

## SPAN: An interactive BASIC program for spectral analysis of time-series data

REBECCA M. WARNER and PAUL G. NEUMANN  
*Laboratory for the Online Acquisition of Data*  
*University of Miami, Coral Gables, Florida 33124*

Spectral analysis is used to evaluate whether time-series data contain regular cycles (periodic components). This type of analysis typically involves two steps: First, the Fourier transform is used as a means of breaking down the variance of the time series into estimates of the contributions made by each of a set of periodic components, and second, a variety of smoothing or average procedures (often referred to as "windows") are used to improve the reliability of these estimates. There are several algorithms available to do Fourier transforms and a wide variety of averaging and smoothing techniques; the choice among these depends in part on computational efficiency and also on the characteristics of the data. The SPAN program uses the Sande-Tukey Fast-Fourier transform algorithm (in Bloomfield, 1976) and includes optional detrending, tapering, and hamming (described below).

Intuitively, the set of spectral estimates for a time series can be interpreted as the result of a set of tests for the presence of cycles with lengths of  $m/1$ ,  $m/2$ ,  $m/3$ , . . . ,  $m/q$  observations (where  $q = m/2$ , and  $m$  is the longest cycle length of interest to the investigator). These  $q$  different periodic components are orthogonal to each other. Each of the  $q$  spectral estimates, therefore, effectively represents the part of the variance of the original time series accounted for by the corresponding periodic component. For the Sande-Tukey algorithm used by the SPAN program, the value of  $m$  must be a power of two that is less than or equal to the number of observations in the original time series. The user selects  $m$  so that the cycle lengths of interest will be included in the set of periodic components.

**Calculations.** Program SPAN reads in a time-series vector of up to 600 data points (easily expandable) from an existing sequential access disk file. The mean of the time series is automatically subtracted out of each observation. The user may request optional removal of linear trend. This means that a regression is run to predict  $X_t$  (the observation at time  $t$ ) from  $t$  (the time index); the detrended data consist of the residuals from this regression. Optional tapering of the time-series data is also available (Bloomfield, 1976). Essentially, tapering involves multiplying the observations near the beginning and end of the time series by fractional weights, in order to prevent the endpoints of the observation period from affecting the power spectrum in undesirable ways. The program supplies prompts to the user, so that the choice of options can be made at execution time. The

program also requests the value of  $m$  (maximum cycle length), and the lagged autocovariances are calculated for lags ranging from 1 to  $m$ . These lagged autocovariances are input to the Fast-Fourier transform routine based on the Sande-Tukey algorithm (Bloomfield, 1976). The resulting spectral estimates may be smoothed by a procedure known as "hamming" (Chatfield, 1975), which involves replacing each spectral estimate by a weighted average of itself and the spectral estimates for the neighboring frequencies. The real and imaginary parts of the complex numbers that are output by the Fast-Fourier transform may optionally be saved in a side file if the user wishes to use these intermediate results as input to other programs.

The final summary output for SPAN is saved on a side file for later printing. It includes a table that lists frequency, period, spectral estimate, and percent of variance accounted for by each of the  $q$  periodic components. The spectral estimates are also plotted on a linear, standardized scale with optional fit to 80- through 132-character print lines. A  $p$  value for the significance of the largest spectral estimates is given. However, since this  $p$  value is based on a test for the size of periodogram ordinates, which typically have larger error variance than spectral estimates, it is actually too conservative (Shimshoni, 1971). However, derivation of exact  $p$  values would be a complicated problem; this  $p$  value is merely intended as a guideline for interpretation of output.

**Interpretation of Output.** When a set of  $q$  periodic components is fitted to a time series that consists of a set of discrete observations, one would expect that each of the  $q$  periodic components would account for a  $1/q$  proportion of the variance. Therefore, any spectral estimate that accounts for much more than this proportion of variance is an indication of significant periodicity. The  $p$  value given for the largest spectral estimate is for a test of the null hypothesis that the spectral estimate did not account for a larger than expected share of the variance. If  $p$  is small (less than .05), then the spectral estimate is judged to be statistically significant. The  $p$  value is based on a joint confidence interval for the whole set of  $q$  spectral estimates. In effect, it tells the user how often he could expect to obtain a power spectrum including a peak as large as or larger than the one obtained if random data were used.

**Program Limitations.** This program is written in Alpha BASIC, a version of BASIC for the Alpha Microsystems AM-100 microcomputer. It runs in less than 5 kbytes of memory. The amount of time required for the calculation of the lagged autocovariances and the Fast-Fourier transform is substantial (approximately 3-5 min for a typical job). This BASIC is a true compiler running on a machine that incorporates arithmetic hardware. Therefore, users who translate the program for use on machines that have arithmetic software can

expect a very substantial increase in run time, as much as 10 to 15 times as long as the estimated run time given above.

**Program Availability.** Annotated program output and a listing of SPAN are available from Rebecca Warner, Laboratory for the Online Acquisition of Data, Department of Psychology, P.O. Box 248185, University of Miami, Coral Gables, Florida 33124.

#### REFERENCES

- BLOOMFIELD, P. *Fourier analysis of time series: An introduction.* New York: Wiley, 1976.
- CHATFIELD, C. *The analysis of time series: Theory and practice.* New York: Halsted, 1975.
- SHIMSHONI, M. On Fisher's test of significance in harmonic analysis. *Geophysical Journal of the Royal Astronomical Society*, 1971, **23**, 373-377.

(Accepted for publication May 19, 1980.)