

# Contributed Articles

## APPLECON: Econometric Analysis with APL-- An Introduction and Demonstration

Stephen D. Lewis

### Abstract

APPLECON is a set of APL workspaces which implement regression techniques commonly encountered in econometrics. Each estimation technique is expressed in a concise and unambiguous manner by using a standard format. To facilitate further analysis, data, key parameters, and estimates are either saved as global variables or can be retrieved by explicit-result functions. APPLECON is viewed as an alternative to existing econometric programs which often take on the character of a "black box". Natural extensions of this work include applications in the area of computer-assisted instruction and development of a comprehensive package for econometric modeling.

- - - - -

### Introduction

With the development of increasingly powerful computers over the past two decades, software packages for dealing with econometric models have become more and more sophisticated. The latest generation of general-purpose econometric packages (e.g., EPLAN [1], TSP [2], and SHAZAM [3]) are impressive tool kits for investigating a wide range of econometric problems. Unfortunately, as economists have increasingly turned to computers for solving their computational problems, there has developed a wide gulf between the programmer and the program user. Typically the econometrician has only limited knowledge of computing, so that one must rely on a set of instructions provided with the program being used in order to obtain and interpret the results. On many occasions, the user experiences difficulty in explaining the results, since it is not always clear what the program is doing. In addition, recent advances in econometrics have made this subject technically complex when compared with classical multiple-regression analysis. These developments have put tremendous demands on econometric packages. Finally, existing programs are usually not flexible enough to incorporate easily major developments occurring in econometrics without considerable modification of the source program.

This paper describes APL workspaces which contain regression techniques commonly

encountered in econometrics. The major objectives in developing an APL-based econometric package have been:

- To provide econometric techniques that can be easily handled in the interactive environment of a computer terminal.
- To present econometric techniques using simple APL notation and functions that can be readily understood by anyone who is familiar with econometrics, but possesses only a limited knowledge of APL.
- To examine the usefulness of APL expressions as an alternate pedagogical technique to classical econometric notation.

While a comprehensive package is still to be developed, enough work has been completed in order for the ideas behind this approach to be properly appreciated and communicated at this time. The first section of the paper discusses the organization and use of the functions. In the second section, the formulation of estimation functions is discussed and the question of using APL as a notation in econometrics is explored. APPLECON is demonstrated in the third section. Summary and future developments form the conclusion.

### I. Workspace Organization and Use

The APPLECON workspaces are organized into groups containing estimation functions and groups containing functions for supplementary analysis. The three basic steps required in order to produce results using APPLECON are data entry, estimation, and supplementary analysis.

In the first step, data are entered by standard APL procedures. Following common practice in econometric work, the data matrix represents one variable per column. Data labels are entered into a character matrix with the rows of this matrix ordered in the same sequence as the columns in the data matrix. If desired, an interactive method can easily be implemented for this step.

The second step is the heart of APPLECON. With the specification of simple function arguments, econometric results are obtainable through a wide range of methods. Current (Version-1) capabilities are listed in Table 1. The argument(s) required consist of numeric vectors and/or scalars which describe the order of the variables used in the equations by reference to the appropriate columns of the data matrix.

Table 1. Estimation Capabilities of Version 1

| Function                        | Estimation Technique               |
|---------------------------------|------------------------------------|
| Least-Squares Group             |                                    |
| OLSQ                            | Ordinary least squares             |
| INST                            | Instrumental variables             |
| TSLS                            | Two-stage least squares            |
| ROLS                            | Linear restricted least squares    |
| Generalized-Least-Squares Group |                                    |
| GLSQ                            | Aitken's generalized least squares |
| OLS3                            | Three-stage least squares          |
| SURE                            | Seemingly unrelated regressions    |
| MGLS                            | Mixed generalized least squares    |

The global variables provide all of the information required for supplementary analysis in the third step. At the present time, APLECON provides for plotting, residual analysis, dependent-variable prediction, and confidence-interval construction.

- - - - -

Table 2. Explicit-Result Functions

|     |  |
|-----|--|
| DW  | Durbin-Watson statistic                    |
| FT  | F-Test (zero-slope coefficients)           |
| R2  | R-square                                   |
| RB2 | R-Square (adjusted for degrees of freedom) |
| SB  | Coefficient standard errors                |
| SSE | Sum-of-squares error                       |
| S2  | Estimated disturbance variance             |

- - - - -

Execution of any of the estimation functions produces a set of global variables and, if desired, a summary table of results. The summary table includes a character representation of the equation(s) estimated along with coefficient estimates. In addition, standard econometric results such as hypothesis tests on coefficients and common goodness-of-fit measures are provided. The global variables can be used to report selectively any of the results in the summary table by means of explicit-result functions listed in Table 2.

## II. APL Notation in Econometrics

The description above and the demonstration below indicate that APLECON produces results similar to existing econometric packages. One of the features of APLECON that distinguishes it from other programs of a similar nature is the adoption of a standard APL format for estimation functions, forming an alternate notation for econometric expressions. While in many cases these functions perform complicated calculations, they have been written in a simple form; anyone familiar with econometric results written in

classical notation will be able with little effort to understand the corresponding APL function. To illustrate this feature, we will show APL translations for the estimation techniques OLSQ, TSLS, and OLS3.

When performing ordinary least squares, the equation estimated is usually expressed in econometrics by:

$$Y = XB + E$$

where the variables' dimensions and interpretations are:

|   |        |                               |
|---|--------|-------------------------------|
| Y | (N, 1) | dependent variable            |
| X | (N, K) | explanatory variables         |
| B | (K, 1) | estimated coefficients        |
| E | (N, 1) | ordinary-least-squares errors |

The formula for the estimated coefficients is commonly expressed as:

$$B = (X'X)^{-1}X'Y$$

The error terms and their estimated variance are calculated by:

$$E = Y - XB$$

$$S2 = E'E / (N - K)$$

The variance-covariance matrix for the estimated coefficients is given by:

$$VC(B) = S2(X'X)^{-1}$$

Using appropriate APL notational changes, the results for ordinary least squares can be expressed as:

```

XPX←(QX)+.×X
XPY←(QX)+.×Y
B←(QXPX)+.×XPY
E←Y-X+.×B
S2←(+/E*2)÷N-K
VCB←S2×QXPX

```

These statements form the basis for the APL function OLSQ and are typical of the format in which the other estimation functions are expressed. Each intermediate result is given in a manner that is easily related to similar results expressed in classical notation. OLSQ produces global variables N, K, Y, X, B, E, S2, and VCB, containing enough information to determine other results commonly required in econometrics. These global variables and the explicit-result functions listed in Table 2 can be combined to produce additional results not provided directly by OLSQ, such as the Student t-statistic based upon the null hypothesis that a coefficient is zero--calculated by  $TB←B÷SB$ .

One modification to OLSQ commonly required in econometrics is the technique called Two-Stage Least Squares, TSLS. APLECON has been designed so that additional techniques can be easily added.

This feature of APLECON is illustrated through the construction of a TSLS function. The estimated equation can be expressed by:

$$Y = ZB + E$$

$$Z = (Y1, X1)$$

$$X = (X1, X2)$$

where:

|    |        |                               |
|----|--------|-------------------------------|
| Y  | (N, 1) | dependent variable            |
| Z  | (N, K) | explanatory variables         |
| Y1 |        | included endogenous variables |
| X1 |        | included exogenous variables  |
| X2 |        | excluded exogenous variables  |
| B  | (K, 1) | estimated coefficients        |
| E  | (N, 1) | error terms                   |

TSLS coefficients are calculated in two steps by:

$$Y1H = X(X'X)^{-1}X'Y$$

$$B = (W'Z)^{-1}W'Y$$

where Y1H is the reduced-form predicted values for Y1 and W = (Y1H, X1) is an (N, K)-matrix.

Similar to OLSQ calculations, the remaining global variables for TSLS are calculated by:

$$E = Y - ZB$$

$$S2 = E'E/(N - K)$$

$$VC(B) = S2(W'Z)^{-1}$$

The APL representation of these calculations is:

```
Y1H←X+.×Y1⊗X
W←Y1H,X1
WPZ←(⊗W)+.×Z
WPY←(⊗W)+.×Y
B←(⊗WPZ)+.×WPY
E←Y-Z+.×B
S2←(+/E*2)÷N-K
VCB←S2×⊗WPZ
```

To consider briefly one further example of the way in which the functions in APLECON can be extended, the Three-Stage Least-Squares estimator OLS3 will be presented. This technique estimates a system of G simultaneous equations where equation I is expressed as:

$$YI = ZI BI + UI$$

where YI and ZI are defined as for TSLS and:

|    |         |                             |
|----|---------|-----------------------------|
| BI | (KI, 1) | OLS3 estimated coefficients |
| UI | (N, 1)  | OLS3 error terms            |
| K  | (G)     | Dimensions K1, K2 ..., KG   |

In matrix form, all G equations can be expressed as:

$$Y = ZB + U$$

where the shape of the matrices Y and U is (NG, 1), of Z is (NG, +/K), and of B is ((+/K), 1). The final stage of OLS3 is an application of Aitken's Generalized Least Squares, and therefore an estimate of the system variance-covariance matrix V is required. The elements of this matrix may be calculated by the Kronecker product:

$$V = V(E) \otimes IN$$

where IN is an N-th order identity matrix and V(E) is a (G, G)-matrix with typical element:

$$EI'EJ/[((N-KI)(N-KJ))]^{1/2}$$

where EI represents the TSLS error terms for Equation I. The OLS3 estimation is now calculated by:

$$VC(B) = (ZH'V^{-1}ZH)^{-1}$$

$$B = V(B)ZH'V^{-1}Y$$

$$U = Y - ZB$$

The matrix ZH is like the matrix Z except that the actual values of the included endogenous variables of each equation are replaced by their estimated reduced-form values.

In terms of APL notation, OLS3 can be formulated by first estimating each equation in the system by TSLS, formulating ZH, and then performing Generalized Least Squares. These calculations are summarized below by:

```
I←1
LOOP:
ZI←Y1I,X1I
YIH←X+.×Y1I⊗X
WI←YIH,X1I
WPZ←(⊗WI)+.×ZI
WPY←(⊗WI)+.×YI
BI←(⊗WPZ)+.×WPY
E←E,YI-ZI+.×BI
→(G≥I+I+1)/LOOP
Ⓜ
VE←((⊗E)+.×E)÷((N-K)∘.×N-K)*.5
V←(2ρNG)ρ 1 3 2 4 ⊗VE∘.×(1N)∘.=1N
VCB←⊗(⊗ZH)+.×(⊗V)+.×ZH
B←VCB+.×(⊗ZH)+.×(⊗V)+.×Y
U←Y-Z+.×B
```

Unlike the APL expressions given for OLSQ and TSLS, the expressions for OLS3 have been abbreviated to make more obvious their comparison with classical notation. The complete function for OLS3 must also include within the loop specifications for constructing Y1I, X1I, X, YI, N, K, ZH, Y, and Z as defined in the text.

### III. Demonstration of APLECON--Three-Stage Least Squares

EXAMPLE: Per-capita demand and supply for food consumption, from Kmenta [4].

This demonstration is meant to illustrate briefly the use of APLECON and the type of output to be expected. Complete details and expanded demonstrations are available with the documentation for APLECON.

#### STEP 1. Data Entry

Q Food consumption per head  
P Ratio of food prices to general consumer prices  
D Disposable income in constant prices  
F Ratio of preceding year's prices received by farmers for products to general consumer prices  
A Time in years  
C Column of 1's for intercept term

Specification of Data Labels:

DL←(6 4)P'...Q...P...D...F...A...C'

Labels are four characters each in length. Periods may be used for input spacing, but are automatically blanked later by the output functions of APLECON.

Specification of Data:

Data are entered by standard APL procedures into an (N, M)-matrix D, where N denotes the number of observations and M the number of variables.

#### STEP 2. Estimation

The demand and supply functions to be estimated are expressed respectively in equation form as  $Q = F(P D C)$  and  $Q = F(P F A C)$ . Variables Q and P are endogenous while D, F, A, and C are exogenous. A common econometric technique for estimating this type of system is OLS3.

Estimation of the demand and supply equations is accomplished by the function OLS3. This function has two numeric arguments, a vector GG and a matrix VG. The elements of these arguments for Equation I are:

GG[I] Number of endogenous variables in Equation I  
VG[I;1] Number of variables in Equation I  
VG[I;2] Dependent variable  
2+VG[I;] First- and second-stage variables  
- included endogenous  
- included exogenous  
- excluded exogenous

OLS3 estimation is performed by the following APL expressions:

GG←2 2  
VG←2 7p0  
VG[1;]←4 1 2 3 6 4 5  
VG[2;]←5 1 2 4 5 6 3  
GG OLS3 VG

yielding these results:

\*\*\* THREE-STAGE LEAST SQUARES \*\*\*

#### EQUATION 1

| Q = F ( P D C ) |         |         |         |
|-----------------|---------|---------|---------|
| LB              | B       | SB      | TB      |
| -----           |         |         |         |
| P               | .2436   | .0965   | 2.5243  |
| D               | .3140   | .0469   | 6.6887  |
| C               | 94.6333 | 7.9208  | 11.9474 |
| R2              | FT      | SSE     | S2      |
| -----           |         |         |         |
| .7548           | 26.1722 | 65.7291 | 3.8664  |

#### EQUATION 2

| Q = F ( P F A C ) |         |          |        |  |
|-------------------|---------|----------|--------|--|
| LB                | B       | SB       | TB     |  |
| -----             |         |          |        |  |
| P                 | .2286   | .0997    | 2.2934 |  |
| F                 | .2282   | .0440    | 5.1861 |  |
| A                 | .3611   | .0729    | 4.9546 |  |
| C                 | 52.1972 | 11.8934  | 4.3888 |  |
| R2                | FT      | SSE      | S2     |  |
| -----             |         |          |        |  |
| .5975             | 7.9175  | 107.9138 | 6.7446 |  |

GLOBAL VARIABLES: OLS3

PARA - NG KG GG  
DATA - YG XG Z  
ESTM - BG EG VC VCBG

### IV. Summary and Further Developments

This paper has introduced the ideas behind an APL-based econometric package entitled APLECON. Econometric analysis is easily handled in the environment of an interactive computer terminal, econometric techniques are presented using APL notation and simple APL functions, and APL expressions provide an alternate pedagogical technique to classical econometric notation. This last feature has been illustrated with APL expressions for ordinary, two-stage, and three-stage least-squares estimation. The use of APLECON has been demonstrated with the estimation of a system of simultaneous equations by three-stage least squares. Current capabilities have been summarized in Table 1.

APLECON is undergoing continued development. Additional capabilities to be issued with the next version (late 1981) include estimation functions for minimum-sum absolute errors, autoregressive techniques (Cochrane-Orcutt, Durbin, Hildreth-Lu, and maximum-likelihood), transformed generalized least squares, indirect least squares, reduced-form least squares, and limited-information maximum likelihood. Functions for supplementary analyses have been expanded to include various hypothesis tests, correlation and moment matrices, matrix ranks, and determinants. Long-range plans for APLECON include expansion into the area of computer-assisted instruction and development of a comprehensive package for econometric modeling. Inquiries and questions concerning APLECON and its distribution should be directed to the author.

#### Acknowledgments

Special thanks for help and suggestions throughout the development of APLECON are due Alan Sharpe, Computer Analyst in the Department of Economics. Others at the University who have provided general comments and particular suggestions are Keith Smillie, Department of Computing Science; Dale Bent and Elizabeth Armitage, Department of Computing Services; and Balder Von Hohenbalken, Department of Economics. Financial and computing support from the University of Alberta and its Humanities and Social Science Research Fund are acknowledged. Comments and suggestions by the editors are greatly appreciated.

Stephen D. Lewis  
Department of Economics  
University of Alberta  
Edmonton, Alberta  
Canada T6G 2H4

#### References

- [1] F. Schober and K. Spielberg. APL econometric planning language (EPLAN), Conference Proceedings of APL76, ACM, New York (Sept. 1976) pp. 359-386.
- [2] Time-Series Processor (TSP), Department of Economics, Princeton University (1967).
- [3] K.J. White. A general computer program for econometric methods--SHAZAM, *Econometrica* 46 (Jan. 1978).
- [4] J. Kmenta. *Elements of Econometrics*, Macmillan, New York (1971) p. 563.

□

#### Uniform Sampling on Simplices and Spheres

Balder Von Hohenbalken

In a recent paper [2] on approximating generalised Shapley values (a solution concept in n-person game theory), my colleague T. Levesque and I faced the problem of drawing uniformly distributed sample points on high-dimensional spheres and simplices. We spent much time exploring various suboptimal approaches, until through the kind help of P. Young and E. Nurminski at IIASA (Laxenburg, Austria) I obtained two excellent procedures which are described and discretely implemented in APL in this note.

Turning to spheres first, a suggestive approach is to select points uniformly randomly from the cube circumscribing the unit ball, discarding those vectors with Euclidean norm greater than 1, then normalizing the rest to unit length. The trouble here is that the volume of the ball relative to that of the circumscribed cube decreases rapidly with increasing dimension, leaving an ever-dwindling proportion of sample points to be projected onto the sphere.

A second approach we tried was to draw uniformly distributed polar angles from  $(-\pi/2, \pi/2)$  and then turn them into Cartesian coordinates; however, the appropriate transformations for higher dimensions contain long multiplicative strings of circular functions, which magnify the notorious numerical fuzziness of these functions at critical angles. Thus the resulting distributions on the sphere become intolerably biased beyond dimension 10.

In contrast, E. Nurminski's solution (oral communication) is both beautiful and free from the curse of dimensionality:

Theorem 1: Let  $K = \{x \in R^D : ||x|| = 1\}$  be the unit  $(D-1)$ -sphere (Kugel) and let the elements of  $y \in R^D$  have independent standard normal distributions. Then  $x = ||y||^{-1} y \in K$  is uniformly distributed on  $K$ .\*

Proof: By independence the joint density for  $x$  is:

$$(2\pi)^{-D/2} \exp(-\sum x_i^2/2) \equiv c \exp(-||x||^2/2)$$

for any region A in K:

-----  
\* This result almost certainly lurks somewhere in the literature; I include it here because of its elegance and to save others our meandering.