

Structural Analysis of Time Series Using the SAS/ETS® UCM Procedure

Rajesh Selukar, SAS Institute Inc., Cary, NC

ABSTRACT

This article introduces the SAS/ETS UCM procedure, which uses structural models to analyze time series data. Structural models provide regression-like decomposition of the response series into latent components (such as trend, seasonal, or other periodic components) and linear and nonlinear regression effects. Apart from the series forecasts, structural modeling provides estimates of these unobserved components; these estimates are very useful in practical decision making. In SAS® 9.2 the UCM procedure contains several new features: incorporation of linear and nonlinear regression effects with time-varying coefficients, approximation of long and complex seasonal patterns by using splines and trigonometric polynomials, detection of structural change, and additional ODS graphics. A few real-life examples illustrate the functionality of the UCM procedure.

INTRODUCTION

An important problem in time series analysis concerns analyzing observations that are collected on a numeric response variable at regular time intervals, such as monthly or quarterly. In addition to the response variable, observations on some predictor variables might also be available. Usually the goals of time series analysis are:

- Forecast the future values of the response series along with some measure of the forecast error.
- Study the impact of predictors on the response series.
- Decompose the response series into constituent factors such as trend, periodic effects, and regressor effects, and be able to estimate and forecast these constituent factors.
- Discover unusual observations or structural changes in the response variable's historic pattern.
- Interpolate the missing response values.
- Obtain an estimate of the response series after noise is removed.

You can use a variety of statistical techniques for achieving some or all of the preceding goals. The structural time series models, also called the unobserved components models (UCMs), constitute a large and flexible class of models that has proved very useful for these purposes. Structural models assume a regression-like decomposition of the response series into latent components (such as trend, seasonal, or other periodic components) and regression effects that can be linear, nonlinear, time-varying, and time-invariant. The latent components, which follow suitable stochastic models, are a key feature of the structural models; they provide a customizable set of patterns to capture the salient movements of the response series. The goal is to include appropriate latent components and regressors and to choose suitable stochastic models of the selected components so that the resulting model explains the observed series well. The UCM procedure implements many of the commonly needed UCMs and provides a rich framework for structural modeling.

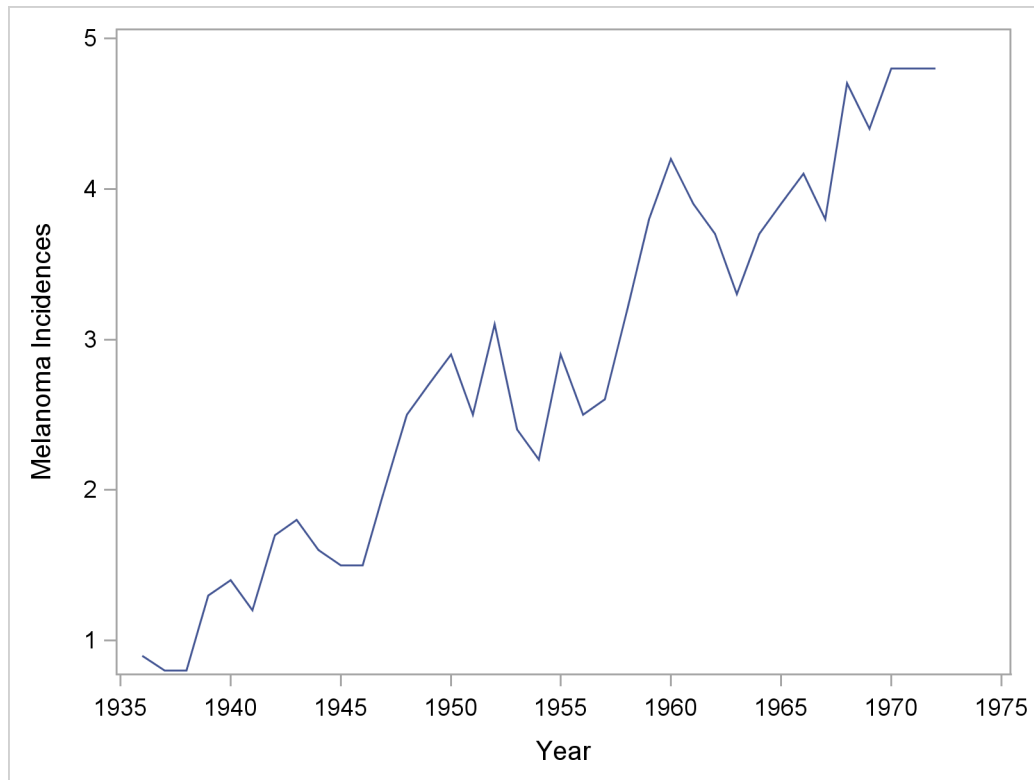
Harvey (1989) is a good reference for UCM-based time series modeling. Traditionally, the integrated autoregressive moving average (ARIMA) models (Box and Jenkins 1976) and, to a limited extent, the exponential smoothing models have been the main tools in the analysis of this type of time series data. It is fair to say that the UCMs capture the versatility of the ARIMA models while possessing the interpretability of the smoothing models. A thorough discussion of the correspondence between the ARIMA models and the UCMs, and the relative merits of UCM and ARIMA modeling, is given in Harvey (1989). The UCMs are also very similar to another set of models, called the *dynamic models*, that are popular in the Bayesian time series literature (West and Harrison 1999).

The remaining sections of this paper are organized as follows: the second section has a simple illustrative example to motivate the main ideas behind structural modeling, the third section provides a brief description of the different UCMs that can be specified, the fourth section describes the available diagnostic tools, and the last section provides two additional examples.

ANALYSIS OF THE MELANOMA SERIES

This example illustrates the main aspects of the structural modeling approach to time series analysis. The time series considered in this example presents the age-adjusted numbers of melanoma incidences per 100,000 people from the Connecticut Tumor Registry for the 37 years from 1936 to 1972 (Houghton, Flannery, and Viola 1980). Figure 1 shows a time series plot of these incidence numbers.

Figure 1 Time Series Plot of the Melanoma Incidences



The plot shows that the melanoma incidences have an increasing trend, nearly linear, and possibly a superimposed cyclical pattern. The following UCM can capture this type of behavior:

$$y_t = \mu_t + \psi_t + \epsilon_t$$

This model contains three latent components: a stochastic trend denoted by μ_t , a stochastic cycle denoted by ψ_t , and the overall error ϵ_t . The stochastic models for these components are discussed next.

The UCM procedure offers two ways to model the trend component μ_t . The first model, called the random walk (RW) model, implies that the trend remains approximately constant throughout the life of the series without any persistent upward or downward drift. The RW model can be described as

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim i.i.d. N(0, \sigma_\eta^2)$$

Note that if $\sigma_\eta^2 = 0$, then the RW model becomes $\mu_t = \text{constant}$. In the second model the trend is modeled as a locally linear time trend (LLT). In the LLT model the trend is modeled using time varying *level* μ_t and time varying *slope* β_t . This model is described by the following evolution equations:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim i.i.d. N(0, \sigma_\eta^2) \\ \beta_t &= \beta_{t-1} + \xi_t, & \xi_t &\sim i.i.d. N(0, \sigma_\xi^2) \end{aligned}$$

The disturbances η_t and ξ_t are assumed to be independent. You can obtain interesting special cases of this model by setting one or both of the disturbance variances σ_η^2 and σ_ξ^2 equal to zero. If you set σ_ξ^2 equal to zero, then the resulting

model has linear trend with fixed slope. If you set σ_η^2 to zero, then the resulting model usually has a smoother trend. If you set both the variances to zero, then the resulting model is the deterministic linear time trend: $\mu_t = \mu_0 + \beta_0 t$. For the melanoma incidences series the LLT model seems suitable.

Similar to the trend, the UCM procedure offers a stochastic cycle component that can capture cyclical patterns of time-varying amplitude and phase. Note that a deterministic cycle ψ_t with frequency λ , $0 < \lambda < \pi$, can be written as

$$\psi_t = \alpha \cos(\lambda t) + \beta \sin(\lambda t)$$

ψ_t is a periodic function with period $2\pi/\lambda$, amplitude $(\alpha^2 + \beta^2)^{1/2}$, and phase $\tan^{-1}(\beta/\alpha)$. In time series situations it is useful to generalize this simple cyclical pattern to a stochastic cycle that has a fixed period but time-varying amplitude and phase. The stochastic cycle considered here is motivated by the following recursive formula for computing ψ_t :

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix}$$

starting with $\psi_0 = \alpha$ and $\psi_0^* = \beta$. Note that ψ_t and ψ_t^* satisfy the relation

$$\psi_t^2 + \psi_t^{*2} = \alpha^2 + \beta^2 \quad \text{for all } t$$

You can obtain a stochastic generalization of the cycle ψ_t by adding random noise to this recursion and by introducing a damping factor ρ for additional modeling flexibility. This model can be described as follows:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda & \sin \lambda \\ -\sin \lambda & \cos \lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} v_t \\ v_t^* \end{bmatrix}$$

where $0 \leq \rho \leq 1$ and the disturbances v_t and v_t^* are independent $N(0, \sigma_v^2)$ variables. The resulting stochastic cycle has a fixed period but time-varying amplitude and phase. This cycle component depends on three parameters: the cycle frequency (λ), the damping factor (ρ), and the variance (σ_v^2) of the disturbance term v_t . Note that if $\rho = 1$ and $\sigma_v^2 = 0$, this cycle model reduces to the usual deterministic cycle.

The overall error term is usually modeled as Gaussian white noise (that is, a sequence of independent, zero mean, normal random variables with fixed variance), but it can be modeled as an autoregressive moving average (ARMA) process also. In this example the error term is taken to be Gaussian white noise.

The following SAS statements show how to specify this model in the UCM procedure and how to request the desired output:

```
proc ucm data=Melanoma;
  id Year interval=year;
  model Incidences;
  irregular plot=smooth;
  level variance=0 noest;
  slope variance=0 noest;
  cycle plot=smooth;
  estimate back=5 plot=panel;
  forecast back=5 lead=5 plot=decomp;
run;
```

The UCM procedure statements can be grouped as the main procedure statement, an ID statement to associate a time index with the observations, statements used for model specification, a statement to control the parameter estimation process, and a statement to control the forecasting phase. The next few subsections explain the preceding UCM procedure statements according to this organization.

Input Data and ID Specification

The PROC UCM statement signifies the start of the UCM procedure and specifies the input data set Melanoma, which contains the response series. The optional ID statement specifies a date, datetime, or time identification variable (Year in this example) to label the observations. The INTERVAL=YEAR option in the ID statement indicates that the measurements were collected on a yearly basis.

Model Specification

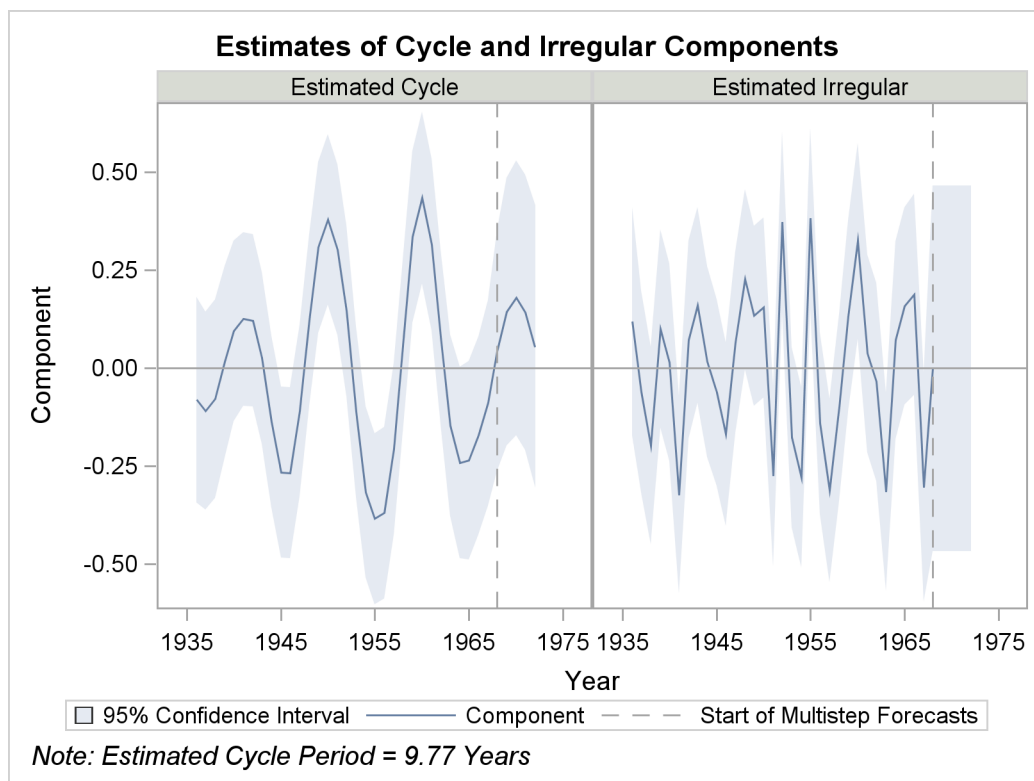
The model specification begins with the MODEL statement, which specifies the response series (Incidences in this case). If the model contains any linear, time-invariant regressors, they are also specified in the MODEL statement.

After this, you can specify the components in the model by using separate statements that enable you to control their individual properties. The IRREGULAR statement specifies the irregular component ϵ_t , and the LEVEL and SLOPE statements specify the trend component μ_t . The CYCLE statement specifies the cycle component ψ_t . Six parameters are associated with this model: the variances of the disturbance terms in the evolution equations of μ_t , β_t , ψ_t and ϵ_t , and the damping factor and the period associated with the cycle ψ_t . You estimate these parameters by maximizing the likelihood of the data. Optionally, you can fix some of these parameters to desired values. For example, the following statements in the program specify the deterministic linear trend by fixing the disturbance variances in the trend evolution equations to zero:

```
level variance=0 noest;
slope variance=0 noest;
```

In addition to model specification, you can use the options in the individual component statements to output the estimated component, either in printed or graphical form. For example, the PLOT=SMOOTH option in the CYCLE and IRREGULAR statements requests the plotting of the smoothed—that is, full sample—estimates of the cycle component ψ_t and the irregular component ϵ_t . Figure 2 shows these plots. In the usual UCM procedure output these plots appear separately; they are paneled here to save space.

Figure 2 Estimates of Cycle and Irregular Components



Note that the estimated cycle has time-varying amplitude.

Estimation Phase

The ESTIMATE statement is an optional statement that controls the overall model-fitting environment. Many useful tables such as the parameter estimates table and a variety of goodness-of-fit statistics are displayed by default. In addition you can control the span of observations used to fit the model, request a variety of other model diagnostic information including different residual diagnostic plots, and save the estimated parameters and other related information in data sets. Figure 3 shows the parameter estimates table for this model.

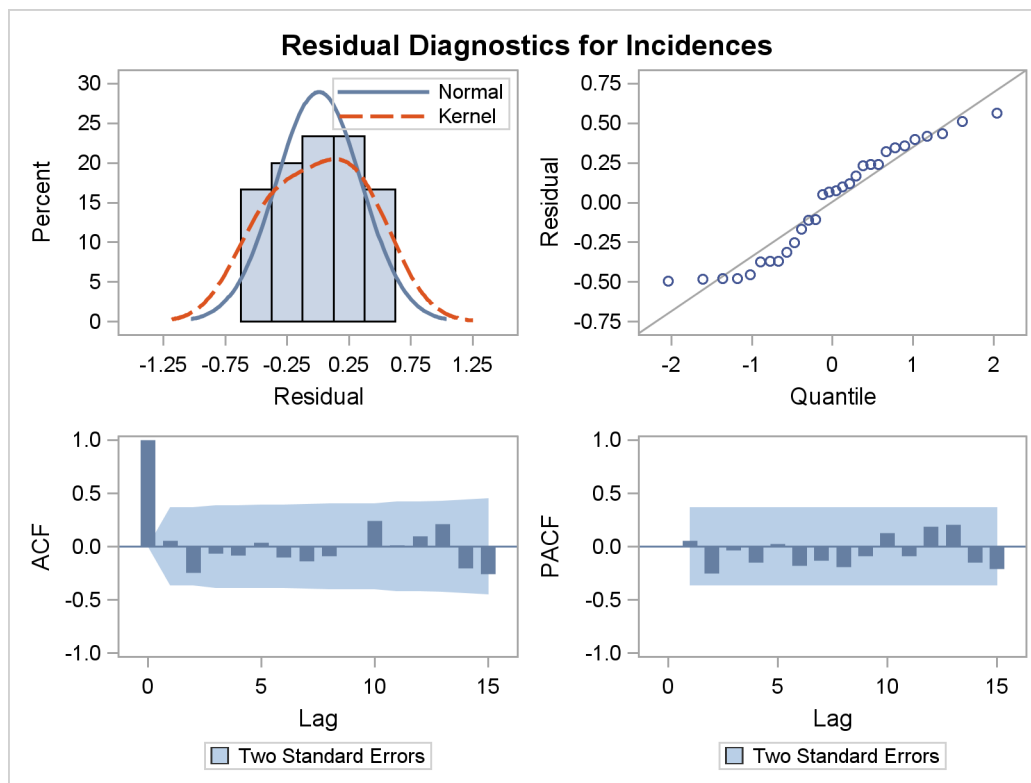
Figure 3 Parameter Estimates for the Trend Plus Cycle Plus Irregular Model

The UCM Procedure					
Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > t
Irregular	Error Variance	0.05675	0.02387	2.38	0.0174
Cycle	Damping Factor	0.94419	0.08743	10.80	<.0001
Cycle	Period	9.76778	0.89263	10.94	<.0001
Cycle	Error Variance	0.00590	0.0045948	1.28	0.1994

The estimated cycle period is 9.77, the damping factor is 0.94, and the disturbance variance is 0.006.

In this example the ESTIMATE statement has two options: BACK=5 and PLOT=ANEL. The BACK=5 option indicates that the last five measurements of the data are to be ignored during the parameter estimation. This can be useful when you want to validate the forecasting performance of a model on the withheld data. The PLOT=ANEL option produces a summary panel of the residual diagnostics, which is shown in Figure 4. The first two plots, the residual histogram and the quantile plot, check the normality assumption. The remaining two plots, the residual autocorrelation plot and the residual partial autocorrelation plot, check whether the residuals resemble white noise. The plots do not indicate significant violations of the model assumptions.

Figure 4 Residual Diagnostics



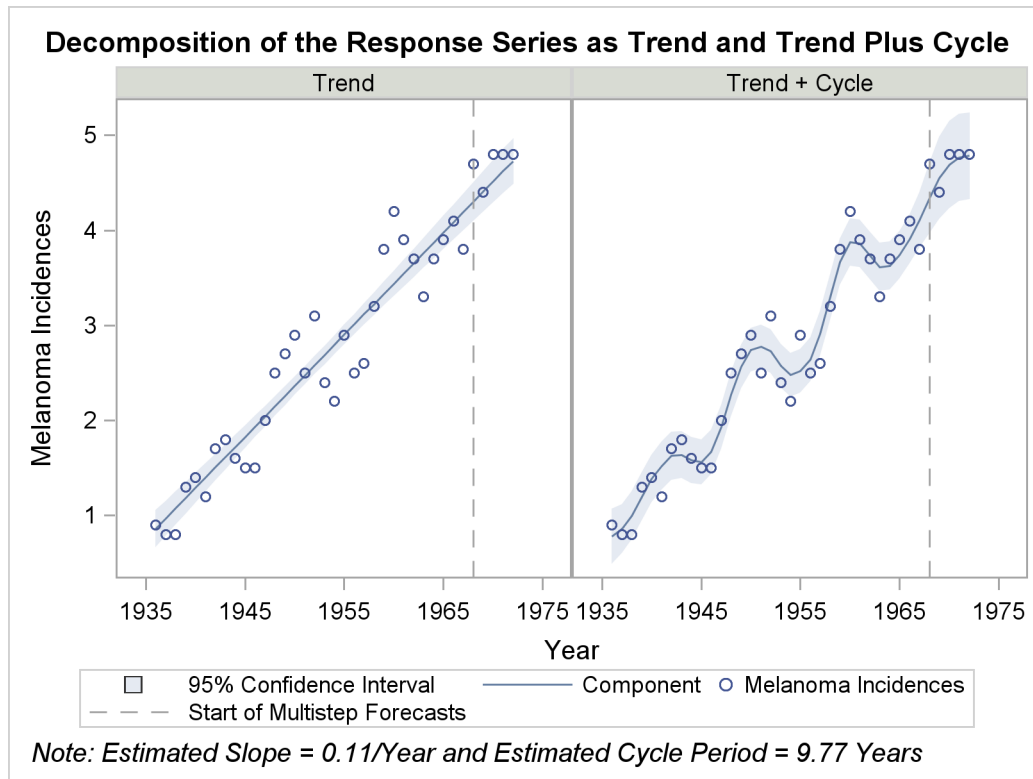
Forecasting Phase

Similar to the ESTIMATE statement, the FORECAST statement specifies the overall forecasting environment for the specified model. You can use the FORECAST statement to specify the span of observations to use to compute the forecasts of the future observations, the number of periods to forecast beyond the historical period, and the significance level of the forecast confidence interval. You can also display or save the series forecasts and the model-based decomposition of the series.

In this example the FORECAST statement has three options: BACK=5, LEAD=5, and PLOT=DECOMP. The option

PLOT=DECOMP requests the plotting of the smoothed trend (μ_t) and the trend-plus-cycle component ($\mu_t + \psi_t$). As in the ESTIMATE statement, the BACK=5 option indicates that the last five measurements of the data are to be ignored during the forecast computations. The LEAD=5 option requests prediction of five future values of the response series. Note that the combination of BACK=5 and LEAD=5 options enable the validation of the forecasting performance of the model on the withheld data. Figure 5 shows the decomposition plots. In the usual UCM procedure output, these plots appear separately; they are paneled here to save space. Note that the estimated trend is perfectly linear because the LEVEL and SLOPE disturbance variances are set to zero. Also note that the trend-plus-cycle plot captures the main aspects of the series quite well.

Figure 5 Decomposition Plots for the Melanoma Incidences



SUPPORTED MODELS

You can specify quite a variety of structural models in the UCM procedure. The structural aspects of a UCM can be broadly categorized as a suitable trend component, zero or more periodic components (usually expressed as superposition of cycles and seasons), zero or more regression effects, and an overall error term. Specifically, suppose that y_t denotes the response series and x_{jt} denote the predictor series. In the UCM procedure you can specify a model of the following type:

$$y_t = \mu_t + \gamma_t + \psi_t + \sum_{j=1}^m \beta_{jt} x_{jt} + \sum_{k=1}^p \delta_k y_{t-k} + \epsilon_t$$

where: μ_t denotes a stochastic trend; ψ_t denotes contribution of zero or more stochastic cycles; γ_t denotes contribution of zero or more stochastic seasons; $\sum_{j=1}^m \beta_{jt} x_{jt}$ denotes, with slightly erroneous notation, contribution of zero or more linear, nonlinear, time-varying, and time-invariant regression effects; $\sum_{k=1}^p \delta_k y_{t-k}$ denotes contribution of zero or more lagged response values; and ϵ_t is the overall error term, usually white noise, but possibly colored (ARMA).

In the remainder of this section some of the modeling features that are newly added in SAS 9.2 are briefly reviewed. The new additions are in four areas: more general regression effect specification, more ways to specify parsimonious models for seasonal effects with long season lengths, more general irregular component, and detection of additive outliers and level shifts. For additional details, see Chapter 29, "The UCM Procedure" (*SAS/ETS User's Guide*).

Modeling Regression Effects

A predictor variable can affect the response variable in a variety of ways. The UCM procedure enables you to model several different types of predictor-response relationships:

- The predictor-response relationship is *linear*, and the regression coefficient does not change with time. This is the simplest kind of relationship, and such predictors are specified in the MODEL statement.
- The predictor-response relationship is *linear*, but the regression coefficient changes with time. Such predictors are specified in the new RANDOMREG statement. Here a regression coefficient—say, β_t —is assumed to evolve as a random walk:

$$\beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim i.i.d. N(0, \sigma^2)$$

Of course, if the random walk disturbance variance σ^2 is zero, then the regression coefficient is not time-varying, and it reduces to the standard regression setting. You can specify multiple RANDOMREG statements, and each statement can contain one or more regressors. The regressors in a particular RANDOMREG statement form a group that is assumed to share the same disturbance variance parameter.

- The predictor-response relationship is *nonlinear*, and the relationship can change with time. This type of relationship can be specified by an appropriate spline. In the time-varying case the regression coefficients of the regressors underlying the spline specification evolve as a random walk. You can specify such predictors in the new SPLINEREG statement.
- A response variable can depend on its own past values—that is, lagged dependent values. You can specify such a relationship in the DEPLAG statement.

Modeling Seasonal Effects with Long Season Lengths

In SAS 9.2 the UCM procedure offers two new ways to specify parsimonious models for long seasonal patterns:

- The new DROPH= and KEEPH= options in the SEASON statement provide complete control over the harmonics included in the trigonometric seasonal model. In the trigonometric seasonal model, you express the seasonal pattern as a sum of stochastic cycles of appropriate frequencies called harmonics. For long seasonal patterns you can specify a parsimonious seasonal model by excluding some insignificant harmonics.
- The new SPLINESEASON statement enables specification of spline approximation to a long seasonal pattern.

Modeling the Irregular Component

The irregular component represents the residual variation remaining in the response series that is modeled using an appropriate selection of structural effects. In most cases, the irregular component can be assumed to be simply Gaussian white noise. However, in some cases the residual variation can be more complicated. In such situations, you might need to model the irregular component as a stationary autoregressive moving average (ARMA) process. As an experimental feature, you can now specify an $ARMA(p,q) \times (P,Q)_s$ irregular component by using the new options in the IRREGULAR statement. Moreover, you can use the ARMA irregular component together with the dependent lag specification (using the DEPLAG statement) to specify an integrated ARMA, $ARIMA(p,d,q) \times (P,D,Q)_s$, model for the response series.

MODEL DIAGNOSIS

The UCM procedure provides extensive tabular and graphical output that is useful for model evaluation. The tabular output includes the following:

- Fit statistics based on log likelihood. These include the log likelihood and the various information criteria such as AIC, BIC, and so on.
- Fit statistics based on one-step-ahead forecast residuals. These include mean squared error (MSE), mean absolute percent error (MAPE), and so on.
- Statistics useful for conducting significance analysis of the components in the model.

The following residual diagnostic plots are available:

- Time series plot of residuals for visual inspection of residuals. Optionally, you can superimpose a LOESS smooth on this plot to examine whether additional patterns remain in the residuals.
- Residual histogram and quantile-quantile plot for normality check.
- Residual autocorrelation and partial autocorrelation plots to check the white noise assumption.
- Time series plots of cumulative residuals and squared cumulative residuals for detecting structural breaks.

EXAMPLES

This section contains two examples that illustrate some of the new features in the UCM procedure. The first example illustrates a time-varying regression situation. The second example illustrates the use of SPLINEREG statement for modeling of nonlinear regression effects and the use of new options in the IRREGULAR statement for the specification of colored noise.

Traffic Accidents Data

In April 1979 the Police Department of Albuquerque, New Mexico, began a special enforcement program aimed at reducing the number of DWI (driving while intoxicated) accidents. The program was administered by a squad of police officers, who used breath alcohol testing (BAT) devices and a van that houses a BAT device (Batmobile). A University of New Mexico study recorded quarterly data on two variables:

- ACC: injuries and fatalities from Wednesday to Saturday nighttime accidents.
- FUEL: fuel consumption (millions of gallons) in Albuquerque. This variable is a proxy for traffic volume.

These data were collected to evaluate the Batmobile program. The first 29 observations were for a control period, and the next 23 observations were for the experimental (Batmobile) period. The variables were measured starting from the first quarter of 1972 up to the last quarter of 1984, covering a span of 13 years.

Figure 6 Time Series Plots of ACC and FUEL

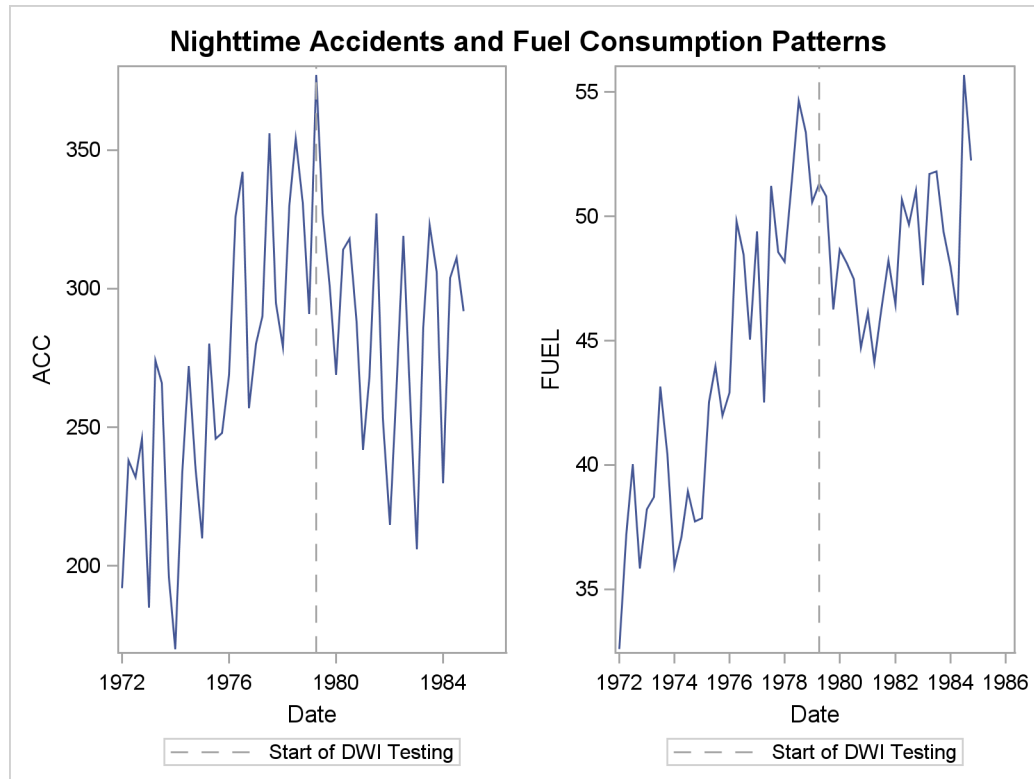


Figure 6 shows the time series plots of ACC and FUEL. The plots indicate that the number of accidents, ACC, are closely related to the traffic volume (measured implicitly using the fuel consumption, FUEL). Moreover, the accidents do seem to decline after the introduction of the BAT program. However, the fuel consumption also seems to decline, at least initially, making the program effectiveness question difficult to answer. In this example this issue is analyzed by fitting a suitable UCM to the data. The following UCM captures the main features of the data:

$$ACC_t = Intercept + \beta_1 FUEL_t + \beta_2 batProgram + \gamma_t + \epsilon_t$$

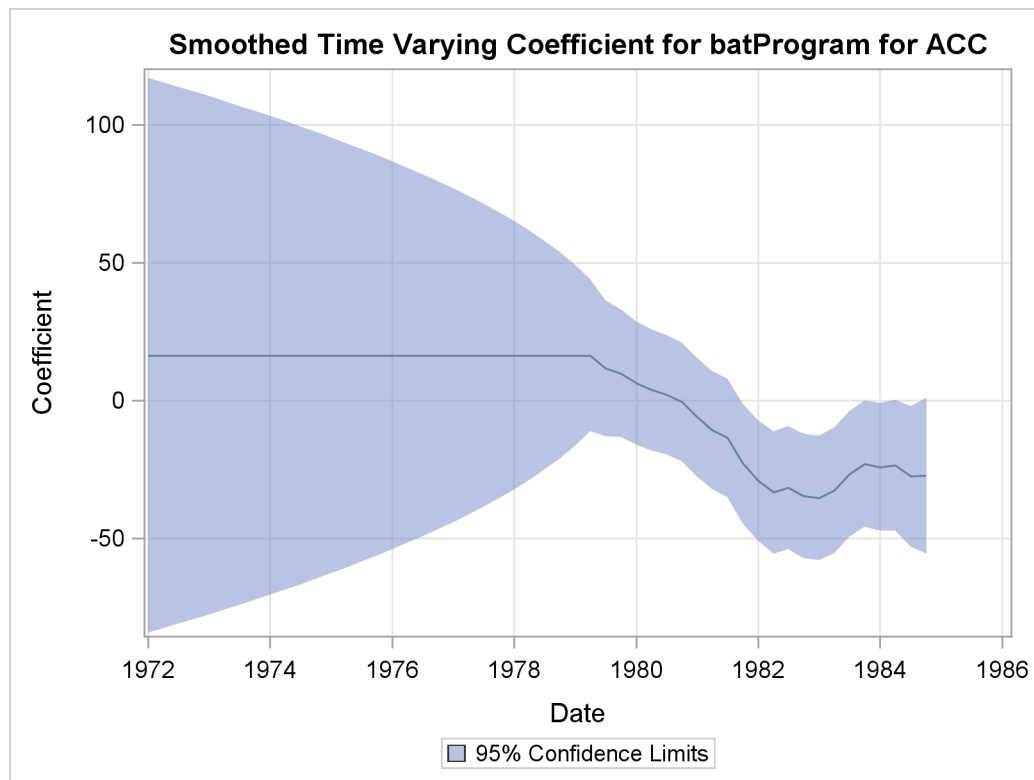
The model consists of a constant term, a fixed quarterly seasonal term (γ_t), a simple regression effect associated with FUEL, a time-varying regression effect associated with a dummy regressor (batProgram) which is zero before the program began and one while the program is in place, and an overall error term (ϵ_t). Note that letting the regression coefficient of batProgram change with time models the possibility that the effect of the Batmobiles might well be cumulative (as awareness of the program becomes dispersed, its effectiveness as a deterrent to driving while intoxicated increases).

The following program specifies such a model:

```
proc ucm data=bat;
  model ACC = FUEL;
  id Date interval=qtr;
  irregular;
  level var=0 noest;
  randomreg batProgram / plot=smooth;
  season length=4 var=0 noest;
run;
```

The LEVEL statement specifies the constant term in the model as a random walk trend with zero disturbance term. The SEASON statement specifies a fixed quarterly seasonality. The MODEL statement specifies the simple regressor, FUEL, and the RANDOMREG statement specifies the variable batProgram that is associated with the time-varying regression effect. This example is discussed more fully in “Example 29.4: Modeling Time-Varying Regression Effects Using the RANDOMREG Statement” (Chapter 29, *SAS/ETS User’s Guide*). There it is shown that the model fits the data adequately. Figure 7 shows the plot of the time-varying coefficient of batProgram. As expected, it shows that the effectiveness of the program increases as awareness of the program becomes dispersed. The effectiveness eventually seems to level off.

Figure 7 Time-Varying Regression Coefficient of batProgram



Electricity Usage Data

The data in this example are created to mirror the electricity demand and temperature data recorded at a utility company in the midwest region of the United States. The data set (not shown), utility, has three variables: Load, Temp, and Date. The Load column contains the daily electricity demand, the Temp column has the average daily temperature readings, and the Date column records the observation date. The data cover a span of four years (January 1, 1999, to December 31, 2002). The goal of the analysis is to develop a statistical model that is useful for predicting daily electricity demand. A preliminary exploration (not shown) of the data reveals the following features:

- The relationship between the electricity load and temperature is nonlinear; the load usually increases when the temperatures are away from the comfortable sixties.
- The electricity load varies by the day of the week, which results in a periodic pattern of length 7.
- The series has no apparent upward or downward trend.

The following two sets of statements fit two competing UCMs to the electricity load series that take into account these observations. In both models the temperature effect is modeled by a deterministic three-degree spline with knots at 30, 40, 50, 60, and 75. The knot locations and the degree were chosen by visual inspection of a scatter plot (not shown) of electricity load versus temperature. The last 60 days of data are withheld for out-of-sample forecast evaluation (note the BACK= option in both the ESTIMATE and FORECAST statements).

The following statements model the trend, season, and the irregular components according to the Airline model (ARIMA(0,1,1)(0,1,1)7), which is commonly used as a baseline model in the Box-Jenkins approach to time series analysis. The Q=, SQ= and S= options in the IRREGULAR statement specify nonseasonal and seasonal moving average orders, and the DEPLAG statement specifies the differencing orders.

```
proc ucm data=utility;
  id date interval=day;
  model load;
  irregular q=1 sq=1 s=7;
  deplag lags=(1) (7) phi=1 1 noest;
```

```

splinereg temp knots=30 40 50 65 75 degree=3
  variance=0 noest;
estimate back=60;
forecast back=60 lead=60;
run;

```

The following statements model the trend, season, and the irregular components by using a minor variation of the basic structural model (BSM), which is commonly used as a baseline model in the structural approach to time series analysis. In this case the LEVEL statement specifies random walk trend, the SEASON statement specifies a deterministic seasonal component, and the IRREGULAR statement specifies an ARMA(1,1) irregular component.

```

proc ucm data=utility;
  id date interval=day;
  model load;
  irregular p=1 q=1;
  level;
  season length=7 variance=0 noest;
  splinereg temp knots=30 40 50 65 75 degree=3
    variance=0 noest;
  estimate back=60;
  forecast back=60 lead=60;
run;

```

In practice you first check any candidate model for adequacy by using various diagnostic procedures. Such diagnostic analysis for the second model (BSM) is done in “Example 29.6: Using Splines to Incorporate Nonlinear Effects” (Chapter 29, *SAS/ETS User’s Guide*), which shows its adequacy. Similar diagnostic analysis (not shown) of the first model (Airline model) reveals some model assumption violations, indicating that the model should be modified. In this illustration these two models are further compared on the basis of two criteria: the information criteria (shown in Figure 8) and their post sample forecast performance (shown in Figure 9). (These tables are produced by combining and summarizing the default output produced by the preceding programs.) In both cases the BSM is preferred over the Airline model for this data set.

Figure 8 Model Comparison Using the Information Criteria

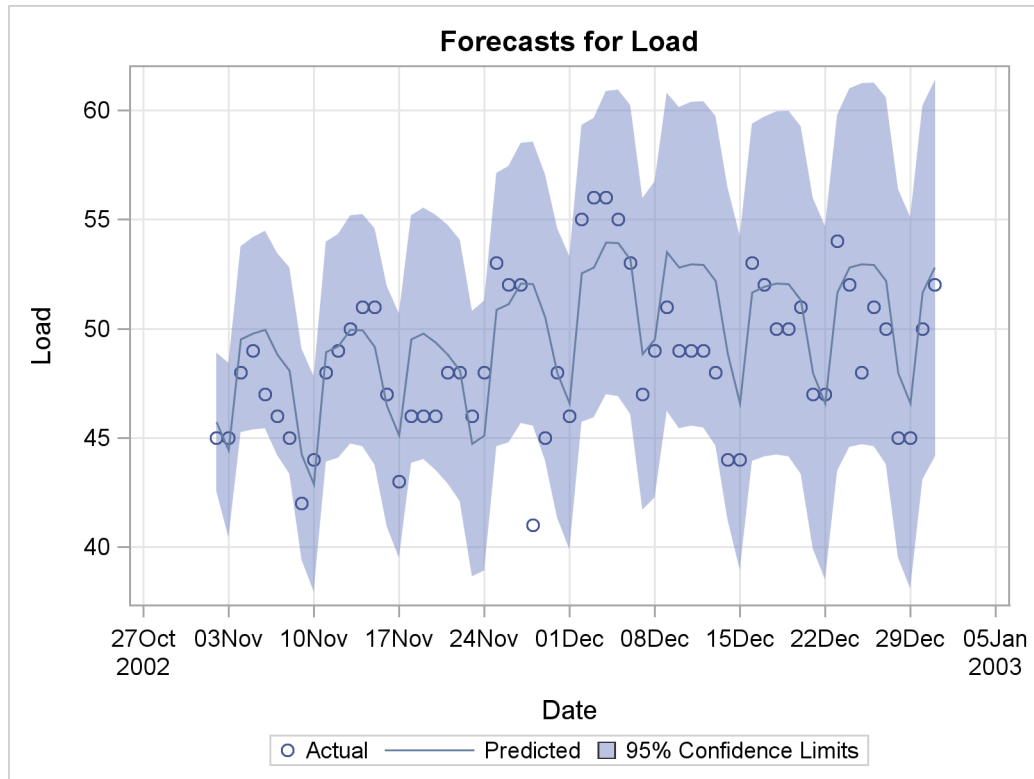
FitStatistic	Airline Model	Basic Structural Model
AIC (smaller is better)	6028.526	5314.223
AICC (smaller is better)	6028.544	5314.252
BIC (smaller is better)	6044.226	5335.160
CAIC (smaller is better)	6047.226	5339.160
HQIC (smaller is better)	6034.398	5322.054

Figure 9 Comparison Based on Sum of Absolute Errors (SAE) for the Post Sample Period

Model	SAE
Basic Structural Model	122.215
Airline Model	320.700

Figure 10 shows the plot of the load forecasts for the withheld data based on the second model (BSM): clearly the forecasts track the demand reasonably well. Note that the forecast validation period is challenging from the prediction point of view—it contains both the Thanksgiving and Christmas periods which are not explicitly modeled.

Figure 10 Electricity Load: Forecast Evaluation of the Withheld Data



REFERENCES

- Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.
- Houghton, A. N., Flannery, J., and Viola, M. V. (1980), "Malignant Melanoma in Connecticut and Denmark," *International Journal of Cancer*, 25, 95–104.
- West, M. and Harrison, J. (1999) *Bayesian Forecasting and Dynamic Models*, Second Edition, New York: Springer-Verlag.

CONTACT INFORMATION

Rajesh Selukar

SAS Institute Inc.

SAS Campus Drive

Cary, NC 27513

Work Phone: 919-531-5183

Fax: 919-677-4444

E-mail: Rajesh.Selukar@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.