



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

E.507

RED TELEFÓNICA Y RDSI

**CALIDAD DE SERVICIO, GESTIÓN DE LA RED
E INGENIERÍA DE TRÁFICO**

**MODELOS PARA LA PREVISIÓN
DEL TRÁFICO INTERNACIONAL**

Recomendación UIT-T E.507

(Extracto del *Libro Azul*)

NOTAS

1 La Recomendación UIT-T E.507 se publicó en el fascículo II.3 del Libro Azul. Este fichero es un extracto del Libro Azul. Aunque la presentación y disposición del texto son ligeramente diferentes de la versión del Libro Azul, el contenido del fichero es idéntico a la citada versión y los derechos de autor siguen siendo los mismos (Véase a continuación).

2 Por razones de concisión, el término «Administración» se utiliza en la presente Recomendación para designar a una administración de telecomunicaciones y a una empresa de explotación reconocida.

© UIT 1988, 1993

Reservados todos los derechos. No podrá reproducirse o utilizarse la presente Recomendación ni parte de la misma de cualquier forma ni por cualquier procedimiento, electrónico o mecánico, comprendidas la fotocopia y la grabación en micropelícula, sin autorización escrita de la UIT.

MODELOS PARA LA PREVISIÓN DEL TRÁFICO INTERNACIONAL

1 Introducción

Para elaborar modelos econométricos y de series cronológicas y formular las previsiones correspondientes hay que dominar diversos métodos y técnicas que tratan una variedad de situaciones diferentes. Así pues, el propósito de esta Recomendación es exponer algunas de las ideas fundamentales sin entrar en explicación de detalles, para los cuales pueden consultarse las publicaciones citadas en la lista de referencias. No pretende, por tanto, constituir una guía completa para la elaboración de modelos econométricos y de series cronológicas y la formulación de las previsiones consiguientes.

La presente Recomendación contiene también algunas directrices para la elaboración de diversos modelos de previsión: identificación del modelo, inclusión de variables explicativas, ajuste de irregularidades, estimación de parámetros, verificación del diagnóstico, etc.

Esta Recomendación describe además diversos métodos para evaluar los modelos de previsión y elegir el modelo.

2 Elaboración del modelo de previsión

A fin de facilitar la descripción de este procedimiento, cabe dividirlo en cuatro etapas. La primera consiste en la búsqueda de una clase de modelos útiles para describir la situación real. Ejemplos de tales clases son los modelos simples, los modelos de suavizamiento, los modelos de autorregresión, los modelos de autorregresión integrados con media móvil (*autoregressive integrated moving average*, ARIMA) o los modelos econométricos. Antes de elegir la clase de modelos ha de analizarse la influencia de las variables externas. Si determinadas variables externas tienen repercusiones importantes en la demanda de tráfico, se las debe incluir en los modelos de previsión, siempre que se disponga de datos anteriores suficientes.

El paso siguiente consiste en adoptar, a título de ensayo un modelo determinado, dentro de la clase de modelos seleccionada. Si la clase es demasiado extensa para que resulte práctico efectuar ajustes directos a los datos, pueden utilizarse métodos aproximados para identificar subclases. Los datos disponibles y el conocimiento del sistema sugerirán una subclase de modelos de extensión apropiada. Asimismo, en algunos casos, puede utilizarse este procedimiento de selección para obtener estimaciones preliminares aproximadas de los parámetros del modelo. Seguidamente, el modelo provisional se ajusta a los datos mediante la estimación de los parámetros. De ordinario, se emplean estimadores de los métodos de los mínimos cuadrados, o de la máxima verosimilitud.

La etapa siguiente es la verificación del modelo. Este procedimiento se denomina a menudo verificación del diagnóstico. Su objeto es establecer la medida en que el modelo se ajusta a los datos y si la discrepancia se considera excesiva, indicar posibles remedios. Así pues, el resultado de esta etapa puede ser la adopción del modelo, si el ajuste fuese aceptable. Si, en cambio, éste fuese insuficiente, ello indicará que deben estimarse los parámetros de nuevos modelos provisionales y efectuar en éstos la verificación del diagnóstico.

¹⁾ El texto de la antigua Recomendación E.506 del *Libro Rojo*, al que se ha añadido un volumen considerable de nuevos textos, se ha convertido en las Recomendaciones E.506 y E.507 actuales.

En la figura 1/E.507 están representadas las etapas del procedimiento de elaboración del modelo.

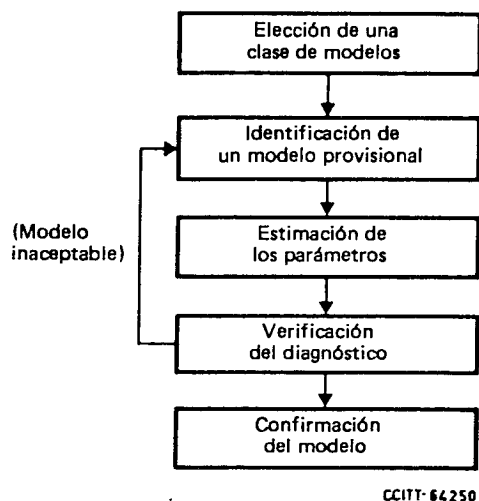


FIGURA 1/E.507

Etapas del procedimiento de elaboración del modelo

3 Diferentes modelos de previsión

El objetivo de este § 3 es dar una breve descripción general de los modelos de previsión más importantes. En el Manual del GAS 10 citado en [5], figura una descripción más detallada de los modelos.

3.1 Modelos de ajuste de curvas

En los modelos de ajuste de curvas la tendencia del tráfico se extrapola calculando los valores de los parámetros de una función que se prevé caracterice el crecimiento del tráfico internacional con el tiempo. Los cálculos numéricos para algunos modelos de ajuste de curvas pueden efectuarse utilizando el método de los mínimos cuadrados.

A continuación se dan ejemplos de modelos de ajuste de curvas corrientes utilizados para la previsión del tráfico internacional:

$$\text{Lineal: } Y_t = a + bt \quad (3-1)$$

$$\text{Parabólica: } Y_t = a + bt + ct^2 \quad (3-2)$$

$$\text{Exponencial: } Y_t = ae^{bt} \quad (3-3)$$

$$\text{Logística: } Y_t = \frac{M}{1 + ae^{bt}} \quad (3-4)$$

$$\text{Gompertz: } Y_t = M(a)^{bt} \quad (3-5)$$

donde

Y_t es el tráfico en el instante t ,

a, b, c son parámetros,

M es un parámetro que describe el nivel de saturación.

En las figuras 2/E.507 y 3/E.507 se muestran las diversas curvas de tendencia.

Las curvas logística y de Gompertz difieren de las curvas lineal, parabólica y exponencial porque tiene un nivel de saturación o asíntotico. Para información más detallada véase [10].

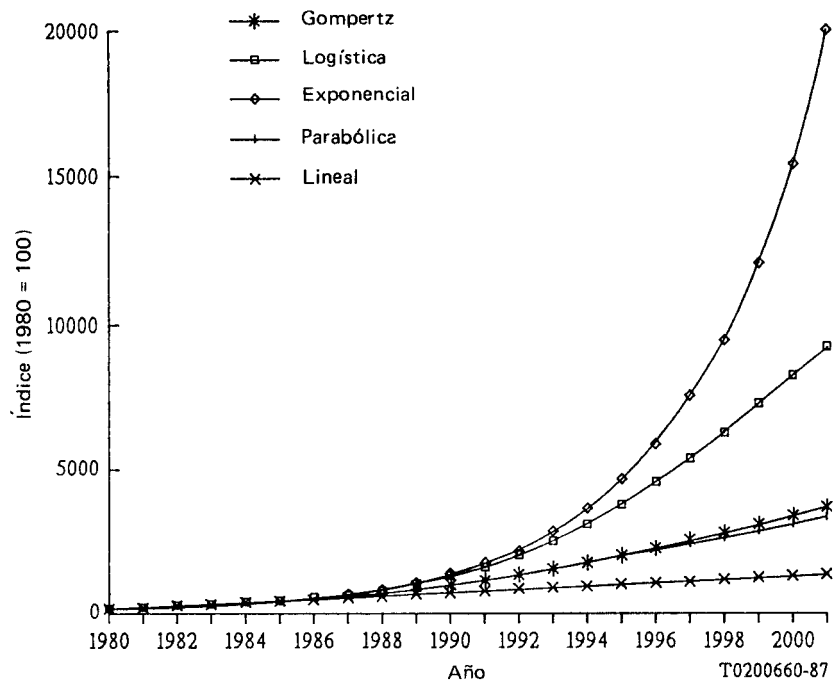


FIGURA 2/E.507

Ejemplo de ajuste del tráfico telefónico internacional utilizando diferentes modelos

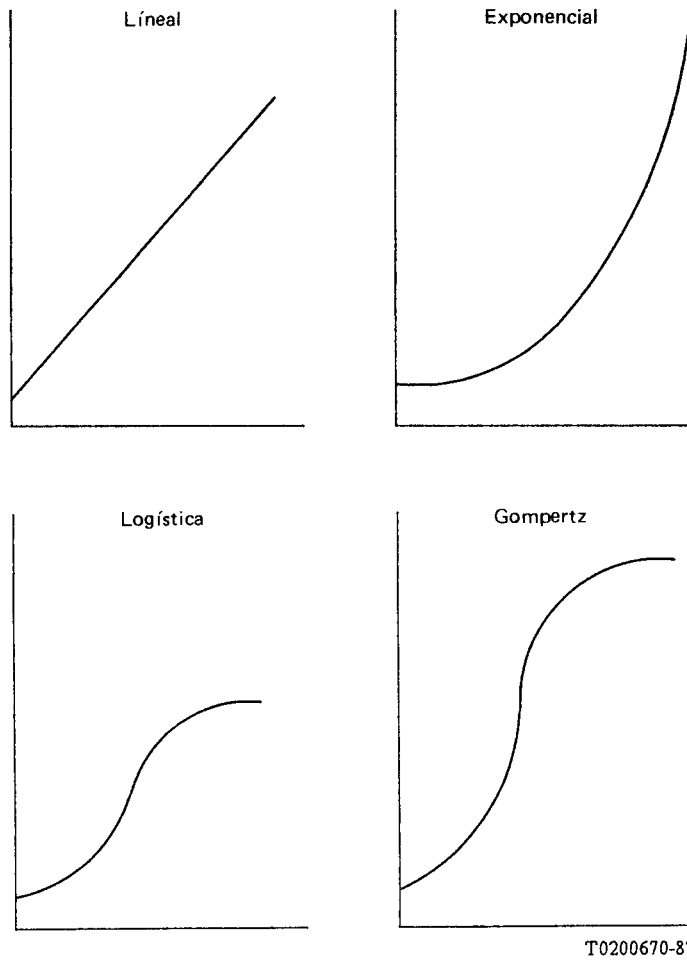


FIGURA 3/E.507

Ejemplos de ajuste de curvas

3.2 Modelos de suavizado

Empleando un proceso de suavizado en el ajuste de curvas, es posible calcular los parámetros del modelo de modo que se ajusten perfectamente a los datos actuales, pero no necesariamente a los datos obtenidos hace ya tiempo.

El proceso de suavizado más conocido es el de la media móvil. El grado de suavizado está controlado por el número de observaciones más recientes incluidas en la media. Todas las observaciones incluidas en la media tienen la misma ponderación.

Además de los modelos de media móvil, existe otro grupo de modelos de suavizado basados en la ponderación de las observaciones. Los modelos más corrientes son:

- suavizado exponencial simple,
- suavizado exponencial doble,
- regresión con descuento,
- método de Holt,
- modelos estacionales de Holt-Winters.

Por ejemplo, en el método del suavizado exponencial, la ponderación de las observaciones anteriores disminuye geoméricamente con el tiempo de acuerdo con la siguiente ecuación:

$$\hat{\mu}_t = (1-a)Y_t + a\hat{\mu}_{t-1} \quad (3-6)$$

donde:

Y_t es el tráfico medido en el instante t ,

μ_t es el nivel estimado en el instante t , y

a es el factor de descuento [y $(1 - a)$ es el parámetro de suavizado].

La repercusión de las observaciones anteriores sobre las previsiones está controlada por la magnitud del factor descuento.

La utilización de modelos de suavizado es especialmente adecuada para previsiones a corto plazo. Para más información véanse [1], [5] y [9].

3.3 Modelos de autorregresión

Si la demanda de tráfico, X_t , en el instante t se puede expresar como una combinación lineal de anteriores observaciones equidistantes de la demanda de tráfico pasada, el proceso es un proceso de autorregresión. En tal caso el modelo viene definido por:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + a_t \quad (3-7)$$

donde

a_t es el ruido blanco en el instante t ;

Φ con $k, k = 1, \dots, p$ son los parámetros de autorregresión.

$AR(p)$ denota el modelo puesto que el orden del modelo es p .

Se pueden estimar los parámetros utilizando análisis de regresión. Debido a tendencias comunes, las variables exógenas ($X_{t-1}, X_{t-2}, \dots, X_{t-p}$) suelen guardar una estrecha correlación. Por consiguiente, las estimaciones de los parámetros estarán correlacionadas. Por otra parte, es difícil hacer pruebas significativas de las estimaciones.

Otra posibilidad consiste en calcular los coeficientes empíricos de autocorrelación y utilizar seguidamente las ecuaciones de Yule-Walker para estimar los parámetros [Φ_k]. Este procedimiento puede aplicarse cuando las series cronológicas [X_t] son estacionarias. Si no lo fuesen, éstas a menudo podrán convertirse en estacionarias, por ejemplo diferenciando las series. El procedimiento de estimación se indica en el § A.1.

3.4 Modelos de autorregresión integrados con media móvil (modelos ARIMA, autoregressive integrated moving average)

Una extensión de la clase de los modelos de autorregresión que incluye los modelos de media móvil se denomina modelos de autorregresión de media móvil (modelos ARIMA, *autoregressive integrated moving average*). Un modelo de media móvil de orden q viene dado por la expresión

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \dots - \theta_q a_{t-q} \quad (3-8)$$

donde

a_t es el ruido blanco en el instante t ;

$[\theta_k]$ son los parámetros de media móvil.

Suponiendo que el término de ruido blanco de los modelos de autorregresión del § 3.3 puede describirse mediante un modelo de media móvil, se obtiene el modelo ARIMA (p, q).

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \dots - \theta_q a_{t-q} \quad (3-9)$$

El modelo ARIMA describe una serie cronológica estacionaria. Si, la serie cronológica es no estacionaria, es necesario diferenciar la serie. Esto se hace como sigue:

Sea Y_t la serie cronológica y B el operador de retroceso

$$X_t = (1 - B)^d Y_t \quad (3-10)$$

donde

d es el número de diferencias para que la serie sea estacionaria.

El nuevo modelo ARIMA (p, d, q) se halla insertando la ecuación (3-10) en la ecuación (3-9).

El método de análisis de las series cronológicas de este tipo fue ideado por G. E. P. Box y G. M. Jenkins [3]. Para analizar dichas series y formular las previsiones correspondientes debe utilizarse generalmente un conjunto de programas destinados al tratamiento de series cronológicas.

Como se indica en la figura 1/E.507, se adopta un modelo provisional. Para ello, se determinan las transformaciones necesarias y el número de parámetros de autorregresión y de media móvil. La elección se basa en la estructura de las autocorrelaciones y de las autocorrelaciones parciales.

El paso siguiente indicado en la figura 1/E.507 es el procedimiento de estimación. Se emplea la estimación por máxima verosimilitud. Desafortunadamente, es difícil hallar estos valores debido a la necesidad de resolver un sistema no lineal de ecuaciones. En la práctica debe disponerse para estos cálculos de un programa de computador. El modelo de previsión se basa en la ecuación (3-9) y el procedimiento para formular previsiones con una anticipación de l unidades de tiempo se indica en el § A.2 del anexo A.

Los modelos de previsión descritos hasta ahora son modelos de una sola variable. También es posible introducir variables explicativas. En este caso el sistema será descrito por un modelo de función de transferencia. Los métodos de análisis de series cronológicas de un modelo de función de transferencia son bastante semejantes a los métodos descritos anteriormente.

En [1], [2], [3], [5], [11], [15] y [17] figuran descripciones detalladas de los modelos ARIMA.

3.5 Modelos espaciales de estados con filtrado de Kalman

Los modelos espaciales de estados constituyen una manera de representar procesos temporales discretos por medio de ecuaciones de diferencias. El método empleado para elaborar los modelos espaciales de estado permite la conversión de cualquier modelo lineal general en una forma adecuada para la estimación recursiva y la previsión. En [1] puede hallarse una descripción más detallada de los modelos espaciales de estado ARIMA.

Para un proceso estocástico, esta representación puede tener la siguiente forma:

$$X_{t+1} = \Phi X_t + Z_t + \omega_t \quad (3-11)$$

y

$$Y_t = H X_t + v_t \quad (3-12)$$

donde

X_t es un vector- s de variable de estado en el instante t ,

Z_t es un vector- s de eventos determinísticos,

Φ es una matriz de transiciones de $s \times s$ que puede, en general depender de t ,

ω_t es un vector- s de errores aleatorios de modelado,

Y_t es un vector- d de medidas en el periodo t ,

H es una matriz de $d \times s$ denominada la matriz de observación, y

v_t es un vector- d de errores de medida.

Tanto ω_t en la ecuación (3-11) como v_t en la ecuación (3-12) son secuencias aleatorias aditivas con estadístico conocido. El valor esperado de cada secuencia es el vector cero y ω_t y v_t satisfacen las condiciones:

$$E[\omega_t \omega_j^T] = Q_t \delta_{ij} \text{ para todos los valores de } t, j, \quad (3-13)$$

$$E[v_t v_j^T] = R_t \delta_{ij} \text{ para todos los valores de } t, j,$$

donde

Q_t y R_t son matrices definidas no negativas,²⁾

y

δ_{ij} es la delta de Kronecker.

Q_t es la matriz de covarianza de los errores de modelado y R_t es la matriz de covarianza de los errores de medida; ω_t y v_t se suponen que no están correlacionadas y se denominan ruido blanco. En otras palabras:

$$E[v_t \omega_j^T] = 0 \text{ para todos los valores de } t, j, \quad (3-14)$$

y

$$E[v_t X_0^T] = 0 \text{ para todos los valores de } t. \quad (3-15)$$

Según las hipótesis formuladas anteriormente, se determinará $X_{t,t}$ de modo que:

$$E[(X_{t,t} - X_t)^T (X_{t,t} - X_t)] = \text{sea mínimo}, \quad (3-16)$$

donde

$X_{t,t}$ es una estimación del vector de estado en el instante t , y

X_t es el vector de variables de estado verdaderas.

La técnica de filtrado de Kalman permite la estimación recursiva de variables de estado para aplicaciones en línea. Esto se hace de la manera siguiente. Suponiendo que hay una variable explicativa Z_t , cuando se dispone de un nuevo punto de datos, se utiliza para actualizar el modelo:

$$X_{t,t} = X_{t,t-1} + K_t (Y_t - HX_{t,t-1}) \quad (3-17)$$

donde

K_t es la matriz de ganancia de Kalman que puede calcularse recursivamente [18].

Intuitivamente, la matriz de ganancia determina qué ponderación relativa se dará al último error de previsión observado para corregirlo. A fin de crear una proyección k pasos más adelante, se utiliza la siguiente fórmula:

$$X_{t+k,t} = \Phi^k X_{t,t} \quad (3-18)$$

donde

$X_{t+k,t}$ es una estimación de X_{t+k} dadas las observaciones Y_1, Y_2, \dots, Y_t .

2) Una matriz A es definida y no negativa si, y sólo si para todos los vectores z , $z^T A z \geq 0$.

Las ecuaciones (3-17) y (3-18) muestran que la técnica de filtrado de Kalman conduce a un procedimiento de previsión conveniente que es recursivo por naturaleza y proporciona estimaciones de varianza mínima, no sesgadas, del proceso temporal discreto de interés.

Para más información, véanse [4], [5], [16], [18], [19] y [22].

El filtro de Kalman funciona bien cuando los datos examinados son estacionales. Los datos de carga de tráfico estacionales pueden representarse con una serie cronológica periódica. De esta manera, puede obtenerse un filtro de Kalman estacional superponiendo un modelo de crecimiento lineal con un modelo estacional. Para un análisis más detallado de las técnicas de filtro de Kalman para datos estacionales, véase [6] y [20].

3.6 Modelos de regresión

Las ecuaciones (3-1) y (3-2) son modelos de regresión típicos. En las ecuaciones, el tráfico, Y_t , es la variable dependiente (o explicativa), mientras que el tiempo t es la variable independiente.

Un modelo de regresión describe una relación lineal entre las variables dependientes y las independientes. Dadas ciertas hipótesis, pueden utilizarse los mínimos cuadrados ordinarios para estimar los parámetros.

Un modelo con varias variables independientes se denomina un modelo de regresión múltiple. El modelo viene dado por:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t \quad (3-19)$$

donde

Y_t es el tráfico en el instante t ,

β_i , $i = 0, 1, \dots, k$ son parámetros,

X_{it} , $i = 1, 2, \dots, k$ es el valor de las variables independientes en el instante t ,

u_t es el término de error en el instante t .

Las variables independientes o explicativas que pueden utilizarse en el modelo de regresión son, por ejemplo, las tarifas, exportaciones, importaciones, grado de automatización. En el § 2 “Datos de base para las previsiones” de la Recomendación E.506 se indican otras variables explicativas.

En [1], [5], [7], [15] y [23] figura una descripción detallada de los modelos de regresión.

3.7 Modelos econométricos

En los modelos econométricos intervienen ecuaciones que relacionan la variable que se desea predecir (la variable dependiente o endógena) con diversas variables socioeconómicas (llamadas variables independientes o explicativas). La forma de las ecuaciones debe reflejar la relación causal esperada entre las variables. Dada una fórmula modelo, se utilizan datos históricos o representativos para estimar los coeficientes de la ecuación. Suponiendo que el modelo sigue siendo válido con el tiempo, pueden utilizarse estimaciones de los valores futuros de las variables independientes para efectuar previsiones de las variables que interesan. En el anexo C se presenta ejemplo de modelo econométrico típico.

Hay una amplia gama de modelos posibles y cierto número de métodos para estimar los coeficientes (por ejemplo, métodos de los mínimos cuadrados, parámetros variables, regresión no lineal, etc.). En muchos aspectos, el conjunto de modelos econométricos disponibles es mucho más flexible que otros modelos. Por mencionar sólo algunos ejemplos, en los modelos econométricos se pueden incorporar efectos retardados, ponderar las observaciones, integrar modelos residuales ARIMA, combinar información de secciones separadas y permitir la variación de los parámetros.

Una de las principales ventajas de la elaboración de un modelo econométrico para la formulación de previsiones, deriva de la necesidad de identificar adecuadamente la estructura o los procesos que generan los datos, y de determinar los trayectos causales correspondientes. Gracias a la identificación explícita de la estructura, las fuentes de error son más fáciles de identificar en los modelos econométricos que en otros tipos de modelos.

Empleando modelos econométricos es fácil detectar los cambios de estructuras y eliminar de los datos históricos los valores atípicos, o bien ponderar debidamente la influencia de éstos. Asimismo, los cambios en los factores que afecten a las variables de que se trate pueden incorporarse fácilmente en la previsión generada a partir de un modelo econométrico.

Con frecuencia pueden elaborarse modelos econométricos bastante fiables con menos observaciones que las necesarias para los modelos de series cronológicas. Con modelos de regresión combinados, un reducido número de observaciones de series cronológicas para varias secciones transversales bastan para elaborar un modelo que puede emplearse para la previsión.

Sin embargo, al estimar el modelo hay que cerciorarse de que se cumplen las hipótesis en que se basan las técnicas que se describen en muchas de las publicaciones citadas en la lista de referencias incluida al final de la presente Recomendación. Por ejemplo, el número de variables independientes que pueden utilizarse viene limitado por la cantidad de datos disponibles para estimar el modelo. Asimismo, hay que evitar las variables independientes intercorrelacionadas. Algunas veces, la correlación entre las variables puede evitarse utilizando datos diferenciados o para los que no se ha considerado la tendencia o aplicando una transformación de variables. Para más detalles, véase [8], [12], [13], [14] y [21].

4 Discontinuidades en el crecimiento del tráfico

4.1 Ejemplos de discontinuidades

Tal vez sea difícil evaluar de antemano la magnitud de una discontinuidad. A menudo, la influencia de los factores que ocasionan las discontinuidades se hace sentir durante un periodo de transición, y la discontinuidad no es tan evidente. Además, resulta difícil identificar con exactitud las discontinuidades derivadas de la implantación, por ejemplo, de un servicio automático internacional, ya que la modificación del modo de explotación va normalmente unida a otras modificaciones (por ejemplo, reducciones de tarifas).

En la figura 4/E.507, se puede observar la influencia que ejercen las discontinuidades en el crecimiento del tráfico.

Se han registrado discontinuidades que han multiplicado por dos, e incluso por más, la intensidad del tráfico cursado. Cabe también observar que pueden registrarse cambios en la tendencia de crecimiento del tráfico después de producirse discontinuidades.

Para las previsiones a corto plazo, tal vez convenga emplear la tendencia del tráfico entre discontinuidades, pero para las previsiones a largo plazo convendrá posiblemente emplear una estimación de la tendencia basada en observaciones a largo plazo, teniendo en cuenta las discontinuidades anteriores.

Además de las fluctuaciones aleatorias debidas a las crestas imprevisibles de tráfico, averías, etc., las medidas de tráfico están sujetas a fluctuaciones sistemáticas, debidas a ciclos diarios o semanales de la intensidad de tráfico, a la influencia de las diferencias horarias, etc.

4.2 Introducción de variables explicativas

La identificación de las variables explicativas para un modelo econométrico es quizás el aspecto más difícil de la elaboración de esos modelos. Las variables explicativas utilizadas en los modelos econométricos representan los principales factores que influyen en la variable cuyo valor se procura determinar. En el § 2 de la Recomendación E.506 figura una lista de variables explicativas.

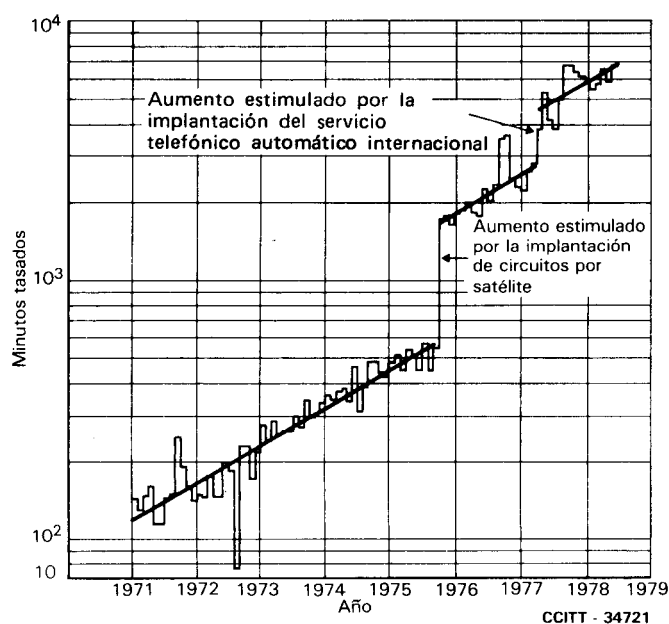


FIGURA 4/E.507

Minutos tasados en comunicaciones telefónicas de salida de Australia hacia Sri Lanka

La teoría económica es el punto de partida para la selección de las variables. Más concretamente, la teoría de la demanda proporciona el marco fundamental para la elaboración del modelo general. No obstante, la descripción de la estructura o el proceso que origina los datos determina a menudo las variables explicativas que se incluyen. Por ejemplo, a fin de describir una estructura de modo apropiado, habrá quizás que incorporar al modelo ciertas relaciones tecnológicas.

Si bien existen algunos criterios para la selección de las variables explicativas (por ejemplo, \bar{R}^2 , el estadístico de Durbin-Watson (D-W), el error cuadrático medio (ECM), y el nivel de previsión *ex-post*, explicados en la lista de referencias), las dificultades estadísticas o la insuficiencia de los datos (ya sean históricos o previstos) limitan el conjunto de variables explicativas posibles, y a menudo hay que recurrir a variables substitutivas. Asimismo, a diferencia de los modelos estadísticos puros, los modelos econométricos admiten variables explicativas, no sólo sobre la base de criterios estadísticos, sino también partiendo de la premisa de que existe una relación de causalidad.

Un modelo econométrico totalmente especificado captará los puntos de inflexión. No habrá discontinuidades en la variable dependiente, a menos que los parámetros del modelo se modifiquen radicalmente en un lapso muy corto. Las discontinuidades del crecimiento del tráfico telefónico constituyen indicaciones de que el mercado o la estructura tecnológica subyacentes han experimentado grandes transformaciones.

Los cambios asociados al crecimiento de la demanda telefónica pueden captarse mediante la regresión de parámetros variables o la introducción de una variable que parezca explicar la discontinuidad (por ejemplo, una variable de publicidad si se estima que la publicidad es la causa del cambio estructural). Las discontinuidades singulares o escalonadas no pueden tratarse mediante la introducción de cambios explicativos: es posible resolver este problema mediante el empleo de variables ficticias.

4.3 Introducción de variables ficticias

En los modelos econométricos, influyen a menudo variables cualitativas. Para medir los efectos de estas variables cualitativas, se utilizan variables ficticias. Con esta técnica, se usa el valor 1 cuando está presente el atributo cualitativo que influye en la variable dependiente y 0 cuando ese atributo está ausente.

Así pues, las variables ficticias son adecuadas para utilizarlas cuando surge una discontinuidad en la variable dependiente. La variable ficticia adoptará, por ejemplo, el valor cero durante el periodo anterior en que las llamadas se cursaban con asistencia de operadora y el valor uno durante el periodo posterior al establecimiento de la marcación directa por el abonado.

Las variables ficticias se utilizan frecuentemente a fin de representar efectos estacionales en la variable dependiente o cuando es menester eliminar el efecto de un valor atípico en los parámetros de un modelo, como un gran salto de la demanda telefónica debido a una huelga en los servicios postales o una aguda disminución debida al deterioro de instalaciones provocado por condiciones climatológicas rigurosas.

El uso indiscriminado de variables ficticias debe desaconsejarse por dos razones:

- 1) las variables ficticias tienden a absorber todo el poder explicativo durante las discontinuidades, y
- 2) esas variables reducen el número de grados de libertad.

5 Evaluación de la especificación del modelo

5.1 Consideraciones generales

Se presentan aquí métodos de prueba de la significación de los parámetros y métodos de cálculo de los intervalos de confianza, para algunos de los modelos de previsión dados en el § 3. En particular, se examinan los métodos relacionados con el análisis de regresión y el análisis de series cronológicas.

Todos los modelos de previsión econométricos que aquí se presentan se consideran modelos de regresión. También pueden considerarse modelos de regresión los modelos de ajuste de curvas explicados en el § 3.1.

Un modelo exponencial dado por

$$Z_t = ae^{bt} \cdot u_t \quad (5-1)$$

puede transformarse a una forma lineal:

$$\ln Z_t = \ln a + bt + \ln u_t \quad (5-2)$$

o

$$Y_t = \beta_0 + \beta_1 X_t + a_t \quad (5-3)$$

donde

$$Y_t = \ln Z_t$$

$$\beta_0 = \ln a$$

$$\beta_1 = b$$

$$X_t = t$$

$$a_t = \ln u_t \text{ (ruido blanco).}$$

5.2 Autocorrelación

Un modelo de previsión adecuado debe dar pequeños residuos autocorrelacionados. Si los residuos están significativamente correlacionados, los parámetros estimados y también las previsiones pueden ser deficientes. Para verificar si los errores están correlacionados, se calcula la función de autocorrelación r_k , $k = 1, 2, \dots$; r_k es la autocorrelación estimada de residuos en el retardo k . Una manera de detectar la autocorrelación entre los residuos es trazar la función de autocorrelación y realizar una prueba de Durbin-Watson. El estadístico de Durbin-Watson es:

$$D - W = \frac{\sum_{t=2}^N (e_t - e_{t-1})^2}{\sum_{t=1}^N e_t^2} \quad (5-4)$$

donde

e_t es el residuo estimado en el instante t ,

N es el número de observaciones.

5.3 Prueba de significación de los parámetros

Un modo de evaluar el modelo de previsión consiste en analizar los efectos de las diferentes variables exógenas. Después de estimar los parámetros del modelo de regresión, se ha de probar la significación de éstos.

En el ejemplo de modelo econométrico del anexo C, se dan los valores estimados de los parámetros. Debajo de estos valores se indica, entre paréntesis, la desviación típica estimada. De acuerdo con una regla práctica, los parámetros se consideran significativos si el valor absoluto de las estimaciones es superior al doble de la desviación típica estimada. Un modo más preciso de probar la significación de los parámetros tiene en cuenta las distribuciones de sus estimadores.

Puede utilizarse el coeficiente de correlación múltiple (o coeficiente de determinación) como criterio para el ajuste de la ecuación.

El coeficiente de correlación múltiple, R^2 , viene dado por:

$$R^2 = \frac{\sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (5-5)$$

Si el coeficiente de correlación múltiple es próximo a 1 el ajuste es satisfactorio. No obstante, un R^2 elevado no implica que la previsión sea exacta.

En los análisis de series cronológicas el examen del modelo se efectúa de otro modo. Como se ha señalado en el § 3.4, el número de parámetros autorregresivos y de media móvil de un modelo ARIMA se determina por un procedimiento de identificación basado en la estructura de la función de autocorrelación y de autocorrelación parcial.

La estimación de los parámetros y de sus desviaciones típicas se lleva a cabo por un procedimiento de estimación no lineal iterativo. Empleando un programa de computador destinado al análisis de series cronológicas, pueden evaluarse las estimaciones de los parámetros mediante el examen de las desviaciones típicas estimadas, del mismo modo que en el análisis de regresión.

Una prueba global del ajuste se basa en el estadístico

$$Q_{N-d} = \sum_{i=1}^N r_i^2 \quad (5-6)$$

donde r_i es la autocorrelación estimada para el periodo i , y d es el número de parámetros del modelo. Cuando el modelo es adecuado, Q_{N-d} presenta aproximadamente la distribución ji-cuadrado (χ^2) con $N - d$ grados de libertad. Para probar el ajuste, el valor Q_{N-d} puede compararse con cuantiles de la distribución ji-cuadrado.

5.4 Validez de las variables exógenas

Los modelos de previsión econométricos se basan en un conjunto de variables exógenas que explican la evolución de la variable endógena (la demanda de tráfico). A fin de efectuar previsiones de la demanda de tráfico, es menester formular previsiones por cada una de las variables exógenas. Es muy importante señalar que una variable exógena no ha de incluirse en el modelo si la previsión de la variable es menos fiable que la previsión de la demanda de tráfico.

Supóngase que la evolución exacta de la variable exógena es conocida, lo que es el caso, por ejemplo, de los modelos sencillos en los cuales la variable explicativa es el tiempo. Si el ajuste del modelo es bueno y el ruido blanco presenta una distribución normal con esperanza matemática nula, es posible calcular límites de confianza para las previsiones. Ello se hace fácilmente mediante un programa de computador.

Sin embargo, los valores de la mayoría de las variables explicativas no pueden predecirse con exactitud. La confianza de la predicción disminuirá en tal caso con el número de periodos. Por consiguiente, las variables explicativas harán que el intervalo de confianza de las previsiones aumente con el número de periodos de previsión. En tales situaciones es difícil calcular un intervalo de confianza en torno a los valores previstos.

Si la demanda de tráfico puede describirse mediante un modelo de autorregresión con media móvil, no se incluyen en el modelo variables explicativas. Por tanto, si el ajuste del modelo no es deficiente, pueden calcularse los límites de confianza de los valores previstos. Ello se hace con el auxilio de un conjunto de programas de análisis de series cronológicas.

5.5 Intervalos de confianza

Los intervalos de confianza, en el contexto de las previsiones, indican construcciones estadísticas de límites o fronteras de previsión o de predicción. Como los modelos estadísticos llevan consigo algunos errores, las estimaciones de los parámetros llevan asociada una cierta variabilidad de sus valores. En otras palabras, aunque se haya identificado el modelo de previsión correcto, la influencia de los factores endógenos causará errores en las estimaciones de los parámetros y en la previsión. Los intervalos de confianza tienen en cuenta la incertidumbre asociada a las estimaciones de los parámetros.

En los modelos causales, otra fuente de incertidumbre en la previsión de la serie que se estudia son las predicciones de las variables explicativas. Este tipo de incertidumbre no puede tratarse por intervalos de confianza y

suele ignorarse, aunque puede ser más importante que la incertidumbre asociada a la estimación de los coeficientes. Por otra parte, la incertidumbre debida a posibles influencias externas no se refleja en los intervalos de confianza.

Para un modelo lineal de regresión estática, el intervalo de confianza de la previsión depende de la fiabilidad de los coeficientes de regresión, de la magnitud de la varianza residual y de los valores de las variables explicativas. El intervalo de confianza del 95% para un valor previsto Y_{N+1} viene dado por:

$$\hat{Y}_N(1) - 2\hat{\sigma} \leq Y_{N+1} \leq \hat{Y}_N(1) + 2\hat{\sigma} \quad (5-7)$$

donde $\hat{Y}_N(1)$ es la previsión del paso siguiente y $\hat{\sigma}$ es el error típico de la previsión.

Lo que quiere decir que esperamos, con una probabilidad del 95%, que el valor real de la serie en el instante $N + 1$ se sitúe dentro de los límites indicados por el intervalo de confianza, suponiendo que no hay errores asociados a la previsión de las variables explicativas.

6 Comparación de modelos de previsión alternativos

6.1 Verificación del diagnóstico - Evaluación de modelos

Las pruebas y verificaciones del diagnóstico son elementos importantes en el procedimiento de elaboración de modelos. La calidad del modelo se caracteriza por los residuos. Los modelos de previsión adecuados deben dar residuos autocorrelacionados pequeños, la varianza de los residuos no debe disminuir ni aumentar y el valor de esperanza matemática de los residuos debe ser cero o próximo a cero. La exactitud de la previsión es afectada por la magnitud del residuo que debe ser pequeño.

Además, los límites de confianza de las estimaciones de parámetros y las previsiones deben ser relativamente pequeños. De la misma manera, el error cuadrático medio debe ser pequeño comparado con los resultados de otros modelos.

6.2 Previsión de los niveles en función de la previsión de los cambios

Muchos modelos econométricos se estiman utilizando los niveles de las variables dependientes e independientes. Dado que las variables económicas evolucionan juntas en el tiempo, se obtienen elevados coeficientes de determinación. La colinealidad entre los niveles de las variables explicativas no plantea un problema cuando el modelo sólo se emplea con fines de previsión ya que las pautas de colinealidad del pasado seguirán existiendo en el futuro. No obstante, cuando se procura medir coeficientes estructurales (por ejemplo, las elasticidades del precio o las de los ingresos) la colinealidad de las variables explicativas (denominada multicolinealidad) hace que los resultados de los coeficientes estimados no sean fiables.

A fin de evitar el problema de multicolinealidad, estimar coeficientes de referencia y formular previsiones, cabe utilizar los cambios de las variables (primera diferencia o primera diferencia logarítmica, que equivale a un cambio porcentual) para estimar un modelo y efectuar predicciones basadas en el mismo. El empleo de cambios de variables para estimar un modelo tiende a eliminar el efecto de multicolinealidad y produce estimaciones más fiables de los coeficientes al suprimir el efecto común de las influencias económicas sobre las variables explicativas.

Si se formulan previsiones a través de los niveles de las variables explicativas y de los cambios de éstas, puede obtenerse una previsión más correcta mediante un proceso de armonización. Este proceso consiste en ajustar los modelos de modo que los dos conjuntos de previsiones den resultados equivalentes.

6.3 Previsiones ex-post

La previsión *ex-post* es la formulación de una predicción empleando un modelo estimado a partir de una submuestra de los datos que comienza con la primera observación y finaliza varios periodos antes de la última observación. En la previsión *ex-post*, se utilizan valores reales de las variables explicativas para formular la predicción. Por otra parte, si se emplean valores previstos de las variables explicativas para efectuar una previsión *ex-post*, puede medirse el error asociado a las variables explicativas incorrectamente previstas.

El propósito de la previsión *ex-post* es evaluar el poder predictivo del modelo mediante la comparación de los valores previstos con los que se hayan registrado realmente en el periodo comprendido entre el fin de la muestra parcial y la última observación. Con la previsión *ex-post*, puede evaluarse la exactitud de las previsiones en lo que se refiere a:

- 1) desviaciones porcentuales de los valores previstos con respecto a los valores reales;
- 2) comportamiento del modelo en los puntos de inflexión;
- 3) comportamiento sistemático de las desviaciones.

Las desviaciones de los valores previstos con respecto a los valores reales proporcionan una idea general de la exactitud del modelo. Las derivas sistemáticas de las desviaciones pueden suministrar información que sirva para modificar el modelo o para reajustar las previsiones teniendo en cuenta esta deriva. De igual importancia para la evaluación del poder predictivo es la correcta indicación del comportamiento en los puntos de inflexión, es decir, la medida en que el modelo es adecuado para predecir los cambios en la variación de la variable dependiente. A continuación se examinan otros criterios de evaluación de la exactitud de las previsiones.

6.4 Criterios de evaluación de la calidad de las previsiones

Un modelo pudiera ajustar muy bien los datos históricos. Sin embargo, cuando se comparan las previsiones con los datos del futuro que no se utilizan para la estimación de parámetros, el ajuste pudiera no ser tan satisfactorio. En consecuencia, la comparación de las previsiones con las observaciones reales puede proporcionar información adicional sobre la calidad del modelo. Supongamos que tenemos la serie cronológica $Y_1, Y_2, \dots, Y_N, Y_{N+1}, \dots, Y_{N+M}$.

Las últimas M observaciones se eliminan de la serie cronológica y del procedimiento de elaboración del modelo. El error de previsión del paso siguiente viene dado por:

$$e_{N+t} = Y_{N+t} - \hat{Y}_{N+t-1} \quad (1) \quad t = 1, 2, \dots, M \quad (6-1)$$

donde

$\hat{Y}_{N+t-1} (1)$ es la previsión del paso siguiente.

Error medio

El error medio, EM, se define por

$$EM = \frac{1}{M} \sum_{t=1}^M e_{N+t} \quad (6-2)$$

El EM es un criterio para evaluar el sesgo de la previsión. Como la esperanza matemática de los residuos debe ser cero, una gran desviación con respecto a cero indica un sesgo en las previsiones.

Error medio en porcentaje

El error medio en porcentaje, EMP, viene definido por

$$EMP = \frac{100}{M} \sum_{t=1}^M \frac{e_{n+1}}{Y_{N+1}} \quad (6-3)$$

También este estadístico indica un posible sesgo de las previsiones. El criterio mide la desviación porcentual en el sesgo. No se recomienda utilizar EMP cuando las observaciones son pequeñas.

Error cuadrático medio

El error cuadrático medio, ECM, de la previsión se define como

$$ECM = \left[\frac{1}{M} \sum_{t=1}^M e_{N+t}^2 \right]^{1/2} \quad (6-4)$$

El ECM es la medida utilizada más corrientemente para evaluar la precisión de las previsiones.

Error medio absoluto

El error medio absoluto, EMA, viene dado por

$$EMA = \frac{1}{M} \sum_{t=1}^M |e_{N+t}| \quad (6-5)$$

Coefficiente de desigualdad de Theil

El coeficiente de desigualdad de Theil se define del modo siguiente:

$$U = \left[\sum_{t=1}^M \frac{e_{N+t}^2}{Y_{N+1}^2} \right]^{1/2} \quad (6-6)$$

Se prefiere utilizar como medida de la exactitud de las previsiones el coeficiente U de Theil ya que la discrepancia entre los valores previstos y los valores reales puede descomponerse en errores debidos a:

- 1) la tendencia central,
- 2) la variación desigual entre los cambios previstos y los reales, y
- 3) la covariación incompleta entre los cambios previstos y los reales.

Este modo de descomponer los errores de previsión puede utilizarse para ajustar el modelo a fin de aumentar su exactitud.

Otra cualidad que debe poseer un modelo de previsión es la capacidad de captar los puntos de inflexión. Es decir, que la variación de los valores previstos cambie de dirección en el mismo periodo en que cambie la dirección de los valores de la serie real estudiada. Si se estima un modelo abarcando un largo periodo de tiempo que contiene varios puntos de inflexión, el análisis de las previsiones *ex-post* permite detectar generalmente la incapacidad del modelo para seguir de cerca la evolución de los valores reales cuando presentan puntos de inflexión.

7 Elección del modelo de previsión

7.1 Calidad de las previsiones

Si bien la elección de un modelo de previsión se basa generalmente en la calidad de las previsiones, debe atenderse también a otras consideraciones. En efecto, han de tenerse en cuenta la longitud del periodo de previsión, la forma funcional del modelo y la exactitud con que pueden preverse las variables explicativas de un modelo econométrico.

La longitud del periodo de previsión influye, junto con las limitaciones de los datos históricos y el propósito con que se utilizará el modelo, en la decisión de utilizar un tipo de modelo en lugar de otro. Por ejemplo, los modelos de autorregresión integrados con media móvil (ARIMA) pueden ser apropiados para las previsiones a corto plazo cuando no se trata de la inestabilidad, se dispone de suficientes datos históricos y las relaciones de causalidad no son de interés. Asimismo, cuando es difícil identificar la estructura que genera los datos, no se dispone de otra opción que utilizar un modelo de previsión basado en los datos históricos de la variable que se desea predecir.

También debe tenerse en cuenta para la elección del modelo de previsión la forma funcional de éste. Aunque es verdad que un modelo más complejo puede reducir el error de especificación del modelo, también es verdad que, en general, aumentará considerablemente el efecto de los errores en los datos. Al elegir la forma del modelo se debe tener en cuenta una solución de compromiso entre estas fuentes de error.

Otra consideración que influye en la elección de un modelo de previsión es la disponibilidad de predicciones de las variables explicativas y la fiabilidad de estas predicciones. Un modelo superior que utilice variables explicativas imposibles de prever con exactitud, puede resultar inferior a un modelo ordinario cuyas variables explicativas se prevean con exactitud.

Si se trata de la inestabilidad del mercado, han de utilizarse para las previsiones modelos econométricos con los que se puedan tener en cuenta las transformaciones estructurales. Cuando interesan las relaciones de causalidad, no cabe emplear como instrumentos de previsión los modelos simples o los modelos ARIMA. Tampoco pueden utilizarse estos modelos si los datos históricos de que se dispone son insuficientes. Por último, cuando el propósito del modelo sea predecir los efectos asociados con los cambios en los factores que influyen en la variable de que se trata, los modelos de series cronológicas pueden ser apropiados (con excepción, por supuesto, de los modelos de función de transferencia y los modelos de múltiples series cronológicas).

7.2 Duración del periodo de previsión

Para la ampliación normal del equipo de conmutación y la adición de circuitos, es necesario un periodo de previsión de unos seis años. Sin embargo, la planificación de nuevos cables u otros medios de transmisión, o de grandes instalaciones de planta, puede requerir un periodo de previsión más dilatado. Las previsiones a largo plazo serán forzosamente menos exactas que las realizadas a corto plazo, pero resultan aceptables.

Al formular previsiones con un modelo estadístico, la duración del periodo abarcado estará enteramente determinada por los siguientes elementos:

- a) los datos históricos disponibles,
- b) la finalidad o la utilización de las previsiones,
- c) la estructura del mercado del que proceden los datos,
- d) el modelo de previsión utilizado, y

e) la frecuencia de los datos.

Los datos históricos disponibles dependen del periodo en el que se han recopilado y la frecuencia de recopilación (o la duración del periodo en el que los datos se han acumulado). Una reducida base de datos históricos sólo servirá para un corto intervalo de previsión. Por ejemplo, si se dispone de 10 ó 20 observaciones, puede utilizarse un modelo para efectuar previsiones que abarquen 4 ó 5 periodos a partir del final de la muestra. En cambio, con 150 ó 200 observaciones, pueden formularse en condiciones análogas previsiones fiables para 30 a 50 periodos.

La finalidad de la previsión influirá evidentemente en el número de periodos que ésta deberá comprender. Para planificar a largo plazo las instalaciones, se necesitan previsiones que abarquen de 15 a 20 años o más. Para la evaluación de las modificaciones de tarifas pueden necesitarse previsiones que abarquen sólo dos o tres años. La modificación de las disposiciones de encaminamiento puede exigir sólo previsiones que se extiendan a algunos meses con posterioridad a la muestra.

La estabilidad o la inestabilidad del mercado influye también en el periodo de previsión. Cuando la estructura del mercado es estable, el periodo de previsión podría llegar a ser igual al periodo histórico considerado. Los mercados muy inestables no proporcionan las mismas facilidades y la previsión sólo puede abarcar un reducido número de periodos.

La naturaleza del modelo utilizado para las previsiones influye en la decisión relativa al tiempo que cabe razonablemente abarcar. Para las previsiones a largo plazo, los modelos estructurales tienden a dar mejores resultados que otros modelos, mientras que a corto plazo todos los modelos parecen dar resultados igualmente buenos.

Debe señalarse que, si bien el propósito de la previsión y el modelo utilizado influyen en el tiempo abarcado, el número de periodos que ha de comprender la previsión es crucial para la elección del modelo de previsión y de las aplicaciones a que se destinará.

ANEXO A

(a la Recomendación E.507)

Descripción de los procedimientos de previsión

A.1 Estimación de los parámetros de autorregresión

La autocorrelación empírica con un retardo k viene dada por la expresión

$$r_k = \frac{v_d}{v_0} \quad (\text{A-1})$$

donde

$$v_k = \frac{1}{N-1} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (\text{A-2})$$

y

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t \quad (\text{A-3})$$

siendo N el número total de observaciones.

La relación entre $[r_k]$ y las estimaciones $[\hat{\Phi}_k]$ de $[\Phi_k]$ se expresa mediante las ecuaciones de Yule-Walker:

$$\begin{aligned} r_1 &= \hat{\Phi}_1 + \hat{\Phi}_2 r_1 + \dots + \hat{\Phi}_p r_{p-1} \\ r_2 &= \hat{\Phi}_1 r_1 + \hat{\Phi}_2 r_2 + \dots + \hat{\Phi}_p r_{p-2} \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned} \quad (\text{A-4})$$

$$r_p = \hat{\Phi}_1 r_{p-1} + \hat{\Phi}_2 r_{p-2} + \dots + \hat{\Phi}_p$$

Por consiguiente, los estimadores $[\hat{\Phi}_k]$ pueden hallarse resolviendo el sistema de ecuaciones.

Para los cálculos, la resolución directa de las ecuaciones puede sustituirse por el siguiente procedimiento recursivo. Sean $[\hat{\Phi}_{k,j}]_j$ estimadores de los parámetros con retardo $j = 1, 2, \dots, k$ cuando el número total de los parámetros es k . En tal caso, se hallan los estimadores $[\hat{\Phi}_{k+1,j}]_j$ mediante

$$\hat{\Phi}_{k+1,k+1} = \frac{r_{k+1} \sum_{j=1}^k \hat{\Phi}_{k,j} r_{k+1-j}}{1 - \sum_{j=1}^k \hat{\Phi}_{k,j} r_j} \quad (\text{A-5})$$

$$\hat{\Phi}_{k+1,j} = \hat{\Phi}_{k,j} - \hat{\Phi}_{k+1,k+1} \hat{\Phi}_{k,k-j+1} \quad j = 1, 2, \dots, k \quad (\text{A-6})$$

Si se define $\hat{\Phi}_{p,j} = \hat{\Phi}_j$, con $j = 1, 2, \dots, p$, la previsión de la demanda del tráfico para el instante $t + 1$ se expresa mediante la fórmula

$$X_{t+1} = \hat{\Phi}_1 X_t + \hat{\Phi}_2 X_{t-1} + \dots + \hat{\Phi}_p X_{t-p} \quad (\text{A-7})$$

A.2 Previsión con modelos ARIMA

La previsión efectuada con l unidades de tiempo de antelación viene dada por la expresión:

$$\begin{aligned} \hat{X}_t(l) = & \hat{\Phi}_1 [X_{t+l-1}] + \hat{\Phi}_2 [X_{t+l-2}] \\ & + \dots + \hat{\Phi}_p [X_{t+l-p}] \\ & + [a_{t+l}] - \hat{\theta}_1 [a_{t+l-1}] \\ & - \hat{\theta}_2 [a_{t+l-2}] - \dots - \hat{\theta}_q [a_{t+l-q}], \end{aligned} \quad (\text{A-8})$$

$$\text{donde } [\hat{X}_j] = \begin{cases} \hat{X}_t(j-t) & \text{si } j > t \\ X_j & \text{si } j \leq t \end{cases} \quad (\text{A-9})$$

$$[a_j] = \begin{cases} 0 & \text{si } j > t \\ X_j - \hat{X}_j & \text{si } j \leq t, \end{cases} \quad (\text{A-10})$$

lo que significa que $[X_j]$ se define como una previsión cuando $j > t$ y como una observación real en los demás casos y que $[a_j]$ se anula por definición cuando $j > t$ dado que el valor de la esperanza matemática del ruido blanco es cero. Si se conocen las observaciones ($j \leq t$), $[a_j]$ es igual al residuo.

ANEXO B

(a la Recomendación E.507)

Filtro de Kalman para un modelo de tendencia lineal

Para elaborar modelos del tráfico telefónico, se supone que no hay cambios determinísticos en el esquema de demanda. Puede elaborarse el modelo de esta situación fijando el componente determinístico Z_t a cero. Entonces el modelo espacial de estados es:

$$\begin{aligned} X_{t+1} &= \varphi X_t + \omega_t \\ Y_t &= H X_t + v_t \end{aligned} \quad (\text{B-1})$$

donde

X_t es un vector-s de variables de estado en el periodo t ,

Y_t es un vector-s de medidas en el año t ,

ϕ es una matriz de transición de $s \times s$ que puede, en general depender de t ,

y

ω_t es un vector-s de errores de modelado aleatorios,

v_t es el error de medidas en el año t .

Para modelar la demanda del tráfico telefónico, adáptese un modelo simple de variable de un dato y dos estados, definido por:

$$X_{t+1} = \begin{bmatrix} x_{t+1} \\ \dot{x}_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} + \begin{bmatrix} \omega_t \\ \dot{\omega}_t \end{bmatrix} \quad (\text{B-2})$$

e

$$y_t = x_t + v_t \quad (\text{B-3})$$

donde

x_t es la carga verdadera en el año t ,

\dot{x}_t es el crecimiento incremental verdadero en el año t ,

y_t es la carga medida en el año t ,

v_t es el error de medida en el año t .

Así pues, como en nuestro modelo

$$\phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \text{y } H = 1.$$

(B-4)

La proyección del escalón siguiente se escribe como sigue:

$$X_{t+1,t} = \begin{bmatrix} x_{t+1,t} \\ \dot{x}_{t+1,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,t} \\ \dot{x}_{t,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,t-1} + \alpha_t(y_t - x_{t,t-1}) \\ \dot{x}_{t,t-1} + \beta_t(y_t - x_{t,t-1}) \end{bmatrix} \quad (\text{B-5})$$

donde

$X_{t+1,t}$ es la proyección de las variables de estado en el periodo $t + 1$ dadas las observaciones a través del año t .

Los coeficientes α_t y β_t son las matrices de ganancia de Kalman en el año t . Se escriben de nuevo las ecuaciones anteriores y se tiene:

$$x_{t,t} = (1 - \alpha_t)x_{t,t-1} + \alpha_t y_t \quad (\text{B-6})$$

y

$$\dot{x}_{t,t} = (1 - \beta_t) \dot{x}_{t,t-1} + \beta_t (y_t - x_{t-1,t-1}) \quad (\text{B-7})$$

El filtro de Kalman crea una tendencia lineal para cada serie cronológica prevista sobre la base de observaciones actuales o medidas de la demanda del tráfico y la previsión del año anterior de dicha demanda. La observación y la carga de tráfico prevista se combinan para producir una carga suavizada que corresponde al nivel del proceso, y un incremento de crecimiento suavizado. Los valores de ganancia de Kalman α_t y β_t pueden ser fijos o adaptativos. En [16] Moreland presenta un método para seleccionar parámetros consistentes fijos que proporcionan una calidad adecuada independiente del ruido del sistema, del error de medida y de las condiciones iniciales. Para más detalles sobre la selección apropiada de estos parámetros véanse [6], [20] y [22].

ANEXO C

(a la Recomendación E.507)

Ejemplos de un modelo econométrico

Para ejemplificar la aplicación de un modelo econométrico, hemos elegido el modelo de minutos tasados de comunicaciones de Estados Unidos de América al Brasil. Se eligió este modelo entre distintos modelos posibles por las siguientes razones:

- a fin de suministrar un ejemplo de la introducción de variables explicativas;
- para poner de relieve las dificultades que plantean los modelos utilizados a la vez para la estimación de la estructura y para fines de previsión y
- para mostrar el modo en que las transformaciones pueden influir en los resultados.

La demanda de minutos tasados de comunicaciones de Estados Unidos al Brasil (MIN) se estima mediante una ecuación log-lineal que incluye como variables explicativas los “mensajes” tasados de Estados Unidos al Brasil (MSG), el índice de las tarifas telefónicas reales (RPI), los ingresos personales en Estados Unidos a precios de 1972 ($YP72$) y el comercio bilateral real entre Estados Unidos y el Brasil (RTR). El modelo está representado por la siguiente expresión:

$$\ln(MIN)_t = \beta_0 + \beta_1 \ln(MSG)_t + \beta_2 \ln(RPI)_t + \beta_3 \ln(YP72)_t + \beta_4 \ln(RTR)_t + u_t \quad (\text{C-1})$$

donde u_t es el término de error de la regresión y donde se prevé que $\beta_1 > 0$, $\beta_2 < 0$, $\beta_3 > 0$ et $\beta_4 > 0$.

Utilizando un tipo de regresión curvilínea (*ridge regression*) para tratar los graves problemas de multicolinealidad, estimamos la ecuación sobre el intervalo 1971:1 (es decir, primer trimestre de 1971) a 1979:4 y obtenemos los resultados siguientes:

$$\ln(MIN)_t = -3,489 + 0,619 \ln(MSG)_t - 0,447 \ln(RPI)_t + 1,166 \ln(YP72)_t + 0,281 \ln(RTR)_t$$

(0,035) (0,095) (0,269) (0,084) (C-2)

$$\bar{R}^2 = 0,985, \quad SER = 0,083, \quad D-W = 0,922, \quad k = 0,10 \quad (\text{C-3})$$

donde \bar{R}^2 es el coeficiente de determinación ajustado, SER es el error típico de la regresión, $D-W$ es el estadístico de Durbin-Watson y k es la constante de la regresión curvilínea. Los valores que figuran bajo la ecuación anterior entre paréntesis son la desviación típica estimada de los parámetros estimados $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ y $\hat{\beta}_4$.

La introducción de los mensajes como variable explicativa en este modelo fue impuesta por el hecho de que la calidad de transmisión ha aumentado desde mediados del decenio de 1970, la relación respuesta/toma (tasa de compleción) se ha elevado y, al mismo tiempo, el fuerte crecimiento de este mercado ha comenzado a disiparse. Asimismo, las tasas de crecimiento de algunos periodos no podrían explicarse por medidas relativas a las tarifas en uno u otro extremo o por cambios en los ingresos personales reales en Estados Unidos. El comportamiento de la variable de mensajes en la ecuación representativa de los minutos tasados puede dar cuenta de todos estos factores.

Dado que el modelo se emplea con un doble propósito, la estimación de la estructura y la previsión, se introduce por lo menos una variable más que las que se emplearían si el modelo se utilizase exclusivamente con fines de previsión. La introducción de variables explicativas adicionales conduce a graves problemas de multicolinealidad y exige el uso de una regresión curvilínea que disminuye \bar{R}^2 y el estadístico de Durbin-Watson. Por tanto, el poder predictivo del modelo se reduce en cierta medida.

El efecto de la transformación de las variables de un modelo puede apreciarse en el análisis de las previsiones *ex-post* realizadas con el modelo de minutos tasados de comunicaciones de Estados Unidos al Brasil. Cuando se utilizan los niveles de las variables, las desviaciones son mayores que las que se registran con el empleo de los logaritmos de las mismas, los que se usaron para obtener un mejor ajuste (el ECM estimado para el modelo de regresión log-lineal es 0,119 827). En el cuadro C-1/E.507 se indican los resultados de las previsiones cuando se emplean los niveles de las variables y cuando se utiliza la forma logarítmica.

CUADRO C-1/E.507

	Logaritmos			Niveles		
	Previsión	Valor real	Desviación (%)	Niveles	Previsión	Desviación (%)
1980: 1	14,858	14,938	-0,540	2 836 269	3 073 697	- 7,725
2	14,842	14,972	-0,872	2 791 250	3 180 334	-12,234
3	14,916	15,111	-1,296	3 005 637	3 654 092	-17,746
4	14,959	15,077	-0,778	3 137 698	3 529 016	-11,089
1981: 1	15,022	15,102	-0,535	3 341 733	3 621 735	- 7,731
2	14,971	15,141	-1,123	3 175 577	3 762 592	-15,601
3	15,395	15,261	0,879	4 852 478	4 244 178	14,333
4	15,405	15,302	0,674	4 901 246	4 421 755	10,844
1982: 1	15,365	15,348	0,110	4 709 065	4 630 238	1,702
2	15,326	15,386	-0,387	4 528 947	4 807 901	- 5,802

Referencias

- [1] ABRAHAM (A.) y LEDOLTER (J.): *Statistical methods for forecasting*, J. Wiley, Nueva York, 1983.
- [2] ANDERSON (O. D.): *Time series analysis and forecasting. The Box-Jenkins approach*. Butterworth, Londres, 1976.
- [3] BOX (G. E. P.) y JENKINS (G. M.): *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976.
- [4] BROWN (R. G.): *Introduction to random signal analysis and Kalman Filtering*. John Wiley & Sons, Nueva York, 1983.
- [5] Manual del CCITT: Datos de planificación y métodos de previsión, Vol. I y II, UIT, Ginebra, 1988.
- [6] CHEMOUIL (P.) y GARNIER (B.): *An Adaptive Short-Term Traffic Forecasting Procedure Using Kalman Filtering*. ITC 11, Tokyo, 1985.
- [7] DRAPER (N.) y SMITH (H.): *Applied Regression Analysis*, Second Edition, John Wiley & Sons, Nueva York, 1981.
- [8] DUTTA (M.): *Econometric Methods*, South-Western Publishing Co., Cincinnati, 1975.
- [9] GARDNER (E. S. Jr.): *Exponential smoothing the state of art*, *Journal of forecasting*, 4, pp. 1-28, 1985.
- [10] GILCHRIST (W.): *Statistical forecasting*, John Wiley & Sons, Nueva York, 1976.
- [11] GRANGER (C. W. J.) y NEWBOLD (P.): *Forecasting Economic Time Series*, Academic Press, Nueva York, 1977.
- [12] JOHNSTON (J.): *Econometric Methods*, Second Edition, McGraw-Hill, Nueva York, 1972.
- [13] JUDGE (G. G.) y otros: *The Theory and Practice of Econometrics*, John Wiley & Sons, Nueva York, 1980.
- [14] KMENTA (J.): *Elements of Econometrics*, Macmillan Publishing Company, Nueva York, 1971.

- [15] MAKRIDAKIS (S.), WHEELWRIGHT (S. C.) y McGEE (V. E.): Forecasting Methods and applications, Second Edition, *John Wiley & Sons*, Nueva York, 1983.
- [16] MORELAND (J. P.): A robust sequential projection algorithm for traffic load forecasting, *The Bell Technical Journal*, Vol. 61, No. 1, 1982.
- [17] NELSON (C. R.): Applied Time Series Analysis for Managerial Forecasting, *Holden-Day*, San Francisco, 1973.
- [18] PACK (C. D.) y WHITAKER (B. A.): Kalman filter models for network forecasting, *The Bell Technical Journal*, Vol. 61, No. 1, pp. 1-9, 1982.
- [19] SORENSON (H. W.): Kalman filtering techniques. Advances in control systems theory and applications, *Academic Press*, Vol. 3, pp. 219-292, 1966.
- [20] SZELAG (C. R.): A short-term forecasting algorithm for trunk demand servicing, *The Bell Technical Journal*, Vol. 61, No. 1, pp. 67-96, 1982.
- [21] THEIL (H.): Principles of Econometrics, *John Wiley & Sons*, Nueva York, 1971.
- [22] TOME (F. M.) y CUNHA (J. A.): Traffic forecasting with a state space model. *ITC 11*, Tokyo, 1985.
- [23] WONNACOTT (T. H.) y WONNACOTT (R. J.): Regression, *John Wiley & Sons*, Nueva York, 1981.

Bibliografía

PINDYCK (R. S.) y RUBINFELD (D. F.): Econometric Models and Econometric Forecasts, *McGraw-Hill*, Nueva York, 1981.

SASTRI (T.): A state space modelling approach for time series forecasting, *Management Science*, Vol. 31, No. 11, pp. 1451-1470, 1985.