NBER COMPUTER RESEARCH CENTER NOTES

The NBER Computer Research Center for Economics and Management Science has been engaged, since its formation in 1971, in developing new software systems for quantitative social science research. Prototype systems for exploratory data analysis, mathematical programming, and econometrics are now in various stages of design and implementation. Following are abstracts of five recent working papers. The complete texts are available at $1.00 per copy from the respective authors at: NBER Computer Research Center, 575 Technology Square, Cambridge, Massachusetts 02139.

Hoaglin, D. C., and S. S. Wasserman, "ROSEPACK Document No. 2: Automating Stem-and-Leaf Displays", NBER Working Paper 109 (November 1975).

The stem-and-leaf display is a natural semigraphic technique to include in statistical computing systems. This paper discusses the choices involved in implementing both automated and flexible versions of the display, develops an algorithm for the automated version, examines various implementation considerations, and presents a set of semiportable FORTRAN subroutines for producing stem-and-leaf displays.

Kaden, N. E., and V. Klema, "ROSEPACK Document No. 1: Semiportability of FORTRAN Programs", NBER Working Paper 103 (September 1975).

Transferring FORTRAN subroutines from one manufacturer's machine to another or from one operating system to another puts certain constraints on the construction of the FORTRAN statements that are used in the subroutines. The reliable performance of this mathematical software should be unaffected by the host environment in which the software is used or by the compiler from which the code is generated. In short, the algorithm is to be independent of the computing environment in which it is run.

The subroutines of the NBER's ROSEPACK (Robust Statistics Estimation Package) are FORTRAN IV source code designed to be semiportable, where semiportable is defined to mean transportable with minimum change.

Marsten, R. E., and T. L. Morin, "Parametric Integer Programming: The Right-Hand-Side Case", NBER Working Paper 106 (October 1975).

A family of integer programs is considered whose right-hand-sides lie on a given line segment L. This family is called a parametric integer program (PIP). Solving a (PIP) means finding an optimal solution for every program in the family. It is shown how a simple generalization of the conventional branch-and-bound approach to integer programming makes it possible to solve such a (PIP). The usual bounding test is extended from a comparison of two point values to a comparison of two functions defined on the line segment L. The method is illustrated on a small example, and computational results for some larger problems are reported.

Wasserman, S. S., "Random Directed Graph Distributions and the Triad Census in Social Networks", NBER Working Paper 113 (November 1975).

This paper uses the concept of the triad census first introduced by Holland and Leinhardt, and describes several distributions on directed graphs.

Methods are presented for calculating the mean and covariance matrix of the triad census for the uniform distribution that conditions on the number of choices made by each individual in the social network. Several complex distributions on digraphs are approximated, and an application of these methods to a sociogram is given.

Welsch, R. E., "Confidence Regions for Robust Regression", NBER Working Paper 111 (November 1975).

This paper describes the results of a Monte Carlo study of certain aspects of robust regression confidence-region estimation for linear models with one, five, and seven parameters. One-step sine estimators (c = 1.42) were used with design matrices consisting of short-tailed, Gaussian, and long-tailed columns. The samples were generated from a variety of contaminated Gaussian distributions.

A number of proposals for covariance matrices were tried, including forms derived from asymptotic considerations and from weighted least squares with data-dependent weights. Comparisons with the Monte Carlo "truth" were made using generalized eigenvalues. In order to measure efficiency and compute approximate t-values, linear combinations of parameters corresponding to the largest eigenvalues of the "truth" were examined.

For design matrices with columns of modest kurtosis, the covariance estimators all give reasonable results, and after adjusting for asymptotic bias, some useful approximate t-values can be obtained. This implies that the standard weighted least squares output using data-dependent weights need only be modified slightly to give useful robust confidence intervals.

When design-matrix kurtosis is high and severe contamination is present in the data, these simple approximations are not adequate.