

Forecasting Time Series Data with the SAS System for Information Delivery

Andrew H. Karp, Sierra Information Services

SouthEast SAS(r) Users Group
September 18-20, 1994
The Omni at Charleston Place
Charleston, South Carolina

**Forecasting Time Series Data
with the
SAS System for Information Delivery**

Andrew H. Karp
President, Sierra Information Services
San Francisco, California
and
Instructor in Computer and Information Systems
University of California-Berkeley
Extension Division

415/441-4702 (voice/fax)
SFBAV001@AOL.COM (Internet)

© 1994 Andrew H. Karp/Sierra Information Services

**Forecasting Time Series Data
with the
SAS System for Information Delivery**

- **Problem:** how to generate "reliable" forecasts of the potential future values of a series of observations which are ordered in time
- What SAS procedures are available to examine a time series and obtain future forecasts of its values?
- How do we know when we have a "good" forecast?

© 1994 Andrew H. Karp/Sierra Information Services

**Forecasting Time Series Data
with the
SAS System for Information Delivery**

■ **Key Terms and Phrases:**

- **Time Series:** a set of observations (data points) ordered in time, with equal periods between the observations. Most frequently a vector, or column of data.
- **Autocorrelation:** the mathematical dependency between a value in the time series and a previous value in the time series
- **Time Series Analysis:** the class of statistical techniques which are applied to data collected in the time domain

© 1994 Andrew H. Karp/Sierra Information Services

**Forecasting Time Series Data
with the
SAS System for Information Delivery**

■ **Analytic Techniques Applied to Time Series Data:**

- **Graphical tools:** visual display of the series
- **Assessment of autocorrelation:** determine the dependency between observations in the data set
- Use the observed dependency to generate forecasts of future values of the time series
- Choose the "best" model from competing potential future forecasts of the same series.

© 1994 Andrew H. Karp/Sierra Information Services

**Forecasting Time Series Data
with the
SAS System for Information Delivery**

■ **Graphical tools available in the BASE SAS module:**

- **PROC PLOT:** Plots the values of two (or more) continuous or interval-level variables. User can override SAS-generated horizontal and/or vertical axis ranges and/or declare unique plotting symbols. Procedure accepts SAS date/time formats as values for the horizontal axis.
- **PROC TIMEPLOT:** Uniquely suited to display of time-ordered data. Like PROC PLOT, the TIMEPLOT procedure can plot two or more continuous-level variables with user-defined plotting symbols. TIMEPLOT also prints the values of the variables, and the JOINREF option permits joining of values occurring at the same time period. TIMEPLOT also generates multi-page output rather than condensing output on one page, which is the default in PROC PLOT.

© 1994 Andrew H. Karp/Sierra Information Services

**Forecasting Time Series Data
with the
SAS System for Information Delivery**

■ **PROC REG: performs ordinary least-squares regression analysis**

- Computes the Durbin-Watson test (d) and coefficient of first order autocorrelation if the DW option is specified in the MODEL statement. Residuals from forecast can be output to a new SAS data set for subsequent analysis using PROC UNIVARIATE and/or PROC PLOT.
- OLS regression requires residuals to be independently and identically distributed or estimated parameters are biased.
- Durbin-Watson test only measures first-order autocorrelation
 - autocorrelation frequently occurs at higher orders, especially in seasonal series.

© 1994 Andrew H. Karp/Sierra Information Services

Forecasting Time Series Data
with the
SAS System for Information Delivery

■ PROC AUTOREG:

- Examines the correlated error structure at lags 1 to p , where p is the user-supplied number of lags to be analyzed.
- Uses detected autocorrelation to produce predicted values of the dependent variable which are less statistically biased than those from OLS regression. Uses Yule-Walker methodology by default, other methods optionally available.
- Non-significant autoregressive parameters can be eliminated in stepwise fashion using the BACKSTEP option in the MODEL statement.

Forecasting Time Series Data
with the
SAS System for Information Delivery

■ PROC AUTOREG: continued

- Optionally produces DW tests and associated p-values at lags 1 to the order specified by the user
- In Version 8.06, supports Generalized Autoregressive Conditional Heteroscedasticity (GARCH) techniques to model and correct for non-constant variance conditions.
- Predicted, actual and forecasted values easily output to a new SAS data set for subsequent inspection and analysis.
- Does not automatically generate forecasts for values of the time-indexed variable beyond those in the input data set. Substantial user invention is required to generate forecasts of future values in the time series

© 1994 Andrew H. Karp/SAS Information Services

Forecasting Time Series Data
with the
SAS System for Information Delivery

■ PROC FORECAST: Produces forecasts of univariate time series using one of three methods:

- Stepwise Autoregressive (STEPAR): Takes the difference between each observation and the estimated trend, and then fits the remaining variation using a backwards-stepping autoregressive process. Essentially the same type of statistical model generated by PROC AUTOREG and used to generate forecasts of future values of the series. The default method in PROC FORECAST.
- Exponential Smoothing (EXPO): fits a trend model with an exponential weighting constant so that more recent observations have greater influence on the forecast than prior observations. User can adjust exponential weights, if they desire.
- Winters Seasonal Method (WINTERS): Similar to exponential smoothing, but includes a seasonal parameter, which is defined by the user, to identify seasonal components in the time series.

© 1994 Andrew H. Karp/SAS Information Services

Forecasting Time Series Data
with the
SAS System for Information Delivery

■ PROC FORECAST (continued):

- Generates two output SAS data sets, but no printed output.
 - OUT= data set contains actual and forecast values, residuals, upper and lower confidence intervals, and standard deviations. Contents can then be displayed using PROC PLOT and/or TIMEPLOT.
 - OUTEST= data set contains parameter estimates generated during execution of PROC FORECAST. Contents can be viewed via PROC PRINT. Useful for comparing competing models of the same series.
- Capable of producing multiple forecasts using the BY statement.
- Limited user intervention is required to generate many forecast series, but at the loss of some statistical precision.

© 1994 Andrew H. Karp/SAS Information Services

Forecasting Time Series Data
with the
SAS System for Information Delivery

■ PROC ARIMA (Autoregressive Integrated Moving Average)

- Implements Box-Jenkins approach to diagnosing a time series, estimating its component parts, generating future values of the series, and checking adequacy of model output.
- Performs differencing operations for both stationarity and/or seasonality at the user's request
- Generates forecasts for univariate series, transfer functions, and intervention models
- Invoked by PROC statement and controlled by a hierarchical set of statements corresponding to the Box-Jenkins forecasting methodology.

© 1994 Andrew H. Karp/SAS Information Services

Forecasting Time Series Data
with the
SAS System for Information Delivery

■ PROC ARIMA (continued):

- **Identify:** reads in the time series data to be used in future steps. Produces correlograms of the ACF, IACF and PACF of the series, as well as a test for white noise. Correlograms identify whether the series is autoregressive, moving average, or both, and the corresponding order of the model, and are also used to provide a visual check for stationarity and/or over-differencing.
- **Estimate:** User states one or more potential models in terms of the AR and/or MA processes uncovered during the IDENTIFY phase. Output includes several key statistics to check for model adequacy and compare competing models for the same time series.
- **Forecast:** Once an appropriate model is found, future forecast values of the series are generated and output to a SAS data set.

© 1994 Andrew H. Karp/SAS Information Services

Forecasting Time Series Data with the SAS System for Information Delivery

■ PROC ARIMA (continued):

- Produces mathematically precise forecasts, depending on the user's ability to interpret output from the IDENTIFY and ESTIMATE stages.
- Requires substantial user expertise and intervention.
- Several iterations of PROC ARIMA are usually required to produce one forecast.
- Cannot produce simultaneous forecasts of subsets of the time series.

© 1994 Andrew H. Karp/SAS Information Services

Forecasting Time Series Data with the SAS System for Information Delivery

■ Other forecasting procedures:

- PROC X11: Implements the U.S. Census Bureau's X-11 Seasonal Adjustment Program.
- PROC SPECTRA: performs analysis of frequency-domain data. Is also capable of applying the Discrete Fourier Transformation to time domain data.
- PROC DATASOURCE: extracts time series data from files available from vendors and stores them in a SAS data set.
- PROC EXPAND: interpolates missing values in a time series, aggregates and interpolates data to different higher/lower frequencies.
- PROC TCSREG: performs time-series cross-sectional regression.
- PROC PDLREG: applies polynomial distributed lag models to data.

© 1994 Andrew H. Karp/SAS Information Services

Forecasting Time Series Data with the SAS System for Information Delivery

■ What's new?

- Substantially enhanced documentation
 - Improved descriptions, examples, and sample SAS code available in the Second Edition of the SAS/ETS module documentation for Version 6.
 - New, two volume, *SAS/ETS Software: Applications Guide*, provides step-by-step instructions as well as both theoretical and applied details underlying correct application of these procedures
- MACROS to perform important statistical tests not implemented in ETS procedures.
 - Box-Cox Transformations
 - Dickey-Fuller unit root test
 - Significance of the Dickey-Fuller unit root test
 - Test appropriateness of a logarithmic transformation to a time series

© 1994 Andrew H. Karp/SAS Information Services