

APLICACIONES A LA ECONOMETRÍA DEL PROGRAMA SHAZAM*

CARLOS M. FERNÁNDEZ JARDÓN

Escuela de Negocios Caixa-Vigo

EDUARDO R. VILELA

Unión de Bancos Suizos

SHAZAM¹ es un paquete econométrico orientado a la práctica con un bajo coste de aprendizaje, tanto para la ejecución de sentencias y cálculos elementales como para la obtención de transformaciones y la construcción de técnicas econométricas relativamente nuevas como la metodología de Johansen-Juselius. Esto se debe a la capacidad de adaptación que presenta este programa, ya que permite la elaboración de nuevos *subprogramas* elaborados por los usuarios, también llamados *procedures*. Esta posibilidad, unida a la facilidad de conectarse en redes de usuarios, como por ejemplo *internet*, hace posible la implementación casi *inmediata* de nuevas técnicas o modificación de las existentes.

La intención de este trabajo es mostrar las principales características (tanto positivas como negativas) de este programa. En primer lugar, se analizarán los posibles entornos de aplicación, los requisitos para su instalación y necesidades de editores de texto. A continuación, tras una rápida descripción de las principales posibilidades de cálculo, se pasará a sus características econométricas. Se analizarán las posibilidades de diagnosis de modelos que presenta el programa y también se examinarán entre los distintos métodos de estimación el lineal y el no lineal, tanto en sus formas dinámicas como estáticas. Para finalizar, se hará un resumen de sus principales limitaciones y se enunciarán las principales conclusiones.

1. ENTORNOS DE APLICACIÓN

Existen muchas versiones del programa, lo que le da una serie de ventajas sobre algunos de sus competidores. Tiene versión de *Mainframe* (IBM, Hewlett Packard, etc.), *Workstations* (Vax, Unix, NeXT etc.), versión para los sistemas operativos DOS, OS/2 y Windows, al igual que para Macintosh. Recientemente se ha incorporado la

* Este trabajo fue elaborado cuando los autores eran profesores de la Universidad de Navarra. Agradecen los comentarios de Antonio Tolsá, Fernando Pérez de Gracia, Enrique Sentana y la información suministrada por Donna Wong. Como es habitual, las opiniones expresadas no representan necesariamente las de la Escuela de Negocios Caixa-Vigo o de la Unión de Bancos Suizos.

(1) SHAZAM, UBC Economics, # 997-1873 East Mall, Vancouver, BC V6T 1Z1, Canada. Fax + 1-604-822-5915 y email info@shazam.econ.ubc.ca.

versión de PowerMac. Además, tiene más de 2000 instituciones con licencia de uso en diferentes universidades y empresas del mundo, creando una red de correo electrónico que permite información y comentarios entre ellos, lo que facilita la resolución de muchos de los problemas que surgen en la utilización habitual de un programa econométrico a la luz de los nuevos avances de la profesión. También posee una dirección de correo electrónico directo para su soporte técnico². Analizaremos la versión 7.0, que es la última con validez en los entornos antes comentados.

Las diferentes versiones permiten trabajar desde ordenadores con 640K de memoria hasta 4Meg de RAM -con o sin co-procesador matemático-, variando en la velocidad de cálculo y el uso de datos.

El SHAZAM, al igual que otros programas, consta de un conjunto de comandos, y cada uno de ellos tiene una serie de opciones que potencian sus posibilidades. Comentaremos, en primer lugar, la estructura de las entradas y salidas de datos para luego hacer referencia a las posibilidades de cálculo y, finalmente, a las limitaciones que encontramos en el paquete.

La comunicación del usuario con el programa se realiza de forma muy similar a otros programas. Existen dos formas de trabajo:

- Un modo interactivo en el que se escribe en pantalla cada comando y se obtiene automáticamente su resultado.
- Y un modo por lotes o *batch* en el que se realiza un fichero de control que se llama desde el programa y se obtienen los resultados de ese fichero.

Para la obtención de resultados es posible tanto el uso de comandos como la notación matricial. Por ejemplo, el comando para una regresión mínimo cuadrática será *ols* en tanto que por medio de la notación matricial se pueden obtener los coeficientes de la regresión por medio de $(x'x)^{-1}x'y$. El usuario habituado a los entornos de *ventanas* como los sistemas operativos Windows o Mac suele hechar en falta este tipo de posibilidad.

2. PROGRAMAS EDITORES

Tiene un editor propio llamado *Edit* bastante sencillo de manejar aunque muy limitado. Asimismo, permite cualquier otro editor que guarde ficheros en formato ASCII, lo que resulta muy cómodo cuando se trabaja en sistemas operativos multiventana. En las versiones PC y Mac se pueden editar los archivos con procesadores de textos habituales. Muchos de los editores de textos que vienen con los distintos sistemas operativos (algunos de ellos detallados en el Cuadro 2) también son aceptados.

Cuadro 2		
Sistema Operativo	Versión	Editor del Sist. Operativo
MS DOS	5.0 +	Edit
DR DOS	5.0 +	Editor
OS/E	-	E
Mac	-	Teach Text
Unix	-	Vi

(2) help@shazam.econ.ubc.ca

3. LECTURA Y TRATAMIENTO DE DATOS

Permite leer datos de diferentes tipos de ficheros: ASCII, TSP, DATABASE, con formato DIF (*Data Interchange Format* –los archivos de Lotus 1-2-3/Symphony pueden ser guardados con este formato–), binarios, Citibase, etc., ..., aunque cuando se incluyen datos alfanuméricos necesita formatos de lectura tipo FORTRAN, lo que limita sus posibilidades. Las versiones para PC y Mac tienen un límite de trescientas variables, en tanto que en las versiones *mainframe* éste aumenta considerablemente. Los límites de la cantidad de datos y variables dependerán del entorno en el que se esté ejecutando el programa. Si bien se puede trabajar con ficheros de gran tamaño, el programa permite trabajar con varios ficheros de datos simultáneamente. En la práctica, por lo general, es recomendable partir los ficheros de abultadas dimensiones en otros más pequeños.

En la práctica, los autores de este trabajo han trabajado con matrices de, aproximadamente, trescientas variables y unos sesenta mil datos en ordenador *mainframe* y no se han encontrado dificultades.

Tiene bastantes problemas para el tratamiento de valores ausentes, puesto que los convierte en ceros, creando confusiones si no se advierte previamente. Un caso especial es al retardar una variable. Si se desea aplicar primeras diferencias el programa restará cero de la primera observación –de la variable no retardada– si no fue corregido el tamaño muestral. En el caso de que el usuario no se de cuenta el programa emite un aviso en el que se señala que se está restando cero.

3.1. Cálculo y transformaciones

En la versión 7.0 se ha mejorado el cálculo matricial con respecto a versiones previas, permitiendo muchas operaciones de forma bastante intuitiva: autovalores, inversas, descomposiciones de Choleski, productos de Hadamard, Productos de Kronecker, concatenación de matrices, etc., con comandos directos. Permite también cálculos de programación lineal, de resolución de ecuaciones, de aproximación numérica de derivadas e integrales, optimización de funciones y análisis de componentes principales. Obtiene fácilmente la descripción de una muestra, y permite tratar tanto los datos temporales como los de corte transversal o de panel, aunque está principalmente orientado a los primeros.

La generación de variables es bastante directa y muy intuitiva. Su utilización es similar a otros programas econométricos. El trabajo con matrices es también simple, aunque en algunos casos es confuso saber si se está trabajando con una matriz o con su traspuesta. Esta situación puede hacer perder mucho tiempo al usuario.

Es programable sin mucha dificultad, permitiendo incluso subprogramas que se definan como comandos, lo cual facilita las implementaciones de nuevas técnicas econométricas.

Se deberá tener en cuenta que se han encontrado “errores” de redondeo entre las versiones Mac (un LCIII 8/250 con co-procesador matemático) y un mainframe HP Serie 8000 en algunos estadísticos y estimadores que requerían el uso de cálculo de cierta complejidad. En algunos casos, la diferencia entre una estimación y la otra llegan a ser el doble. Uno de los principales inconvenientes de estos errores de redondeo está en la construcción de contrastes –como para verificar los valores del *test de la traza* de Johansen-Juselius– que se encuentren cerca de los valores críticos. Para un análisis empírico de este punto se puede ver Vilela (1995).

4. DIAGNOSIS

Las observaciones atípicas e influyentes no son necesariamente *malas*; de hecho pueden contener la información muestral más interesante. Suelen ser consecuencia de errores o de circunstancias atípicas en relación con el resto de los datos que se están analizando. Tan sólo después de que esos puntos hayan sido identificados podrán ser calificados. Con posterioridad a su calificación se tomará la decisión *de qué hacer* con ellos. El disponer de elementos de diagnosis para una calificación correcta hará aumentar la confianza en los resultados obtenidos. De hecho, este procedimiento ayudará al usuario a valorar como legítima la gran influencia de ciertos puntos-aunque concluya que no hace falta ninguna medida correctiva.

Las bases para las técnicas de diagnosis radican en la observación del comportamiento de los resultados de las regresiones ante algún tipo de perturbación *controlada* –como, por ejemplo, quitar una fila de datos o una observación. Con este comportamiento se verá si afecta al *ajuste a los datos*, a los valores de los parámetros estimados, a los residuos o a alguna forma funcional de ellos como, por ejemplo, el coeficiente de determinación, las desviaciones típicas, autocorrelaciones, etc. Este programa permite conocer los *residuos estudentizados*, DFFIT y DFFITS, los valores de la diagonal de la matriz *hat*, el estadístico *covratio* y los DFBETAS. Al igual que en otros programas, como el SAS, sigue las directrices del libro sobre diagnosis de Belsley, Kuh y Welsch (1980).

Uno de los aspectos positivos y distintivo a destacar de este programa es, sin lugar a dudas, las herramientas para la diagnosis de observaciones atípicas e influyentes.

Dentro de los distintos contrastes para la diagnosis de los modelos se debe mencionar que permite identificar la heterocedasticidad desde contrastes simples, como el de Chow o detectar su forma funcional como, por ejemplo, cuando:

- Se presenta como consecuencia de la variable dependiente o,
- Cuando es la desviación estándar una función lineal de las variables explicativas y
- La varianza es una función lineal de las variables exógenas.

Asimismo, entre otros elementos de diagnosis se pueden mencionar test de autocorrelación, test CUSUM y CUSUMQ, test de especificación (*Reset* de Ramsey), y técnicas de *bootstrap* y *jackknife*. También tiene tests de normalidad –simetría, curtosis, bondad de ajuste, Jarque-Bera, etc.– y de residuos recursivos –hacia adelante y atrás–, Durbin-Watson, etc. Sin embargo, sólo son válidos cuando se realizan regresiones por mínimos cuadrados. Dentro de las diferentes posibilidades para diagnosis de la multicolinealidad se puede mencionar los del Factor de Incremento de la Varianza, número de condición, índices de condición, proporción de la varianza explicada por cada componente principal, etc.

Una posible solución de este problema es el de la regresión *Ridge* (o *alomada*) o por componentes principales.

Las salidas son bastante claras, y en general son amplias, aunque dependen de las opciones que se utilicen en cada comando. Una excepción a la claridad antes mencionada es que, en la mayoría de los casos, las salidas más completas de cada comando no se encuentran en el manual. Se echa de menos un análisis detallado de todos los algoritmos usados para su cálculo. No obstante, la diferencia entre esta versión y la anterior es notable.

Sin duda, uno de los principales fallos es la deficiente capacidad gráfica del SHAZAM si se lo compara con las posibilidades del módulo GRAPH del SAS. Para el usuario que desee mejorar sus presentaciones gráficas la opción más rápida y eficaz suele ser la exportación de los datos o salidas a otros programas de gráficos como, por ejemplo, el Harvard Graphics o similar. Para algunas versiones de ordenadores compatibles ofrece la alternativa de utilizar el programa de gráficos GNPLOT.

Los resultados se pueden enviar a la pantalla, a un fichero de resultados, forma típica del trabajo por lotes, o a la impresora directamente. Los envíos a un fichero de resultados son leídos con facilidad por los procesadores de textos habituales.

Un párrafo especial merecen los avisos de error, ya que SHAZAM, a diferencia de la mayoría de los programas del mercado, advierte al usuario *desprevenido* de los múltiples errores que se suelen cometer. En la enseñanza es increíblemente útil por la capacidad *pedagógica* de estos avisos.

Entre los múltiples errores que avisa, se pueden enumerar la inclusión de la misma variable en ambos lados de la ecuación, la existencia de un número menor de observaciones que de variables, mayor cantidad de ecuaciones que de parámetros, restricciones equivocadas, subidentificación en sistemas de ecuaciones, que la función de verosimilitud no alcance máximo, *hessianos* no definidos positivos y una larga lista de etcéteras.

Un problema habitual en las regresiones es la distorsión que provoca la no existencia de término independiente en el modelo en algunos contrastes estadísticos básicos, como por ejemplo la F y el R^2 . El SHAZAM, a diferencia de otros modelos considerados superiores como el SAS³ u otros relativamente similares como el Micro-TSP (7.0) o el Limdep⁴, alerta al usuario sobre los inconvenientes de omitir la ordenada de origen y da al usuario tres R^2 para el caso de que exista o no término independiente en forma implícita. Asimismo, al igual que el SPSS y el SYSTAT, avisa cuando la descomposición de la varianza es calculada centrada o no en cero .

5. COMANDOS

5.1. Métodos lineales de estimación

Permite estimar diferentes versiones del modelo de regresión lineal, tanto por mínimos cuadrados, como por máxima verosimilitud. En el primer caso incluye modelos heterocedásticos, estimaciones consistentes de matrices de covarianza, tanto en autocorrelación como heterocedasticidad, Newey-West y White respectivamente, modelos de ecuaciones simultáneas en dos y tres etapas, variables instrumentales, mínimos cuadrados generalizados y el método de los momentos generalizados.

En el caso de los modelos heterocedásticos permite su estimación máximoverosímil de la denominada heterocedasticidad de la variable dependiente (Theil 1971), la de tipo multiplicativo (Harvey 1976, 1990), heterocedasticidad condicional generalizada ARCH (Engle 1982), GARCH (Bollerlev 1986), ARCH en media o ARCH-M Engle, Lilien y Robins (1987).

También tiene regresión robusta, máxima verosimilitud con diferentes distribuciones : Exponencial, Gamma, etc. En ambos casos permite la inclusión de restriccio-

(3) Cfr. Uyar y Erdem (1990).

(4) Cfr. Okunade *et. al.* (1993).

nes. Asimismo, ofrece posibilidades de estimaciones máximo verosímiles en algunos modelos con errores no-normales. Para ello cuenta con diferentes distribuciones como la de Weibull, Gama, Beta, Gama generalizada, etc. Si el usuario se encuentra con una ley de distribución que no está incluida en el programa podrá incluir la función de densidad que desee en el comando de estimación no lineal.

5.2. Modelos dinámicos

Para el análisis dinámico tiene muchas e interesantes posibilidades, incluyendo los modelos ARIMA, con los tres pasos de identificación, inferencia y predicción. El usuario deberá especificar cuál de los tres pasos antes mencionados desea en cada caso. También se pueden obtener estimación con heterocedasticidad en modelos GARCH, modelos autocorrelados con diferentes técnicas -hasta cuatro distintas-, modelos autorregresivos, retardos distribuidos, etc.

Permite realizar predicciones y evaluarlas con métodos cuadráticos, coeficiente de Theil, etc., tanto en modelos estáticos como dinámicos y alguno de los métodos avanzados, que se indican más adelante.

5.3. Otros métodos

Por otro lado tiene posibilidad de calcular modelos de datos de panel, siguiendo el manual de Kmenta (1986), y los más clásicos de modelos *logit*, *probit*, *tobit*.

Permite imponer desigualdades dentro de las regresiones lineales por medio de la metodología *bayesiana* descrita por Geweke (1986). Este método parte de un conjunto de coeficientes y de una matriz de covarianzas, ambos provenientes de una estimación previa, y genera una gran cantidad de iteraciones utilizando la denominada Iteración de Montecarlo con una distribución aleatoria de coeficientes. El usuario deberá especificar las restricciones.

Presenta, en su última versión, test de raíces unitarias de Dickey-Fuller (en su versión simple DF y aumentada ADF), Phillips-Perron, y los correspondientes test de cointegración, pero no tiene los de Johansen, y otros más modernos, aunque indica procedimientos para llevarlos a la práctica. Presenta, junto con los contrastes de DF y ADF, sus valores críticos al 10% con los valores de los criterios de selección/información de Akaike y Schwarz.

6. APLICACIONES A LA DOCENCIA

Otras de las cualidades de este programa está en sus posibilidades para la enseñanza. Tiene una gran capacidad para la simulación de distribuciones, incluyendo muchas de las más comunes en Economía, lo que permite explicar y facilitar la comprensión por parte de los alumnos de propiedades de los estimadores.

En primer lugar, su fácil manejo matricial permite la explicación de los problemas econométricos primero, por medio de la elaboración por parte de los alumnos de los estimadores y estadísticos para, luego, comprender mejor el funcionamiento de los comandos.

Por otro lado, es extremadamente versátil para distintos niveles de enseñanza de la Econometría por medio del uso de las distintas opciones en los respectivos comandos. Brinda excelentes posibilidades tanto para la enseñanza de cursos básicos o avanzados de licenciatura como, también, para cursos de post-grado. La facilidad de programar permite al investigador aplicar nuevas técnicas.

Las prestaciones de la versión *mainframe* para trabajar en red son ampliamente satisfactorias, ya que son mucho más rápidas que las versiones PC o Mac. Los autores de este comentario han trabajado en los últimos cinco años la versión *mainframe* en un HP Serie 8000 con 60 ordenadores conectados en red y no han tenido mayores problemas ni con la versión 7.0 ni con la anterior.

Otras de las ventajas que se encuentran desde el punto de vista docente es que los datos y los comandos de algunos libros de Econometría se pueden conseguir en *diskettes* a un coste muy bajo. Como ejemplo se pueden mencionar los *diskettes* para los libros de Berndt (1991), Greene (1990), Judge (1985 y 1988), Maddala (1988), Gujarati (1988), etc.

7. LIMITACIONES

Las principales limitaciones vienen dadas por el tratamiento de datos multivariantes, que deja bastante que desear. Faltan muchas posibilidades de métodos estadísticos multivariantes, que cada vez tienen más uso en Econometría. Por ejemplo, el análisis factorial que posee es muy limitado, ya que, por ejemplo, no calcula las communalidades (aunque da los *loadings*); de las rotaciones, tan sólo tiene la *Varimax* y ninguna de las oblicuas, etc.

En el análisis de las series temporales únicamente permite el uso de modelos Box-Jenkins para datos univariantes, sin incluir estimación espectral ni modelos VAR, etc., aunque permite programar estos casos.

En cuanto al tratamiento de datos ausentes, ya hemos comentado la necesidad de incluir comandos complementarios, lo que complica la simplicidad habitual del programa.

Tiene el problema, similar al TSP, de la dificultad para particionar las matrices o trabajar con matrices por bloques. Sin embargo, es mejor respecto a éste la forma de programar directamente con matrices que permite obtener procedimientos de forma sencilla.

También presenta problemas cuando se quiere trabajar con modelos que necesitan aspectos más estadísticos, como Análisis de la varianza, etc., en que no tiene resultados directos.

Se nota una mejora en los métodos de programación anexos al programa: por ejemplo, se ha ampliado -de dos a nueve- la posibilidad de bucles DO (*do loops*); no obstante, aún faltan otras opciones por mejorar.

Tiene un problema adicional que es la poca explicación de las opciones en los manuales de uso del programa, aunque en la última versión es manifiestamente mejor que en la anterior.

8. CONCLUSIONES

Al compararlo con otros programas similares presenta algunas ventajas e inconvenientes. Por un lado, tiene menos implantación en las universidades españolas que otros programas como el TSP, RATS, GAUSS, etc., lo que hace que su uso a veces parezca muy limitado, aunque ya hemos comentado que en el extranjero está muy difundido.

Su manual, si bien ha mejorado sensiblemente con respecto a la versión anterior, aún dista mucho de manuales muy completos como los del SAS. La deficiencia de no explicar las salidas más completas de cada comando hace que el usuario pierda tiempo por un lado, buscando trabajos que traten el tema y, posteriormente, verificando de forma manual el algoritmo usado por el programa.

El servicio técnico de post-venta, si bien tiene el inconveniente de tener que recurrir a la Universidad de British Columbia (Canadá), por lo general es rápido y eficaz para solucionar problemas técnicos de instalación o econométricos. Esta es una diferencia a tener en cuenta en la comparación con otros programas econométricos más difundidos en España.

Es un programa orientado fundamentalmente al estudio econométrico, por lo que no tiene métodos multivariantes de uso común en las ciencias sociales como correlación canónica, análisis discriminante, análisis factorial (completo), etc...; en todo ello es superado por otros paquetes como SAS, BMDP, SPSS, etc.

Al no ser un programa únicamente para trabajar con series temporales en la fase de identificación, sería de desear la inclusión de algún tipo de sistema experto que complemente la labor del usuario en este punto.

No permite un uso de ratón con menús como ocurre, por ejemplo con el Microfit, lo que lo hace menos *amigable*, pero permite programación y tiene pocas limitaciones en cuanto al número de variables. Los autores de este trabajo han realizado estimaciones con comparaciones internacionales o regionales con matrices de datos que tenían más de sesenta mil observaciones en un *mainframe* y el programa no tuvo problemas en realizar las operaciones requeridas.

Tiene menos posibilidades en el uso de variables dependientes y de datos censurados que programas como el LIMDEP; no obstante, es bastante completo en cuanto a las técnicas básicas.

Desde el punto de vista pedagógico es útil básicamente por dos motivos:

1. debido a que tiene *diskettes* con las soluciones a la mayoría de los problemas propuestos en libros básicos como el Pindyck y Rubinfeld (1991) hasta libros útiles para cursos avanzados o de post-grado como Berndt (1991) o el Judge (1985).
2. Por otro lado tiene la ventaja de los avisos de error, a los que se hiciera referencia con anterioridad, y que ayudan en los cursos teórico-prácticos de econometría de cualquier nivel a la mejor comprensión de los problemas de la materia.



REFERENCIAS BIBLIOGRÁFICAS

- Belsley, D., E. Kuh y R. Welsch (1980): *Regression diagnostics*, Wiley.
- Berndt, E.R. (1991): *The practice of econometrics*, Addison-Wesley.
- Bollerslev, T. (1986): "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics*, nº 31, págs. 307-327.
- Engle, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance in the United Kingdom inflation", *Econometrica*, nº 50, págs. 987-1.007.
- Engle, R., Lilien, D.M. y Robins, R.P. (1987): "Estimating time varying risk premia in the term structure: The ARCH-M model", *Econometrica*, nº 55, págs. 391-407.

- Geweke, J. (1986): "Exact inference in the inequality constrained normal linear regression model", *Journal of Applied Econometrics*, nº 1, págs. 127-141.
- Greene, W.H. (1990): *Econometric Analysis*, MacMillan.
- Gujarati, D. (1988): *Basic Econometric*, Segunda Edición, McGraw-Hill.
- Harvey, A. (1976): "Estimating regression models with multiplicative heteroscedasticity", *Econometrica*, nº 44, págs. 461-465.
- Harvey, A. (1990): *The econometric analysis of time series*, Segunda Edición, MIT Press.
- Judge, G., Griffiths, W., Hill, R., Lütkepohl, H. y Lee, T. (1985): *The theory and practice of Econometrics*, Segunda Edición, Wiley.
- Judge, G., Griffiths, W., Hill, R., Lütkepohl, H. y Lee, T. (1988): *Introduction to the theory and practice of Econometrics*, Segunda Edición, Wiley.
- Maddala, G.S. (1988): *Introduction to econometrics*, MacMillan.
- Okunade, A., Chang, C. y Evans, R. (1993): "Comparative analysis of regression output summary statistics in common statistical packages", *The American Statistician*, nº 47, págs. 298-303.
- Theil, H. (1971): *Principles of Econometrics*, Wiley.
- Uyar, B. y Erdem, O. (1990): "Regression procedures in SAS: Problems?", *The American Statistician*, nº 44, págs. 296-301.
- Vilela, E. (1995): *Inflation rate convergence in the European Union*, Springer-Verlag, Contributions in Economics, próxima aparición.