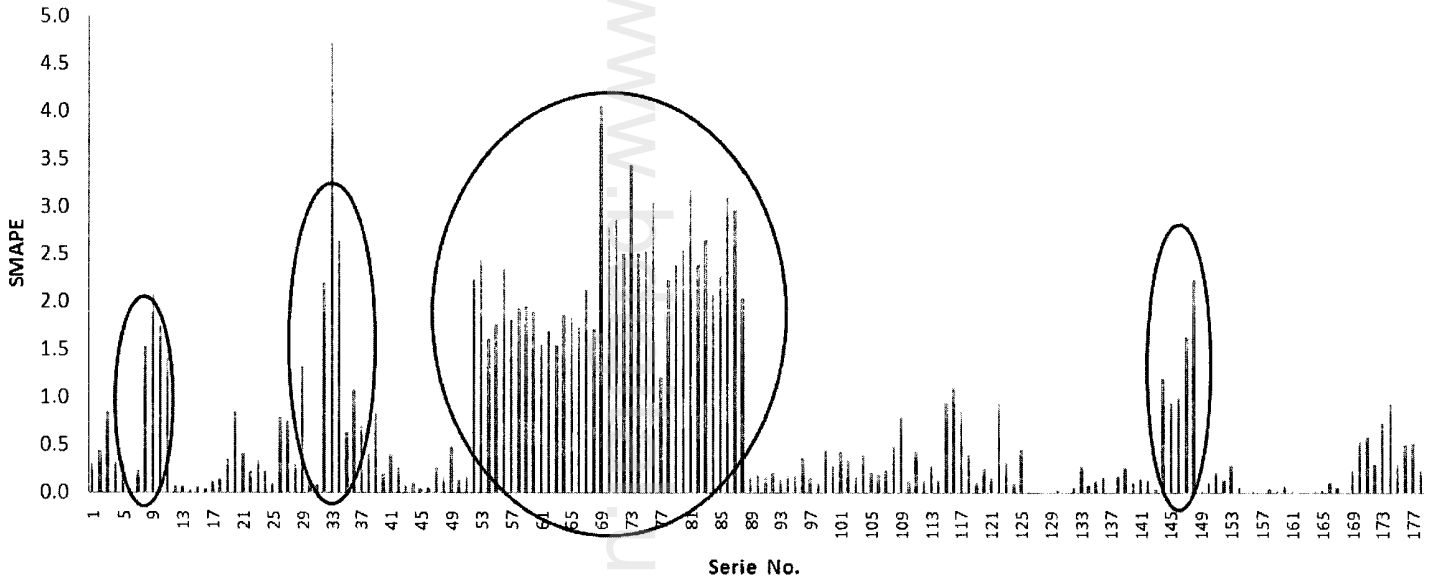


Figura 4.2: Series Problemáticas. El eje x muestra cada una de las 178 series, mientras que en el eje y se muestra el error SMAPE promedio de los cuatro horizontes estimados. Para cada serie, los SMAPEs promedio han sido apilados en una barra por serie. Los elipses rojos muestran posibles zonas dificultosas. Los distintos colores corresponden a los errores SMAPE de los siete métodos de estimación.

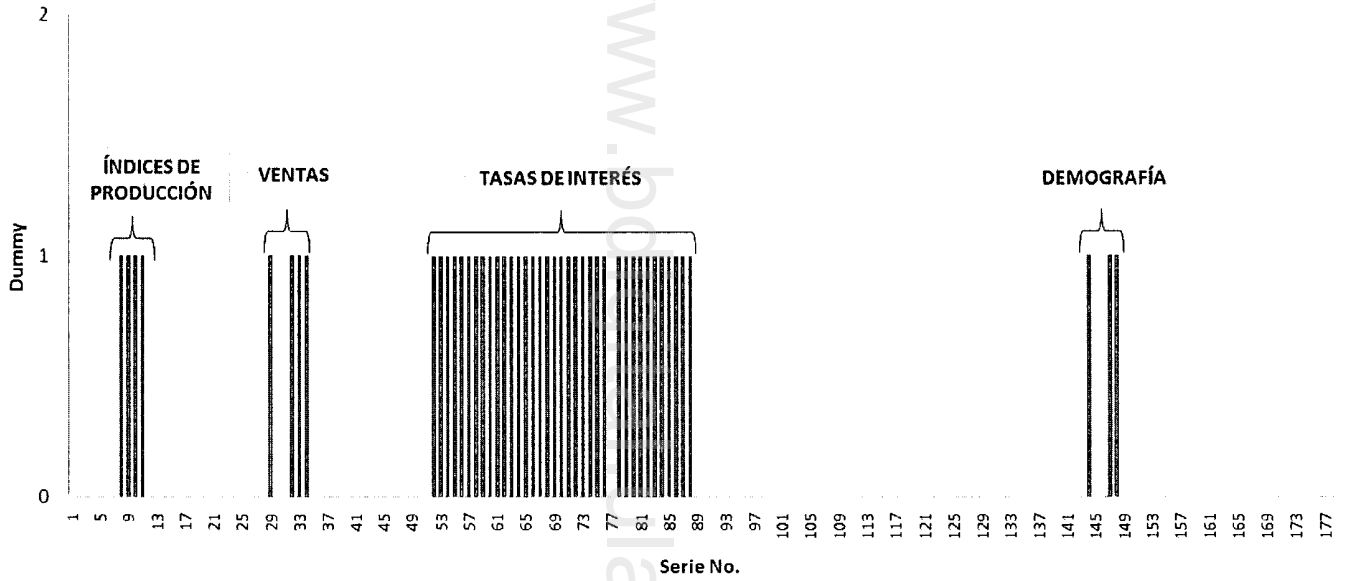


Figura 4.3: Series Problemáticas. El eje x muestra cada una de las 178 series, mientras que en el eje y se muestra una variable categórica (dummy) que toma valor 1 si la serie ha sido definida como problemática y valor 0 en caso contrario.

Conclusiones y Comentarios Finales

En este trabajo hemos utilizado una serie de métodos para estimar modelos univariantes con datos venezolanos. El objetivo ha sido elaborar pronósticos fuera de muestra con cada uno de ellos y comparar los resultados obtenidos. Hemos procurado el uso de pruebas estadísticas para verificar la significancia de las diferencias observadas. Los resultados y conclusiones presentados son válidos para las series, métodos y diseños propuestos acá. En ningún caso deben interprestarse como resultados definitivos de aplicación general. Este tabajo agrega a la literatura al considerar la estimación de pronósticos con métodos no convencionales[1] para un nuevo grupo de series de tiempo.

En primer lugar, sí existen diferencias significativas entre los distintos métodos de estimación. Resulta conveniente dividir a los siete métodos, en tres grupos de desempeños estadísticamente similares. El primer grupo conformado por los métodos RROB, SVM y OLS; representan el grupo de mejor desempeño. El grupo con el segundo mejor desempeño lo conforman REGT y NN y por último, un tercer grupo que contiene a GXSM y NCH. Esta clasificación, sin embargo, se hace borrosa en la medida en que el horizonte de pronóstico se hace más grande; para éste escenario los resultados favorecen

[1] Especialmente en la predicción de series de tiempo económicas.

más la opción de dos grandes grupos: I) RROB, SVM, REGT, NN, OLS; II) GXSM, NCH.

Los métodos más simples (i.e. NCH. GXSM), en general, tienen un desempeño pobre. Considerando los distintos tipos de series, muestran una mejora al punto de igualarse con el resto de los métodos de estimación para las categorías (a) índices de producción, (b) producción física, (c) ventas, (h) variables demográficas, (i) otras, en horizontes medios y largos.

Para el horizonte más corto posible (horizonte = 1), se observan diferencias significativas entre los distintos métodos tomando en cuenta las distintas categorías de datos. Las SVM presentan los mejores resultados con diferencias estadísticamente significativas para las categorías (b) producción física, (f) agregados monetarios y (g) ingresos y gastos públicos. Este resultado sugiere que la importancia de las no linealidades varía entre las series estudiadas y que conviene seguir probando este método (y otros no lineales) con series de ese tipo. La RROB parece ser la mejor opción de estimación para el resto de las categorías. Las NN presentan resultados intermedios en la clasificación, superando a los métodos más simples, y en el mejor de los casos, sobrepasando por muy poco el desempeño del REGT.

Los resultados indican además que los procedimientos robustos son una buena alternativa y valdría la pena *ajustar* algunos de los métodos como NN y SVM para incorporar esta característica en sus procesos de estimación. Sacando de las comparaciones al RROB, el modelo no lineal SVM muestra el mejor desempeño.

Existen muchos resultados relacionados con comparaciones que involucran NN, y pueden resultar positivos, negativos y mixtos; probablemente por la gran gama de especificaciones a las que están sujetas (para un estudio extenso vea Zhang et al. 1998). En lo que concierne a las SVMs, no se consigue la misma cantidad de estudios y muchos de ellos tienen propósitos ilustrativos más que comparativos. Por otro lado, luego de la revisión bi-

bliográfica podemos pensar que el uso de REGT en el pronóstico de series de tiempo tampoco parece ser una práctica muy común.

Adicionalmente, basados en los errores SMAPE y sencillos cálculos utilizando una medida de tendencia central como la mediana, y la desviación absoluta mediana, hemos identificado un conjunto de *series problemáticas*. Las tasas de interés bancarias merecen especial atención en lo que se refiere a este particular.

En cuanto al diseño general del estudio conviene mencionar algunas posibles mejoras que pueden introducirse. El primer punto concierne la naturaleza estática del experimento. Como se mencionó en la sección 3.1, se ha utilizado un esquema de muestra fija para la estimación de los parámetros. El uso de una ventana móvil o expansiva se corresponde más con la realidad del pronosticador: cada vez que se hace disponible un nuevo dato, este se incorpora al proceso de estimación. Sin embargo el costo computacional de un experimento como ese, con tal cantidad de series de tiempo, los métodos aquí propuestos y varios horizontes, es extremadamente alto.

En segundo lugar, la estimación de más horizontes permitiría estudiar la evolución de los errores con respecto a esta variable (i.e. verificar si es lineal o no). Esta mejora también enfrenta el problema de limitaciones computacionales.

Un tercer punto tiene que ver con la forma en que se especifica el número de rezagos en los modelos. Ésto aplica especialmente para los métodos SVM, NN y REGT, donde aún no se alcanza un concenso sobre cuál será la mejor forma de hacerlo. Aquí decidimos utilizar el mismo número de rezagos óptimos basados en el criterio Akaike (Akaike 1974), que se sugiere para una regresión lineal múltiple estimada por OLS. Pensamos que esto aportaría consistencia en términos de comparabilidad de los resultados; introducir distintas variables explicativas para los distintos métodos (y series) no nos permitiría distinguir entre los efectos del método y los de la especifi-

cación. Pero por supuesto, hay otros posibles criterios. Por ejemplo, calcular un criterio Akaike para cada método de estimación (y serie) aún cuando las especificaciones pudiesen terminar siendo distintas; o por ejemplo, apelar a algún procedimieto como la validación cruzada (Kohavi 1995, Cawley 2006). Dorta (2007) menciona una gama de criterios estadísticos de selección para el caso líneal, pero el caso no lineal se encuentra relativamente desatendido en la literatura.

Una cuarta mejora está relacionada con la función de error. Usamos una sóla función, el SMAPE, pero es posible incorporar otras medidas de error y verificar si los resultados se mantienen. La introducción de otros errores no utilizados frecuentemente, como los no simétricos, constituye una indiscutible mejora. Al final, las funciones de error se deben corresponder con los costos que enfrenta el decisor. Algunos estudios sobre este tema los ofrecen Christoffersen y Diebold (1996), Christoffersen y Diebold (1997) y Pope et al. (2006).

Aún otra mejora es el cálculo de intervalos de confianza o densidades para los pronósticos puntales. Esto implica asignar probabilidades a los resultados y abrir la mente del decisor a un posible rango de resultados en lugar de un dato en particular (vea por ejemplo Giordano et al. 2007, Boero y Marrocu 2001, Clements y Smith 1998).

Por último, se pueden combinar distintas predicciones para obtener un resultado combinado que puede incluirse en las comparaciones de los métodos. Clemen (1989), Aiolfi y Timmermann (2006), Deutsch et al. (1994) son tres ejemplos que tratan el tema.

Todas estas consideraciones son posibles extensiones para próximos y mejores estudios.

Bibliografía

Ahmed, N., A. Atiya, N. El Gayar, y H. El-Shishiny. An empirical comparison of machine learning models for time series forecasting, 2007. URL `http://alumnus.caltech.edu/~amir/compar-final-9-2007.pdf`. 57

Aiolfi, M. y A. Timmermann. Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1-2):31–53, 2006. 80

Akaike. H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. 42, 79

Boero, G. y E. Marrocu. Evaluating non-linear models on point and interval forecasts: an application with exchange rate returns. Working Paper CRENoS 200110, Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia, 2001. URL `http://ideas.repec.org/p/cns/cnscwp/200110.html`. 80

Box, G. y G. Jenkins. *Time Series Analysis, Forecasting and Control.* Holden Day, San Francisco, 1970. 7

Breiman, L., J. Freidman, R. Olshen, y C. Stone. *Classification and regression trees.* Wadsworth, 1984. 13, 47

Burns, T., A. Britton, R. Quirk, P. Mathias, y J. Mason. The Interpretation and Use of Economic Predictions [and Discussion]. *Proceedings of the*

Royal Society of London. Series A, Mathematical and Physical Sciences (1934-1990), 407(1832):103–125, 1986. 3

Cawley, G. Leave-one-out cross-validation based model selection criteria for weighted ls-svms. En *Proceedings of the International Joint Conference on Neural Networks*, páginas 2970–2977, 2006. 80

Chalimourda, A., B. Schölkopf, y A. Smola. Experimentally optimal ν in support vector regression for different noise models and parameter settings. *Neural Networks*, 17(1):127–141, 2004. 56

Chatfield. C. *The Analysis of Time Series: An Introduction.* Chapman & Hall/CRC, 2004. 4, 9, 40, 41, 44

Chevillon, G. Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4):746–785, 09 2007. URL `http://ideas.repec.org/a/bla/jecsur/v21y2007i4p746-785.html`. 37

Christoffersen, P. F. y F. X. Diebold. Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics*, 11(5):561–71, Sept.-Oct. 1996. URL `http://ideas.repec.org/a/jae/japmet/v11y1996i5p561-71.html`. 80

Christoffersen, P. F. y F. X. Diebold. Optimal prediction under asymmetric loss. Working Papers 97-11, Federal Reserve Bank of Philadelphia, 1997. URL `http://ideas.repec.org/p/fip/fedpwp/97-11.html`. 80

Chu, W., S. Keerthi, y C. Ong. Bayesian support vector regression using a unified loss function. *Neural Networks, IEEE Transactions on Neural Networks*, 15(1):29–44, 2004. 28

Clemen, R. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583, 1989. 80

Clements, D. y M. Hendry. *A Companion to Economic Forecasting*, capítulo An Overview of Economic Forecasting, páginas 1–18. Blackwell Publishers Ltd., 2002. 2

Clements, M. y D. Hendry. Evaluating a Model by Forecast Performance. *Oxford Bulletin of Economics & Statistics*, 67(s1):931–956, 2005. 37

Clements, M. y J. Smith. Evaluating the forecast of densities of linear and non-linear models: Applications to output growth and unemployment. The Warwick Economics Research Paper Series (TWERPS) 509, University of Warwick, Department of Economics, 1998. URL `http://ideas.repec.org/p/wrk/warwec/509.html`. 80

Cooper, J. y C. Nelson. The Ex-ante Prediction Performance of the St. Louis and FRB-MIT-Penn Econometric Models and Some Results on Composite Predictors. *Journal of Money, Credit and Banking*, 7(1):1–32, 1975. 2

Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, 1989. 24

Dallal, G. Multiple comparison procedures, 2001. URL `http://www.tufts.edu/~gdallal/mc.htm`. 59

De Gooijer, J. y D. Zerom. Kernel-based multistep-ahead predictions of the US short-term interest rate. *Journal of Forecasting*, 19(4):335–353, 2000. 59

Demuth, H., M. Beale, y M. Hagan. Neural network toolbox 6, user guide. Reporte técnico, The Math Works Inc., 2008. 52, 54

Deutsch, M., C. W. J. Granger, y T. Terasvirta. The combination of forecasts using changing weights. *International Journal of Forecasting*, 10

(1):47–57, June 1994. URL http://ideas.repec.org/a/eee/intfor/v10y1994i1p47-57.html. 80

Dhrymes, P., E. Howrey, S. Hymans, J. Kmenta, E. Leamer, R. Quandt, J. Ramsey, H. Shapiro, y V. Zarnowitz. Criteria for evaluation of econometric models. *Annals of Economic and Social Measurement*, 1(3):291–324, 1972. 2

Dorta, M. Selección de variables por muestreo para regresión múltiple basado en el risk inflation criterion. Serie Documentos de Trabajo 77, Banco Central de Venezuela, 2007. 80

Duan, K., S. Keerthi, y A. Poo. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing*, 51(4):41–59, 2003. 56

Dyn, N. Interpolation of scattered data by radial functions. *Topics in multivariate approximation. Academic Press, New York*, 216, 1987. 35

Elliott, G., T. Rothenberg, y H. James. Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813–836, 1996. 41

Fleitas, C., M. Mirabal, E. Roo, y G. Sánchez. Modelo de simulación de programación financiera. Serie Documentos de Trabajo 35, Banco Central de Venezuela, 2002. 5

Foresee, F. y M. Hagan. Gauss-Newton Approximation to Bayesian Learning. *Proceedings of the International Joint Conference on Neural Networks*, 3, 1997. 54

Fox, J. *An R and S-Plus Companion to Applied Regression.* Sage Publications, 2002. 11, 46

Franses, P. y F. Kleibergen. Unit roots in the Nelson-Plosser data: Do they matter for forecasting? *International Journal of Forecasting*, 12(2):283–288, 1996. 41

Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701, 1937. 58

Friedman, M. A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 34(205):109–109, 1939. 58

Funahashi, K. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183–192, 1989. 24

Geisser, S. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350):320–328, 1975. 49

Giordano, F., M. La Rocca, y C. Perna. Forecasting nonlinear time series with neural network sieve bootstrap. *Computational Statistics & Data Analysis*, 51(8):3871–3884, May 2007. URL `http://ideas.repec.org/a/eee/csdana/v51y2007i8p3871-3884.html`. 80

Gómez, V. y A. Maravall. Programs tramo and seats: Instructions for the user. Reporte técnico, Banco de España, 1997. URL `http://www.bde.es/servicio/software/tramo/manualdos.pdf`. 40

Gómez, V. y A. Maravall. *A Course in Time Series Analysis*, capítulo 8, páginas 202–247. Wiley New York, 2001. URL `http://www.bde.es/servicio/software/tramo/sasex.pdf`. 40

Gonzalez, S. Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models. Working paper 2000-07, Department of Finance, Government of Canada, 2000. 43

Guerra, J., G. Sánchez, y B. Reyes. Modelos de series de tiempo para predecir la inflación en Venezuela. Serie Documentos de Trabajo 13, Banco Central de Venezuela, Diciembre 1997. 5

Gujarati, D. *Basic econometrics*. McGraw-Hill, 2004. 40, 41

Hagan, M., H. Demuth, y M. Beale. *Neural Network Design*. PWS Publishing, 1996. 24

Hagiwara, K. y K. Kuno. Regularization learning and early stopping in linear networks. *IEEE-INNS-ENNS International Joint Conference on Neural Networks*, 04:4511, 2000. ISSN 1098-7576. URL `http://www2.computer.org/portal/web/csdl/doi/10.1109/IJCNN.2000.860822`. 51

Halperin, M., S. Greenhouse, J. Cornfield, y J. Zalokar. Tables of percentage points for the studentized maximum absolute deviate in normal samples. *Journal of the American Statistical Association*, 50(269):185–195, 1955. 61

Harter, H. Tables of range and studentized range. *Annals of Mathematical Statistics*, 31:1122–1147, 1960. 60

Heiberger, R. y R. Becker. Design of an S function for robust regression using iteratively reweighted least squares. *Journal of Computational and Graphical Statistics*, 1(3):181–196, 1992. 13

Hendry, D. Unpredictability and the Foundations of Economic Forecasting. *Economics Working Papers, Nuffield College, Oxford University*, 2004. 2

Hendry, D. y M. Clements. Economic forecasting: some lessons from recent research. *Economic Modelling*, 20(2):301–329, 2003. 2

Herbrich, R., M. Keilbach, T. Graepel, P. Bollmann-Sdorra, y K. Obermayer. Neural networks in economics: background, applications and new

developments. En *Advances in Computational Economics: Computational Techniques for Modelling Learning in Economics*, páginas 169–196. Kluwer Academics, 2000. 25

Hogg, R. Statistical robustness: One view of its use in applications today. *The American Statistician*, 33(3):108–115, 1979. 11

Hollander, M. y D. Wolfe. *Nonparametric statistical methods.* Wiley New York, 1973. 59

Hollander, M. y D. Wolfe. *Nonparametric statistical methods.* Wiley New York. segunda edición, 1999. 60

Hornik, K., M. Stinchcombe, y H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. 24

Hsu, C., C. Chang, C. Lin, et al. A practical guide to support vector classification. Reporte técnico, National Taiwan University, 2003. 54, 55

Hsu, J. *Multiple Comparisons: Theory and Methods.* Chapman & Hall, 1996. 60

Huber, P. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101, 1964. 11

Jenkins, G. y D. Watts. *Spectral analysis and its applications.* Holden-Day San Francisco, 1968. 45

Karush, W. Minima of functions of several variables with inequalities as side constraints. *Department of Mathematics. University of Chicago*, 1939. 31

Kasabov, N. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering.* Bradford Books, 1996. 16, 17, 19, 24, 25

Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. En *International Joint Conference on Artificial Intelligence*, páginas 1137–1143. Morgan Kaufmann, 1995. 80

Kolb, R. y H. Stekler. How well do analysts forecast interest rates? *Journal of Forecasting*, 15(5):385–394, 1996. 59

Koning, A., P. Franses, M. Hibon, y H. Stekler. The M3 competition: Statistical tests of the results. *International Journal of Forecasting*, 21(3): 397–409, 2005. 57, 58, 59, 60

Kuan, C.-M. y H. White. Artificial neural networks: an econometric perspective. *Econometric Reviews*, 13(1):1–91, 1994. URL `http://ideas.repec.org/a/taf/emetrv/v13y1994i1p1-91.html`. 53

Kuhn, H. y A. Tucker. Nonlinear programming. *Proc. Second Berkeley Sympos. on Math. Statist. Probab*, páginas 481–492, 1951. 31

Levenberg, K. A method for the solution of certain nonlinear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944. 53

Lewis. R. An Introduction to Classification and Regression Tree (CART) Analysis. *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California*, páginas 1–14, 2000. 13

MacKay, D. Bayesian Interpolation. *Neural Computation*, 4(3):415–447, 1992. 52

Makridakis, S. y M. Hibon. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4):451–476, 2000. 57

Mann. H. y A. Wald. On the statistical treatment of linear stochastic differential equations. *Econometrica*, 11:173–200, 1943. 44

Marcellino, M., J. Stock, y M. Watson. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135(1-2):499–526, 2006. 37

Marget, A. Morgenstern on the Methodology of Economic Forecasting. *The Journal of Political Economy*, 37(3):312, 1929. 4

Marquardt, D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441, 1963. 54

McCulloch, W. y W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 5(4):115–133, 1943. 16, 18

McDonald, B. y W. Thompson. Rank sum multiple comparisons in one and two way classifications. *Biometrika*, 54:487–497, 1967. 60

McDonald, B. y W. Thompson. Corrections and amendments: Rank sum multiple comparisons in one and two way classifications. *Biometrika*, 59: 699, 1972. 60

McNees, S. The accuracy of macroeconomic forecasts. *Analyzing Modern Business Cycles*, páginas 143–173, 1990. 3

Medeiros, M., T. Teräsvirta, y G. Rech. Building Neural Network Models for Time Series: A Statistical Approach. *Journal of Forecasting*, 25(1):49–75, 2006. 42

Mercer, J. Functions of Positive and Negative Type, and Their Connection with the Theory of Integral Equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Math. or Phys. Character (1896-1934)*, 209(-1):415–446, 1909. 33

Micchelli, C. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2(1):11–22, 1986. 35

Mincer, J. y V. Zarnowitz. *Economic Forecasts and Expectations*, capítulo The Evaluation of Economic Forecasts. NBER, 1969. 2

Moody, J. The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. En *Advances in Neural Information Processing Systems*, volume 4, páginas 847–854, 1992. 42

Moshiri, S. y N. Cameron. Neural Network vs. Econometric Models in Forecasting Inflation. *Journal of Forecasting*, 19:210–217, 2000. 43

Murata, N., S. Yoshizawa, y S. Amari. Learning curves, model selection and complexity of neural networks. En Hanson, S. J., J. D. Cowan, y C. L. Giles, editores, *Advances in Neural Information Processing Systems*, volume 5, páginas 607–614. Morgan Kaufmann, San Mateo, CA, 1993. URL `citeseer.ist.psu.edu/murata93learning.html`. 42

Ormerod, P. *Butterfly Economics: A New General Theory of Social and Economic Behavior*. Basic Books, 2001. 38

Pindyck, R. y D. Rubinfeld. *Econometric models and economic forecasts*. McGraw-Hill New York, 1998. 10

Pope, P., D. Peel, y M. Clatworthy. Are analysts loss functions asymmetric? Working Papers 003094, Lancaster University Management School, Economics Department, 2006. URL `http://ideas.repec.org/p/lan/wpaper/003094.html`. 80

Quinlan, J. Simplifying decision trees. *International Journal of Man-Machine Studies*, 27(3):221–234, 1987. 47

Riesz, F. y B. Szőkefalvi-Nagy. *Functional Analysis.* Dover Publications, 1990. 33

Rumelhart, D., G. Hintont, y R. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. 23

Said, S. y D. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984. 41

Sampson, M. The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models. *Journal of Applied Econometrics*, 6(1):67–76, Enero-Marzo 1991. URL `http://ideas.repec.org/a/jae/japmet/v6y1991i1p67-76.html`. 63

Sarle, W. Stopped training and other remedies for overfitting. *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, 352:360, 1995. 51, 52

Sarle, W. Neural Network FAQ, periodic posting to the Usenet newsgroup comp. ai. neural-nets, 1997. URL `ftp://ftp.sas.com/pub/neural/FAQ.html`. 53, 54

Scholkopf, B., P. Bartlett, A. Smola, y R. Williamson. Support Vector Regression with Automatic Accuracy Control. *Proceedings of the 8th International Conference on Artificial Neural Networks*, páginas 111–116. 1998. 34

Scholkopf, B., A. Smola, R. Williamson, y P. Bartlett. New Support Vector Algorithms. *Neural Computation*, 12(5):1207–1245, 2000. 27

Sindelar, R. y R. Babuska. Input selection for nonlinear regression models. *IEEE Transactions on Fuzzy Systems*, 12(5):688–696, 2004. 42

Smola, A. y B. Scholkopf. A tutorial on support vector regression. Reporte Técnico NC2-TR-1998-030, NeuroCOLT2, 1998. 27, 28, 30, 34, 35, 54

Stekler, H. Macroeconomic forecast evaluation techniques. *International Journal of Forecasting*, 7(3):375–384, 1991. 59

Stekler, H. The M3-competition: the need for formal statistical tests. *International Journal of Forecasting*, 17(4):576–577, 2001. 59

Stevens, G. Economic forecasting and its role in making monetary policy. *Reserve Bank of Australia Bulletin*, páginas 1–9, September 1999. 5

Stevens, G. Better Than A Coin Toss? The Thankless Task of Economic Forecasting. *Reserve Bank of Australia Bulletin*, páginas 6–14, September 2004. 4, 5

Stock, J. y M. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *Cointegration, Causality and Forecasting: A Festschrift in honour of Clive WJ Granger*, páginas 1–44, 1999. 37, 43

Stone, M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, 36(2):111–133, 1974. 49

Teräsvirta, T. Forecasting economic variables with nonlinear models. Working Paper Series in Economics and Finance 598, Stockholm School of Economics, May 2005. URL `http://ideas.repec.org/p/hhs/hastef/0598.html`. 37

Theil, H. *Applied Economic Forecasting*. North-Holland Publishing Company, 1966. 2

Tkacz, G. y S. Hu. Forecasting gdp growth using artificial neural networks. Working Papers 99-3, Bank of Canada, 1999. URL `http://ideas.repec.org/p/bca/bocawp/99-3.html`. 53

Vapnik, V. y A. Lerner. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24(6):774–780, 1963. 27

Vapnik, V., S. Golowich, y A. Smola. Support vector method for function approximation, regression estimation, and signal processing. *Cambridge, MA*, páginas 169–184, 1997. 28

Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, 2000. 28, 29

Wahlstrom, S. Comparing forecasts from lstar models and linear ar models. Reporte técnico, Stockholm University, Mathematical Statistics, 2004. 37

Wallis, K. Macroeconomic Forecasting: A Survey. *Economic Journal*, 99 (394):28–61, 1989. 3

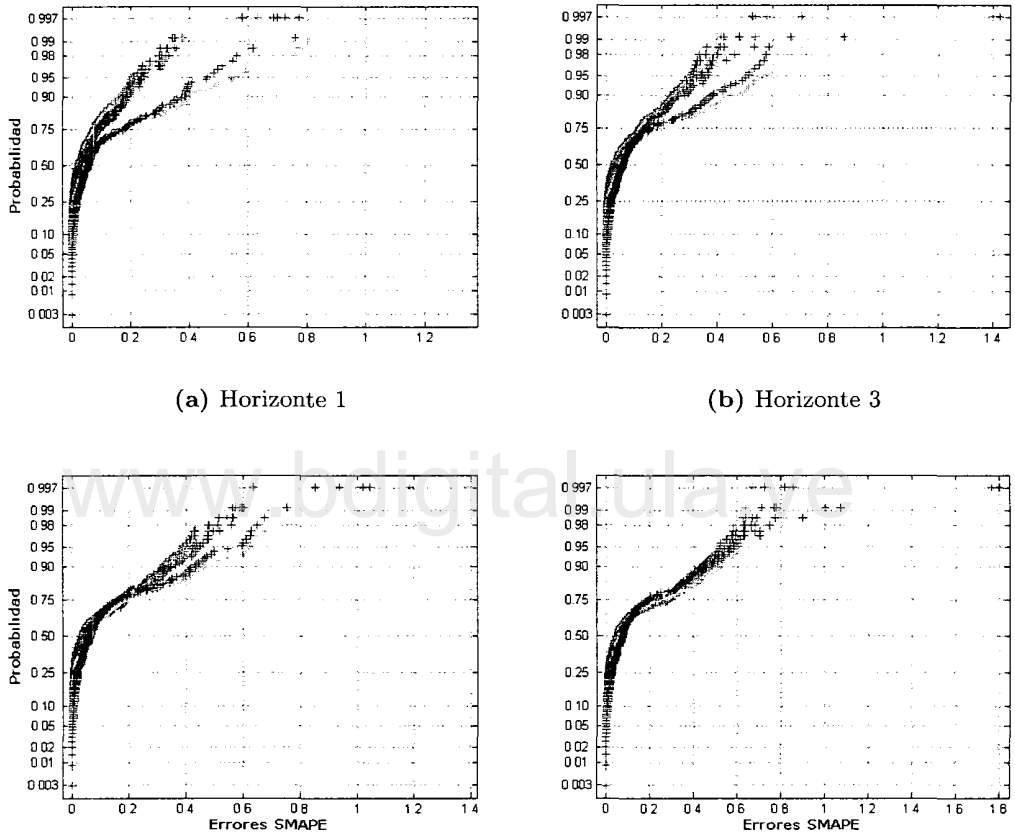
West, K. Tests for Forecast Encompassing When Forecasts Depend on Estimated Regression Parameters. *Journal of Business & Economic Statistics*, 19(1):29, 2001. 37

Yohannes. Y. y P. Webb. Classification and regression trees, CART: A user manual for identifying indicators of vulnerability to famine and chronic food insecurity. Microcomputers in Policy Research Series 3, International Food Policy Research Institute (IFPRI), 1999. 13, 14, 16, 46

Zhang, G., B. Eddy Patuwo, y M. Y. Hu. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14 (1):35–62, 1998. 54, 78

Apéndice A

Figuras



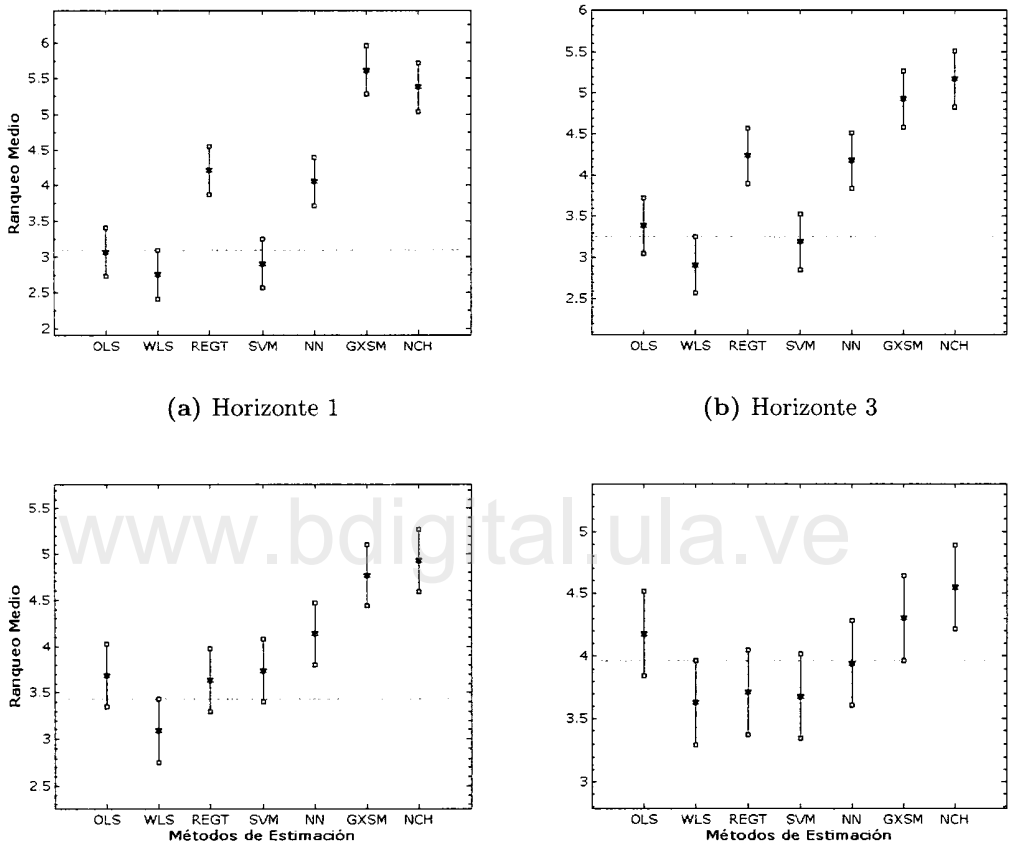
(a) Horizonte 1

(b) Horizonte 3

(c) Horizonte 6

(d) Horizonte 12

Figura A.1: Gráficos de Probabilidad Normal para SMAPEs. Los distintos colores representan cada uno de los siete métodos de estimación.



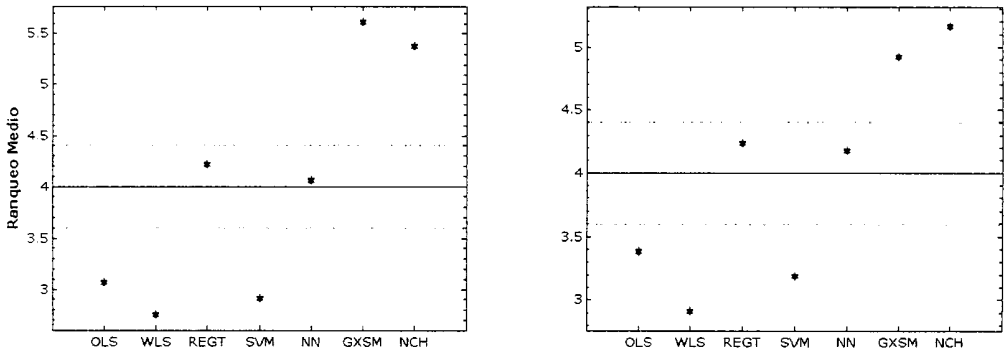
(a) Horizonte 1

(b) Horizonte 3

(c) Horizonte 6

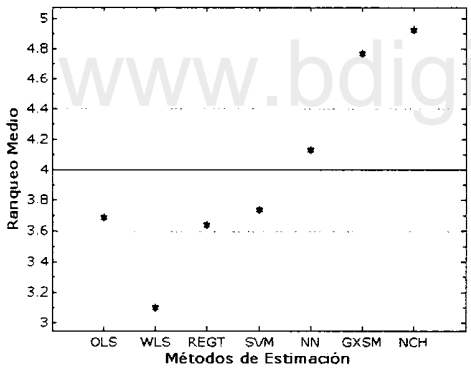
(d) Horizonte 12

Figura A.2: Gráficos de Pruebas Multicomparativa con el Mejor. La línea horizontal punteada representa el límite superior del método con mejor desempeño. Los métodos cuyos rangos no tocan la línea tienen un desempeño (estadístico) significativamente peor que el mejor método.

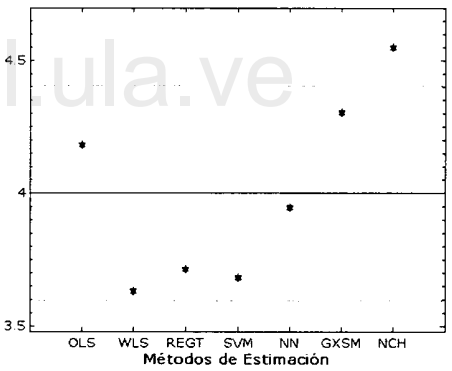


(a) Horizonte 1

(b) Horizonte 3

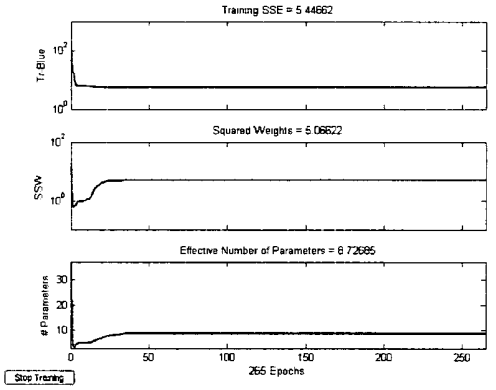


(c) Horizonte 6

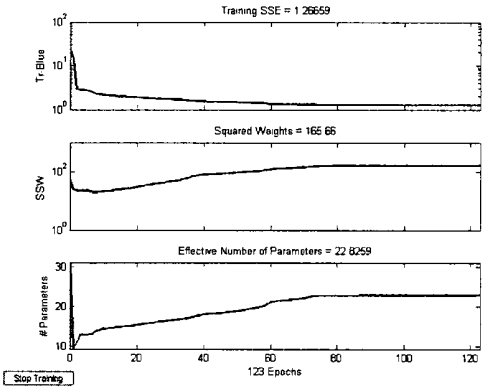


(d) Horizonte 12

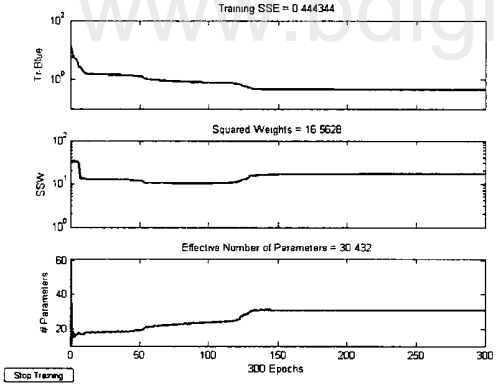
Figura A.3: Gráficos de Pruebas Multicomparativa con la Media. La línea central de trazo oscuro representa el ranqueo medio de todos los métodos. Las líneas punteadas forman un intervalo de confianza indicando cuáles métodos tienen un desempeño (estadístico) significativamente distinto al medio. Los * denotan los ranqueos promedios de cada uno de los métodos.



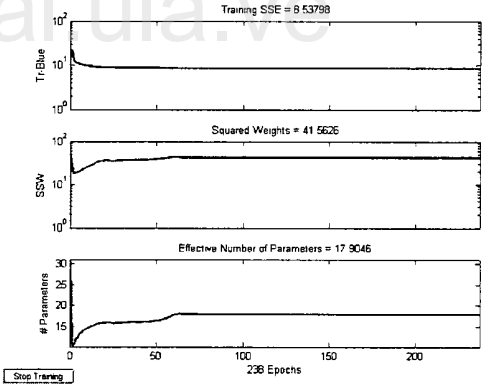
(a) Serie 28, horizonte 1



(b) Serie 15, horizonte 3

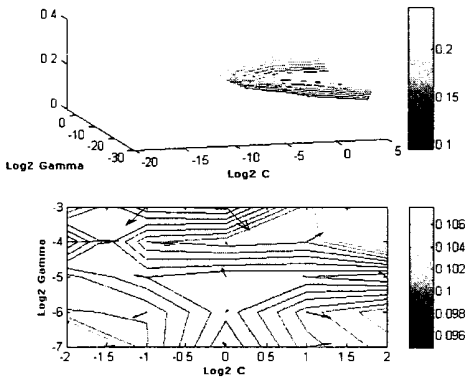


(c) Serie 69, horizonte 6

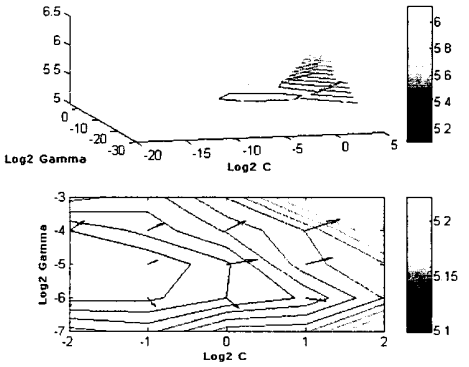


(d) Serie 177, horizonte 12

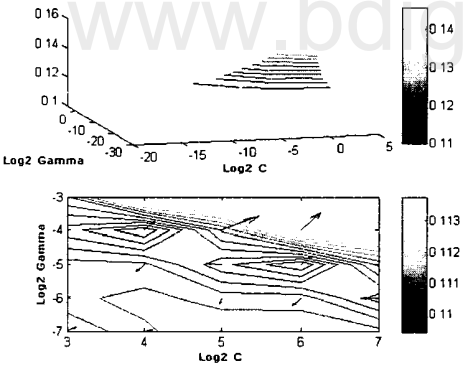
Figura A.4: Entrenamiento de Redes Neuronales con Regularización Bayesiana.



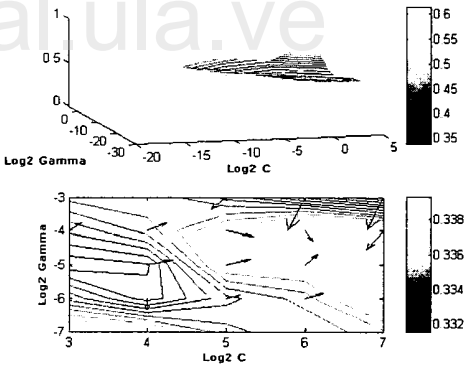
(a) Serie 158, horizonte 1



(b) Serie 170, horizonte 3



(c) Serie 20, horizonte 6



(d) Serie 99, horizonte 12

Figura A.5: Búsqueda de Malla con $v-$validación Cruzada ($v = 10$) para Máquinas de Soporte Vectorial.

Apéndice B

Cuadros

Abreviaturas

1. OLS: mínimos cuadrados ordinarios
2. RROB: regresión robusta
3. REGT: árboles de regresión
4. SVM: máquinas de soporte vectorial
5. SVMR: máquinas de soporte vectorial con datos reescalados
6. NN: redes neuronales
7. NNR: redes neuronales con datos reescalados
8. GXSM: suavización exponencial
9. NCH: modelo de cambio cero

	SMAPE	Percentiles						
	Promedio	5	10	25	50	75	90	95
OLS	0.058	0.001	0.001	0.004	0.024	0.078	0.172	0.213
	0.081	0.001	0.002	0.006	0.029	0.128	0.238	0.314
	0.110	0.001	0.003	0.008	0.037	0.181	0.339	0.390
	0.160	0.001	0.004	0.013	0.052	0.281	0.478	0.541
RROB	0.058	0.001	0.001	0.004	0.023	0.077	0.173	0.210
	0.079	0.001	0.002	0.006	0.029	0.121	0.238	0.309
	0.105	0.001	0.002	0.007	0.035	0.166	0.322	0.389
	0.155	0.001	0.004	0.012	0.049	0.275	0.469	0.563
REGT	0.068	0.001	0.001	0.005	0.030	0.098	0.189	0.241
	0.089	0.001	0.002	0.007	0.036	0.138	0.255	0.340
	0.112	0.001	0.003	0.008	0.039	0.187	0.320	0.428
	0.148	0.001	0.004	0.012	0.053	0.260	0.440	0.589
SVM	0.058	0.001	0.001	0.004	0.023	0.082	0.169	0.206
	0.083	0.001	0.002	0.005	0.028	0.131	0.252	0.307
	0.114	0.001	0.003	0.008	0.038	0.184	0.345	0.395
	0.145	0.001	0.004	0.012	0.053	0.225	0.488	0.532
NN	0.064	0.001	0.001	0.005	0.030	0.091	0.187	0.234
	0.094	0.001	0.002	0.008	0.033	0.141	0.277	0.365
	0.121	0.001	0.003	0.009	0.041	0.195	0.349	0.446
	0.146	0.001	0.004	0.012	0.058	0.248	0.436	0.522
GXSM	0.139	0.001	0.004	0.015	0.054	0.174	0.397	0.549
	0.146	0.001	0.004	0.014	0.058	0.190	0.452	0.564
	0.159	0.001	0.004	0.016	0.068	0.203	0.463	0.618
	0.176	0.002	0.005	0.023	0.067	0.264	0.526	0.600
NCH	0.122	0.001	0.003	0.016	0.056	0.181	0.382	0.460
	0.133	0.001	0.005	0.018	0.060	0.158	0.386	0.509
	0.140	0.002	0.005	0.018	0.064	0.160	0.430	0.577
	0.154	0.002	0.006	0.023	0.076	0.199	0.454	0.581

Cuadro B.1: Errores SMAPE y Percentiles. Las cuatro filas por cada método representan los horizontes 1, 3, 6, 12 (de arriba hacia abajo).

	Ranqueo	Prop. Acum. por Posición de Llegada					
	Promedio	1	2	3	4	5	6
OLS	3.067	0.140	0.399	0.635	0.815	0.949	0.994
	3.388	0.124	0.298	0.522	0.770	0.916	0.983
	3.691	0.096	0.247	0.449	0.652	0.899	0.966
	4.180	0.034	0.140	0.337	0.579	0.815	0.916
RROB	2.753	0.242	0.472	0.697	0.876	0.961	1.000
	2.910	0.247	0.444	0.646	0.826	0.938	0.989
	3.096	0.208	0.382	0.601	0.798	0.933	0.983
	3.629	0.129	0.264	0.455	0.680	0.854	0.989
REGT	4.213	0.124	0.219	0.343	0.506	0.758	0.837
	4.236	0.101	0.253	0.348	0.489	0.719	0.854
	3.640	0.213	0.337	0.511	0.624	0.798	0.876
	3.713	0.225	0.371	0.522	0.624	0.713	0.831
SVM	2.910	0.236	0.427	0.669	0.837	0.949	0.972
	3.191	0.174	0.371	0.635	0.764	0.899	0.966
	3.742	0.101	0.298	0.466	0.618	0.803	0.972
	3.680	0.101	0.264	0.449	0.680	0.876	0.949
NN	4.056	0.073	0.180	0.320	0.539	0.871	0.961
	4.180	0.090	0.208	0.320	0.545	0.781	0.876
	4.135	0.062	0.174	0.343	0.618	0.798	0.871
	3.944	0.079	0.264	0.438	0.573	0.798	0.904
GXSM	5.618	0.079	0.124	0.152	0.197	0.236	0.596
	4.927	0.163	0.242	0.287	0.343	0.410	0.629
	4.770	0.185	0.287	0.326	0.360	0.404	0.669
	4.303	0.258	0.348	0.404	0.449	0.500	0.736
NCH	5.382	0.107	0.180	0.185	0.230	0.275	0.640
	5.169	0.101	0.185	0.242	0.264	0.337	0.702
	4.927	0.135	0.275	0.303	0.331	0.365	0.663
	4.551	0.174	0.348	0.393	0.416	0.444	0.674

Cuadro B.2: Ranqueos y Proporciones Acumuladas por posición de Llegada. Las cuatro filas por cada método representan los horizontes 1, 3, 6. 12 (de arriba hacia abajo).

Método	$\bar{R}$	Intervalo de $\bar{R}$	Peor que el mejor
OLS	3.067	(2.73,3.405)	no
	3.388	(3.05,3.725)	no
	3.691	(3.353,4.029)	no
	4.180	(3.842,4.517)	no
RROB	2.753	(2.415,3.09)	no
	2.910	(2.573,3.248)	no
	3.096	(2.758,3.433)	no
	3.629	(3.292,3.967)	no
REGT	4.213	(3.876,4.551)	sí
	4.236	(3.898,4.574)	sí
	3.640	(3.303,3.978)	no
	3.713	(3.376,4.051)	no
SVM	2.910	(2.573,3.248)	no
	3.191	(2.853,3.529)	no
	3.742	(3.404,4.079)	no
	3.680	(3.342,4.017)	no
NN	4.056	(3.719,4.394)	sí
	4.180	(3.842,4.517)	sí
	4.135	(3.797,4.472)	sí
	3.944	(3.606,4.281)	no
GXSM	5.618	(5.28,5.956)	sí
	4.927	(4.589,5.265)	sí
	4.770	(4.432,5.107)	sí
	4.303	(3.966,4.641)	no
NCH	5.382	(5.044,5.72)	sí
	5.169	(4.831,5.506)	sí
	4.927	(4.589,5.265)	sí
	4.551	(4.213,4.888)	sí

Cuadro B.3: Prueba Multicomparativa con el Mejor. Las cuatro filas por cada método representan los horizontes 1, 3, 6, 12 (de arriba hacia abajo). $\bar{R}$ es el ranqueo promedio.

Método	Ranqueo	Relación con la Media
OLS	3.067	mejor
	3.388	mejor
	3.691	igual
	4.180	igual
RROB	2.753	mejor
	2.910	mejor
	3.096	mejor
	3.629	igual
REGT	4.213	igual
	4.236	igual
	3.640	igual
	3.713	igual
SVM	2.910	mejor
	3.191	mejor
	3.742	igual
	3.680	igual
NN	4.056	igual
	4.180	igual
	4.135	igual
	3.944	igual
GXSM	5.618	peor
	4.927	peor
	4.770	peor
	4.303	igual
NCH	5.382	peor
	5.169	peor
	4.927	peor
	4.551	peor
Ranqueo Medio Promedio: 4		
Intervalo: (3.597,4.403)		

Cuadro B.4: Prueba Multicomparativa con la Media. Las cuatro filas por cada método representan los horizontes 1, 3, 6, 12 (de arriba hacia abajo).

	Categorías de Series de Tiempo								
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
OLS	0.083	0.024	0.103	0.002	0.134	0.003	0.038	0.015	0.036
	0.107	0.030	0.108	0.003	0.222	0.007	0.043	0.022	0.050
	0.131	0.034	0.121	0.006	0.319	0.011	0.048	0.033	0.064
	0.156	0.039	0.141	0.012	0.499	0.016	0.054	0.060	0.080
RROB	0.080	0.020	0.102	0.001	0.137	0.003	0.039	0.015	0.034
	0.104	0.028	0.107	0.003	0.213	0.007	0.042	0.021	0.046
	0.128	0.032	0.119	0.005	0.304	0.011	0.048	0.029	0.061
	0.149	0.035	0.137	0.012	0.482	0.015	0.052	0.060	0.078
REGT	0.101	0.033	0.114	0.002	0.146	0.003	0.052	0.027	0.035
	0.111	0.035	0.126	0.003	0.229	0.007	0.051	0.031	0.047
	0.131	0.036	0.126	0.006	0.323	0.011	0.054	0.033	0.061
	0.159	0.037	0.139	0.012	0.422	0.016	0.057	0.073	0.085
SVM	0.084	0.021	0.103	0.001	0.134	0.003	0.036	0.019	0.034
	0.106	0.028	0.111	0.003	0.225	0.007	0.047	0.021	0.045
	0.135	0.029	0.119	0.006	0.326	0.011	0.058	0.041	0.064
	0.148	0.035	0.131	0.011	0.427	0.019	0.057	0.062	0.090
NN	0.099	0.023	0.110	0.002	0.143	0.004	0.045	0.018	0.068
	0.174	0.030	0.116	0.003	0.239	0.011	0.048	0.035	0.052
	0.133	0.031	0.126	0.007	0.330	0.012	0.051	0.055	0.121
	0.155	0.038	0.135	0.015	0.419	0.019	0.055	0.067	0.097
GXSM	0.105	0.024	0.165	0.041	0.402	0.052	0.063	0.022	0.242
	0.108	0.031	0.139	0.088	0.458	0.052	0.065	0.023	0.091
	0.111	0.032	0.142	0.128	0.491	0.091	0.066	0.028	0.092
	0.120	0.035	0.151	0.059	0.552	0.093	0.069	0.050	0.081
NCH	0.106	0.026	0.134	0.051	0.352	0.065	0.061	0.022	0.134
	0.108	0.026	0.173	0.052	0.381	0.066	0.068	0.023	0.086
	0.108	0.025	0.156	0.055	0.413	0.068	0.078	0.026	0.074
	0.112	0.032	0.159	0.064	0.453	0.072	0.082	0.037	0.095

Cuadro B.5: SMAPEs por Categoría de Datos. (a) índices de producción, (b) producción física, (c) ventas, (d) índices de precios, (e) tasas de interés, (f) agregados monetarios, (g) ingresos y gastos públicos, (h) variables demográficas, (i) otras. Las cuatro filas por cada método representan los horizontes 1, 3, 6, 12 (de arriba hacia abajo).

	Categorías de Series de Tiempo								
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
OLS	2	1	2	2	3	3	4	5	3
	0	0	3	3	2	4	3	3	1
	0	0	1	3	2	2	3	1	0
	0	0	1	3	0	3	1	0	0
RROB	3	2	2	6	3	3	3	5	3
	2	2	3	4	4	3	5	3	1
	0	2	0	5	5	2	3	3	0
	1	1	1	3	1	2	2	0	0
REGT	0	0	0	3	2	3	1	0	3
	0	0	0	2	2	2	1	0	1
	0	0	2	3	3	2	1	1	0
	0	0	1	2	1	5	1	0	1
SVM	3	3	2	4	3	4	5	3	3
	0	3	3	3	2	5	5	3	1
	0	2	0	3	2	3	3	1	0
	0	2	1	3	0	2	1	0	0
NN	0	1	1	2	2	2	3	2	1
	0	0	1	2	2	2	2	0	0
	0	0	0	2	2	2	2	0	0
	0	0	1	2	4	2	2	0	0
GXSM	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	1	0	1	0
	1	0	0	0	0	0	1	1	0
	2	0	0	0	0	0	1	0	0
NCH	0	0	0	1	1	0	0	0	0
	0	0	0	1	1	0	0	0	0
	0	0	0	1	1	1	0	0	0
	0	0	0	1	0	0	0	0	0

Cuadro B.6: Desempeño Relativo de cada Método: Prueba Binomial por Categoría de Datos. No. de métodos que tienen un desempeño significativamente más pobre que el nombrado. (a) índices de producción, (b) producción física, (c) ventas. (d) índices de precios. (e) tasas de interés, (f) agregados monetarios, (g) ingresos y gastos públicos, (h) variables demográficas, (i) otras. Las cuatro filas por cada método representan los horizontes 1, 3, 6, 12 (de arriba hacia abajo).

Cuadro B.7: Códigos Numéricos de Series de Tiempo.

No.	Nombre de serie
1	ÍNDICE DE PRODUCCIÓN FÍSICA DE ACERO (AÑO BASE = 1997)
2	ÍNDICE DE PRODUCCIÓN FÍSICA DE CABILLAS (AÑO BASE = 1997)
3	ÍNDICE DE PRODUCCIÓN FÍSICA DE CAUCHOS: AUTOMÓVILES (AÑO BASE = 1997)
4	ÍNDICE DE PRODUCCIÓN FÍSICA DE CAUCHOS: CAMIONETA (AÑO BASE = 1997)
5	ÍNDICE DE PRODUCCIÓN FÍSICA DE CEMENTO (AÑO BASE = 1997)
6	ÍNDICE DE PRODUCCIÓN FÍSICA DE ELECTRICIDAD (AÑO BASE = 1997)
7	ÍNDICE DE PRODUCCIÓN FÍSICA DE HIERRO (AÑO BASE = 1997)
8	ÍNDICE DE PRODUCCIÓN FÍSICA DE VEHÍCULOS: BUSES Y MINIBUSES (AÑO BASE = 1997)
9	ÍNDICE DE PRODUCCIÓN FÍSICA DE VEHÍCULOS: DE CARGA (AÑO BASE = 1997)
10	ÍNDICE DE PRODUCCIÓN FÍSICA DE VEHÍCULOS: DE PASAJEROS (AÑO BASE = 1997)
11	ÍNDICE DE PRODUCCIÓN FÍSICA DE VEHÍCULOS: RÚSTICOS (AÑO BASE = 1997)
12	CONSUMO ELÉCTRICO CADAFE (MILES DE KW/H)
13	CONSUMO ELÉCTRICO EDELCA (MILES DE KW/H)
14	CONSUMO ELÉCTRICO ELECAR (MILES DE KW/H)
15	CONSUMO ELÉCTRICO ENELVEN (MILES DE KW/H)
16	CONSUMO ELÉCTRICO TOTAL (MILES DE KW/H)
17	PRODUCCIÓN DE HIERRO (MILES DE TONELADAS)
18	PRODUCCIÓN DE ACERO (MILES DE TONELADAS)
19	PRODUCCIÓN DE PRODUCTOS LARGOS DEL ACERO (MILES DE TONELADAS)
20	PRODUCCIÓN DE CABILLAS (MILES DE TONELADAS)
21	PRODUCCIÓN TOTAL DE PRODUCTOS LARGOS DEL ACERO (MILES DE TONELADAS)
22	PRODUCCIÓN TOTAL DE PRODUCTOS PETROQUÍMICOS (MILES DE TONELADAS)
23	VENTAS EXTERNAS DE HIERRO (MILES DE TONELADAS)
24	VENTAS INTERNAS DE HIERRO (MILES DE TONELADAS)
25	VENTAS TOTALES DE HIERRO (MILES DE TONELADAS)
26	VENTAS EXTERNAS DE ALUMINIO (MILES DE TONELADAS)

Continúa en la próxima página ...

Cuadro B.7 Continuación

No.	Nombre de serie
27	VENTAS INTERNAS DE ALUMINIO (MILES DE TONELADAS)
28	VENTAS TOTALES DE ALUMINIO (MILES DE TONELADAS)
29	VENTAS EXTERNAS DE CEMENTO (MILES DE TONELADAS)
30	VENTAS INTERNAS DE CEMENTO (MILES DE TONELADAS)
31	VENTAS TOTALES DE CEMENTO (MILES DE TONELADAS)
32	VENTAS EXTERNAS DE PRODUCTOS LARGOS DEL ACERO, SIN CABILLAS (MILES DE TONELADAS)
33	VENTAS EXTERNAS DE CABILLAS (MILES DE TONELADAS)
34	VENTAS EXTERNAS DE PRODUCTOS LARGOS DEL ACERO
35	VENTAS INTERNAS DE PRODUCTOS LARGOS DEL ACERO, SIN CABILLAS (MILES DE TONELADAS)
36	VENTAS INTERNAS DE CABILLAS (MILES DE TONELADAS)
37	VENTAS INTERNAS DE PRODUCTOS LARGOS DEL ACERO (MILES DE TONELADAS)
38	VENTAS TOTALES DE PRODUCTOS LARGOS DEL ACERO (MILES DE TONELADAS)
39	VENTAS TOTALES DE CABILLAS (MILES DE TONELADAS)
40	VENTAS DE VEHÍCULOS NACIONALES (UNIDADES)
41	VENTAS DE VEHÍCULOS IMPORTADOS (UNIDADES)
42	VENTAS TOTALES DE VEHÍCULOS (UNIDADES)
43	IPC: ALIMENTOS, BEBIDAS Y TABACOS (AÑO BASE = 1968)
44	IPC: GASTOS DIVERSOS (AÑO BASE = 1968)
45	IPC: GASTOS DEL HOGAR (AÑO BASE = 1968)
46	IPC: VESTIDO Y CALZADO (AÑO BASE = 1968)
47	IPM: PRODUCCIÓN AGROPECUARIA (AÑO BASE = 1984)
48	IPM: GENERAL (AÑO BASE = 1984)
49	IPC: GENERAL (AÑO BASE = 1997)
50	IPM: IMPORTADO (AÑO BASE = 1984)
51	IPM: NACIONAL (AÑO BASE = 1984)
52	TASA DE INTERÉS ACTIVA DE LOS SEIS PRINCIPALES BCU DEL PAÍS (%)
53	TASA DE INTERÉS ACTIVA DE LOS BCU: SECTOR AGRÍCOLA (%)

Continúa en la próxima página ...

Cuadro B.7 Continuación

No.	Nombre de serie
54	TASA DE INTERÉS ACTIVA DE LOS BCU: SECTOR COMERCIO Y SERVICIOS (%)
55	TASA DE INTERÉS ACTIVA DE LOS BCU: DESCUENTOS (%)
56	TASA DE INTERÉS ACTIVA DE LOS BCU: SECTOR INDUSTRIAL (%)
57	TASA DE INTERÉS ACTIVA DE LA BCU (%)
58	TASA DE INTERÉS ACTIVA DE LA BH (%)
59	TASA DE INTERÉS ACTIVA DE LOS BH: NO PREFERENCIALES PARA ADQUISICIÓN DE VIVIENDAS (%)
60	TASA DE INTERÉS ACTIVA DE LOS BH: NO PREFERENCIALES PARA LA CONSTRUCCIÓN DE VIVIENDAS (%)
61	TASA DE INTERÉS ACTIVA DE LOS BH: PREFERENCIALES PARA ADQUISICIÓN DE VIVIENDAS (%)
62	TASA DE INTERÉS ACTIVA DE LA BANCA DE INVERSIÓN (%)
63	TASAS DE INTERÉS ACTIVA DE LOS BI: SECTOR COMERCIO Y SERVICIOS (%)
64	TASA DE INTERÉS ACTIVA DE LOS BI: PRÉSTAMOS HIPOTECARIOS (%)
65	TASA DE INTERÉS ACTIVA DE LOS BI: PRÉSTAMOS INDUSTRIALES (%)
66	TASA DE INTERÉS ACTIVA DE LOS BI: PRÉSTAMOS NO HIPOTECARIOS (%)
67	TASA DE INTERÉS ACTIVA DE LOS BI: ADQUISICIÓN DE VEHÍCULOS (%)
68	TASA DE INTERES PASIVA: DEPÓSITOS DE AHORRO DE LOS SEIS PRINCIPALES BCU DEL PAÍS (%)
69	TASA DE INTERES PASIVA DE LOS BCU: DEPÓSITOS DE AHORRO (%)
70	TASA DE INTERES PASIVA DE LOS BCU: DPF A MÁS DE 180 DÍAS (%)
71	TASA DE INTERES PASIVA DE LOS BCU: DPF 180 DÍAS (%)
72	TASA DE INTERES PASIVA DE LOS BCU: DPF 90 DÍAS (%)
73	TASA DE INTERES PASIVA DE LOS BH: DEPÓSITOS DE AHORRO (%)
74	TASA DE INTERES PASIVA DE LOS BH: CERTIFICADOS DE AHORRO MÁS DE 180 DÍAS (%)
75	TASA DE INTERES PASIVA DE LOS BH: CERTIFICADOS DE AHORRO 180 DÍAS (%)
76	TASA DE INTERES PASIVA DE LOS BH: CERTIFICADOS DE AHORRO 90 DÍAS (%)
77	TASA DE INTERES PASIVA DE LOS BH: DPF 90 DÍAS (%)
78	TASA DE INTERES PASIVA DE LOS BH (%)
79	TASA DE INTERES PASIVA DE LOS BI: CERTIFICADOS DE AHORRO MÁS DE 180 DÍAS (%)
80	TASA DE INTERES PASIVA DE LOS BI: CERTIFICADOS DE AHORRO 180 DÍAS (%)

Continúa en la próxima página ...

Cuadro B.7 Continuación

No.	Nombre de serie
81	TASA DE INTERES PASIVA DE LOS BI: CERTIFICADOS DE AHORRO 90 DÍAS (%)
82	TASA DE INTERES PASIVA DE LOS BI: DPF 90 DÍAS (%)
83	TASA DE INTERES PASIVA DE LOS BI (%)
84	TASA DE INTERES PARA EL CÁLCULO DE LAS PRESTACIONES SOCIALES (%)
85	TASA DE INTERES PASIVA DE LOS BCU: DPF 30 Y 60 DIAS (%)
86	TASA DE INTERES PASIVA DE LOS BH: CERTIFICADOS DE AHORRO 30 Y 60 DIAS (%)
87	TASA DE INTERES PASIVA DE LOS BI: DPF 30 Y 60 DIAS (%)
88	TASA DE INTERES PASIVA DE LOS BI: CERTIFICADOS DE AHORRO 30 Y 60 DIAS (%)
89	MONEDAS Y BILLETES (MM DE Bs. ANTIGUOS)
90	DEPOSITOS A LA VISTA (MM DE Bs. ANTIGUOS)
91	M1 (MM DE Bs. ANTIGUOS)
92	DEPOSITOS DE AHORRO (MM DE Bs. ANTIGUOS)
93	DEPOSITOS A PLAZO (MM DE Bs. ANTIGUOS)
94	CUASIDINERO (MM DE Bs. ANTIGUOS)
95	M2 (MM DE Bs. ANTIGUOS)
96	CÉDULAS HIPOTECARIAS (MM DE Bs. ANTIGUOS)
97	M3 (MM DE Bs. ANTIGUOS)
98	RECAUDACIÓN ISLR: TECNOLOGÍA PETROLERA (MM DE Bs. ANTIGUOS)
99	RECAUDACIÓN ISLR: HIERRO (MM DE Bs. ANTIGUOS)
100	RECAUDACIÓN ISLR: OTRAS ACTIVIDADES (MM DE Bs. ANTIGUOS)
101	RECAUDACIÓN ISLR: SUCESIONES Y DONACIONES (MM DE Bs. ANTIGUOS)
102	RECAUDACIÓN TRIBUTOS DIRECTOS: TOTAL (MM DE Bs. ANTIGUOS)
103	RECAUDACIÓN TRIBUTOS INDIRECTOS: IMPORTACIÓNES ORDINARIAS (MM DE Bs. ANTIGUOS)
104	RECAUDACIÓN TRIBUTOS INDIRECTOS: SERVICIOS DE ADUANA (MM DE Bs. ANTIGUOS)
105	RECAUDACIÓN TRIBUTOS INDIRECTOS: ADUANAS (MM DE Bs. ANTIGUOS)
106	RECAUDACIÓN TRIBUTOS INDIRECTOS: IVA (MM DE Bs. ANTIGUOS)
107	RECAUDACIÓN TRIBUTOS INDIRECTOS: RENTA DE LICORES (MM DE Bs. ANTIGUOS)

Continúa en la próxima página ...

Cuadro B.7 – Continuación

No.	Nombre de serie
108	RECAUDACIÓN TRIBUTOS INDIRECTOS: UTILIDAD POR OPERACIONES CAMBIARÍAS (MM DE Bs. ANTIGUOS)
109	RECAUDACIÓN TRIBUTOS INDIRECTOS: OTROS (MM DE Bs. ANTIGUOS)
110	RECAUDACIÓN TRIBUTOS INDIRECTOS: TOTAL (MM DE Bs. ANTIGUOS)
111	RECAUDACIÓN TOTAL DE TRIBUTOS (MM DE Bs. ANTIGUOS)
112	INGRESOS PÚBLICOS: DIVIDENDOS PDVSA (MM DE Bs. ANTIGUOS)
113	INGRESOS PÚBLICOS: REMANENTE UTILIDADES BCV (MM DE Bs. ANTIGUOS)
114	INGRESOS PÚBLICOS: TOTAL INGRESOS ORDINARIOS (MM DE Bs. ANTIGUOS)
115	INGRESOS PÚBLICOS EXTRAORDINARIOS: UTILIZACION TOTAL DE CREDITO PUBLICO EXTERNO (MM DE Bs. ANTIGUOS)
116	INGRESOS PÚBLICOS EXTRAORDINARIOS: UTILIZACIÓN TOTAL DE CRÉDITO PUBLICO (MM DE Bs. ANTIGUOS)
117	INGRESOS PÚBLICOS EXTRAORDINARIOS: UTILIDADES DEL BCV (MM DE Bs. ANTIGUOS)
118	INGRESOS PÚBLICOS EXTRAORDINARIOS: TOTAL (MM DE Bs. ANTIGUOS)
119	INGRESOS PÚBLICOS: TOTAL (MM DE Bs. ANTIGUOS)
120	EGRESOS ORDINARIOS DEL GOBIERNO CENTRAL: TRANSFERENCIAS AL FIEM (MM DE Bs. ANTIGUOS)
121	EGRESOS ORDINARIOS DEL GOBIERNO CENTRAL: TOTAL (MM DE Bs. ANTIGUOS)
122	EGRESOS EXTRAORDINARIOS DEL GOBIERNO CENTRAL: AMORTIZACION DE DEUDA CENTRALIZADA TOTAL (MM DE Bs. ANTIGUOS)
123	EGRESOS EXTRAORDINARIOS DEL GOBIERNO CENTRAL: TOTAL (MM DE Bs. ANTIGUOS)
124	EGRESOS DEL GOBIERNO CENTRAL: TOTAL (MM DE Bs. ANTIGUOS)
125	RECAUDACIÓN TRIBUTOS INDIRECTOS: IDB (MM DE Bs. ANTIGUOS)
126	POBLACIÓN VENEZOLANA DE 15 AÑOS Y MÁS
127	POBLACIÓN VENEZOLANA DE 15 24 AÑOS
128	POBLACIÓN VENEZOLANA DE 25 44 AÑOS
129	POBLACIÓN VENEZOLANA DE 45 64 AÑOS
130	POBLACIÓN VENEZOLANA DE 65 Y MÁS AÑOS
131	POBLACIÓN ECONÓMICANTE ACTIVA
132	TASA DE ACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA (%)
133	TASA DE ACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA DE 15 24 AÑOS (%)
134	TASA DE ACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA DE 25 44 AÑOS (%)

Continúa en la próxima página ...

Cuadro B.7 – Continuación

No.	Nombre de serie
135	TASA DE ACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA DE 45 64 AÑOS (%)
136	TASA DE ACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA DE 65 Y MÁS AÑOS (%)
137	POBLACIÓN OCUPADA
138	TASA DE OCUPACIÓN (%)
139	TASA DE OCUPACIÓN DE LA POBLACIÓN OCUPADA DE 15 24 AÑOS (%)
140	TASA DE OCUPACIÓN DE LA POBLACIÓN OCUPADA DE 25 44 AÑOS (%)
141	TASA DE OCUPACIÓN DE LA POBLACIÓN OCUPADA DE 45 64 AÑOS (%)
142	TASA DE OCUPACIÓN DE LA POBLACIÓN OCUPADA DE 65 Y MÁS AÑOS (%)
143	POBLACIÓN DESOCUPADA
144	TASA DE DESOCUPACIÓN (%)
145	TASA DE DESOCUPACIÓN DE LA POBLACIÓN DESOCUPADA DE 15 24 AÑOS (%)
146	TASA DE DESOCUPACIÓN DE LA POBLACIÓN DESOCUPADA DE 25 44 AÑOS (%)
147	TASA DE DESOCUPACIÓN DE LA POBLACIÓN DESOCUPADA DE 45 64 AÑOS (%)
148	TASA DE DESOCUPACIÓN DE LAS POBLACIÓN DESOCUPADA DE 65 Y MÁS AÑOS (%)
149	POBLACIÓN ECONÓMICAMENTE INACTIVA
150	TASA DE INACTIVIDAD (%)
151	TASA DE INACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE INACTIVA DE 15 24 AÑOS (%)
152	TASA DE INACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE INACTIVA DE 25 44 AÑOS (%)
153	TASA DE INACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE INACTIVA DE 45 64 AÑOS (%)
154	TASA DE INACTIVIDAD DE LA POBLACIÓN ECONÓMICAMENTE INACTIVA DE 65 Y MÁS AÑOS (%)
155	POBLACIÓN OCUPADA DEL SECTOR FORMAL
156	POBLACIÓN OCUPADA DEL SECTOR INFORMAL
157	TRABAJADORES POR CUENTA PROPIA NO PROFESIONALES DEL SECTOR INFORMAL
158	PATRONOS O EMPLEADORES DEL SECTOR INFORMAL
159	EMPLEADOS Y OBREROS DEL SECTOR INFORMAL
160	AYUDANTES FAMILIARES NO REMUNERADOS DEL SECTOR INFORMAL
161	TOTAL OCUPADOS (EMPLEADOS Y OBREROS) DEL SECTOR PÚBLICO

Continúa en la próxima página ...

Cuadro B.7 – Continuación

No.	Nombre de serie
162	TOTAL OCUPADOS DEL SECTOR PRIVADO
163	EMPLEADOS Y OBREROS DEL SECTOR PRIVADO
164	TRABAJADORES POR CUENTA PROPIA DEL SECTOR PRIVADO
165	PATRONOS Y EMPLEADORES DEL SECTOR PRIVADO
166	MIEMBROS DE COOPERATIVA DEL SECTOR PRIVADO
167	AYUDANTES FAMILIARES DEL SECTOR PRIVADO
168	POBLACIÓN VENEZOLANA TOTAL
169	PROPORCIÓN DE ESTUDIANTES (%)
170	PROPORCIÓN DE LA POBLACIÓN EN QUE-HACERES DEL HOGAR (%)
171	PROPORCIÓN DE INCAPACITADOS PARA TRABAJAR (%)
172	PROPORCIÓN DE LA POBLACIÓN EN OTRA SITUACIÓN (%)
173	COEFICIENTE DE PREFERENCIA DEL PUBLICO POR EL DINERO M1 (COCIENTE)
174	COEFICIENTE DE PREFERENCIA DEL PUBLICO POR MONEDAS Y BILLETES (COCIENTE)
175	COEFICIENTE DE RESERVAS BANCARIAS (COCIENTE)
176	MULTIPLICADOR DE LA LIQUIDEZ MONETARIA M2
177	MULTIPLICADOR DEL DINERO M1
178	PRECIO PROMEDIO CESTA PETROLERA VENEZOLANA (US$)

Cuadro B.8: Características de los Datos.

Rgo. = Rango, Estac. = presenta estacionalidad, Log. = fue transformada a logaritmos, Tend. = presenta tendencia, R.U. = presenta raíz unitaria, H1-H3-H6-H12 = rezagos especificados en horizonte(H), SPr. = es una "serie problemática", DsvE. = desviación estándar.

No.	Categoría	Rgo. (# de obs.)	Est.	Log.	Tend.	R.U.	H1	H3	H6	H12	SPr.	DsvE.
1	ÍNDICE DE PRODUCCIÓN	Ene1988 - Dic2004 (204)	Sí	Sí	No	No	2	2	1	2	No	0.17
2	ÍNDICE DE PRODUCCIÓN	Ene1988 - Dic2004 (204)	Sí	Sí	No	No	2	2	1	2	No	0.26
3	ÍNDICE DE PRODUCCIÓN	Ene1988 - Feb2005 (206)	Sí	Sí	Sí	No	2	1	10	11	No	0.23
4	ÍNDICE DE PRODUCCIÓN	Ene1988 - Feb2005 (206)	Sí	Sí	Sí	No	1	1	1	1	No	0.26
5	ÍNDICE DE PRODUCCIÓN	Ene1988 - Feb2005 (206)	Sí	Sí	No	Sí	2	2	1	1	No	0.30
6	ÍNDICE DE PRODUCCIÓN	Ene1988 - Feb2005 (206)	Sí	Sí	Sí	Sí	2	4	1	1	No	0.19
7	ÍNDICE DE PRODUCCIÓN	Ene1988 - Oct2004 (202)	Sí	Sí	Sí	No	4	2	1	1	No	0.19
8	ÍNDICE DE PRODUCCIÓN	Ene1988 - Mar2005 (207)	Sí	Sí	No	Sí	3	1	1	5	Sí	0.92
9	ÍNDICE DE PRODUCCIÓN	Ene1988 - Mar2005 (207)	Sí	Sí	No	Sí	1	1	1	1	Sí	1.86
10	ÍNDICE DE PRODUCCIÓN	Ene1988 - Mar2005 (207)	Sí	Sí	No	Sí	1	2	11	3	Sí	1.01
11	ÍNDICE DE PRODUCCIÓN	Ene1988 - Mar2005 (207)	Sí	Sí	No	No	2	2	1	12	Sí	1.04
12	PRODUCCIÓN FÍSICA	Ene1990 - Jul2007 (211)	Sí	Sí	No	Sí	1	1	1	1	No	0.16
13	PRODUCCIÓN FÍSICA	Ene1990 - Jul2007 (211)	Sí	Sí	Sí	No	1	1	1	12	No	0.19
14	PRODUCCIÓN FÍSICA	Ene1990 - Jul2007 (211)	Sí	Sí	Sí	Sí	2	1	1	1	No	0.09
15	PRODUCCIÓN FÍSICA	Ene1990 - Jul2007 (211)	Sí	Sí	No	Sí	5	3	1	1	No	0.18
16	PRODUCCIÓN FÍSICA	Ene1990 - Jul2007 (211)	Sí	Sí	No	Sí	1	1	1	1	No	0.14
17	PRODUCCIÓN FÍSICA	Abr1996 - Jul2007 (136)	Sí	Sí	No	Sí	3	2	2	1	No	0.20
18	PRODUCCIÓN FÍSICA	Ene1995 - Ago2007 (152)	Sí	Sí	No	Sí	1	1	1	1	No	0.64
19	PRODUCCIÓN FÍSICA	Ene1998 - Ago2007 (116)	Sí	Sí	Sí	No	1	1	1	1	No	0.24
20	PRODUCCIÓN FÍSICA	Ene1995 - Ago2007 (152)	Sí	Sí	No	No	1	2	1	1	No	0.35
21	PRODUCCIÓN FÍSICA	Ene1998 - Ago2007 (116)	Sí	Sí	Sí	No	1	2	1	1	No	0.25
22	PRODUCCIÓN FÍSICA	Ene1991 - Ago2002 (140)	Sí	Sí	No	Sí	3	1	1	1	No	0.69

Continúa en la próxima página ...

Cuadro B.8 – Continuación

No.	Categoría	Rgo. (# de obs.)	Est.	Log.	Tend.	R.U.	H1	H3	H6	H12	S.Pr.	Dsv.E.
23	VENTAS	Ene1994 - Jul2007 (163)	Sí	Sí	No	No	6	4	4	2	No	0.31
24	VENTAS	Ene1994 - Jul2007 (163)	Sí	Sí	Sí	No	1	1	1	1	No	0.39
25	VENTAS	Ene1994 - Jul2007 (163)	Sí	Sí	No	No	1	1	1	1	No	0.18
26	VENTAS	Ene1990 - Ago2007 (212)	Sí	Sí	No	No	3	1	1	1	No	0.21
27	VENTAS	Ene1990 - Ago2007 (212)	Sí	Sí	No	No	2	1	1	1	No	0.28
28	VENTAS	Ene1990 - Ago2007 (212)	Sí	Sí	No	No	4	3	6	1	No	0.16
29	VENTAS	Ene1998 - Ago2007 (116)	Sí	Sí	No	No	1	1	1	1	Sí	0.48
30	VENTAS	Ene1998 - Ago2007 (116)	Sí	Sí	No	Sí	1	1	1	1	No	0.26
31	VENTAS	Ene1998 - Ago2007 (116)	Sí	Sí	No	Sí	2	1	1	1	No	0.13
32	VENTAS	Ene1998 - Ago2007 (116)	Sí	Sí	No	No	2	4	1	1	Sí	0.54
33	VENTAS	Feb1990 - Ene2007 (204)	No	Sí	No	No	2	1	1	1	Sí	1.23
34	VENTAS	Ene1998 - Sep2007 (117)	Sí	Sí	No	No	1	1	1	1	Sí	0.65
35	VENTAS	Ene1998 - Ago2007 (116)	Sí	Sí	No	No	2	2	1	1	No	0.41
36	VENTAS	Ene1990 - Ago2007 (212)	Sí	Sí	Sí	Sí	3	7	3	6	No	0.47
37	VENTAS	Ene1998 - Sep2007 (117)	Sí	Sí	No	No	2	2	1	1	No	0.38
38	VENTAS	Ene1998 - Sep2007 (117)	Sí	Sí	No	No	1	1	1	1	No	0.27
39	VENTAS	Ene1990 - Ago2007 (212)	Sí	Sí	Sí	Sí	9	7	7	1	No	0.35
40	VENTAS	Ene1990 - Ago2007 (212)	Sí	Sí	No	Sí	2	1	7	2	No	0.47
41	VENTAS	Mar1991 - Ago2007 (198)	Sí	Sí	No	Sí	8	6	3	1	No	1.61
42	VENTAS	Ene1990 - Ago2007 (212)	Sí	Sí	No	Sí	3	2	1	5	No	0.64
43	ÍNDICE DE PRECIO	Ene1962 - Dic1985 (288)	Sí	Sí	Sí	Sí	1	1	1	1	No	0.63
44	ÍNDICE DE PRECIO	Ene1962 - Dic1985 (288)	No	Sí	No	Sí	12	10	7	1	No	0.46
45	ÍNDICE DE PRECIO	Ene1962 - Dic1985 (288)	Sí	Sí	No	Sí	4	4	1	6	No	0.31
46	ÍNDICE DE PRECIO	Ene1962 - Dic1985 (288)	No	Sí	Sí	Sí	2	1	12	12	No	0.65
47	ÍNDICE DE PRECIO	Ene1962 - Dic2002 (492)	No	Sí	No	Sí	1	11	12	7	No	2.40
48	ÍNDICE DE PRECIO	Ene1962 - Dic2002 (492)	No	Sí	No	Sí	3	2	12	11	No	2.29
49	ÍNDICE DE PRECIO	Ene1958 - Jun2007 (594)	No	Sí	Sí	Sí	5	3	2	1	No	2.71

Continúa en la próxima página ...

Cuadro B.8 - Continuación

No.	Categoría	Rgo. (# de obs.)	Est.	Log.	Tend.	R.U.	H1	H3	H6	H12	S.Pr.	Dsv.E.
50	ÍNDICE DE PRECIO	Ene1962 - Dic2002 (492)	No	Sí	No	Sí	2	2	12	12	No	2.19
51	ÍNDICE DE PRECIO	Ene1962 - Dic2002 (492)	No	Sí	No	Sí	2	1	12	11	No	2.33
52	TASA DE INTERÉS BANCARIA	Ene1990 - Jun2007 (210)	Sí	No	Sí	No	6	1	1	1	Sí	14.58
53	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	Sí	No	No	Sí	1	1	6	3	Sí	14.08
54	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	1	1	1	1	Sí	15.43
55	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	1	1	1	1	Sí	15.60
56	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	Sí	No	No	Sí	3	1	1	1	Sí	15.81
57	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	Sí	No	No	Sí	1	1	1	1	Sí	15.39
58	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2000 (202)	Sí	No	No	Sí	1	1	1	1	Sí	9.57
59	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2000 (202)	Sí	No	Sí	Sí	1	1	1	1	Sí	14.81
60	TASA DE INTERÉS BANCARIA	Mar1984 - Abr1994 (122)	No	No	Sí	Sí	1	1	1	8	Sí	16.34
61	TASA DE INTERÉS BANCARIA	Ene1993 - Dic2000 (96)	No	No	Sí	Sí	1	1	12	9	Sí	4.15
62	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	1	1	1	1	Sí	14.56
63	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	1	1	1	1	Sí	14.58
64	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	Sí	No	No	Sí	1	1	1	1	Sí	15.88
65	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	Sí	No	No	Sí	1	1	1	1	Sí	15.29
66	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	1	1	1	1	Sí	14.50
67	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	1	1	1	1	Sí	14.46
68	TASA DE INTERÉS BANCARIA	Ene1991 - Jun2007 (198)	No	No	Sí	Sí	3	12	12	12	Sí	11.51
69	TASA DE INTERÉS BANCARIA	Ene1992 - Dic2001 (120)	Sí	No	Sí	Sí	1	12	7	10	Sí	12.11
70	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2000 (202)	Sí	No	No	Sí	1	1	1	1	Sí	12.70
71	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2000 (202)	No	No	No	No	6	3	1	1	Sí	13.81
72	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	1	1	1	1	Sí	13.40
73	TASA DE INTERÉS BANCARIA	Ene1992 - Dic2001 (120)	Sí	No	Sí	Sí	1	12	11	8	Sí	11.55
74	TASA DE INTERÉS BANCARIA	Mar1984 - Oct1994 (128)	No	No	Sí	Sí	3	1	1	1	Sí	11.90
75	TASA DE INTERÉS BANCARIA	Mar1984 - Nov1994 (129)	No	No	Sí	Sí	1	1	1	1	Sí	14.90
76	TASA DE INTERÉS BANCARIA	Mar1984 - May2001 (207)	No	No	No	Sí	2	4	1	1	Sí	13.24

Continúa en la próxima página ...

Cuadro B.8 – Continuación

No.	Categoría	Rgo. (# de obs.)	Est.	Log.	Tend.	R.U.	H1	H3	H6	H12	S.Pr.	Dsv.E.
77	TASA DE INTERÉS BANCARIA	Mar1984 - Dic1993 (118)	Sí	No	Sí	Sí	1	1	1	1	No	14.08
78	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2000 (202)	No	No	No	No	3	3	1	1	Sí	12.72
79	TASA DE INTERÉS BANCARIA	Mar1984 - Oct1995 (140)	Sí	No	Sí	Sí	1	1	1	1	Sí	9.66
80	TASA DE INTERÉS BANCARIA	Mar1984 - Ene1999 (179)	Sí	No	No	Sí	6	1	1	1	Sí	9.97
81	TASA DE INTERÉS BANCARIA	Mar1984 - Oct1999 (188)	No	No	No	Sí	1	1	1	1	Sí	12.40
82	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	1	1	1	1	Sí	12.50
83	TASA DE INTERÉS BANCARIA	Mar1984 - Dic2001 (214)	No	No	No	Sí	3	1	1	1	Sí	11.87
84	TASA DE INTERÉS BANCARIA	Jun1975 - Oct2006 (377)	No	No	No	Sí	5	4	7	1	Sí	12.69
85	TASA DE INTERÉS BANCARIA	Ene1990 - Dic2001 (144)	Sí	No	No	No	3	6	6	1	Sí	12.37
86	TASA DE INTERÉS BANCARIA	Ene1990 - Dic2001 (144)	Sí	No	Sí	No	3	7	6	4	Sí	12.32
87	TASA DE INTERÉS BANCARIA	Ene1990 - Dic2001 (144)	Sí	No	No	No	2	1	1	1	Sí	11.32
88	TASA DE INTERÉS BANCARIA	Ene1990 - Dic2001 (144)	Sí	No	No	Sí	3	1	1	1	Sí	12.00
89	AGREGADO MONETARIO	Ene1976 - May2007 (377)	No	Sí	No	Sí	12	10	7	12	No	2.42
90	AGREGADO MONETARIO	Ene1976 - May2007 (377)	No	Sí	Sí	Sí	1	12	11	5	No	2.42
91	AGREGADO MONETARIO	Ene1976 - May2007 (377)	No	Sí	Sí	Sí	12	12	11	4	No	2.42
92	AGREGADO MONETARIO	Ene1976 - May2007 (377)	No	Sí	Sí	Sí	12	12	11	5	No	2.37
93	AGREGADO MONETARIO	Ene1976 - May2007 (377)	Sí	Sí	Sí	Sí	1	4	1	7	No	2.19
94	AGREGADO MONETARIO	Ene1976 - May2007 (377)	Sí	Sí	Sí	Sí	5	4	1	1	No	2.35
95	AGREGADO MONETARIO	Ene1976 - May2007 (377)	Sí	Sí	Sí	Sí	5	3	1	2	No	2.37
96	AGREGADO MONETARIO	Ene1976 - May2007 (377)	No	Sí	No	Sí	1	1	1	1	No	0.91
97	AGREGADO MONETARIO	Ene1976 - May2007 (377)	Sí	Sí	Sí	Sí	5	4	1	2	No	2.30
98	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Dic2001 (192)	Sí	Sí	No	Sí	3	1	1	2	No	1.75
99	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Dic2001 (192)	Sí	Sí	Sí	Sí	4	7	1	2	No	1.52
100	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	No	Sí	No	Sí	12	7	2	1	No	2.09
101	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	No	Sí	1	1	1	1	No	1.60
102	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	Sí	Sí	2	5	2	1	No	2.22
103	INGRESOS/GASTOS PÚBLICOS	Oct1993 - Oct2006 (157)	No	Sí	No	Sí	1	1	6	1	No	0.95

Continúa en la próxima página ...

Cuadro B.8 Continuación

No.	Categoría	Rgo. (# de obs.)	Est.	Log.	Tend.	R.U.	H1	H3	H6	H12	S.Pr.	Dsv.E.
104	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	No	Sí	No	Sí	11	10	1	1	No	1.69
105	INGRESOS/GASTOS PÚBLICOS	Ene1987 - Oct2006 (238)	Sí	Sí	Sí	Sí	2	1	1	1	No	2.15
106	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	No	Sí	Sí	Sí	6	7	1	1	No	3.18
107	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	No	Sí	2	1	1	1	No	1.80
108	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Dic2001 (192)	No	Sí	Sí	Sí	5	7	1	1	No	2.16
109	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	No	Sí	No	Sí	11	7	2	1	No	2.19
110	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	Sí	Sí	1	1	1	1	No	2.67
111	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	Sí	No	3	1	1	1	No	2.11
112	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Dic2001 (192)	No	Sí	Sí	Sí	12	12	8	2	No	1.40
113	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Dic2001 (192)	Sí	Sí	No	Sí	2	1	1	1	No	1.40
114	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	Sí	Sí	2	4	1	1	No	2.18
115	INGRESOS/GASTOS PÚBLICOS	Abr1994 - Oct2006 (151)	Sí	Sí	No	Sí	6	4	1	2	No	2.07
116	INGRESOS/GASTOS PÚBLICOS	Jul1994 - Oct2006 (148)	Sí	Sí	No	Sí	4	1	1	1	No	3.14
117	INGRESOS/GASTOS PÚBLICOS	Ene1995 - Mar2002 (87)	Sí	Sí	Sí	Sí	3	1	1	1	No	1.34
118	INGRESOS/GASTOS PÚBLICOS	Dic1990 - Oct2006 (191)	No	Sí	Sí	Sí	4	4	1	9	No	3.35
119	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	Sí	Sí	2	4	1	1	No	2.22
120	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Dic2001 (192)	Sí	Sí	Sí	Sí	4	4	7	1	No	2.33
121	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	No	Sí	4	5	1	3	No	2.12
122	INGRESOS/GASTOS PÚBLICOS	Abr1992 - Jul2006 (172)	Sí	Sí	No	Sí	5	1	1	1	No	2.38
123	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	No	Sí	Sí	Sí	7	2	7	1	No	2.90
124	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Oct2006 (250)	Sí	Sí	Sí	Sí	4	5	1	3	No	2.18
125	INGRESOS/GASTOS PÚBLICOS	Ene1986 - Dic2001 (192)	Sí	Sí	Sí	No	2	2	1	1	No	1.91
126	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	No	Sí	No	Sí	8	4	4	2	No	0.06
127	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	No	Sí	No	Sí	7	5	2	1	No	0.03
128	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	No	Sí	Sí	Sí	4	3	6	4	No	0.04
129	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	No	Sí	No	Sí	3	1	2	12	No	0.08
130	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	Sí	2	2	4	1	No	0.08

Continúa en la próxima página ...

Cuadro B.8 – Continuación

No.	Categoría	Rgo. (# de obs.)	Est.	Log.	Tend.	R.U.	H1	H3	H6	H12	S.Pr.	Dsv.E.
131	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	Sí	Sí	No	Sí	2	2	1	1	No	0.07
132	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	Sí	No	No	Sí	2	2	1	1	No	1.84
133	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	No	No	2	10	9	11	No	3.37
134	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	No	No	1	1	1	12	No	5.53
135	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	Sí	Sí	4	1	2	1	No	1.50
136	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	Sí	Sí	9	4	7	3	No	1.57
137	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	Sí	Sí	Sí	Sí	1	2	1	8	No	0.08
138	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	Sí	No	No	No	2	11	7	10	No	2.65
139	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	Sí	No	2	12	12	9	No	4.65
140	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	Sí	No	2	2	12	12	No	2.56
141	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	No	No	2	1	8	9	No	2.17
142	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	No	No	5	3	1	1	No	2.11
143	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	Sí	Sí	No	Sí	1	8	12	6	No	0.20
144	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	Sí	No	No	No	2	11	7	6	Sí	2.66
145	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	Sí	No	2	12	12	9	No	4.65
146	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	Sí	No	2	2	12	12	No	2.56
147	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	No	No	2	1	8	9	Sí	2.17
148	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	No	No	3	1	1	1	Sí	2.05
149	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	Sí	Sí	Sí	Sí	2	2	3	6	No	0.08
150	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	Sí	No	No	Sí	2	2	1	1	No	1.84
151	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	No	No	2	10	9	11	No	3.37
152	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	No	Sí	1	1	1	1	No	1.26
153	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	Sí	Sí	4	1	2	1	No	1.50
154	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	No	Sí	Sí	9	4	7	3	No	1.57
155	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	Sí	Sí	2	12	12	9	No	0.11
156	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	Sí	1	1	2	1	No	0.03
157	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	No	1	1	1	1	No	0.03

Continúa en la próxima página ...

Cuadro B.8 – Continuación

No.	Categoría	Rgo. (# de obs.)	Est.	Log.	Tend.	R.U.	H1	H3	H6	H12	S.Pr.	Dsv.E.
158	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	Sí	Sí	2	2	9	3	No	0.10
159	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	Sí	2	1	1	4	No	0.08
160	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	Sí	Sí	2	1	12	12	No	0.27
161	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	Sí	Sí	8	6	1	1	No	0.13
162	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	Sí	Sí	1	4	1	1	No	0.05
163	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	Sí	1	1	1	12	No	0.07
164	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	No	1	1	1	1	No	0.03
165	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	Sí	2	2	1	11	No	0.11
166	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	Sí	2	1	1	1	No	0.56
167	DEMOGRÁFICA	Nov2000 - Jun2007 (80)	Sí	Sí	No	Sí	2	6	2	1	No	0.25
168	DEMOGRÁFICA	Ene1999 - Jun2007 (102)	No	Sí	No	Sí	5	7	1	1	No	0.05
169	DEMOGRÁFICA	May1999 - Jun2007 (98)	Sí	No	No	Sí	1	1	1	12	No	1.05
170	DEMOGRÁFICA	May1999 - Jun2007 (98)	Sí	No	Sí	Sí	3	4	1	1	No	2.04
171	DEMOGRÁFICA	May1999 - Jun2007 (98)	Sí	No	No	Sí	1	1	12	7	No	1.06
172	DEMOGRÁFICA	May1999 - Jun2007 (98)	Sí	No	No	Sí	1	1	12	7	No	1.27
173	OTRA	Ene1989 - Ene2006 (205)	Sí	No	No	Sí	2	1	1	1	No	0.05
174	OTRA	Ene1989 - Ene2006 (205)	Sí	No	Sí	Sí	1	4	7	1	No	0.03
175	OTRA	Ene1989 - Ene2006 (205)	Sí	No	Sí	Sí	4	2	1	1	No	0.10
176	OTRA	Ene1989 - Ene2006 (205)	Sí	No	No	Sí	2	1	1	1	No	0.71
177	OTRA	Ene1989 - Ene2006 (205)	Sí	No	No	Sí	2	1	10	3	No	0.24
178	OTRA	Ene1998 - Sep2007 (117)	Sí	Sí	Sí	Sí	2	1	12	12	No	0.55