

CRITICA DE PROGRAMAS DE ORDENADOR

SCA Statistical System

Por

FERNANDO ALVAREZ GARRIDO

DOLORES GARCIA MARTOS

(Instituto Nacional de Estadística)

1. INTRODUCCION

SCA Statistical System es un paquete de análisis estadístico, econométrico y de series temporales producido por Scientific Computing Associates, que surge en 1981 por iniciativa de un grupo de conocidos especialistas en el terreno del análisis de series de tiempo y la computación (G.E.P. Box, G. B. Hudak, L. M. Liu, M. E. Muller y G. C. Tiao). Vino a cubrir una importante laguna existente en el software estadístico y econométrico de aquel entonces, que había incorporado sólo parcialmente y con bastantes restricciones los rápidos avances de la teoría y técnicas de análisis de series temporales producidos durante la década anterior. A partir de la primera versión, comercializada en 1983, y con el fin de constituirse en un gran paquete estadístico integrado, ha ido ampliando el espectro de sus posibilidades a otros campos del análisis estadístico.

2. CARACTERISTICAS GENERALES

SCA esta conformado actualmente por dos paquetes de programas, cada uno de los cuales está constituido por varios módulos independientes. Estos paquetes y módulos son los siguientes:

QSCA.- Paquete de análisis estadístico en general y control de calidad y productividad, integrado por los módulos GSA (análisis estadístico general) y QPI (control de calidad y productividad).

SCA.- Paquete de análisis de series temporales y modelos econométricos, constituido por los módulos UTS (análisis univariante y uniecuacional de series temporales), MTS (análisis multivariante de series temporales) y ECOM/M. (modelos econométricos y procedimientos de desestacionalización).

SCA esta disponible en versiones para grandes ordenadores y PCs. Puede utilizarse de forma interactiva o en batch. En la versión para PCs sólo están disponibles los módulos GSA, QPI y UTS, y la identificación de modelos multivariantes. Los dos restantes requieren unos recursos informáticos que no alcanzan a cubrir los actuales PCs de 16 bits (desconocemos si los nuevos PCs de 32 bits permitirán la incorporación de estos módulos). Los requisitos informáticos mínimos para la utilización de SCA-PCs son el disponer del sistema operativo MS-DOS 2.1 o posterior, de memoria RAM de al menos 512K y de disco duro.

Tres aspectos son, a nuestro entender, los que resumen los rasgos más característicos de SCA:

a) Es, en el campo del análisis de series temporales (univariante y multivariante), y en parte en el de modelos econométricos clásicos, el paquete de programas más completo, avanzado y versátil.

b) Es de fácil uso y su asimilación por parte de usuarios principiantes es muy rápida. La sintaxis del lenguaje del paquete se acerca más a un esquema lingüístico (el del inglés básico) que a uno informático, salvo en las típicas sentencias de asignación (funciones matemáticas, por ejemplo) y de control (instrucciones habituales de los lenguajes de programación). Las instrucciones para la realización de un procedimiento concreto se articulan en un "párrafo" compuesto de una o varias "frases". La primera frase de cada párrafo, que hace referencia al tipo de análisis o tratamiento a realizar y a las variables o modelos que intervienen en el mismo, es siempre obligatoria, mientras que las restantes suelen ser opcionales. La selección de opciones por defecto (DEFAULT) que toma SCA cuando el usuario no especifica las suyas propias es, a nuestro entender, bastante acertada y coherente con las recomendaciones de la teoría y práctica econométrica.

c) Cumple los requerimientos exigibles a un paquete de análisis estadístico avanzado: entrada y salida suficientemente flexible de datos de y hacia ficheros externos, manejo de ficheros externos sin necesidad de cerrar la sesión de trabajo, dispone de interfases para el intercambio de datos con otros conocidos paquetes estadísticos, permite la creación dinámica de variables y matrices, realiza las funciones algebraicas escalares habituales y las principales operaciones matriciales, contempla la utilización de MACROS, dispone de posibilidades de programación, etc.

3. SCA EN EL ANALISIS ECONOMETRICO Y DE SERIES TEMPORALES

Como se apuntaba en el apartado anterior, uno de los rasgos distintivos de SCA respecto a otros paquetes de análisis estadístico es la atención especial que en él se presta a los procedimientos de modelización y análisis econométrico y de series temporales. En efecto, SCA contempla exhaustivamente todas las posibilidades de análisis ARIMA de series temporales propuestas en la literatura científica y que hasta ahora

han sido posible implantar computacionalmente, así como unas muy amplias posibilidades de modelización desde un punto de vista econométrico "clásico".

Los procedimientos de análisis econométrico y de series temporales que incorpora comparten, en general, un esquema común. Su punto de partida, una vez definidas en la sesión de trabajo las variables a analizar, es formular para ellas una tentativa de especificación de modelo o modelos, que se va a fundamentar, si es de series temporales, en el completo panel de posibilidades estadísticas y gráficos de identificación que ofrece SCA. La tentativa de especificación de modelo se plasma en una sentencia específica para este fin, a la cual se asigna un nombre que será el que se utilice en todos los procedimientos posteriores que hagan referencia a dicho modelo.

Quisieramos resaltar la importancia de las sentencias de especificación de modelos, muy características de SCA, por varios motivos:

En primer lugar, todos los procedimientos que hacen referencia a algún modelo (estimación, predicción, simulación, filtrado, etc.) requieren que este, durante la sesión de trabajo y mediante la correspondiente sentencia, haya sido especificado y se le haya asignado un nombre. Las sentencias de los procedimientos mencionados requerirán citar, como dato obligatorio, el nombre del modelo o modelos a tratar. Un ejemplo de la utilidad de este enfoque es el de los procedimientos de predicción para funciones de transferencia (uniecuaionales o multiecuaionales), los cuales requieren predicciones de las variables input que, en SCA, son generadas sin pasos intermedios con solo introducir el nombre de algún modelo definido anteriormente para ellas en la sentencia de predicción de la función de transferencia.

En segundo lugar, las sentencias de especificación de modelos son muy sencillas y cómodas, siendo su esquema básico la secuencia: Nombre de la sentencia de especificación de modelo-Nombre del modelo-Formulación del modelo. Aunque hay varias formas de explicitar la estructura de un modelo, la algebraica es la que resulta cómoda para el analista que tiene que modelizar muchas series o probar múltiples especificaciones, pues le evita el trabajo de "descifrar" la codificación del modelo en que se incurre cuando se trabaja con paquetes de programas que utilizan una notación más compacta (SCA permite también definir modelos con una notación simplificada). Para aclarar más este aspecto, comentaremos a continuación la siguiente sentencia de especificación de un modelo de función de transferencia:

UTSMODEL NAME IS FT.

MODEL IS IPI((1-B) (1-B**12)) = C + (W0+W1*B)NPROD((1-B**12)) +
 (1-TH1*B-TH4*B**4) / (1-PH1*B) (1-PH12*B**12-PH24*B**24) NOISE.

En ella, FT es el nombre del modelo, IPI y NPROD son las variables output e input respectivamente. Las diferenciaciones aparecen expresamente, entre paréntesis, incorporando el operador de retardos "B" elevado a diferentes potencias. Igualmente aparecen los distintos polinomios que definen la función de transferencia y el esquema ARMA

del ruido. NOISE se incluye para indicar que el filtro al que multiplica es precisamente el esquema ARMA del ruido.

Por lo que respecta a las posibilidades concretas de análisis econométrico y de series temporales, a continuación se detallan las más relevantes:

- **Análisis univariante de series temporales:** Identificación mediante ACF, PACF, IACF y EACE, estimación máximo-verosímil exacta o condicional, detección de valores anómalos y medidas espectrales univariantes.

- **Análisis de intervención:** Creación de variables impulso, escalón y rampa y estimación máximo-verosímil exacta o condicional de su filtro ARMA conjuntamente con el del ruido, tanto con un esquema aditivo como multiplicativo. Similarmente, y a fin de estimar el efecto de los días laborables y de Pascua, SCA crea las variables a introducir en los modelos.

- **Análisis de función de transferencia:** Identificación mediante CCF y CORNER, estimación máximo-verosímil exacta o condicional de modelos multi-input y medidas espectrales entre dos variables.

- **Análisis multivariante de series temporales:** Identificación mediante CCM, ECCM y matrices de autocorrelación parcial obtenidas por regresión BACKSTEP, medidas de dimensionalidad de las matrices de variables, residuos y parámetros, análisis canónico dinámico y estimación máximo-verosímil exacta o condicional sin limitaciones en las restricciones de las matrices de parámetros.

- **Modelos de regresión uniecuacionales:** Estimación OLS y GLS. Estimación de modelos con errores autorregresivos (Cochrane-Orcutt y Hildreth-Lu).

- **Modelos multiecuacionales de ecuaciones aparentemente no relacionadas:** Estimación máximo-verosímil de sistemas de funciones de transferencia aparentemente no relacionadas, admitiéndose esquemas ARMA para los inputs y los ruidos de cada ecuación.

- **Modelos multiecuacionales simultáneos:** Estimación FIML de sistemas de funciones de transferencia simultáneas, admitiéndose esquemas ARMA para los inputs y los ruidos de cada ecuación.

- **Filtrado y simulación en el ámbito temporal mediante filtros ARMA o funciones de transferencia,** tanto univariantes como multivariantes, de ruidos generados por SCA o variables definidas en la sesión de trabajo.

- **Desestacionalización por un procedimiento basado en modelos (enfoque de Hilmer-Tiao) e interfase para la creación de los ficheros de entrada del X11-ARIMA.**

- **Cálculo de los parámetros phi y psi asociados al esquema ARMA de un modelo. Verificación de las condiciones de estacionariedad e invertibilidad de los polinomios AR y MA y cálculo de sus raíces.**

- Predicción basada en modelos estimados o preespecificados por el analista: univariantes y multivariantes de series temporales, con análisis de intervención, de función de transferencia uniecuacional o multiecuacional y modelos de regresión en general.

- Predicción por procedimientos de alisado: Alisado exponencial simple y doble, método de Holt, método de Winters (aditivo y multiplicativo) y alisado exponencial con indicadores estacionales y funciones armónicas.

4. ANALISIS ESTADISTICO GENERAL Y DE CONTROL DE CALIDAD Y PRODUCTIVIDAD EN SCA

Las posibilidades de análisis estadístico general de SCA son, a grandes rasgos, las siguientes:

- Cálculo de medidas habituales de estadística descriptiva: media, varianza, coeficientes de variación, asimetría y kurtosis, mediana, cuartiles, correlación, gráficos de Box, etc.

- Cálculo de estadísticos no paramétricos para una muestra y para varias muestras (dependientes o independientes): contrastes binomiales, de recorridos, de la chi-cuadrado, de Kolmogorov-Smirnov, de la mediana, U de Mann-Whitney, H de Kruskal-Wallis, Q de Cochran y Friedman, coeficientes de concordancia de Kendall, etc.

- Construcción de diagramas, gráficos e histogramas.

- Construcción de tablas de frecuencias y estadísticos asociados: de la chi-cuadrado, V de Cramer, Lambda, Tau B y C de Kendall, de Somer, coeficientes de contingencia, incertidumbre simétrica y asimétrica, etc.

- Generación de números aleatorios y simulación de procesos.

- Análisis de la varianza.

- Regresión.

- Análisis univariante de series temporales.

En el campo del control de calidad y de la productividad SCA incorpora los más avanzados métodos seguidos en la industria internacional, no sólo de control, sino también de mejora, para lo cual se hace un énfasis especial en el diseño experimental de productos. Entre las posibilidades de análisis de SCA en este área son de destacar las siguientes:

- Diseño factorial flexible a dos niveles: replicaciones completas y fraccionarias, aleatorización, diseño de Plackett-Burman y del punto central, estructura de alias y experimentación secuencial.

– Análisis de diseños experimentales: exámen de efectos, diagramas cúbicos, de dispersión, de probabilidad, de Bayes, análisis de diseños factoriales equilibrados y no equilibrados y análisis de varianza y covarianza.

– Diseño de productos en ausencia de efectos del medio ambiente: ratios señal-ruido de Taguchi, media y función de pérdida ponderada, diseño robusto de productos y otros estadísticos.

– Análisis de la transformación de Box-Cox.

5. OTRAS CARACTERISTICAS DE SCA

– *Entorno operacional.*— Con el fin de optimizar el consumo de recursos informáticos SCA permite especificar, en cada sesión de trabajo, el tamaño máximo de memoria de ordenador a utilizar. Es posible también establecer el nivel de respuestas, salidas y mensajes emitidos por el paquete durante su ejecución.

– *Entrada y salida de datos.*— SCA contempla, de una forma flexible, la lectura y escritura de datos de o hacia ficheros externos secuenciales —en formato libre, fijo y binario—, así como la creación de ficheros externos durante su ejecución. De cara a su utilización conjunta con otros conocidos paquetes de análisis estadístico y graficado (SAS, SPEAKEASY y EMS en grandes ordenadores y STARTGRAPHICS y SYSTAT en PCs.), incorpora interfases que permiten el intercambio de ficheros de datos con ellos.

– *Definición de variables y matrices.*— La variables que se definen en un programa se almacenan en vectores $n \times 1$ (variables unidimensionales) o matrices $n \times m$ (variables m -dimensionales). La creación de variables puede originarse por alguno de los cauces siguientes:

- a) Como resultado de la lectura de datos procedentes de ficheros externos.
- b) Como output de procedimientos estadísticos (residuos, predicciones, parámetros, etc.).
- c) A través de instrucciones que modifican variables ya existentes (diferenciación, retardos, concatenación, etc.) o crean directamente otras nuevas de acuerdo al esquema que se establezca (dummies, filtrado de series, generación de variables aleatorias, simulación, etc.).
- d) Mediante las operaciones algebraicas habituales (tanto escalares como matriciales).
- e) Utilizando sentencias lógicas (DO, IF-THEN-ELSE).

SCA presenta un problema importante, ya que no permite pasar de variables a matriz y viceversa. En la actualidad, sólo pueden definirse matrices en el paso de lectura de datos o como output de procedimientos multivariantes.

– *Programación de procedimientos.*– Las principales posibilidades de programación de procedimientos que se han observado son las siguientes:

- a) Aplicación repetida de un mismo procedimiento a un gran número de series mediante MACROS, los cuales están constituidos por el conjunto de sentencias a ejecutar repetidamente en las cuales el nombre de las variables, de los ficheros de entrada de datos y los caracteres de los argumentos a utilizar son simbólicos. Durante una sesión de SCA se puede llamar en veces sucesivas al MACRO, debiéndose consignar en cada llamada solamente los nombres concretos que tomen las variables, ficheros de entrada y los valores de los argumentos del procedimiento.
- b) Programación de bucles y utilización de instrucciones de bifurcación para la ejecución repetida, pero con variaciones en cada vuelta, de un procedimiento que se aplique a la misma variable.

– *Observaciones perdidas (missing values).*– En todos aquellos procedimientos en los que la aparición de valores retardados dan lugar a una pérdida de observaciones, las variables generadas por los mismos tiene igual longitud que las inicialmente tratadas, pero con valores nulos en el lugar de las observaciones que se pierden. En general, los distintos procedimientos estadísticos de SCA pueden ser ejecutados sin necesidad de eliminar estos valores.

– *Gráficos.*– SCA contempla la confección de varios tipos de gráficos para el análisis de variables y resultados: Gráficos de una variable en el tiempo, gráficos de una variable contra otra, diagramas probabilísticos e histogramas, existiendo opciones para regular su tamaño, caracteres, intervalo entre marcas de los ejes, etc. Una limitación importante de SCA es que no permite el graficado directo, durante una sesión de trabajo, en plotter o pantalla gráfica. La realización de graficar de estas características requiere la creación de ficheros externos con las series a gráficos y la utilización de software gráfico independiente.

– *Ayudas.*– Durante una sesión interactiva de SCA se puede pedir información al paquete sobre sintaxis de procedimientos, opciones, errores, etc.

6. REQUERIMIENTOS FUTUROS DE SCA

De cara al futuro, es de esperar que SCA siga incrementando sus posibilidades de análisis y de tratamiento de datos, aparte de pulir algunos procedimientos, como son los de estimación con modelos complicados, que a veces dan problemas. Las líneas de desarrollo deseables serían, a nuestro entender, las siguientes:

- Aumentar las posibilidades de manejo y tratamiento de ficheros de datos, variables y matrices, así como de sentencias lógicas para programación.
- Completar el catálogo de procedimientos de estimación para modelos simultáneos

multiecuacionales. Concretamente, se echan en falta los métodos LIML, 3SLS, 2SLS y métodos de variables instrumentales en general.

– Ampliar las posibilidades de análisis a nuevos enfoques de las series de tiempo: Modelos de espacio de los estados y filtro de Kalman.

– Incorporar procedimientos automáticos de modelización ARIMA univariante con análisis de intervención.