

# EL LENGUAJE GAUSS Y SUS APLICACIONES: UNA PANORÁMICA

VÍCTOR AGUIRREGABIRIA  
CÉSAR ALONSO

*CEMFI y Universidad Complutense de Madrid*

**E**n este trabajo pretendemos mostrar una panorámica del lenguaje Gauss, aportando una valoración de las ventajas e inconvenientes que, a nuestro juicio, Gauss presenta con respecto a otros paquetes econométricos convencionales. La discusión se centra en la versión 2.2, si bien comentaremos brevemente algunas novedades que aparecen en la versión 3.0, recientemente aparecida.

El lenguaje Gauss está concebido fundamentalmente para resolver problemas de tipo matemático y estadístico. Gauss<sup>(1)</sup> se compone de tres partes: la correspondiente al lenguaje propiamente dicho, la biblioteca de gráficos y un conjunto de módulos de aplicaciones (que se venden por separado). Existen dos versiones de GAUSS, ambas compatibles con DOS 3.3 o posterior<sup>(2)</sup>. La versión de 16 bits requiere un ordenador con microprocesador de la generación 8086 o posterior, coprocesador matemático y un mínimo de 640 Kb de memoria RAM. La versión de 32 bits (GAUSS-386) requiere microprocesador 80386 ó 80486, coprocesador matemático compatible y un mínimo de 2 Mb de memoria RAM (4 Mb recomendadas); esta versión permite utilizar memoria virtual, es decir, emular memoria RAM utilizando parte del disco duro del ordenador. La versión de 32 bits supera ampliamente los límites del programa con respecto a la versión de 16 bits, de manera que el tamaño de los elementos del programa queda limitado casi únicamente por la disponibilidad de memoria RAM.

Su principal novedad con respecto a otro tipo de lenguajes es que es un lenguaje matricial, en el sentido de que la estructura de datos básica es la matriz (siendo vectores y escalares casos particulares), que puede contener tanto números como letras. Esto dota a Gauss de gran flexibilidad y rapidez a la hora de escribir expresiones matemáticas con matrices. Una de las principales virtudes de Gauss es que, sin dejar de ser un potente lenguaje de programación —disponiendo de las instrucciones típicas de un lenguaje estructurado—, incorpora potentes instrucciones propias de programas diseñados para el usuario final.

Las bases de la programación estructurada con Gauss están, al igual que en otros lenguajes, en el uso de procedimientos. Un procedimiento es un conjunto

(1) Creado y distribuido por Aptech y Systems Inc., 23804 S.E. Kent-Kangley Road, Maple Valley, Washington 98038. Tel. (206) 4327855, Fax (206) 4327832.

(2) Recientemente ha aparecido una versión de Gauss para usuarios del entorno Unix. La versión 2.2 de Gauss para DOS puede utilizarse con Windows.

de instrucciones que realizan ciertas operaciones y proporcionan unos resultados. A todo procedimiento se le debe asignar un nombre, y una vez definido se puede utilizar dentro del programa principal como si se tratase de una instrucción propia del lenguaje. Los procedimientos pueden escribirse en un fichero distinto al que contiene el programa a ejecutar, para llamarlos después durante la ejecución de un programa, siempre que se hayan incorporado previamente a una biblioteca de programas y procedimientos. Las bibliotecas permiten personalizar el uso de Gauss, ya que procedimientos escritos por el usuario pueden utilizarse de igual modo que las instrucciones originales de Gauss.

Cada uno de los módulos de aplicaciones suministrados en el paquete Gauss se compone de procedimientos escritos en Gauss, e incluidos en su biblioteca correspondiente, que pueden llamarse desde un programa escrito en Gauss. Estas aplicaciones han sido escritas por usuarios de Gauss especializados en el campo de la aplicación respectiva. Además, cada una de las subrutinas que componen una aplicación puede ser editada y modificada a conveniencia del usuario.

Gauss dispone de un excelente editor propio en el que se pueden escribir tanto programas en Gauss como ficheros ASCII (de texto) en general, y desde el que es posible ejecutar el programa que se esté editando. Es posible utilizar alternativamente cualquier otro editor modificando las opciones de inicio de Gauss.

Por último, Gauss dispone también de una biblioteca de gráficos que permite realizar gráficos de varios tipos —incluyendo la posibilidad de hacer gráficos tridimensionales— con una calidad muy alta y con la posibilidad de determinar el tamaño del gráfico y los tipos de caracteres, etc., así como de editar el gráfico incluyendo comentarios y combinar gráficos en una sola pantalla, etc.

Gauss permite trabajar interactivamente (escribiendo sentencias desde el teclado) y en *batch* (ejecutando desde DOS un programa escrito en un editor), si bien la forma de sacar partido de todas las posibilidades de este lenguaje es mediante ejecución en *batch*.

## 1. EL LENGUAJE Y LA GESTIÓN DE DATOS

El lenguaje de programación Gauss combina la sencillez de las instrucciones en los lenguajes estructurados de nivel medio (Pascal, QuickBasic, Fortran) con la potencia de instrucciones propias de paquetes econométricos como TSP, RATS, etc.

Así, Gauss dispone de los elementos típicos de un lenguaje de programación: instrucciones de control de flujo (it-then-else, do-while, do-if, goto, gosub), operadores aritméticos y lógicos, funciones y procedimientos, e instrucciones de entrada y salida de datos a pantalla, impresora o fichero.

Pero, junto con estas características típicas de un lenguaje de nivel medio, Gauss dispone también de operadores e instrucciones muy potentes para trabajar con matrices y cadenas. Entre los operadores cabe destacar los de suma, multiplicación, división y exponenciación de matrices elemento a elemento; multiplicación de matrices, matriz traspuesta, producto de Kronecker, concatenación vertical y horizontal de matrices, función factorial, etc. En lo que respecta a las instrucciones existe una gran diversidad, y pueden clasificarse en los siguientes grupos: funciones científicas (trigonométricas y logarítmicas), diferenciación e integración (gradiente, hesiano, integrales de funciones de hasta tres variables), álgebra lineal

(descomposiciones de Cholesky, Crout y QR, determinante, autovalores y autovectores, inversa, inversa generalizada, etc.), generadores de números aleatorios, operaciones con números complejos, funciones estadísticas (estadísticos descriptivos, matrices de varianzas-covarianzas y matrices de correlaciones, mínimos cuadrados ordinarios), distribuciones estadísticas (función de distribución y densidad de la normal estándar y distribuciones chi-cuadrado, gamma, beta, t de Student), series y sucesiones, manipulación de matrices (suma de filas o columnas, vectorización, selección de filas según una expresión lógica, etc.), gestión de datos y creación de variables (sustitución de observaciones no disponibles *-missing values-* por valores, creación de variables ficticias, etc.).

Gauss puede leer y escribir datos en formato ASCII, lo que permite compartir datos y resultados con otros programas. Sin embargo, la forma más eficiente de almacenar datos para utilizarlos en Gauss es utilizando el formato Gauss (*Gauss Datasets*), ya que de esta forma la lectura y escritura es mucho más rápida. Gauss incluye una utilidad llamada ATOG.EXE que transforma muy fácilmente ficheros en ASCII a ficheros en formato Gauss. Además el programa STAT/TRANSFER<sup>(3)</sup> permite traducir datos en formato Gauss a otros formatos convencionales (Stata, DBase y hojas de cálculo) o viceversa.

En cuando a los límites del programa, en la versión 2.2 de Gauss para microprocesador de 32 bits se han ampliado bastante, pasando en la mayoría de los casos a depender tan sólo de la memoria RAM instalada en el ordenador.

## 2. LOS MÓDULOS DE APLICACIONES

Existen doce módulos de aplicaciones, consistentes en procedimientos escritos en Gauss que pueden ser llamados desde cualquier programa Gauss. Estas aplicaciones abarcan desde la gestión de datos hasta la estimación de diversos modelos econométricos. Los programas están escritos por usuarios experimentados en Gauss. Un comentario general sobre estos procedimientos es que son bastante lentos. A continuación describiremos sus principales características, basándonos en la versión 2.01 de estas aplicaciones<sup>(4)</sup>.

Los módulos DTRAN y DUTIL incorporan procedimientos que permiten gestionar y modificar los ficheros de datos en formato Gauss; tanto eliminación de observaciones y transformación o creación de variables en el caso de DTRAN; como eliminación de variables, descripción y listado de fichero de datos, retardos de las variables, renombrar variables, fusión y clasificación de ficheros de datos, etc., en el caso de DUTIL. Una de las ventajas de este módulo es que es posible explorar interactivamente el contenido de un fichero en formato Gauss, sin necesidad de escribir un programa. La obtención de estadísticos descriptivos de las variables (columnas) de un fichero de datos en formato Gauss es posible mediante el módulo BSTAT. Los procedimientos incluyen el cálculo de coeficientes de correlación entre variables, frecuencias relativas, medias y varianzas condicionadas y no condicionadas, así como estimación por MCO, MC2E y MC3E de un sistema de ecuaciones lineales. Sin embargo, existen actualmente programas mucho más

(3) Distribuido por Circle Systems Inc., 1001 Fourth Avenue, Suite 3200, Seattle, WA 98154. Tel. (206) 6823783, Fax (206) 3284788. También puede obtenerse a través del distribuidor de Gauss, Aptech Systems, Inc.

(4) Recientemente ha aparecido la versión 3.0 de algunos de estos módulos.

adecuados para este fin, como por ejemplo Stata. Las aplicaciones de Gauss disponibles para este propósito son claramente inferiores en términos de tiempo y de dificultades de manejo.

Los módulos QUANTAL y COUNT permiten la estimación de modelos microeconómicos con variables dependientes cuyo rango de variación es distinto de la recta real. El módulo QUANTAL está diseñado para estimar modelos de variable dependiente cualitativa, incluyendo estimación máximo verosímil de modelos de elección discreta con distribuciones normal o logística y regresión de Poisson para datos de tipo entero. Los procedimientos del módulo COUNT permiten la estimación por máxima verosimilitud de modelos de duración y modelos estadísticos para datos de tipo entero. Los modelos que pueden estimarse son: modelos de duración con función de riesgo exponencial, gamma o Pareto, pudiendo estar censurada la variable dependiente, modelos de regresión con distribuciones Poisson, binomial negativa y otras, para la variable dependiente. El módulo LOGLIN permite el análisis log-lineal de tablas de contingencia.

Existen otros dos módulos econométricos, TSCS y AUTOREG. El primero permite obtener algunos estimadores para modelos lineales de datos de panel, incluyendo el estimador MCO de efectos fijos (estimador de análisis de covarianza) y el estimador MCG de efectos aleatorios. Sin embargo, TSCS está un tanto desfasado con respecto a la práctica econométrica actual con datos de panel. Así, no incorpora la posibilidad de estimar por el método generalizado de los momentos (MGM) o de realizar contrastes de especificación. En este sentido el programa DPD (Dynamic Panel Data), escrito en Gauss por M. Arellano y S. Bond, permite la estimación por MGM de modelos lineales utilizando datos de panel. AUTOREG estima por máxima verosimilitud modelos de regresión con perturbaciones autorregresivas de cualquier orden y calcula autocorrelaciones y autocovarianzas. De nuevo, las aplicaciones de este módulo son relativamente limitadas. En concreto, no incluye la estimación de modelos media móvil, análisis espectral, o análisis de cointegración. En lo que respecta al análisis de cointegración, recientemente ha aparecido un programa, escrito en Gauss, denominado COINT, y que distribuye también la empresa Aptech Systems.

Existen dos módulos de optimización, uno general (OPTMUM) de minimización de funciones reales de variable vectorial y otro específico (MAXLIK) para la estimación por máxima verosimilitud. Los algoritmos de optimización disponibles son: Descenso Máximo, Newton, Broyden-Fletcher-Goldfarb-Shanno y Davidson-Fletcher-Powell. Todos estos algoritmos tienen la posibilidad de calcular el paso óptimo en cada iteración según distintos métodos (Regla de oro, Newton o método de Berndt). El procedimiento de MAXLIK se centra en la minimización del logaritmo de una función de verosimilitud definida por el usuario. Además de los algoritmos incluidos en OPTMUM, también se incluye el método de Berndt-Hall-Hall-Hausman (BHHH) para la maximización del logaritmo de una función de verosimilitud. Una limitación de los algoritmos de optimización incluidos en estos procedimientos es que son bastante lentos<sup>(5)</sup>. El módulo SÍMPLEX resuelve problemas de programación lineal de pequeña escala mediante el algoritmo del símplex.

---

(5) Las nuevas versiones 3.0 de Optmum y Maxlik utilizan un algoritmo numérico que no requiere la inversión de matrices ya que utiliza el método de factorización de Choleski. Esto podría mejorar la velocidad de estas aplicaciones, si bien no hemos utilizado hasta el momento estas nuevas versiones.

Finalmente, con el módulo NLSYS es posible obtener la solución de un sistema de ecuaciones simultáneas no lineales utilizando un algoritmo cuasi-Newton, pudiendo bien especificar el Jacobiano analíticamente, o bien obtenerlo numéricamente en cada iteración.

La principal ventaja de los módulos de Gauss es que permiten un uso más asequible para usuarios que no estén interesados en especializarse en la programación con Gauss. Además, el hecho de que todos los procedimientos contenidos en los módulos puedan editarse permite ampliar las posibilidades de estos módulos, si bien para ello sí es necesario un conocimiento a fondo del programa Gauss. Todos los módulos disponen de ficheros con ejemplos muy útiles para facilitar su utilización.

Además, es posible incluir dentro de un programa Gauss subrutinas compiladas en Fortran, Ensamblador o C. La versión 2.2 de Gauss para 32 bits permite utilizar estas subrutinas si previamente han sido compiladas utilizando Ensamblador de Microsoft o de Phar Lap, Fortran de Microsoft versión 5.0, C de Microsoft versión 6.0, o Metaware High C versión 2.3.

### 3. COMENTARIOS ADICIONALES

El principal problema de Gauss es el mismo que el de cualquier otro lenguaje de programación de nivel medio: el coste de aprendizaje es relativamente alto. A este problema contribuye el manual de instrucciones, que es bastante árido, especialmente si no se tienen conocimientos de programación en general. Si bien el manual describe de forma clara el uso del editor y dedica dos breves capítulos a los procedimientos y las bibliotecas, se limita a una enumeración alfabética de sus instrucciones. La explicación a cada instrucción es concisa pero clara; sin embargo, se echa de menos un capítulo que describa el diseño de programas en Gauss con algunos ejemplos. Afortunadamente, este inconveniente se supera pronto dado que el número de operadores e instrucciones básicas es relativamente pequeño y éstos son bastante intuitivos. Además, la existencia de instrucciones muy potentes —más propias de un paquete informático que de un lenguaje— y de los módulos de aplicaciones facilitan la labor.

El coste del aprendizaje de Gauss es muy superior al de paquetes económicos convencionales como TSP, RATS, Limdep, Microfit, etc. Por tanto, el uso de Gauss no es muy recomendable si el usuario está interesado sólo en la estimación de modelos de regresión más o menos sencillos. Si por el contrario se está interesado en la programación de métodos econométricos de estimación relativamente novedosos, posiblemente es una de las mejores alternativas, por su flexibilidad y potencia y por su relativa sencillez frente a lenguajes de programación más abstractos como Fortran.

Si el usuario está interesado en el uso de programas econométricos interactivos por menú, Gauss no es una buena opción frente a Microfit, MicroTSP o PcGive, si bien existe una aplicación llamada GAUSSX que simula a MicroTSP.

Una ventaja clara de Gauss frente a otros programas es la posibilidad de trabajar con bases de datos de gran tamaño, lo que le hace especialmente apropiado para aplicaciones econométricas con datos de panel o con datos de sección cruzada. En concreto, todos los módulos de aplicaciones pueden trabajar con ficheros de datos tan grandes como permita la capacidad del disco duro. Esto

puede justificar el coste de aprendizaje de este lenguaje, pero su uso puede no estar tan justificado si se va a realizar econometría de series temporales.

Un problema bastante importante de Gauss, sobre todo cuando se empieza a programar en este lenguaje, es que sus mensajes de error son bastante confusos. En ocasiones resulta difícil identificar el origen del fallo en la ejecución del programa. Además, Gauss señala la línea donde se produce el error sólo si el programa escrito en Gauss se ejecuta desde el editor propio de Gauss, no si el programa se ejecuta desde un editor alternativo o desde DOS.

Por último, la opción de memoria virtual debe usarse sólo si es estrictamente necesario para ejecutar un programa por limitaciones de la memoria RAM: el uso de disco duro como memoria RAM hace la ejecución enormemente lenta. No es éste un problema específico de Gauss, sino del uso de memoria virtual en general.

