

EL SISTEMA SAS: REVISIÓN DE SUS POSIBILIDADES DE APLICACIÓN EN ECONOMÍA

SERGIO JIMÉNEZ

Universitat Pompeu Fabra

JOSÉ M. LABEAGA

UNED y Fundación Empresa Pública

El Sistema SAS no es un simple programa informático para economistas. De hecho, es un paquete completo y versátil que se utiliza en diversos ámbitos para el manejo y análisis de datos y para la preparación y presentación de informes. La simplicidad de esta primera definición necesita una explicación más amplia. Primero, es un sistema lo que implica la existencia de un conjunto de programas (módulos en la terminología SAS) interrelacionados, que, a su vez, se estructuran en procesos simples (procedimientos). Segundo, permite la lectura y combinación de ficheros de datos de casi cualquier forma imaginable. Tercero, la capacidad de análisis se extiende más allá del tratamiento de modelos económicos o econométricos complejos. Por último, el mayor potencial de SAS reside, a nuestro juicio, en el hecho de que es un sistema integrado que permite tanto la preparación como la presentación de informes de forma simple. A este respecto, cabe resaltar que los datos en formato SAS pueden ser utilizados sin cambios en prácticamente cualquier entorno informático (PC, estaciones de trabajo, grandes ordenadores); ello representa una gran ventaja cara al intercambio de información entre grupos de trabajo [con este fin el programa incluye un módulo específico de comunicaciones (SAS/CONNECT)].

El origen del sistema SAS se remonta a finales de los años sesenta en el Instituto de Estadística de la Universidad de Carolina del Norte¹ y tal como su nombre indica (Statistical Analysis System) fue desarrollado inicialmente para su uso en tareas de manipulación de datos y análisis estadístico. Por tanto, no está pensado desde un principio para su uso específico en Economía. Desde nuestro punto de vista existen varias razones que han contribuido a que la utilización de SAS no esté muy extendida entre los economistas (al menos en nuestro país). Primera, el coste de aprendizaje es muy superior al de la gran mayoría de paquetes estadístico-econométricos de uso común, estando más cerca de lo que se requiere para familiarizarse con un lenguaje de programación (aunque dispone de un interesante módulo tutorial, SAS/TUTOR). Segunda, el voluminoso y en muchas ocasiones intimidatorio carácter de su documentación. Tercera, su coste, ya que

(1) SAS Institute, S.A. (España). Marbella, 19. 28034 MADRID. Tel: (91) 3721500. Fax: (91) 3721689.

no se adquiere el programa sino la licencia para su uso, que se ha de renovar anualmente². De hecho, para evitar un excesivo coste el programa está organizado en módulos independientes, con lo que el usuario puede adquirir aquellos que necesita. Estas dos últimas cuestiones hacen que sea interesante cuando el grupo de usuarios potenciales es numeroso.

En el lado positivo de la balanza, el programa ofrece más de lo que se puede encontrar en un paquete estándar. No vamos a efectuar un repaso exhaustivo de todas estas posibilidades sino describirlas sucintamente, dejando para el lector interesado la tarea de profundizar en las capacidades que proporciona. Por ello, la nota va dirigida principalmente a una audiencia de "principiantes" que desean usarlo como instrumento de trabajo y, en todo caso, explorar el potencial del sistema sobre la base del comentario de las posibilidades que cada uno de los principales módulos proporciona.

Describiremos la última versión a nuestra disposición, la 6.07 (aunque se debe señalar que es un programa en permanente mejora). Los comentarios están basados en su uso en estaciones de trabajo con entorno UNIX, que es el que consideramos más adecuado. Esto no constituye, en modo alguno, una restricción a nuestra exposición, ya que el programa es homogéneo en todos los entornos habituales de trabajo. Aun así queremos resaltar el atractivo de la combinación SAS-estación de trabajo en entorno UNIX, ya que permite, amén de una alta velocidad de proceso, trabajar con varios ficheros de gran tamaño simultáneamente.

Por otro lado, la alternativa clásica de ordenador personal en entorno MS-DOS tampoco es desdeñable, aunque la productividad del sistema es ostensiblemente menor. En éste último caso la recomendación sería configurar el PC con el máximo de memoria RAM que permita y un disco duro de gran capacidad. Hay dos claros motivos para ello: primero, obtener velocidad de proceso y segundo, poder gestionar y almacenar los ficheros internos temporales que el lenguaje genera al tiempo que se realizan las tareas. En todo caso, los requerimientos mínimos para su adecuada utilización en ordenadores personales son: disponer de coprocesador matemático, un mínimo de 4 Mb de memoria RAM y un disco duro de al menos 120Mb, ya que el programa ocupa en sí mismo una amplia parte del mismo³.

Por último, como en casi todos los paquetes estándar las instrucciones de SAS para el uso de los diversos instrumentos de análisis son sencillas, siendo posible ejecutarlas tanto interactivamente explotando el sistema de menús, como mediante el proceso por lotes haciendo servir conjuntos de instrucciones predefinidos por el sistema (procedimientos) o por el usuario (MACROS).

(2) El coste anual de una configuración básica (en versión PC) es francamente prohibitivo (más de 400.000 ptas), si bien se ha de tener en cuenta que para las instituciones universitarias hay un descuento del 75 por ciento, lo que redundará en que el precio final para este colectivo sea accesible.

(3) Una configuración básica de cara al análisis econométrico ocupa casi 25Mb.

1. PROCESO DE DATOS, SISTEMAS DE PRESENTACIONES Y DE PROCEDIMIENTOS

La primera necesidad básica para la utilización de un programa es que permita acceder de forma cómoda, sencilla y flexible a los datos que constituyen la base de cualquier aplicación, bien sea para su simple manipulación bien para la realización de análisis posteriores. Como en la mayoría de programas los datos pueden ser externos o estar generados internamente por el propio paquete. Si son externos, las posibilidades de lectura son por acceso al fichero o mediante la introducción dentro del propio programa de instrucciones. Entre los tipos de ficheros a los que SAS puede acceder se encuentran los generados por otros productos informáticos como programas (por ejemplo, SPSS) u hojas de cálculo (LOTUS) y en los dos códigos más comunes ASCII y EBCDIC. Asimismo, los datos pueden ser numéricos o de tipo carácter y de formato libre o fijo.

Una vez leído el fichero de datos [tanto por variables como por matrices (ARRAYS)], el programa los transforma a su propio formato interno, a los que es posible acceder de modo más rápido en aplicaciones posteriores si se guardan de forma permanente. Destacamos, además, la facilidad con la que el programa puede concatenar diversas series de datos o seleccionar submuestras de análisis mediante el uso de restricciones y/o operadores lógicos. Otra tarea que SAS lleva a cabo satisfactoriamente es la identificación de valores erróneos o de series incompletas, lo que no impide la realización de cualquier tipo de operación ulterior (transformación de datos, análisis o estimación).

La segunda fase del proceso consiste en transformar los datos para su uso posterior. La transformación de variables, el empleo de funciones, la posibilidad de todos los operadores algebraicos, relacionales o lógicos, la generación de números aleatorios o el cálculo de valores para las distribuciones más comunes se ven completadas por funciones especiales, siendo de particular interés la diversidad de funciones de fecha (20 funciones), financieras (17, con mención especial a las de depreciación, descuento y valor presente), de carácter (que permiten reorganizar la información de variables carácter de una forma sencilla) y, finalmente, las de retardos y diferencias. Todas las operaciones descritas se pueden realizar sobre escalares o variables. No obstante, la combinación de operaciones con escalares, series y matrices está prevista en base al lenguaje matricial, que tal vez constituye uno de los módulos más completos y potentes dentro de SAS. Su uso es sencillo y similar al conocido GAUSS⁴.

Para finalizar con los datos, es conveniente hacer referencia a las posibilidades contempladas para la descripción y manejo de series temporales. El programa permite identificar de forma muy sencilla las observaciones correspondientes a series de periodicidad anual, semestral, trimestral, mensual, de medio mes, de diez días, semanal, diaria, horaria, de minutos y de segundos. Es decir, no es posible completar más las posibilidades para la frecuencia de las observaciones regularmente distribuidas. Asimismo, son interesantes los procedimientos para ordenar (ilimitado), extremadamente útil para organizar paneles de datos o describir y tabular un conjunto de datos controlando por las características deseadas.

(4) Una interesante descripción de su funcionamiento se puede encontrar en Aguirregabiria y Alonso (1993).

En relación al carácter integrado del programa, existen dos subsistemas dentro de SAS que ayudan al uso y aprendizaje del mismo así como a la mejora en la presentación de los resultados. El primero, conocido como el "Sistema Manejador de Presentaciones", cuenta con numerosas opciones para su uso por medio de ventanas y menús, un editor completo y un buen sistema de ayudas que se puede utilizar desde cualquier posición (ventana). El segundo subsistema (SAS/ASSIST) permite, mediante un sistema de menús interactivos, ir construyendo y ejecutando un cierto conjunto de instrucciones de forma sencilla, con la particularidad de que automáticamente crea el programa necesario para realizar todo el proceso (esto es especialmente útil en los modos gráficos).

2. ORGANIZACIÓN DE SAS: LOS MÓDULOS, LOS PROCEDIMIENTOS Y LAS MACROS

SAS está organizado en módulos (a su vez organizados en procedimientos) que permiten realizar desde el más sencillo manejo de los datos (BASE/SAS), asistencia para la realización de tareas (SAS/ASSIST), gráficos (SAS/GRAPH), estimación de modelos (SAS/STAT y SAS/ETS), utilización de hojas de cálculo (SAS/CALC) o uso de un potente lenguaje matricial (SAS/IML). Son también interesantes, aunque para un público mucho más especializado, los módulos de comunicaciones (SAS/CONNECT), uso simultáneo de datos (SAS/SHARE), programación (SAS/EIS), investigación operativa (SAS/OR), ..., hasta un total de diecinueve módulos, que no detallaremos.

El módulo base (BASE/SAS) permite el manejo y posterior análisis (estadística descriptiva) de grandes conjuntos de datos. De nuestra experiencia, reseñamos que es capaz de organizar y tabular un conjunto de cien mil datos en cuestión de minutos (dependiendo del ordenador que se utilice). Asimismo es útil para organizar una base de datos sobre bases ya existentes puesto que las posibilidades de etiquetado y descripción de variables son muy completas.

Respecto al módulo SAS/GRAPH destacamos la gran variedad de gráficos que incorpora, desde el más simple histograma a complejas proyecciones en tres dimensiones (por ejemplo, representación de funciones de densidad bivariantes, muy útiles en el análisis de máxima verosimilitud y en la simulación por el método de Montecarlo). Las posibilidades son realmente espectaculares. Sin embargo debemos matizar que optimizar el uso de éste módulo no es, en modo alguno, sencillo y requiere que el usuario sea bastante experto en el manejo del sistema o, en todo caso, de una buena asistencia técnica (en determinados casos puede bastar con utilizar el módulo ASSIST para la construcción del programa).

El módulo estadístico básico (SAS/STAT) es muy completo y, quizás suficiente, para aquellos usuarios que no pretenden analizar modelos sofisticados. Incorpora, de partida, procedimientos para regresión lineal (con un procedimiento específico para cualquier modelo de estas características) y no lineal (que proporciona hasta cinco posibilidades de estimación –Gauss, Marquardt, Newton, Gradiente y Dud–). Sin embargo, no dispone, a diferencia por ejemplo de TSP, de cálculo automático de derivadas analíticas. En cualquier caso, es posible programar mediante sencillas instrucciones tanto el gradiente, como el *hessiano* del modelo analizado.

Por otro lado, dicho módulo también permite realizar análisis discriminante, factorial, de varianza (sencilla o múltiple), de correspondencias y de componentes principales; posibilidad de estimar modelos con variables cualitativas (aunque es

un apartado muy incompleto), análisis de duración flexible (con disponibilidad de hasta ocho distribuciones), y así hasta más de treinta procedimientos estadísticos predefinidos. Cada procedimiento permite, en general, multitud de opciones, destacando la posibilidad de estimar con restricciones, efectuar los test de hipótesis más corrientes y guardar los resultados generados de cara a su posterior reutilización (para por ejemplo programar tests menos habituales).

El complemento del módulo estadístico es el módulo específico de series temporales (SAS/ETS). Este último incorpora análisis de series temporales, análisis de datos de panel (procedimiento añadido muy recientemente, aunque creemos que todavía no es lo suficientemente flexible para las aplicaciones más habituales en Economía), análisis espectral, estimación de modelos autorregresivos y predicción (incluyendo el conocido procedimiento X11 de desestacionalización).

La resolución de sistemas de ecuaciones es bastante completa. Permite resolver sistemas estáticos o dinámicos usando los métodos Newton, Gauss-Siedel o Jacobiano. Tanto el número de períodos de resolución (o predicción) como el punto de partida pueden ser fijados de forma flexible. En relación a los métodos de estimación las posibilidades son numerosas: mínimos cuadrados ordinarios, bietápicos, SURE y trietápicos, todos ellos con posibilidad de estimación iterativa o no lineal (siguiendo el método de Gauss o el de Marquardt). Sin embargo el procedimiento adolece del mismo inconveniente que ya se mencionó para la estimación no lineal uniecuacional: no dispone de cálculo automático de derivadas analíticas. Sin duda en este aspecto queda lejos de su equivalente en TSP. En cualquier caso el programa permite proveer gradiente y *hessiano*. Ello simplifica enormemente el proceso de estimación en algunos problemas altamente no lineales pero cuyas funciones son fáciles de derivar.

De los procedimientos de estimación comentados, lo más interesante, desde nuestro punto de vista, reside en la estimación y simulación de modelos lineales. Dichos procedimientos son extremadamente flexibles de cara a la estimación de los mismos, con o sin restricciones (cruzadas o no), usando cualquiera de los métodos usuales (desde MCO a FIML). En relación a la simulación son asimismo flexibles, ya que se puede aprovechar la facilidad con que SAS almacena los resultados de otros procesos (coeficientes, valores ajustados y estadísticos). Realizar simulación según el método de Montecarlo es sencillo y económico en términos de tiempo. Por otra parte, el procedimiento para modelos no lineales permite estimar, resolver, predecir o simular sistemas de ecuaciones también con considerable flexibilidad. Por otro lado, echamos en falta un procedimiento flexible para el análisis de datos de panel, para el tratamiento de modelos VARMA, para el tratamiento más explícito de la estimación usando el método generalizado de momentos y la máxima verosimilitud, que queda, quizás, un poco difuminada entre los diferentes procedimientos mencionados.

El módulo IML está especialmente preparado para resolver problemas haciendo uso del álgebra matricial. Destacan en él un buen número de funciones (incluyendo todas las disponibles para el módulo base), procesos preprogramados (además de los habituales como la resolución de ecuaciones lineales, destacamos que tiene programada la estimación y simulación de modelos ARMA). El módulo se completa con un aspecto esencial que es la facilidad de escritura de los programas (en esto se parece a GAUSS) y la posibilidad de disponer de un buen número de ejemplos (aplicación del método de Newton para resolver sistemas de ecuaciones no lineales o estimación de modelos de respuesta binaria). La programación

en IML permite resolver problemas no explícitamente contemplados en los módulos precedentes (por ejemplo, la estimación de modelos con datos de panel y la estimación máximo verosímil de modelos con variables dependientes cualitativas o limitadas) y evitar las rigideces que alguno de los procedimientos considerados pudiera tener, a coste de un esfuerzo no desdeñable de programación.

Adicionalmente, SAS permite el uso y almacenamiento de conjuntos de instrucciones programadas por el usuario bajo el nombre de MACROS que pueden ser utilizadas en cualquier fase del trabajo. De hecho, la continua mejora del sistema deviene de la incorporación a los módulos del programa de las MACROS programadas por algunos usuarios (una vez contrastadas). Así, es de esperar que algunas de las deficiencias señaladas sean, en breve, subsanadas. Por otra parte, la facilidad con que en cada procedimiento dentro de un mismo programa se puede plantear el uso de conjuntos de observaciones diferentes facilita, por ejemplo, la tarea de estimar y predecir con conjuntos de datos alternativos.

3. CONCLUSIONES: A MODO DE COMPARACIÓN

Resulta difícil comparar globalmente los programas estadístico-económétricos estándar y el sistema SAS. No es sencillo tampoco hacerlo con un lenguaje de programación como GAUSS. Sin embargo, abstrayendo una gran parte de las posibilidades que brinda SAS, la conclusión de nuestro comentario se centra en la comparación de sus posibilidades en relación a otros programas de uso común entre economistas.

Para empezar, el aprendizaje de SAS no es tan sencillo como el de los paquetes de uso más generalizado como LIMDEP, RATS o TSP. Tampoco es tan complicado como un lenguaje de programación, a pesar de que su abundante documentación así parece indicarlo. Sin embargo, no se puede pretender acceder al cien por cien de las capacidades que proporciona el programa porque algunas partes no serán útiles para la mayoría de problemas que tengamos planteados y porque, en definitiva, nuestro interés básico consiste en analizar el potencial de SAS como herramienta de trabajo y no como lenguaje de programación. La documentación, además de ser completa, provee las referencias más importantes en cada caso concreto, que son útiles en no pocas ocasiones para saber lo que internamente está haciendo cada programa.

Por contra, las capacidades de SAS se comparan de forma ventajosa en casi todos los apartados respecto a muchos de los paquetes. Tanto en el proceso por lotes, como de forma interactiva o mediante el sistema de menús, es probablemente más completo que la mayoría (incluidos en este apartado PC-GIVE o MLCROFIT) si bien su uso es también más complejo. Dispone, además, de un buen editor propio y de un eficaz sistema de ayudas. La lectura y procesamiento de datos es muy flexible (puesto que permite acceder a casi cualquier tipo de datos así como combinar diferentes bases) y rápida (ya que dispone de una eficiente gestión interna de ficheros SAS). De hecho creemos que uno de los fuertes del sistema es la posibilidad de gestionar bases de datos para un grupo de usuarios especialmente numeroso.

La transformación de variables es sencilla y no impone ningún tipo de limitación en cuanto a la realización de operaciones, transformaciones especiales o empleo de funciones. El módulo de operaciones con matrices es probablemente el más completo de entre todos los paquetes y es comparable al que provee un

lenguaje concebido para trabajar de ese modo como es GAUSS. La capacidad para describir y representar datos es otro de sus fuertes, aunque en lo primero no llega a las cotas de programas como STATA.

Las posibilidades en cuanto a estimación de modelos quedan incompletas en varios apartados. No dispone de programación automática de modelos para datos de panel, aunque se pueda programar mediante el módulo de operaciones matriciales. Permite muy pocas posibilidades en modelos con variables cualitativas y limitadas, a pesar de que de nuevo esta deficiencia se puede subsanar mediante la propia programación. No tiene incorporado casi ningún desarrollo de técnicas de regresión no paramétrica. Adicionalmente, creemos que en la solución de algunos problemas econométricos SAS es superado por TSP especialmente en estimación de modelos no lineales, por RATS en tratamiento de modelos de series temporales y por LIMDEP en estimación de modelos de variable dependiente cualitativa o limitada. Lo anterior es lógico si pensamos que los tres están diseñados explícitamente para economistas. Creemos, no obstante, que las limitaciones quedan compensadas por otras ventajas puestas de manifiesto y de entre las que destacamos la enorme flexibilidad de SAS para interrelacionar procesos de estimación, conjuntos de datos y resolver problemas o programar modelos mediante su módulo de álgebra matricial.

Finalmente, queremos hacer referencia a tres apartados que, si bien para la mayoría de los economistas aplicados pueden ser marginales, no dejan de tener su importancia. Primero, en relación a técnicas estadísticas el programa es muy completo tanto para el análisis univariante, como multivariante, permitiendo la realización de simples programas de frecuencias, estadísticos descriptivos, correlaciones, contrastes de asociación no paramétricos, análisis de la varianza o de componentes principales. En este sentido, podemos compararlo favorablemente con paquetes más especializados como BMDP o SPSS, si bien estos últimos son más completos en algunos temas como análisis discriminante o *cluster*. Segundo, en relación a la posibilidad de realizar gráficos baste decir que SAS tiene incorporado su propio módulo (aunque no dentro del programa base). Tercero, es adecuado para producir y presentar informes. Esta capacidad puede ser importante en algunos ámbitos del análisis económico y se completa con numerosas opciones en temas financieros o incluso módulos completos como el de investigación operativa.



REFERENCIAS BIBLIOGRÁFICAS

Aguirregabiria, V. y C. Alonso (1993): "El lenguaje GAUSS y sus aplicaciones: Una panorámica", *Revista de Economía Aplicada*, vol. 1, n.º 1, págs. 201-206.