The Unobservable Components Model[1]

Prof. Tom Fomby
Department of Economics
Southern Methodist University
Dallas, TX 75275
May 2008

I. Introduction and Motivation of UCM

In this section we are going to be presenting the Unobserved Components time series model. This model was first introduced to the econometrics and statistics fields by A.C. Harvey (1989) in his book Forecasting, Structural Time Series Models and the Kalman Filter (Cambridge University Press). Here we will introduce the model as an "organizing model" for the discussion of time series modeling in general but also to compare and contrast the UCM model with other time series models to come, namely, the very basic Deterministic trend / Deterministic seasonal model to be discussed next, and the exponential smoothing and Box-Jenkins models that follow.

The UCM can be considered to be a multiple regression model with time-varying coefficients. It is based on the principles that (i) it is useful to view time series as being decomposable into trend, seasonal, and cycle components and (ii) time series models that give equal weight to both near and far distant observations (as in the deterministic trend model to be discussed later) are often not very useful. With respect to point (i) inefficient and inaccurate forecasting is likely to arise for anyone who ignores the salient characteristics of the time series to be forecast. For example, if one builds a time series model that has no allowance for seasonal variation yet the time series has significant seasonal variation in it, then the forecasting accuracy of such a naïve model is likely to be poor. With respect to point (ii), in many time series the adjacent observations are more closely correlated with each other than observations that are far apart. As a result time series models that are "local" in nature and weight recent observations more than observations in the far past, tend to predict better when applied to economic and business time series than models that treat time series data "globally" as in the deterministic time trend model. Apart from the deterministic time trend model that treats all observations as equally important when constructing forecasts, the other models that we study (UCM, exponential smoothing, and Box-Jenkins models) are local in nature in that the more distant an observation is from the point of forecast, the less weight the distant observation carries in determining forecasts of the time series in question.

Notation of UCM

The fully specified Unobserved Components Model is written as

[1] This presentation relies heavily on the material contained in the SAS HELP file under the keyword "Proc UCM".

$$y_t = \mu_t + \gamma_t + \psi_t + r_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{m} \beta_j x_{jt} + \varepsilon_t \, . \tag{1}$$

In equation (1) y_t represents the time series to be modeled and forecast, μ_t the **trend component**, γ_t the **seasonal component**, ψ_t the **cyclical component**, r_t the **autoregressive component**, and ε_t the **irregular component**. All of these components are assumed to be unobserved and must be estimated given the time series data on y_t and x_{jt}, hence the title **unobserved components model**. In addition, (1) allows the inclusion of the autoregressive regression terms $\sum_{i=1}^{p} \phi_i y_{t-i}$ and the explanatory regression terms $\sum_{j=1}^{m} \beta_j x_{jt}$, the former representing the "momentum" of the time series as it relates to its past observations and the latter representing the causal factors that one is willing to suppose affects the time series in question.

In traditional time series analysis it is often assumed that a time series y_t can be **additively decomposed** into four components, namely, **trend**, **season**, **cycle**, and **irregular** components as in

$$y_t = T_t + S_t + C_t + I_t \tag{2}$$

where T_t represents the trend in y_t at time t, S_t the seasonal effect at time t, C_t the cyclical effect at time t and I_t the irregular effect at time t. Obviously, the UCM model (1) does employ this decomposition but, in addition, allows unobserved autoregressive effects and explanatory regression effects making it a very powerful model indeed. One of the major advantages of the UCM (1) is its **interpretability** and, as we will see, PROC UCM in SAS provides some very nice graphical representations of this decomposition.

To demonstrate this decomposition, consider the following characterizations of trend, cycle, seasonal, and irregular components that have been encoded in a SAS program entitled Decomposition.sas.

$T_t = 100 + 4.0 * t$ (deterministic trend: intercept = 100, slope = 4)
$C_t = 50 * \cos(3.1416 * t / 10)$ (deterministic cycle: amplitude = 50, Period = 20 months, phase = 0)
S_t = {fixed seasonal effects: -50, -25, 25, -25, -50, 50, 75, 50, 5, -25, -50, 20}
(i.e. Jan. effect = -50, Feb. effect = -25, … , Dec. effect = 20)
$I_t \rightarrow NIID(0,100)$

Therefore, the actual population model can be written as

$$
\begin{aligned}
y_t &= T_t + C_t + S_t + I_t \\
&= 100 + 4.0t + 50\cos(0.31416t) - 50*dum1 - 25*dum2 \\
&\quad + 25*dum3 - 25*dum4 - 50*dum5 + 50*dum6 \\
&\quad + 75*dum7 + 50*dum8 + 5*dum9 - 25*dum10 \\
&\quad - 50*dum11 + 20*dum12 + \varepsilon_t \,.
\end{aligned}
$$

The above seasonal dummies are defined $dumi$ = 1 if the observation is in the i-th month and zero otherwise. Obviously the intercept of the month varies by month. For example, the intercept for all of the January months is (100 – 50 = 50), the intercept for the February months is (100 – 25 = 75), etc. The irregular component is represented by the unobserved error ε_t which is normally and independently distributed with mean zero and variance of 100.

Each of these components is plotted in order in the following figures below:

Figure 1

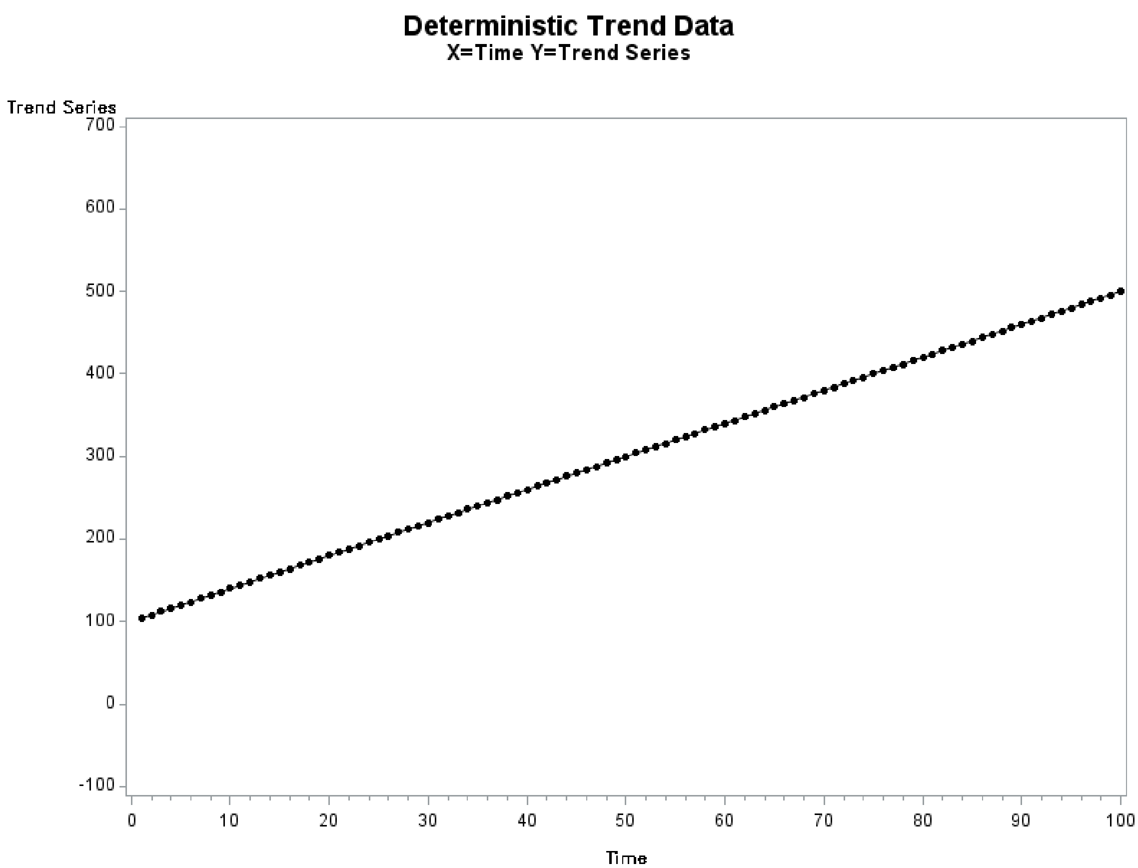


Figure 2

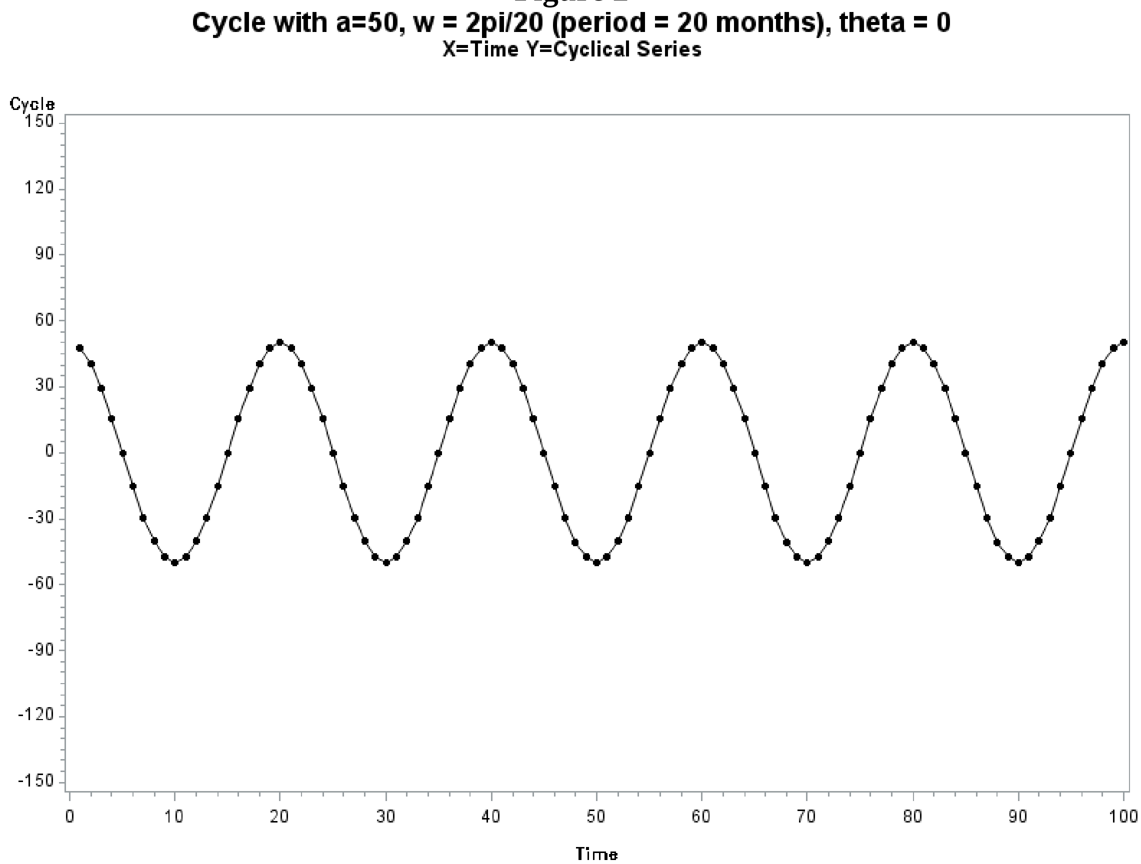


Figure 3

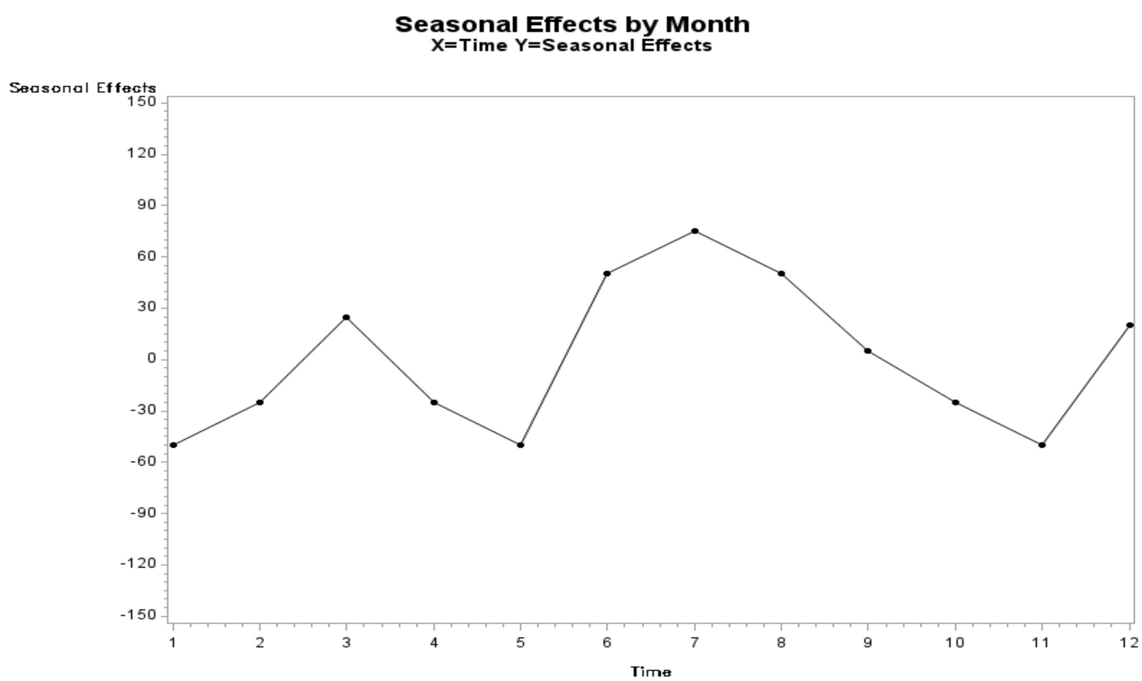
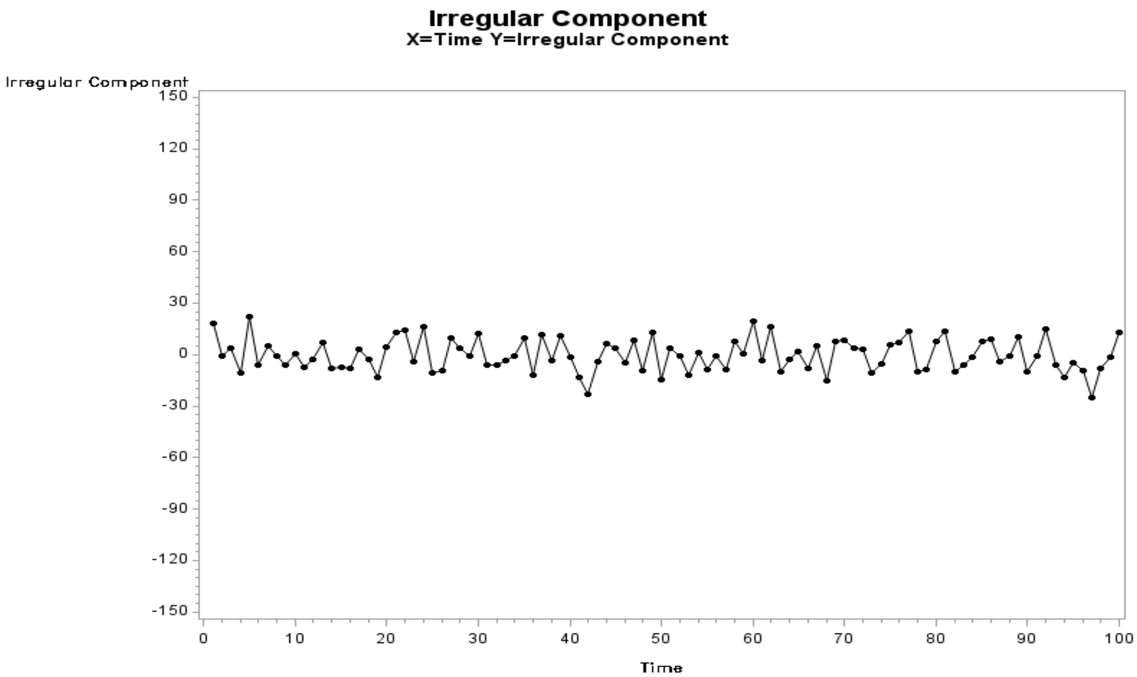


Figure 4



In the following graphs we sum these components up as is intended in the additive decomposition (2)

.**Figure 5**

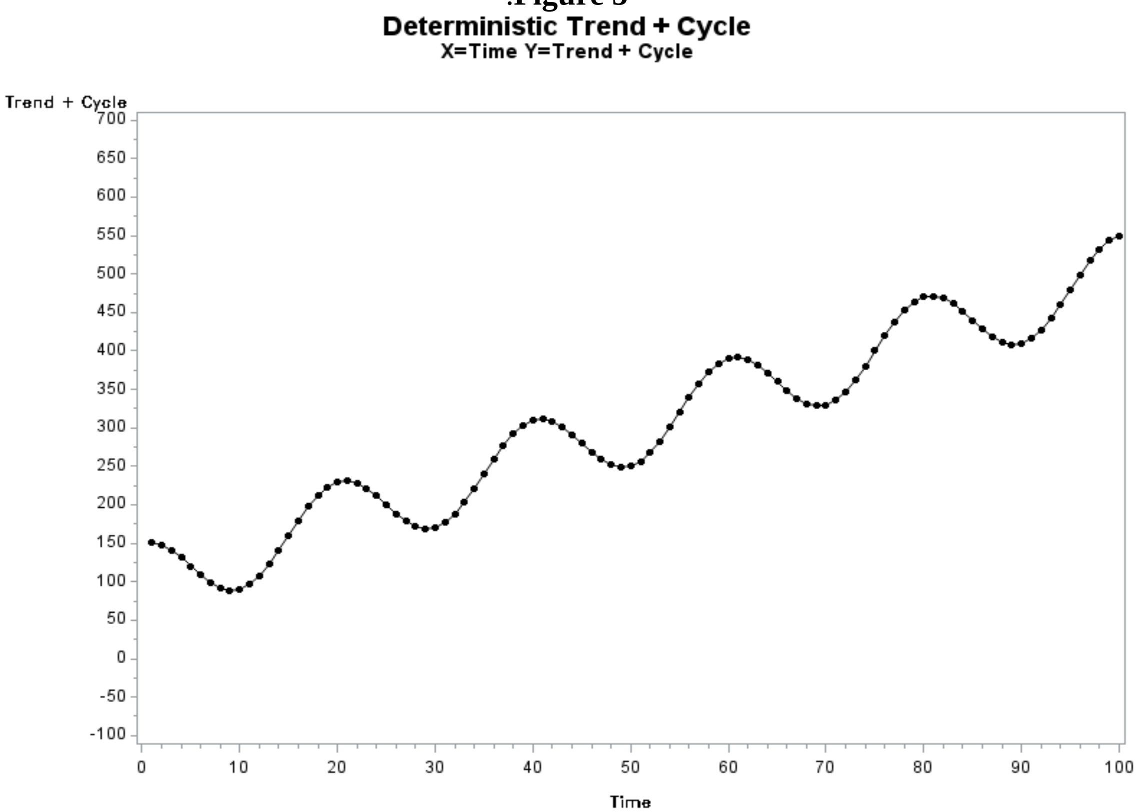


Figure 6

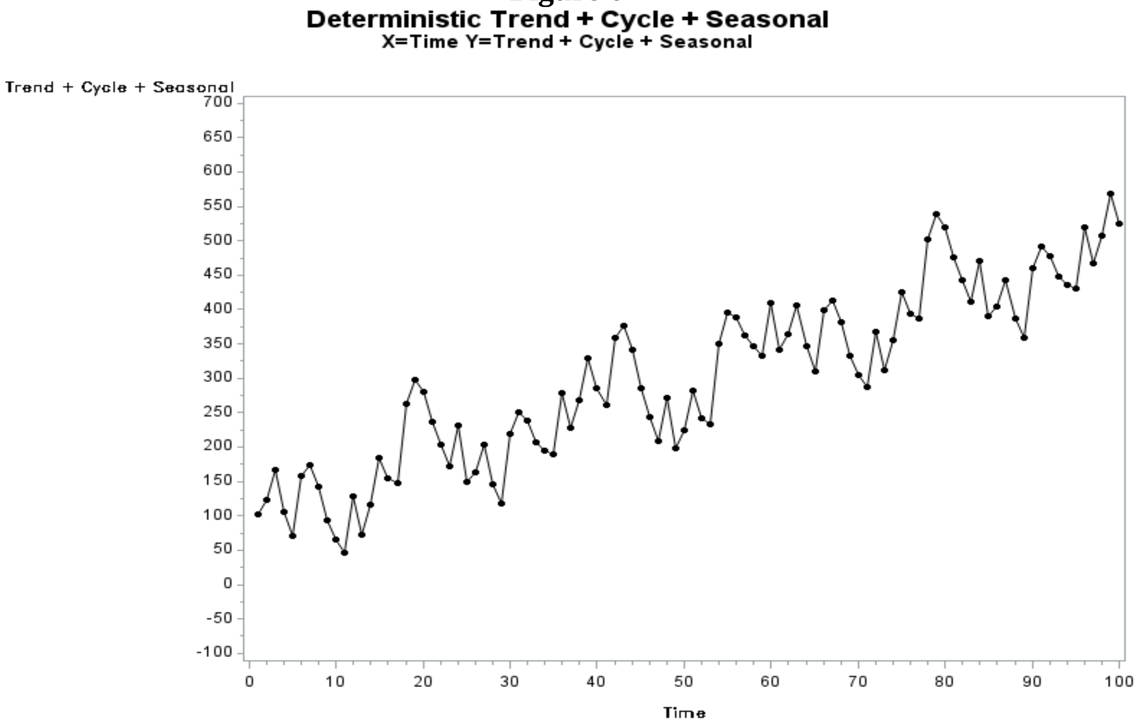
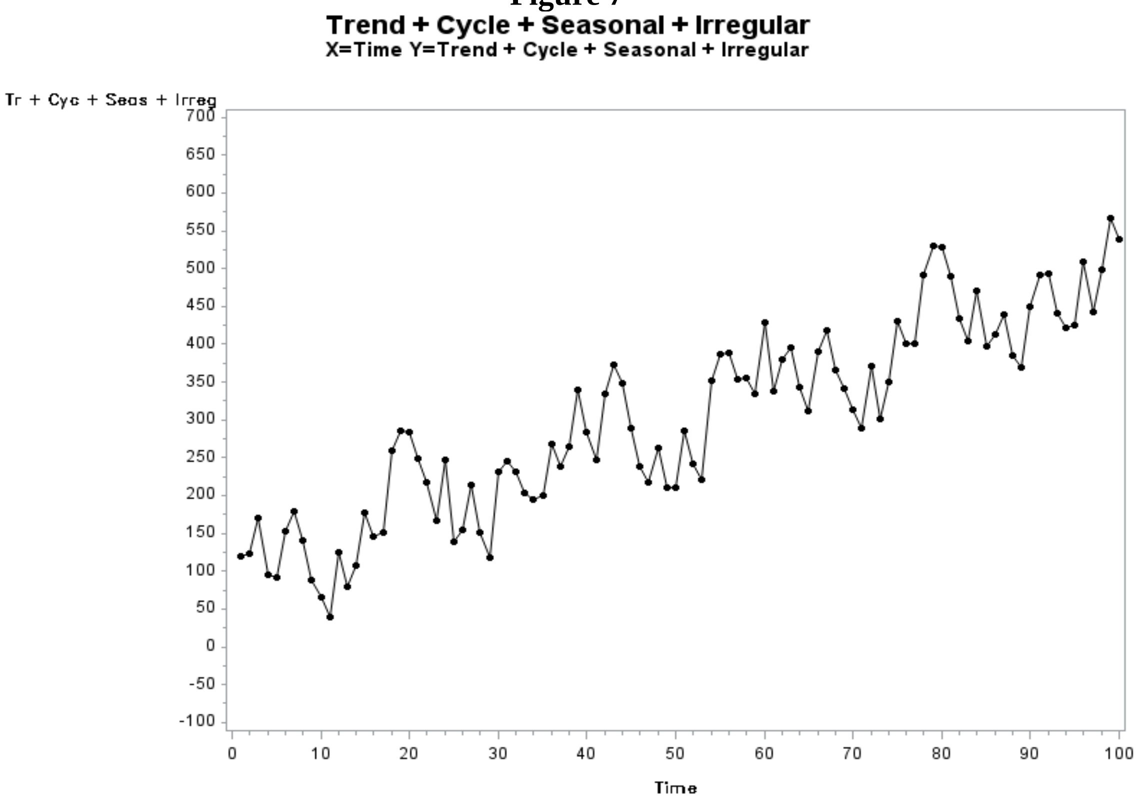
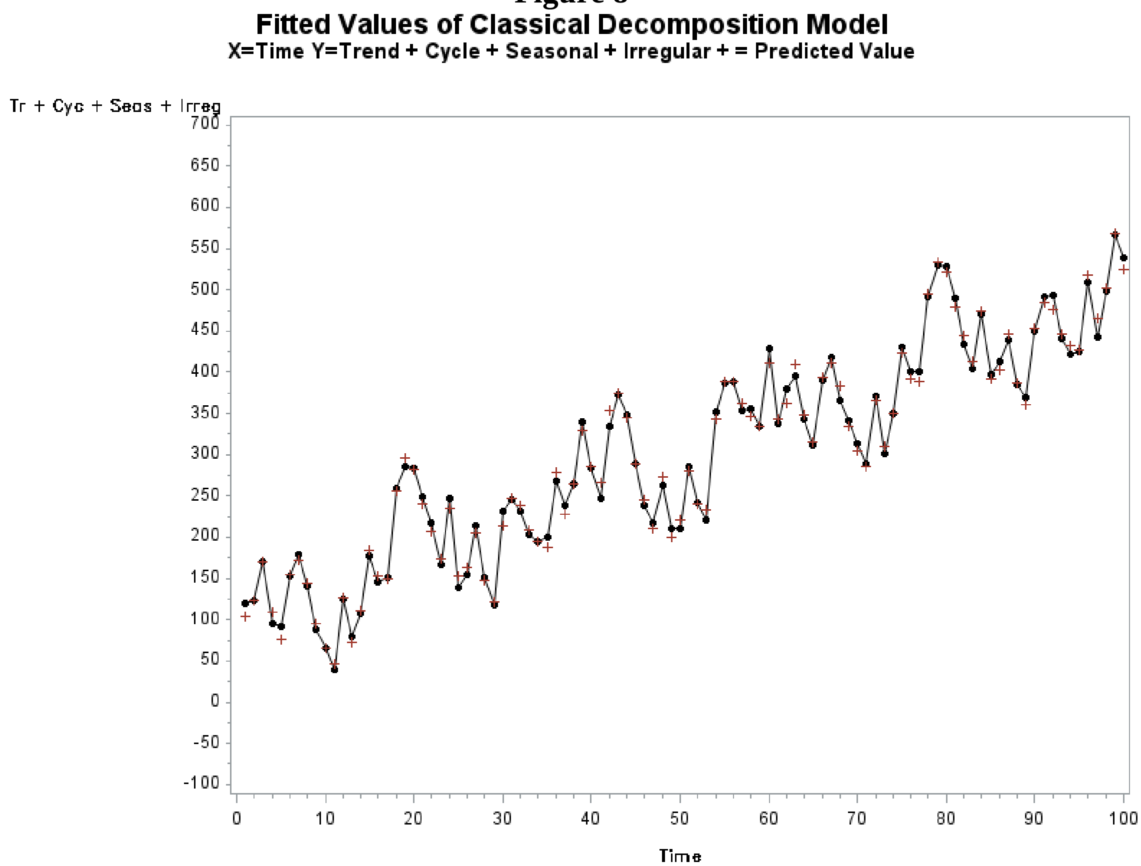


Figure 7



And in Figure 8 below we have the representation of the fitted model obtained by Proc Nlin (a Nonlinear Least Squares procedure).

Figure 8



The fitted model is

$$
\begin{aligned}
y_t &= 51.74 + 3.98t + 50.76\cos(0.3140t - 0.0314) + 21.36 * dum2 + 75.21 * dum3 \\
&+ 23.99 * dum4 + 2.21 * dum5 + 93.26 * dum6 + 120.3 * dum7 \\
&+ 100.6 * dum8 + 56.16 * dum9 + 24.69 * dum10 - 0.58 * dum11 \\
&+ 69.97 * dum12 + \hat{\varepsilon}_t
\end{aligned}
$$

which closely matches the population parameters. Obviously the January intercept is estimated as 51.74, the February intercept is 51.74 + 21.36 = 73.1, etc. The slope of the trend is 3.98 while the amplitude of the cycle is 50.76, the phase is -0.0314, and the period is $p = 2\pi / w = 2 * 3.1416 / 0.3140 = 20.01$ months.

In the spirit of the above empirical model, we now move to cover the parts of the UCM (1) in more detail.

II. Modeling the Trend

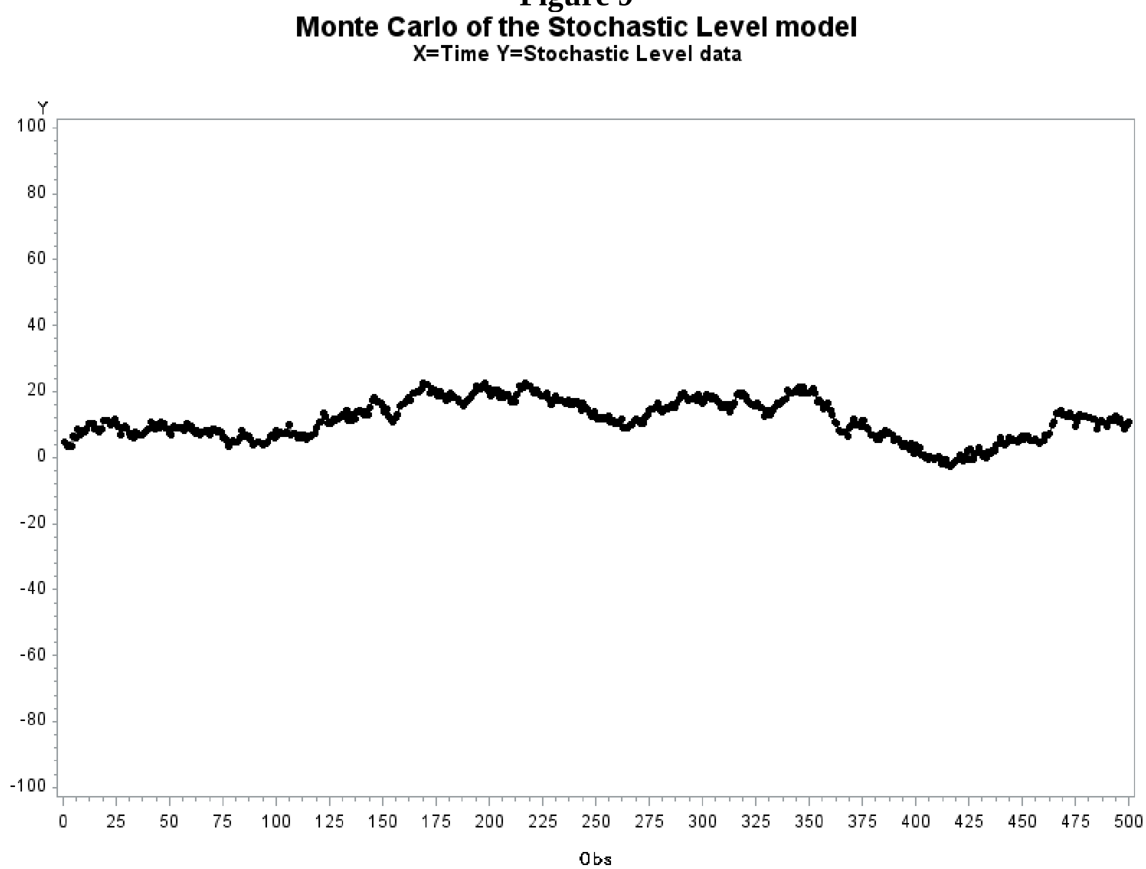
(a) Random Walk (without drift) Model for Trend

When the trend component is modeled as

$$\mu_t = \mu_{t-1} + \eta_t \quad , \quad \eta_t \mapsto NIID(0, \sigma_\eta^2) \tag{3}$$

where "$\mapsto NIID(0,\sigma_\eta^2)$" reads "distributed as a normal, independently and identically distributed random variable, it referred to as the **Random Walk** (RW) **trend model** in the UCM. This model is especially appropriate for time series data that are **flat and slow-turning**. Notice if $\sigma_\eta^2 = 0$ then $\mu_t = \mu_{t-1}$ for all t and therefore $\mu_t = \mu_0 =$ constant. In this special case, the data are expected to **revert back to the mean** μ_0 in fairly short order. A simulation of the Random Walk trend model is presented in the following graph as produced by the SAS program Stochastic Level Model.sas with $\mu_0 = 0$ and $\sigma_\eta^2 = 1.0$.

Figure 9

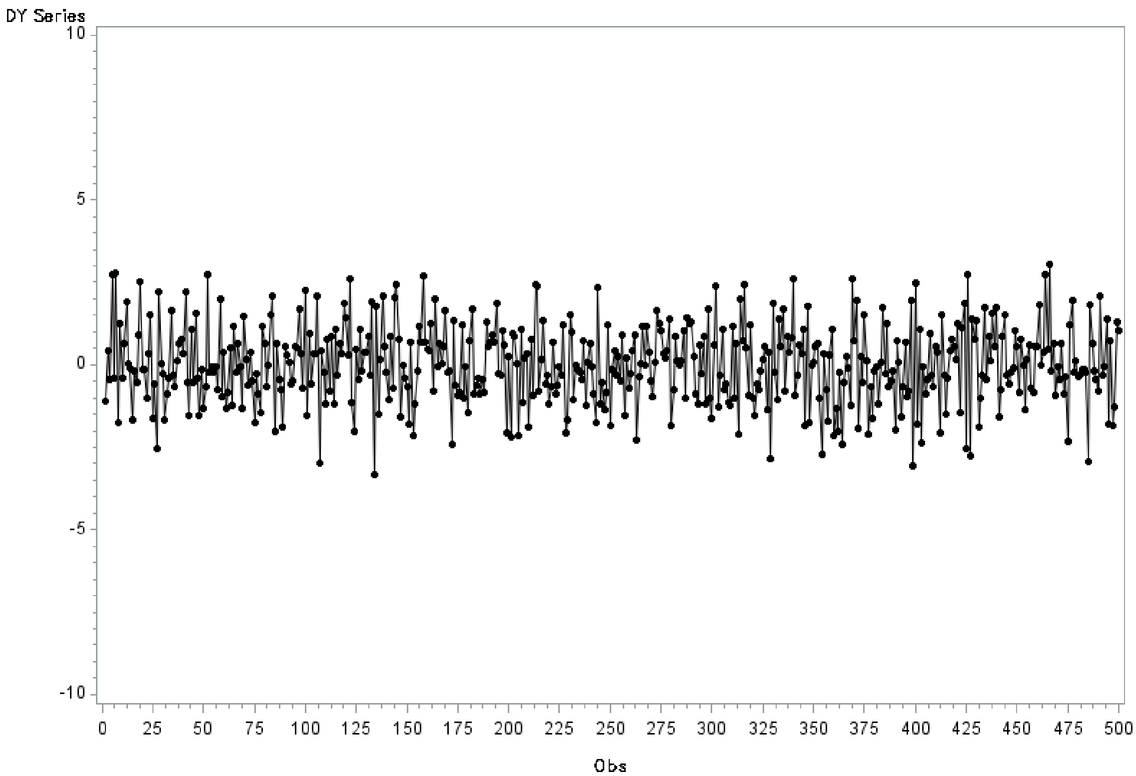


Notice that the data, as expected, is **flat and slow-turning**. Just as a point of note, we will see later that **when a time series has a random walk level (3), it can be thought of**

as following an ARIMA(0,1,1) Box Jenkins model hence the "loose" relationship between Box-Jenkins models and this simple UCM model we previously alluded to.

In contrast, when we set $\sigma_{\eta}^{2} = 0$ we have a constant mean. If the time series is then of the form of $y_t = \mu_0 + \varepsilon_t$ where $\mu_0 = 0$ and $\varepsilon_t \rightarrow NIID(0,1)$ we have the following graph that has strong mean reversion in it.

Figure 10
Differenced Stochastic Level Data
X=Time Y=DY Series



(b) Locally Linear Trend (LLT) Model

In this model the trend is characterized by the following **level** and **slope** equations

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \rightarrow NIID(0, \sigma_{\eta}^{2}) \qquad \textbf{(level)} \qquad (4)$$

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \rightarrow NIID(0, \sigma_{\xi}^{2}). \qquad \textbf{(slope)} \qquad (5)$$

Here μ_t represents the **stochastic level** of the trend while β_t represents the **stochastic slope** of the trend. Furthermore, it is assumed that η_t and ξ_t are independent of each other. In the presence of both stochastic level and stochastic slope the data need not have a linear trend but can have a trend with the curvature (slope) of the data slowly evolving as well. See Figure 11 below that has been generated by the SAS program Stochastic Level _ Stochastic Slope.sas with $\mu_0 = 0, \sigma_\eta^2 = 1, \sigma_\xi^2$, and $\sigma_\varepsilon^2 = 4$. Again, as a point of note, we will see later that **when a time series has a stochastic level and stochastic slope making up the trend as in (4) and (5), it can be thought of as following an ARIMA(0,2,2) Box Jenkins model** and hence, again, the “loose” relationship between Box-Jenkins models and this UCM model we previously alluded to.

If $\sigma_\xi^2 = 0$ then essentially you have a random walk with fixed drift, β_0, which would be appropriate for data that has a linear appearing trend. See Figure 12 below that has been generated by the SAS program Stochastic Level _ Fixed Slope.sas with $\sigma_\eta^2 = \sigma_\varepsilon^2 = 16$ and by setting the slope error variance to zero ($\sigma_\xi^2 = 0$) and setting $\beta_0 = 2$, we get the fixed slope of 2.0. Notice the data is now slowly turning around a fixed trend (drift) of 2.0. If, in addition, $\beta_0 = 0$ then we are back to the Random Walk without drift model of trend in (3) above and depicted in Figure 9 above. If both error variances are equal to zero ($\sigma_\xi^2 = \sigma_\eta^2 = 0$), the resulting model becomes the deterministic linear time trend model: $y_t = \mu_0 + \beta_0 t + \varepsilon_t$. See Figure 13 below that depicts a deterministic trend time series generated by the SAS program Deterministic Trend.sas where the trend equation is given by $y_t = 100 + 4 * t + \varepsilon_t$ with $\varepsilon_t \to NIID(0,400)$. In this latter case the data is expected to revert fairly quickly back to the deterministic trend. In the example to be examined later Proc UCM in SAS provides some very useful diagnostics for determining whether components are stochastic or non-stochastic, the significance of the components, and the goodness-of-fit of proposed UCM models.

Figure 11

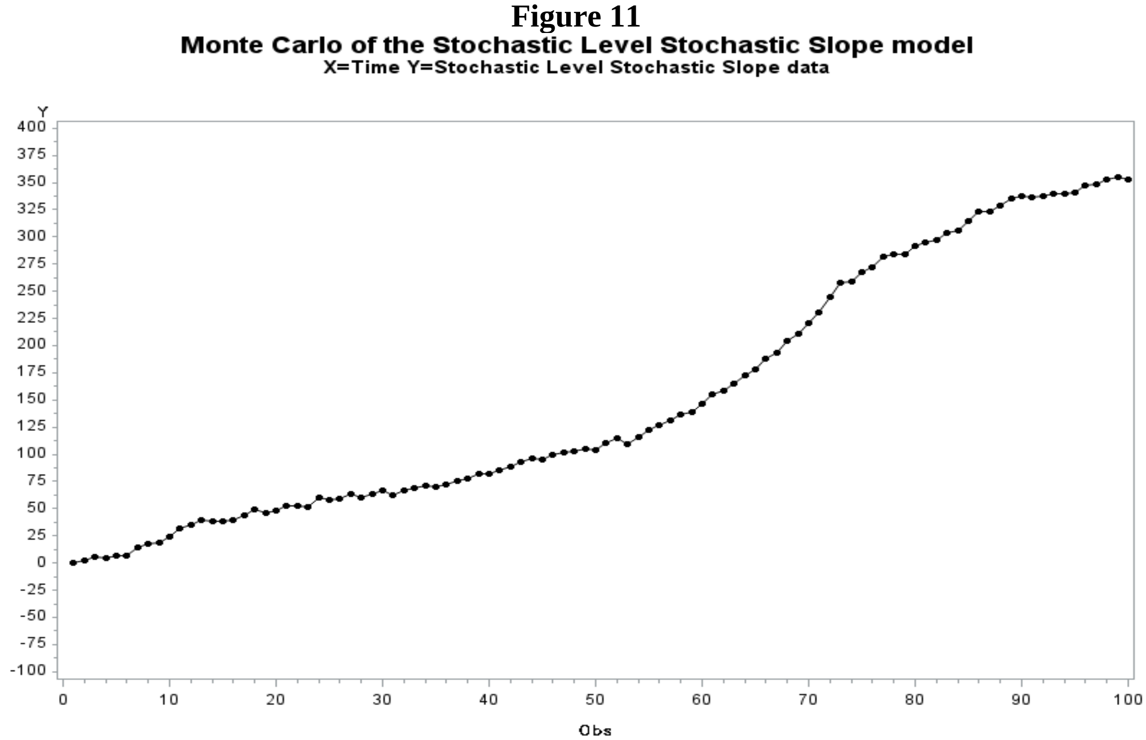


Figure 12

Monte Carlo of the Stochastic Level Fixed Slope model

X=Time Y=Stochastic Level Fixed Slope data

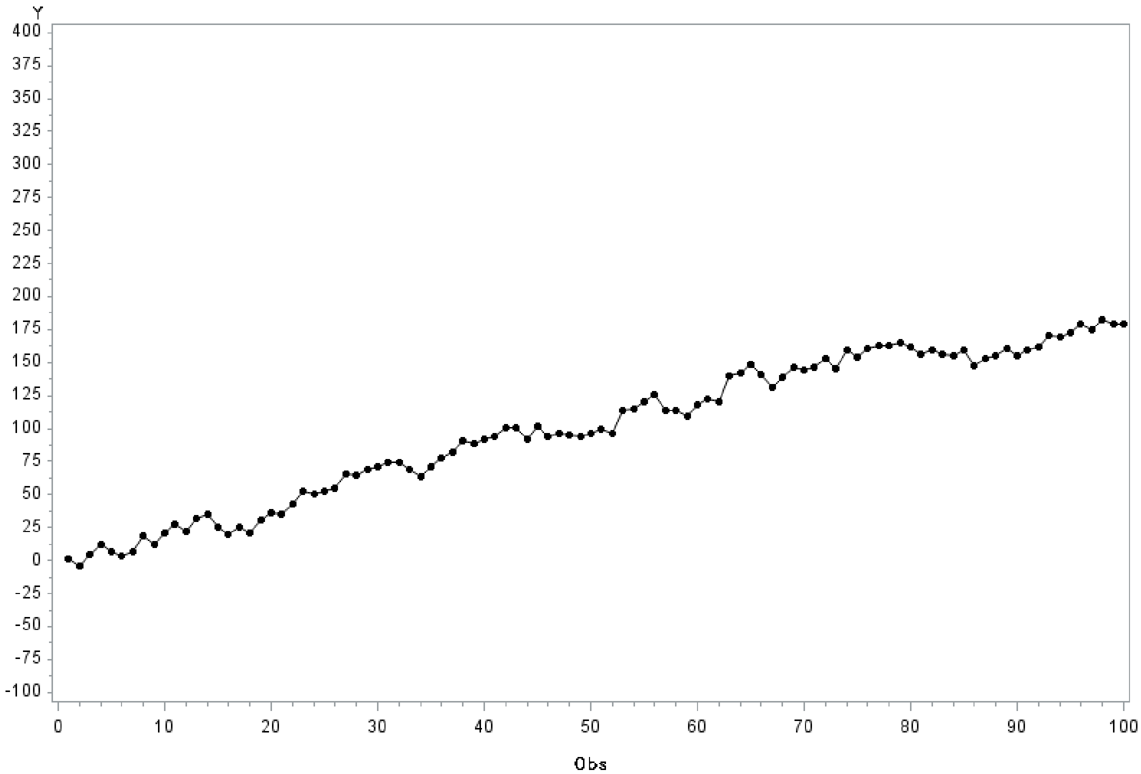
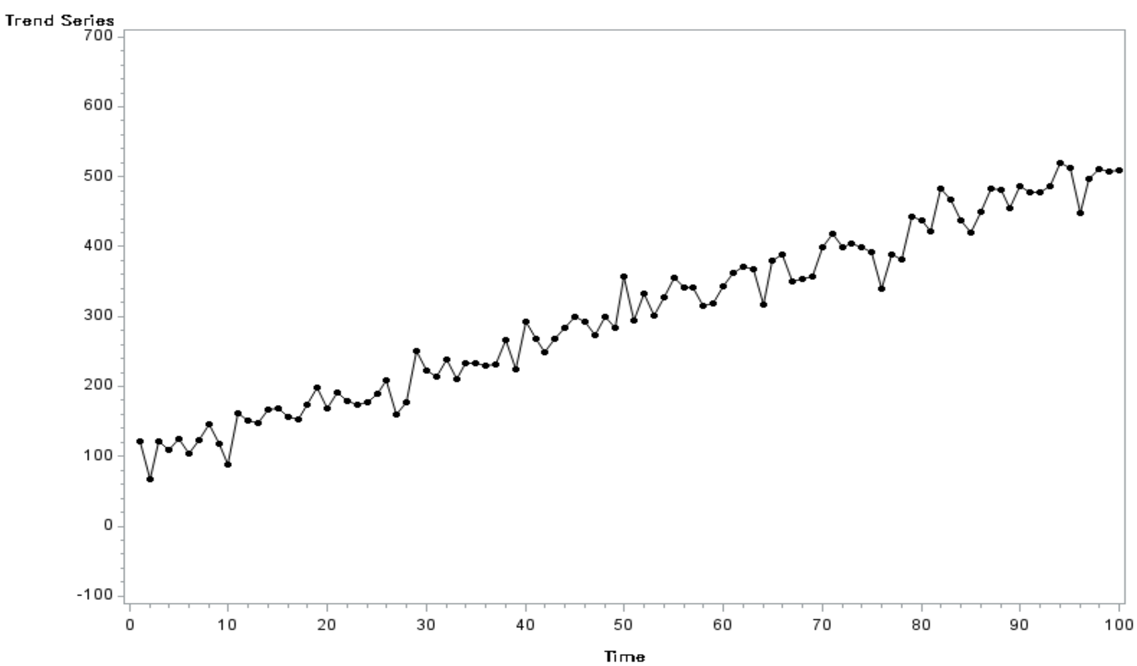


Figure 13

Deterministic Trend Data

X=Time Y=Trend Series



Thus we can see that the Stochastic Level _ Stochastic Slope model can be specialized in several different ways as represented by Figures 9 – 13 simply by setting some of the variances of the level and slope components (σ_η^2 and σ_ξ^2) equal to zero. In fact, Proc UCM in SAS provides t-tests of the significance of these error variances and based upon them the level and slope components can be modeled as being fixed as compared to being stochastic. See the below example for an illustration of these tests.

III. Modeling the Cycle

Proc UCM provides two basic ways of modeling the unobserved cyclical component, ψ_t : a **deterministic** (non-stochastic) trigonometric cycle (or cycles) and a **stochastic** trigonometric cycle. We will discuss these two representations in turn.

(a) A Deterministic Cyclical Model

Let the deterministic cycle with frequency λ, $0 < \lambda < \pi$, be written as

$$\psi_t = \alpha \cos(\lambda t) + \beta \sin(\lambda t) \quad . \tag{6}$$

If t is observed continuously, ψ_t is a periodic function with period $2\pi/\lambda$, amplitude $(\alpha^2 + \beta^2)^{1/2}$, and phase $\tan^{-1}(\beta/\alpha)$. If ψ_t is measured only at integer values of t, then ψ_t is not exactly periodic unless $\lambda = (2\pi j/k)$ for some integers j and k. Unfortunately, the cycles in economic and business time series data are scarcely ever as systematic as would be depicted in any one deterministic periodic function (6). However, from Fourier analysis we know that fairly complex cyclical data can be written as a **sum** of a finite number of sinusoidals like (6). In Proc UCM in SAS one can specify more than one such sinusoidal for the purpose of capturing some of the more complex cyclical patterns in economic and business time series data.

(b) A Stochastic Cyclical Model

As an alternative to specifying one or more of the deterministic cycles (6) and introducing a multitude of α, β, and λ parameters, one can specify a stochastic cyclical model as in

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \nu_t \\ \nu_t^* \end{bmatrix} \tag{7}$$

where $0 \le \rho \le 1$ is a damping factor and the disturbances ν_t and ν_t^* are independently distributed as $N(0, \sigma_\nu^2)$ random variables. This model can capture quite complex cyclical patterns in economic and business time series without introducing an abundance of parameters. If $\rho < 1$, ψ_t has a stationary distribution with mean zero and variance $\sigma_\nu^2/(1-\rho^2)$. If $\rho = 1$, ψ_t is non-stationary. Of course if $\sigma_\nu^2 = 0$ we revert to the deterministic cyclical model (6).

IV. Modeling a Seasonal Effect

Most economic and business time series exhibit seasonality. Assume that we are analyzing monthly time series data. **A rough definition of seasonality** can be expressed as follows: For any given month, deviations from trend tend to be of the same sign from one year to the next. For example, December toy sales tend to be above the secular trend of toy sales because of the holiday buying habits of consumers. In contrast, January toy sales tend to be below the secular trend because of the lack of a child-oriented holiday in January. The same can be said for seasonal quarterly data and the similar year-over-year variation in the data compared to the secular trend in the data. One model for such seasonal variation is called the **Stochastic Dummy Variable Seasonal model** discussed below.

(a) Stochastic Dummy Variable Seasonal Model

Let there be s seasons during the year, s = 12 for monthly data, s = 4 for quarterly data, and s = 2 for bi-annual data. Consider the following model for the seasonal effect γ_t at time t:

$$\sum_{i=0}^{s-1} \gamma_{t-i} = \omega_t \quad , \quad \omega_t \to NIID(0, \sigma_\omega^2). \tag{8}$$

In this model the sum of the seasonal effects has a zero mean although their stochastic nature allows them to evolve either slowly over time (when σ_ω^2 is small) or quickly over time (when σ_ω^2 is large).

(b) Deterministic Dummy Variable Seasonal Model

In the special case where $\sigma_\omega^2 = 0$ in (8), we have the following the so-called Deterministic Dummy Variable Seasonal model. In this model the seasonal effects $\gamma_1, \gamma_2, \cdots, \gamma_s$ are fixed and do not vary over time in contrast to the stochastic specification in (8). In this case a test of the **absence of seasonality** in the time series data being analyzed amounts to testing the null hypothesis $H_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_{s-1} = 0$, where the sum constraint $\sum_{i=1}^{s} \gamma_i = 0$ implies that $\gamma_s = 0$ as well.

Proc UCM in SAS also accommodates a trigonometric form of seasonality much like the trigonometric cyclical form previously specified in (7). We do not pursue this model further. For more information on this specification, one can consult Harvey (1989) or the Procedure documentation of Proc UCM in SAS HELP.

V. Modeling Additional Effects

(a) Modeling an Unobserved Autoregressive Component

Rather than modeling the cyclical nature of a time series via either the deterministic cyclical model (6) or the stochastic cyclical model (7), one can use the rather straight-forward specification

$$r_t = \rho r_{t-1} + \upsilon_t \ , \quad \upsilon_t \to NIID(0, \sigma_\upsilon^2) \tag{9}$$

where the unobserved autoregressive component r_t follows a first-order autoregression with $-1 < \rho < 1$. This autoregression, despite its simplicity, can capture many of the movements in time series data that represent business cycle inertia and that are present in many business and economic time series.

(b) Including Regression Terms

Moving beyond univariate time series modeling, one can specify regression terms for adding additional explanatory power: $\sum_{j=1}^{m} \beta_j x_{jt}$ and $\sum_{i=1}^{p} \phi_i y_{t-i}$. The inputs x_{jt} are intended to provide economic and business "causes and effects" that might help one in deriving more accurate forecasts of y_t. We will study the possibility of such causal variables when we turn to modeling multivariate time series later in this course.

VI. Obtaining the Estimates of the Unobservable Components and Other Statistics – the Kalman Filter

It is well known in the advanced time series literature that Unobservable Components models can be thought of as being special cases of more general models called **Gaussian State Space Models** (GSSM). Once the specific Unobservable Components model has been cast in State Space form the various unobserved components can be estimated using something called the **Kalman Filter**. We will leave the detailed explanation of these matters to the references Harvey (1989), Durbin and Koopman (2001), and the SAS HELP for Proc UCM. Needless to say, Proc UCM in SAS produces, via the Kalman filter, graphs of the various unobserved components as a function of time. These graphs contain either the "filtered" estimates of the unobserved components over time or the "smoothed" estimates of the unobserved components. In the filtered case the estimates are obtained recursively as one-step-ahead forecasts of the components whereas the smoothed estimates are derived using the entire time series and a smoothing of the filtered estimates. These graphs are very informative. As in the additive decomposition of (2), a time series observation at time t can then be additively decomposed into the estimated trend component, $\hat{\mu}_t$, estimated seasonal component, $\hat{\gamma}_t$, estimated cyclical component, $\hat{\psi}_t$, and estimated irregular component, $\hat{\varepsilon}_t$, represented in the graphs produced by Proc UCM. Therefore, at time t an observation on y can be additively decomposed as in $y_t = \hat{\mu}_t + \hat{\gamma}_t + \hat{\psi}_t + \hat{\varepsilon}_t$ assuming the presence of unobserved components for trend, season, cycle, and irregular effects. These estimated unobserved component graphs will be presented in our example below. Other statistics such as estimated error variances of the unobserved components, t-statistics of the significance of the error variances, Chi-square statistics for gauging the significance of the various components, and goodness-of-fit statistics are also produced by Proc UCM using the Kalman filter and are useful for building a UCM model that fits the data well.

VII. Getting Started in Building a UCM Model – the BSM model

One approach for building a good UCM for a given time series is to build a "Basic" structural model of the time series and then add to the Basic model as necessary. In this spirit, Harvey (1989) has defined the following UCM as the **Basic Structural model** (BSM):

$$y_t = \mu_t + \gamma_t + \varepsilon_t \tag{10}$$

where

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$

$$\beta_t = \beta_{t-1} + \xi_t .$$

Thus, the BSM consists of the locally linear trend (LLT) model for trend and a seasonal component of either the stochastic dummy variable form or the trigonometric form (not discussed here).

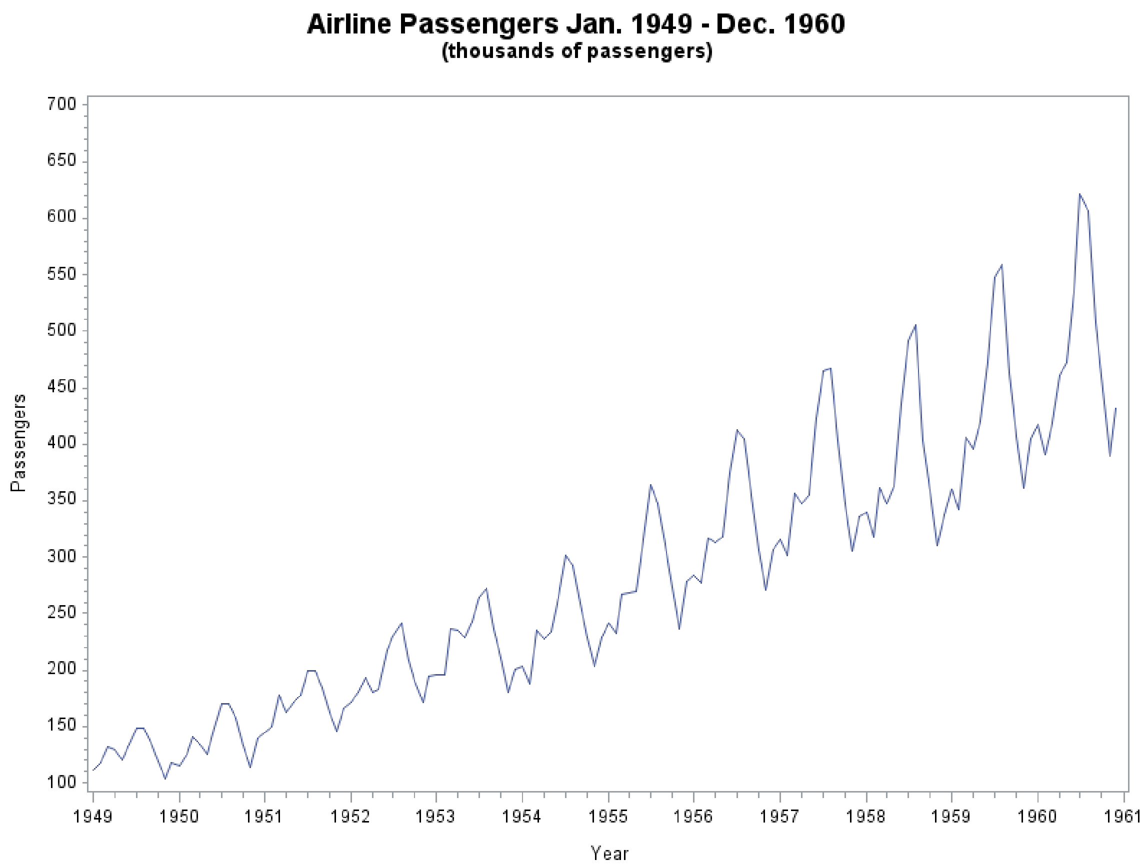
The syntax for the BSM model in SAS is as follows:

```
Proc UCM;
  model y;
     irregular;
     level;
     slope;
     season length = s type = dummy;
```

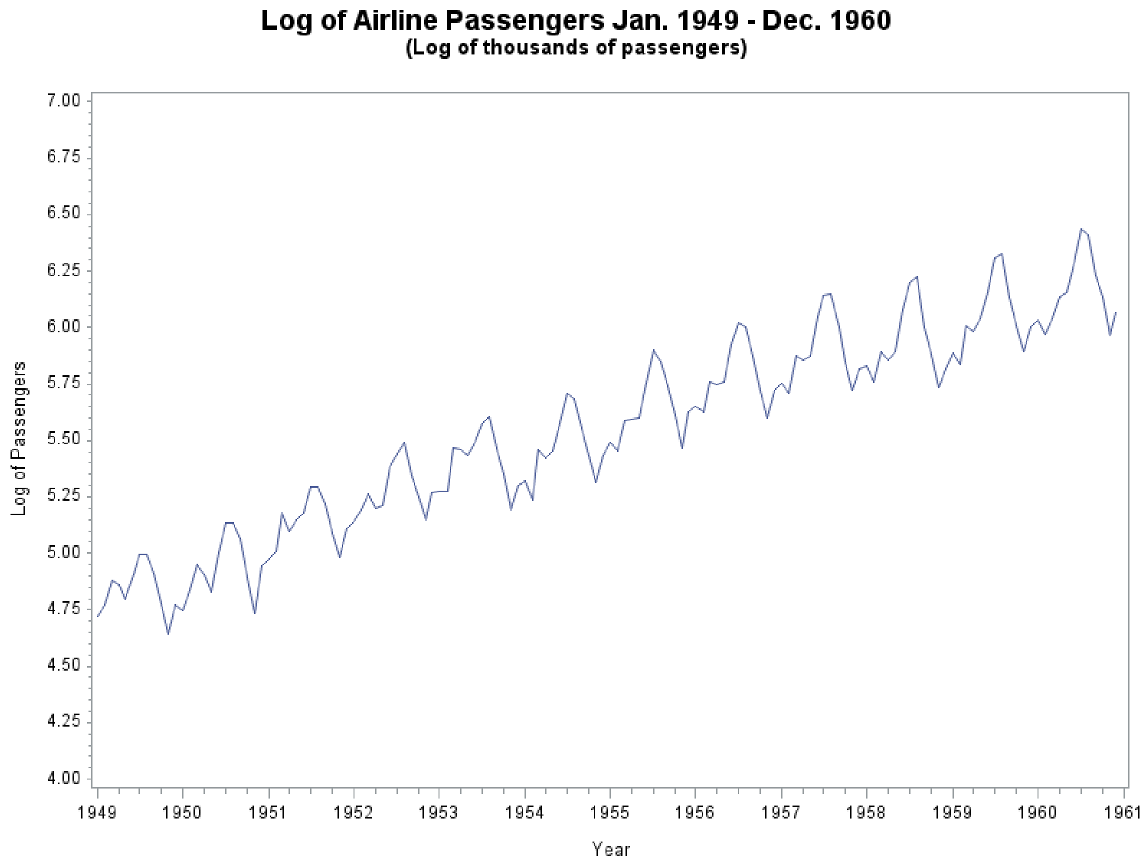
The first line of the SAS code indicates the move to a procedure step in the SAS program using the procedure UCM. The dependent variable to be modeled is y and the Unobservable Components model is to have an irregular component, a level and slope component in the trend, and the stochastic dummy variable seasonal specification (8) is chosen for the seasonal component. Of course, in the above, s is replaced with 12 if the data is monthly, 4 if the data is quarterly, and 2 if the data is bi-annual in nature. If this model fits the time series data at hand well, then additional components can be added by way of cyclical and autoregressive unobservable components and adding regression terms. In contrast, if the data does not contain a stochastic trend and or if the data does not have a seasonal component, the BSM can be correspondingly simplified.

VIII. An Example – the Box-Jenkins Airline Data

A very frequently used time series to demonstrate the nature of a time series with linear trend and seasonality is the so-called Airline Passenger data originally published in Box and Jenkins (1970). This series is a monthly series involving the number of airline passengers that traveled per month over the time period January 1949 through December 1960. As can be seen from the SAS graph below, the data has a linear trend in it and has reoccurring seasonal deviations from trend. These SAS graphs have been generated in a SAS program called BSM.sas that can be found in the appendix to this document.



In order to stabilize the variability of the data around the trend in the latter years, Box and Jenkins (1970) recommended taking the natural logarithms of the data and analyzing them instead of the original series. This transformation of the data is plotted in the graph below.



Later in this course we will study statistical methods for determining the desirability of this transformation. The statistical methods are contained in the SAS macro called %logtest. At any rate we are going to take the logarithmic form to be the preferred form to analyze this data in.

The BSM Model of the Airline Passenger Data

Now let us turn to the analysis of the Airline data vis-à-vis the SAS program BSM.sas contained in the appendix. First consider the output produced by the following SAS code that fits a BSM to the Airline data.

```
title 'BSM: stochastic level, stochastic slope, stochastic dummy
seasonal';

 proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope plot=smooth;
        season length = 12 type=dummy plot=smooth;
        estimate plot=(residual normal acf);
        run;
```

The first line of this SAS Procedure step simply provides SAS with a title for the output to be produced by this part of the BSM.sas program. The ID for the plots is “date” and the frequency of the data is monthly. The model statement tells SAS that “logpass” is the time series to be modeled. The program then specifies the trend to have both the level and slope unobservable components and a stochastic dummy variable seasonal. When plotting the level, slope, and seasonal components we want the smoothed versions. Finally, when estimating the model the program asks SAS to provide us with a plot of the residuals, a histogram of the residuals with a overlaid plot of the normal distribution implied by the variance of the residuals, and a plot of the autocorrelation function (acf) of the residuals with 95% confidence intervals of the various autocorrelations. We will discuss the details of the autocorrelation function later in this course.

So the first major output we get is the following:

BSM: stochastic level, stochastic slope, stochastic dummy seasonal

The UCM Procedure

Input Data Set	
Name	WORK.AIRLINE
Time ID Variable	date

Estimation Span Summary									
Variable	**Type**	**First Obs**	**Last Obs**	**NObs**	**NMiss**	**Min**	**Max**	**Mean**	**Standard Deviation**
logpass	**Dependent**	JAN1949	DEC1960	144	0	4.64439	6.43294	5.54218	0.44146

So the data set we inputted wound up in the “work” subdirectory of SAS and the time ID variable is “date”. SAS also recognizes that “logpass” is the dependent variable to be analyzed by PROC UCM and its time span is given along with minimum and maximum values and the like.

Skipping through some of the output that Proc UCM produces, the next output of interest is the likelihood-based goodness-of-fit measures of the fitted model.

Likelihood Based Fit Statistics	
Full Log-Likelihood	217.42040
Diffuse Part of Log-Likelihood	-4.96981
Normalized Residual Sum of Squares	131.00000
Akaike Information Criterion	-400.84079

Bayesian Information Criterion	-350.35397
Number of non-missing observations used for computing the log-likelihood = 144	

The full likelihood function is the function that Procedure UCM maximizes in getting the parameters (error variances) and unobserved components of the model. Its fitted value is 217.42040. As the next two items, "Diffuse Part of Log-Likelihood" and "Normalized Residual Sum of Squares" are not necessary for our discussion, we will leave their definitions for the reader to read in SAS HELP and its detailed description of Proc UCM. The next two goodness-of-fit statistics, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), are defined as follows:

$$AIC = (-2L + 2k)/T \qquad (11)$$

$$BIC = (-2L + k*\log(T))/ T \qquad (12)$$

where L denotes the full likelihood value of the fitted model, k is the number of free parameters that are estimated in the chosen model, and T is the number of observations used to estimate the candidate model. These goodness-of-fit criteria are useful for discriminating among various competing UCM models. **The specification that minimizes these two measures is to be recommended over its competitors.**

Now let's consider some additional output produced by the above SAS program.

Likelihood Optimization Algorithm Converged in 19 Iterations.

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	0.00012951	0.0001294	1.00	0.3167
Level	Error Variance	0.00069945	0.0001903	3.67	0.0002
Slope	Error Variance	2.64778E-12	1.24107E-9	0.00	0.9983
Season	Error Variance	0.00006413	0.00004383	1.46	0.1435

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.05	0.8320
Level	1	132433	<.0001
Slope	1	17.86	<.0001

Season	11	772.21	<.0001

Summary of Seasons			
Name	**Type**	**Season Length**	**Error Variance**
Season	**DUMMY**	12	0.00006413

In the BSM you will recall that the error variances of the irregular, level, slope, and season components are, respectively, σ_ε^2, σ_η^2, σ_ξ^2, and σ_ω^2 of the irregular, level, slope, and stochastic dummy variable seasonal components, respectively. These are the "free parameters" of the model and their estimates are reported in the table labeled "Final Estimates of the Free Parameters" and are determined by considering all of the data that is available on the dependent variable logpass. These estimates and their corresponding t-values allow one to infer whether the corresponding component is non-stochastic (the null hypothesis) or is stochastic (the alternative hypothesis). The results reported here indicate that the slope component of the trend may be suitably modeled as being non-stochatic (fixed) rather than stochastic. This seems logical given the linear shape (as compared to having some curvature) of the logpass data. In the subsequent model we examine here, labeled BSM2, we will take this suggestion to heart. That is, we will compare the BSM model with stochastic slope with BSM2 with fixed slope. The other error variances range from middling significance to highly significant so we assume, for now, that these components are best modeled as being stochastic.

The next important diagnostic output is contained in the Table labeled "Significance Analysis of Components (Based on the Final State)". Here the Chi-square statistics test the null hypothesis that the given component is not significant while the alternative hypothesis implies the opposite. Given the output of the table, one concludes that all three components analyzed (level, slope, and season) appear to be significant as determined at the final state (observation) of the Kalman filter. The two-sided p-values of the components are less than the conventional levels of significance. (We retain the irregular component of any UCM model as a matter of principle.) The last table of course reports that the type of seasonal component chosen is the stochastic dummy variable type.

The BSM Model with Fixed Slope

As suggested by the results reported in the "Final Estimates of the Free Parameters" Table above, we try the BSM Model with Fixed Slope as invoked by the following Proc UCM statement in SAS:

```
title1 'BSM2: stochastic level, fixed slope, stochastic dummy
seasonal,';
title2 'stochastic cycle';

        proc ucm data = airline;
```

```
id date interval = month;
model logpass;
irregular;
level plot=smooth;
slope var = 0 noest plot=smooth;
season length = 12 type=dummy plot=smooth;
    estimate plot=(residual normal acf);
run;
```

This procedure step is just like the previous one except the slope statement now includes "var = 0 and noest" which specifies that the error variance for the slope component is to be set to zero and not estimated implying that the slope coefficient in the trend is to be estimated as a fixed parameter. The major diagnostic tables for this model are reproduced below.

BSM2: stochastic level, fixed slope, stochastic dummy seasonal

Likelihood Based Fit Statistics	
Full Log-Likelihood	217.42040
Diffuse Part of Log-Likelihood	-4.96981
Normalized Residual Sum of Squares	130.99998
Akaike Information Criterion	-402.84080
Bayesian Information Criterion	-355.32379
Number of non-missing observations used for computing the log-likelihood = 144	

Likelihood Optimization Algorithm Converged in 19 Iterations.

Final Estimates of the Free Parameters

Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	0.00012951	0.0001294	1.00	0.3167
Level	Error Variance	0.00069945	0.0001903	3.67	0.0002
Season	Error Variance	0.00006413	0.00004383	1.46	0.1435

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.05	0.8320
Level	1	132433	<.0001
Slope	1	17.86	<.0001
Season	11	772.21	<.0001

Again all three of the components (level, fixed slope, and season) are statistically significant. Moreover the fit of the BSM2 model is better than the fit of the initial BSM model because the BSM2 model's goodness-of-fit measures (AIC = -402.84 and BIC = -355.32) are smaller than the corresponding goodness-of-fit measures for the BSM model (AIC = -400.84 and BIC = -350.350. We have achieved an improvement in our original specification in going from a stochastic slope specification to a fixed slope specification in the trend.

Should the Seasonal Component be modeled as stochastic or non-stochastic (fixed) over time?

In the BSM2 Table labeled "Final Estimates of the Free Parameters" you will notice that the "season" error component is not highly significant (p = 0.1435). Thus one might question whether or not we should specify, not only the slope coefficient, but also the seasonal coefficients (there are 11 of them) to be non-stochastic (fixed). We consider this issue by entertaining the BSM3 model which is estimated by the following Proc UCM statements:

```
title 'BSM3: stochastic level, fixed slope, non-stochastic dummy
seasonal';

       proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy var = 0 noest plot=smooth;
        estimate plot=(residual normal acf);
        run;
```

Here you will notice that the "season" statement now has the additional specification "var = 0 noest" which specifies that the dummy variable seasonal component should now be treated as fixed and estimated accordingly. The only detail we provide from the output produced by this model is the following goodness-of-fit table.

BSM3: stochastic level, fixed slope, non-stochastic dummy seasonal

Likelihood Based Fit Statistics	
Full Log-Likelihood	217.42040
Diffuse Part of Log-Likelihood	-4.96981
Normalized Residual Sum of Squares	131.00000
Akaike Information Criterion	-394.92959
Bayesian Information Criterion	-350.38239
Number of non-missing observations used for computing the log-likelihood = 144	

In comparing the AIC = -394.93 and BIC = -350.38 produced by this model with the AIC and BIC measures when the seasonal component was treated as being stochastic (AIC = -402.84 and BIC = -355.32) we see that the stochastic specification is to be preferred. Evidently, the seasonal component in the Airline passenger data evolved slightly over time (as we will see in a graph of the seasonal component presented in the final model we choose later).

Should a Cyclical Component be added to the model?

There are, of course, several different possible specifications of a cyclical component for an UCM. Here we try a stochastic cyclical model of the form (7) above. The SAS code used to implement this model is as listed below.

```
title1 'BSM4: stochastic level, fixed slope, stochastic dummy
seasonal,';
title2 'one stochastic cycle';

       proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy plot=smooth;
        cycle plot=smooth;
        estimate plot=(residual normal acf);
```

```
run;
```

Notice that the "cycle" statement has been added to the previous code resulting in a model we label as BSM4. The essential diagnostic tables for this model are presented below.

BSM4: stochastic level, fixed slope, stochastic dummy seasonal, one stochastic cycle

Likelihood Based Fit Statistics	
Full Log-Likelihood	228.29612
Diffuse Part of Log-Likelihood	-4.96981
Normalized Residual Sum of Squares	131.00000
Akaike Information Criterion	-418.59223
Bayesian Information Criterion	-362.16578
Number of non-missing observations used for computing the log-likelihood = 144	

Likelihood Optimization Algorithm Converged in 43 Iterations.

Final Estimates of the Free Parameters

Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	**Error Variance**	0.00032039	0.0001206	2.66	0.0079
Level	**Error Variance**	0.00028734	0.0001017	2.83	0.0047
Season	**Error Variance**	0.00005436	0.00003033	1.79	0.0731
Cycle	**Damping Factor**	1.00000	3.88275E-6	257549	<.0001
Cycle	**Period**	12.07340	0.21719	55.59	<.0001
Cycle	**Error Variance**	1.085318E-8	6.41319E-8	0.17	0.8656

Significance Analysis of Components (Based on the Final State)

Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.00	0.9490
Level	1	125871	<.0001

Slope	1	46.93	<.0001
Cycle	2	38.12	<.0001
Season	11	356.74	<.0001

From these tables we see that included the cycle in our model has helped in that we have a better fit of our data (AIC = -418.59 and BIC = -362.166). Moreover, the seasonal component is highly significant although from the t-statistic on the cycle error variance it is not clear whether we should be modeling our cycle to be stochastic as in (7) or non-stochastic as in (6).

Should there be one cycle or two cycles in the Cyclical Component?

First, before addressing whether our cycle should be stochastic or non-stochastic, let us see if adding a second cycle (in the spirit of Fourier analysis) in explaining the cyclical component would be useful. The SAS code that allows us to investigate this issue is contained below.

```
title1 'BSM5: stochastic level, fixed slope, stochastic dummy
seasonal,';
title2 'two stochastic cycles';

       proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy plot=smooth;
            cycle plot=smooth;
            cycle plot=smooth;
            estimate plot=(residual normal acf);
        run;
```

Here a second "cycle" statement has been added to the SAS code. This model is referred to as the BSM5 model. The relevant output for considering the second cycle is provided below.

BSM5: stochastic level, fixed slope, stochastic dummy seasonal,
two stochastic cycles

Likelihood Based Fit Statistics	
Full Log-Likelihood	228.29612
Diffuse Part of Log-Likelihood	-4.96981
Normalized Residual Sum of Squares	131.00000
Akaike Information Criterion	-413.92702

Bayesian Information Criterion	-348.59113
Number of non-missing observations used for computing the log-likelihood = 144	

Likelihood Optimization Algorithm Converged in 42 Iterations.

Final Estimates of the Free Parameters

Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	**Error Variance**	0.00036207	0.0001257	2.88	0.0040
Level	**Error Variance**	0.00022676	0.00009692	2.34	0.0193
Season	**Error Variance**	0.00005395	0.00002949	1.83	0.0673
Cycle_1	**Damping Factor**	1.00000	4.23693E-6	236020	<.0001
Cycle_1	**Period**	12.07678	0.21534	56.08	<.0001
Cycle_1	**Error Variance**	9.699839E-9	5.43887E-8	0.18	0.8585
Cycle_2	**Damping Factor**	1.00000	0.0001823	5486.93	<.0001
Cycle_2	**Period**	50.47848	6.27949	8.04	<.0001
Cycle_2	**Error Variance**	1.889762E-9	2.90126E-9	0.65	0.5148

Significance Analysis of Components (Based on the Final State)

Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.00	0.9601
Level	1	89818.2	<.0001
Slope	1	57.39	<.0001
Cycle_1	2	41.97	<.0001
Cycle_2	2	3.95	0.1389

As we can see from the significance analysis of the components, the second cycle is not significant at conventional levels (p=0.1389). Moreover, the goodness-of-fit of the model has deteriorated somewhat (AIC = -413.93, BIC = -348.59) with the addition of the second cycle.

Is the Cyclical Component Stochastic or Non-Stochastic?

Now that we have settled on one cycle for the cyclical component instead of two, let us turn to the issue of whether that one cycle should be modeled as being stochastic or non-stochastic. The SAS code that addresses this issue is presented below. The model with the non-stochastic cycle is labeled the BSM6 model.

```
title1 'BSM6: stochastic level, fixed slope, stochastic dummy
seasonal,';
title2 'non-stochastic cycle';

      proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy plot=smooth;
        cycle noest=variance var = 0 plot=smooth;
        estimate plot=(residual normal acf);
        run;
```

To make the cycle non-stochastic the code "noest = variance var = 0" has been added to the "cycle" statement. The major result is that the non-stochastic specification of the cycle does not result in an improvement in the fit of the data vis-à-vis the AIC and BIC measures. (AIC = -398.84 and –345.38). Therefore we adopt the stochastic specification (7) of the cycle for this data.

Considering an Autoregressive Component as an alternative to the stochastic cyclical component

It has been mentioned that the Autoregressive Unobserved Component (9) might be considered as an alternative to the Fourier type of specification for the cycle in time series data. The SAS code that allows us to try out this specification (labeled BSM7) is listed below.

```
title1 'BSM7: stochastic level, fixed slope, stochastic dummy
seasonal,';
title2 'autoregressive component for cycle';

       proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy plot=smooth;
        autoreg plot=smooth;
        estimate plot=(residual normal acf);
        run;
```

Although the output of this procedure indicates that the autoregressive component is statistically significant, it does not improve on the fit of the data (AIC = -403.11, BIC = -349.66) vis-à-vis the previous BSM4 model specification. **Therefore, we select the BSM4 specification of a stochastic level and fixed slope for the trend, a stochastic dummy variable seasonal, and a single stochastic cycle along with the irregular component.** A summary of the Goodness-of-Fit values for the various models is contained in the following table.

Goodness-of-Fits for Various UCMs

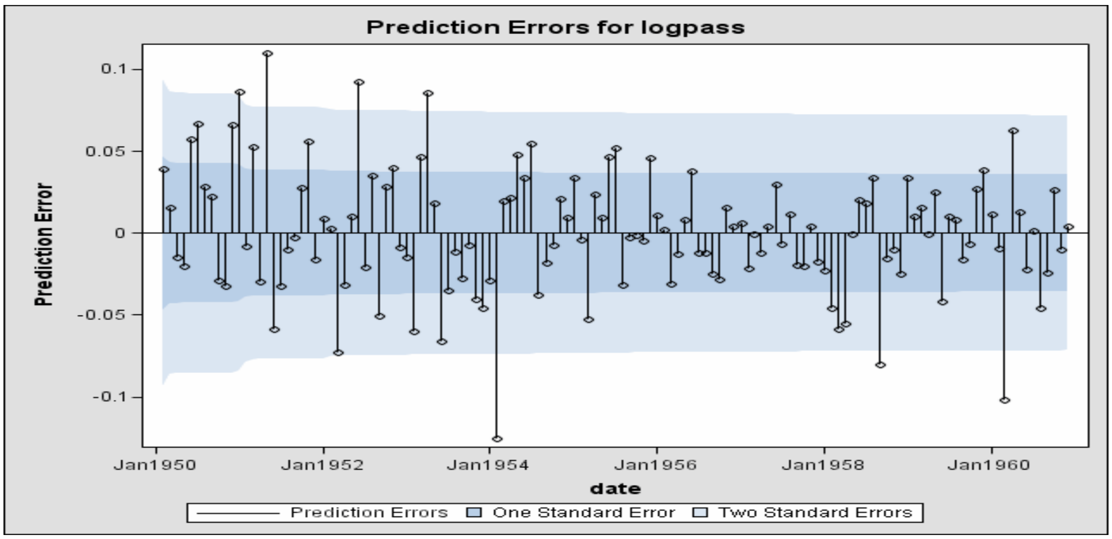
* Number of autocorrelations that are outside of their 95% confidence intervals.

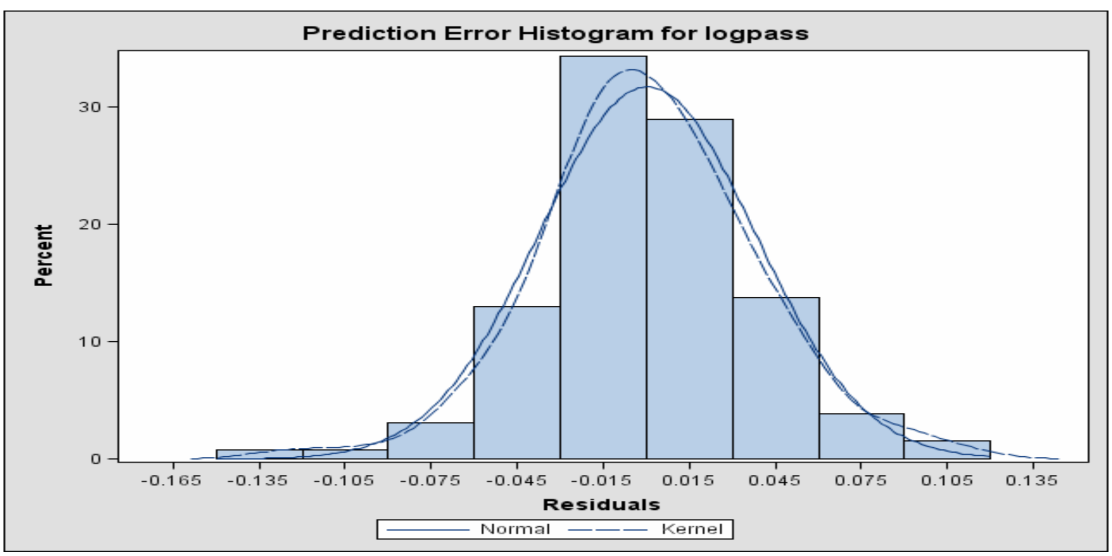
Model	**Description**	**Variance Components**	**AIC**	**BIC**	**Residuals w. n. ***
BSM	stochastic level	$\sigma_\eta^2 > 0$	-400.84	-350.350	2
	stochastic slope	$\sigma_\xi^2 > 0$			
	stochastic dummy season	$\sigma_\omega^2 > 0$			
BSM2	stochastic level	$\sigma_\eta^2 > 0$	-402.84	-355.32	2
	fixed slope	$\sigma_\xi^2 = 0$			
	stochastic dummy season	$\sigma_\omega^2 > 0$			
BSM3	stochastic level	$\sigma_\eta^2 > 0$	-394.93	-350.38	4
	fixed slope	$\sigma_\xi^2 = 0$			
	non-stochastic dummy season	$\sigma_\omega^2 = 0$			
BSM4	stochastic level	$\sigma_\eta^2 > 0$	-418.59	-362.166	1
	fixed slope	$\sigma_\xi^2 = 0$			
	stochastic dummy season	$\sigma_\omega^2 > 0$			
	one stochastic cycle	$\sigma_\nu^2 > 0$			
BSM5	stochastic level	$\sigma_\eta^2 > 0$	-413.93	-348.59	2
	fixed slope	$\sigma_\xi^2 = 0$			
	stochastic dummy season	$\sigma_\omega^2 > 0$			
	two stochastic cycles	$\sigma_{\nu 1}^2 > 0$, $\sigma_{\nu 2}^2 > 0$			
BSM6	stochastic level	$\sigma_\eta^2 > 0$	-398.84	–345.38	2
	fixed slope	$\sigma_\xi^2 = 0$			
	stochastic dummy season	$\sigma_\omega^2 > 0$			
	one non-stochastic cycle	$\sigma_\nu^2 = 0$			
BSM7	stochastic level	$\sigma_\eta^2 > 0$	-403.11	-349.66	2
	fixed slope	$\sigma_\xi^2 = 0$			
	stochastic dummy season	$\sigma_\omega^2 > 0$			
	Autoregressive component	$\sigma_\upsilon^2 > 0$			

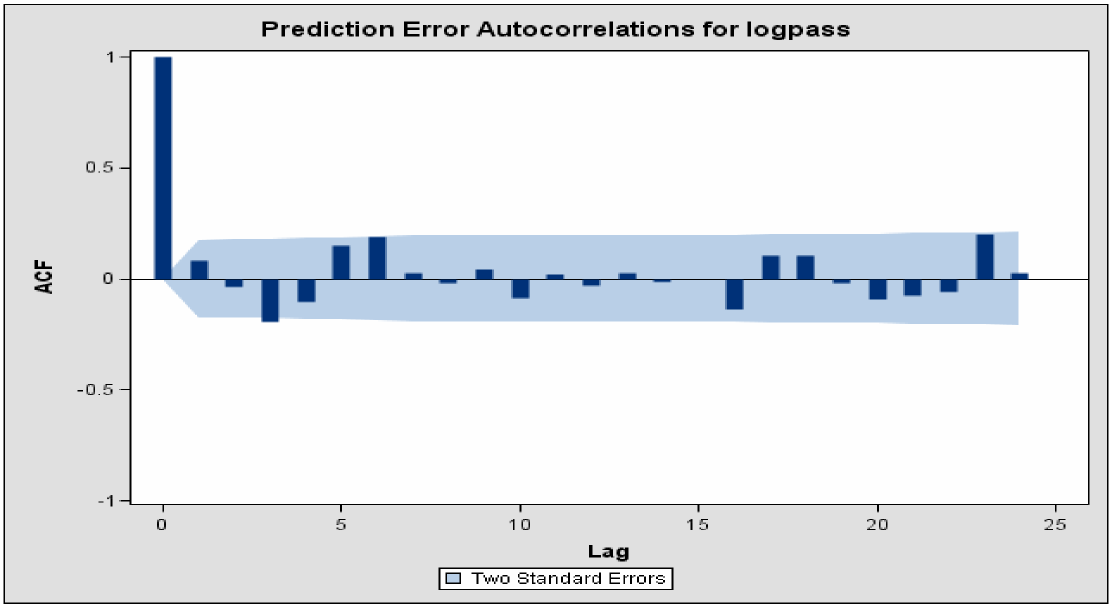
Remember, don't count the autocorrelation at lag = 0 because it is equal to one **by definition**. Any number above 3 occurrences would indicate the lack of white noise residuals.

Further Verification of the BSM4 model

Additional verification of the tentatively chosen BSM4 model can be obtained by reviewing the output produced by the statement "estimate plot = (residual normal act)". This statement produces (1) a residual plot of the model (i.e. plot of the irregular component), (2) a plot comparing a histogram of the residuals of the model with a superimposed normal distribution having the same mean (zero) and variance as the residuals themselves, and (3) a plot of the autocorrelation function of the model's residuals with a 95% confidence interval of the autocorrelations under the assumption that the residuals are independent of each other. These 3 plots are reproduced below.



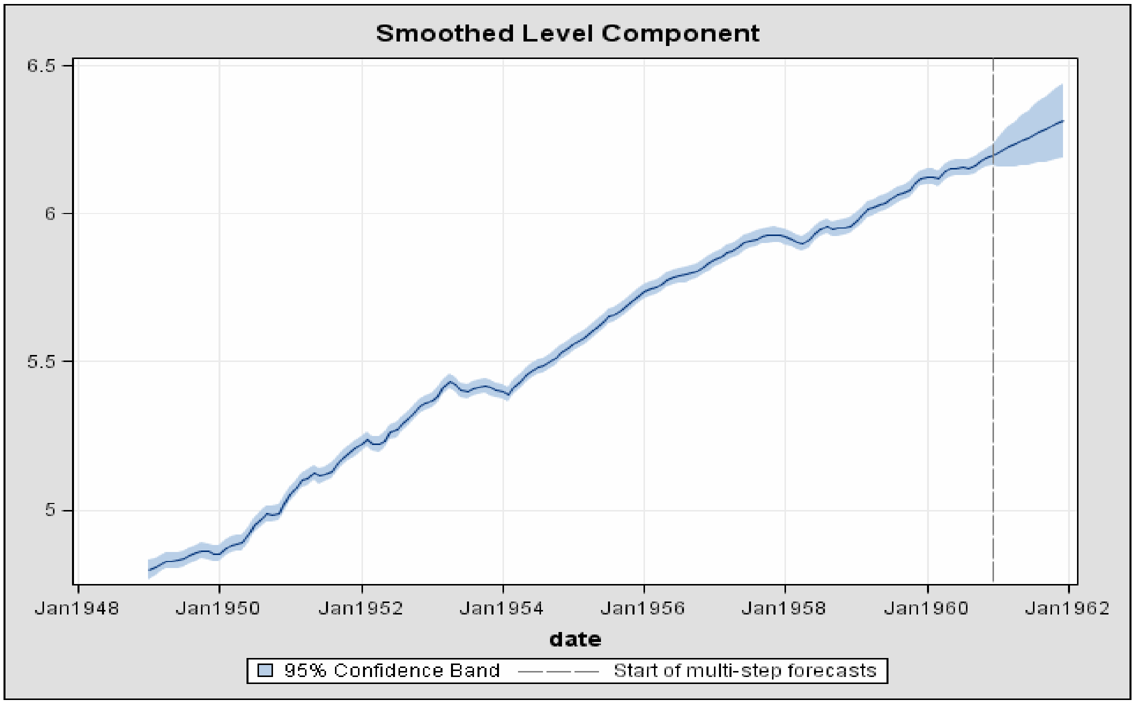


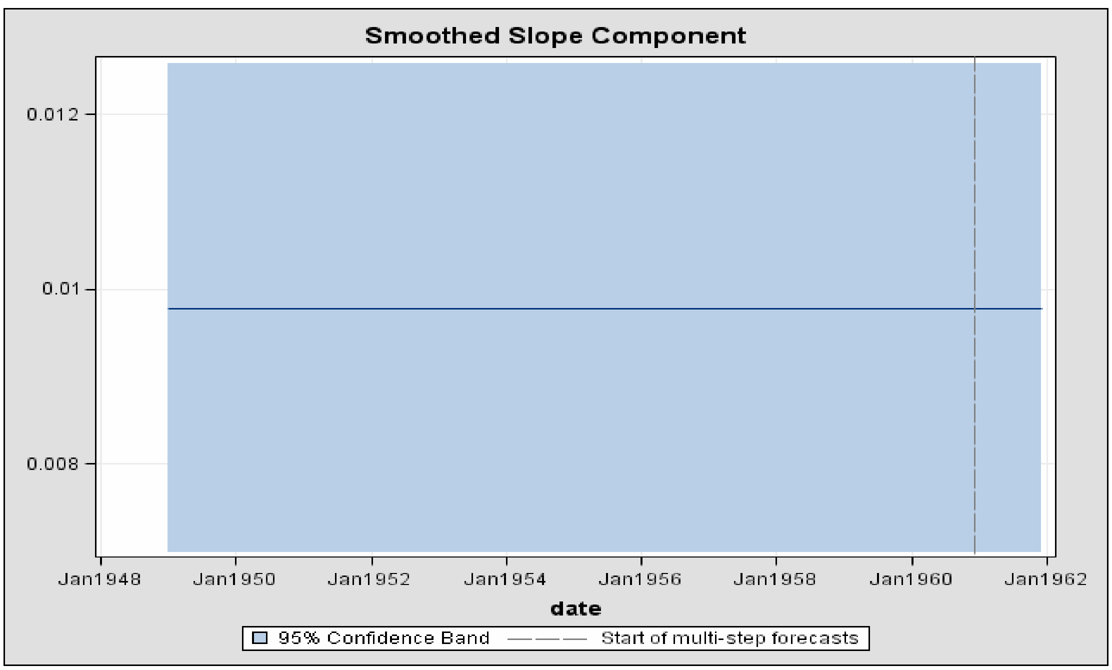


In terms of the “Prediction Errors for logpass” graph, the residuals of the model (prediction errors) appear to be varying in an unsystematic way. That is, there do not appear to be “stretches” in the data whereby the residuals have positive runs for a little while and then negative runs for a little while, etc. They seem to be pretty unsystematically varying which is indicative of the independence of the residuals over time and indicative that there is “nothing systematic left in the data” to be described by additional components and parameters that we might entertain adding to our model. This is verified in the third graph labeled “Prediction Error Autocorrelations for logpass” where all of the autocorrelations at the various lags are inside the blue-shaded 95% confidence intervals of zero autocorrelation. (More about autocorrelations and the autocorrelation function later. The autocorrelation at lag zero is by definition 1.0 so its value is not of importance in the graph.) Finally, the graph labeled “Prediction Errors Histogram for logpass” shows the residual (prediction error) histogram pretty closely resembling a normal distribution which is also a desirable trait of a good UCM model. Therefore, all three graphs support the previous contention that model BSM4 is adequate for explaining the variation in the Airline passenger data we set out to model.

The Component Graphs associated with model BSM4

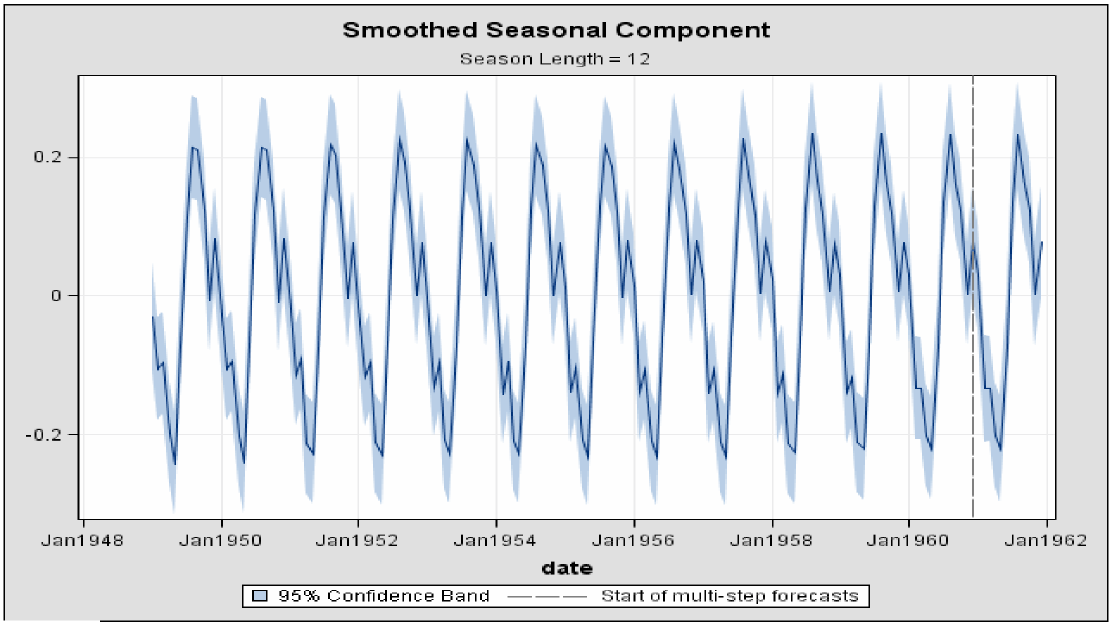
In this section we are going to report the component graphs for the BSM4 model. They are obtained by applying the Kalman filter and constructing the “smoothed” estimates as described previously. The first graph is the smoothed level component graph representing a time plot of $\hat{\mu}_t$ and the forecasts of it 12 periods ahead along with its 95% confidence interval. The second graph is the smoothed slope component graph ($\hat{\beta}_t$) which happens to be fixed and estimated to be approximately 0.01.

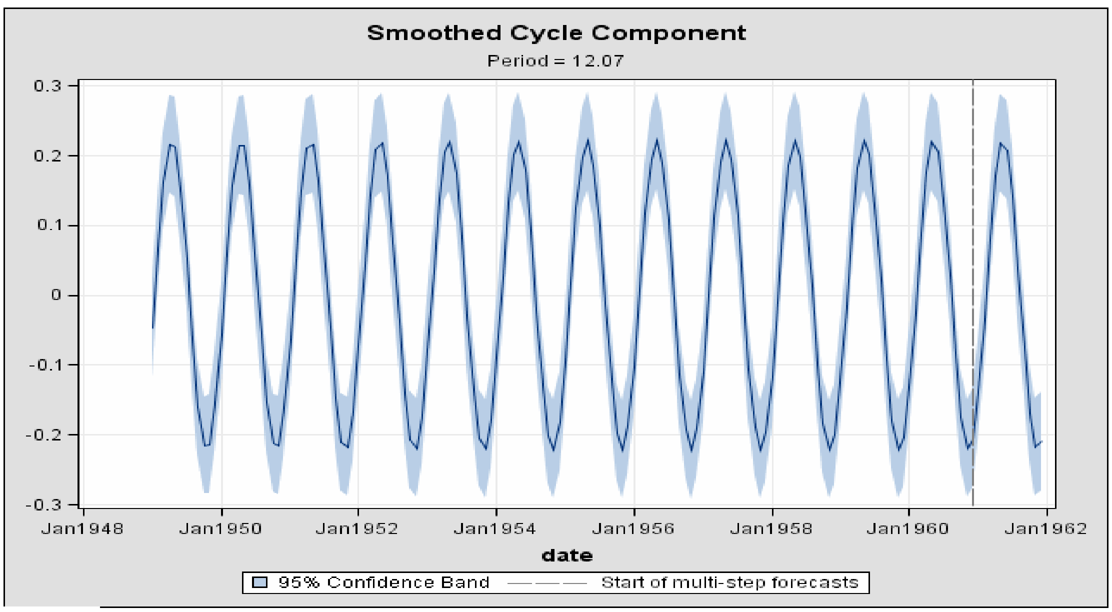




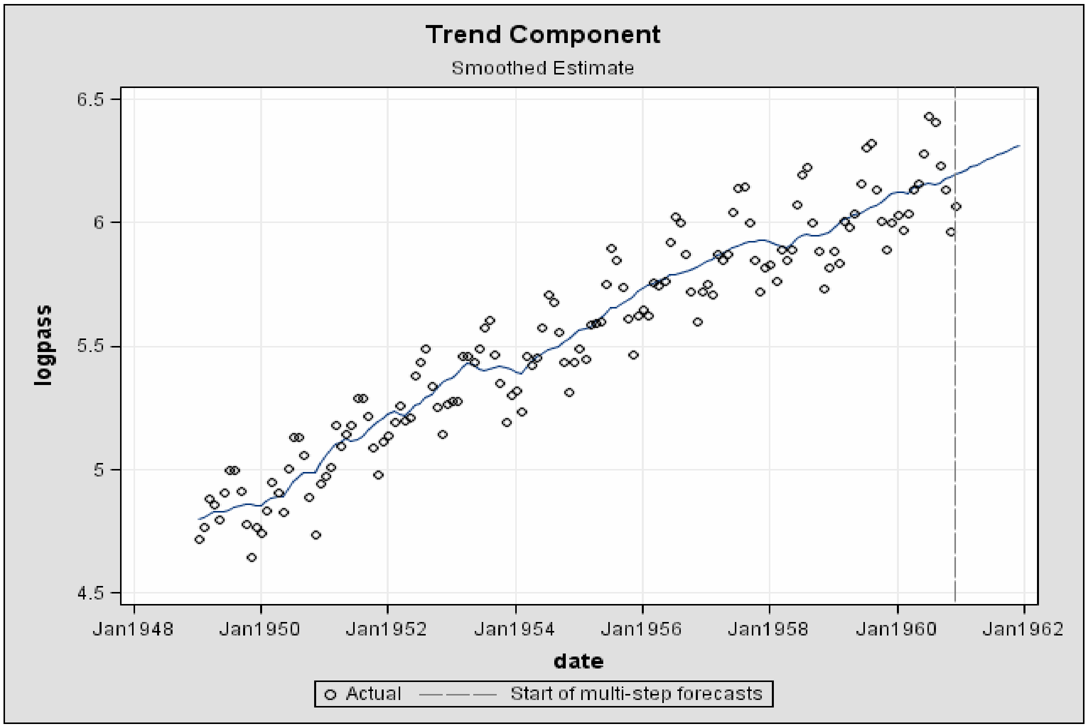
The third and fourth graphs below depict the smoothed seasonal component ($\hat{\gamma}_t$) and smoothed cyclical component ($\hat{\psi}_t$), respectively, as a function of time. In all of these graphs the blue shading indicates the 95% confidence intervals of the components. Also notice that each graph is extended for 12 months beyond the end of the data

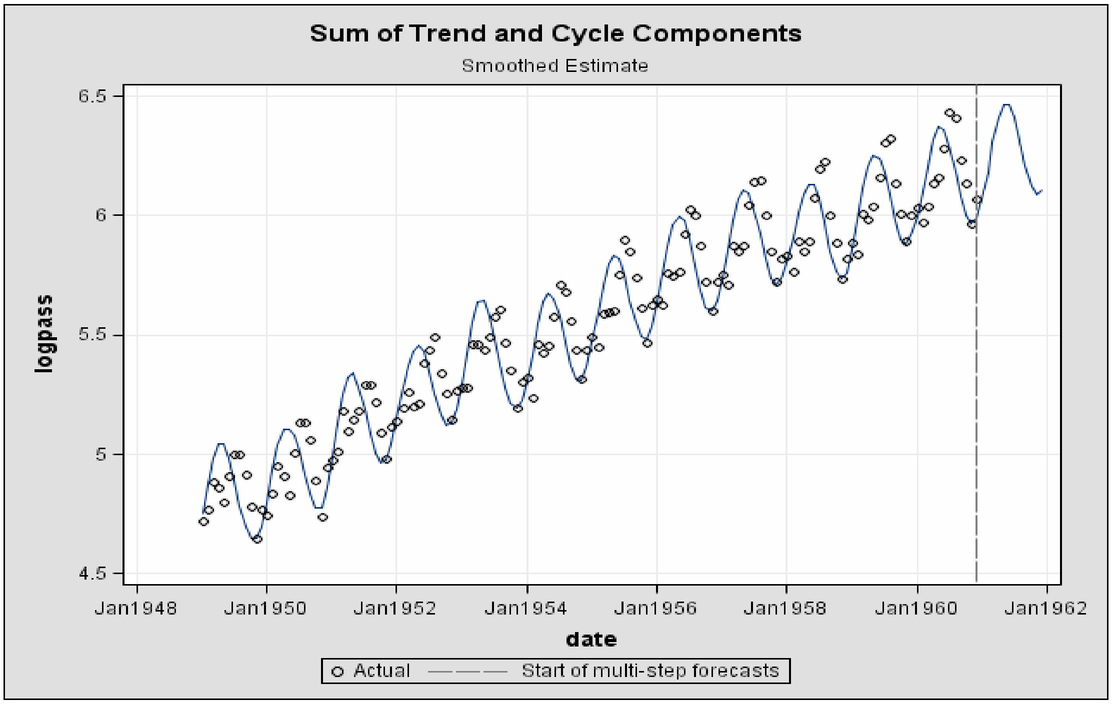
(December 1960) representing the forecasted values of the components. Of course these forecasted components, once they are added together, form the forecasted values of logpass reported below.

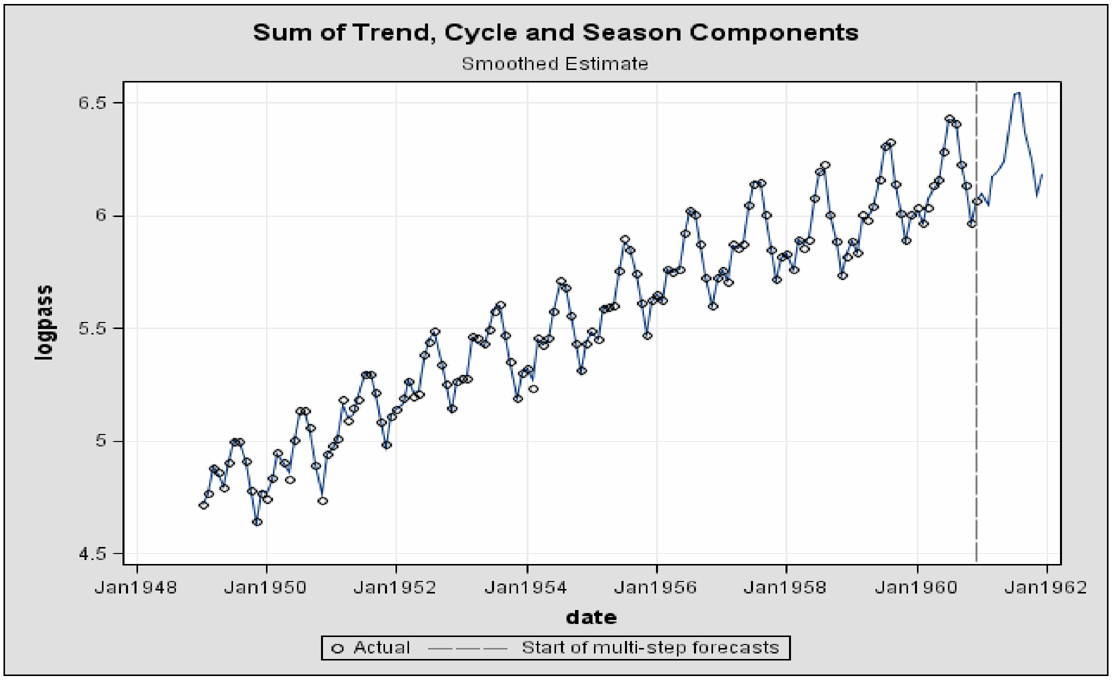




The following three graphs depict the "adding" up of the components of trend, season and cycle into the fitted values of the Airline series. Notice the extension of these series into the forecast period.

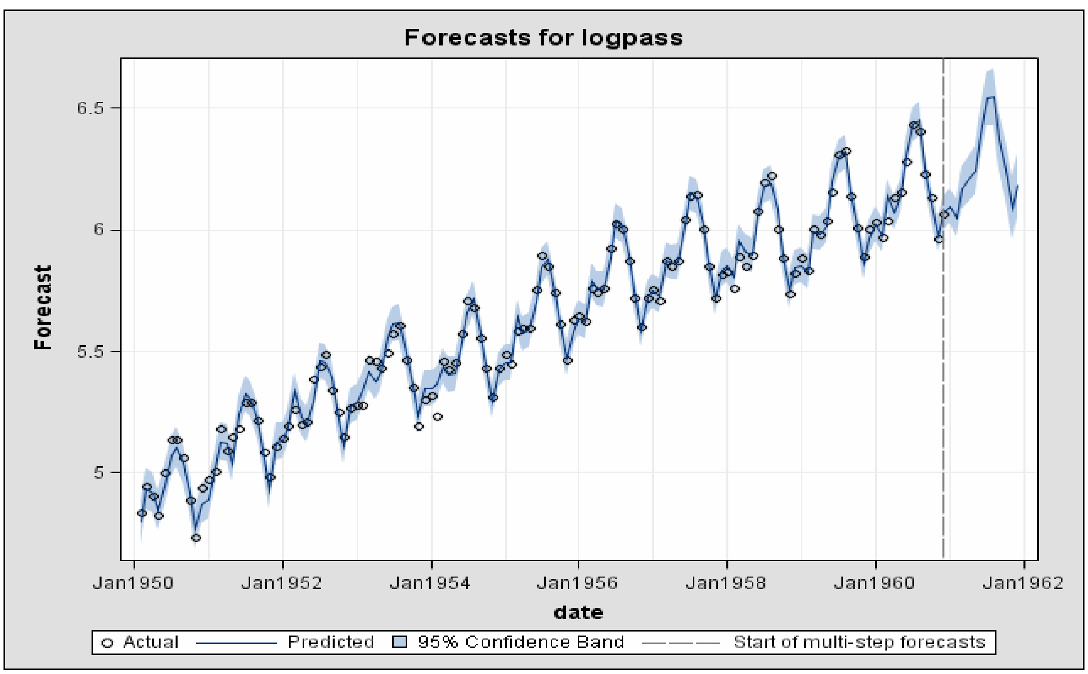

Trend Component
Smoothed Estimate
logpass
6.5
6
5.5
5
4.5
Jan1948
Jan1950
Jan1952
Jan1954
Jan1956
Jan1958
Jan1960
Jan1962
date
Actual
Start of multi-step forecasts


Sum of Trend and Cycle Components
Smoothed Estimate
logpass
6.5
6
5.5
5
4.5
Jan1948
Jan1950
Jan1952
Jan1954
Jan1956
Jan1958
Jan1960
Jan1962
date
Actual
Start of multi-step forecasts



Predicting with model BSM4

The 12 out-of-sample forecasts and their 95% confidence intervals are depicted in the graph below in the last UCM procedure step in the BSM.sas program.



The numerical values of the forecasts are produced by the last statements in BSM.sas. Since the interest of the user is more likely to involved obtaining the forecasts of the number of passengers per month in the subsequent 12 months as compared to the log of the number of passengers expected the SAS statements at the end properly transform the log forecasts into the "level" forecasts of passengers (in thousands). As we will see in later discussion, an unbiased forecast of passengers (pass) is obtained by the anti-log transformation

$$\text{Passf} = \exp(\text{forecast} + 0.5*\text{std}**2) \tag{13}$$

and the upper and lower 95% confidence limits of passengers forecasted are obtained by the transformations

$$\text{passlcl} = \exp(\text{lcl}) \tag{14}$$

and

$$\text{passucl} = \exp(\text{ucl}). \tag{15}$$

In (13) "passf" denotes the desired passenger forecast, exp(.) is the exponential (anti-log) function, "forecast' is log of passenger forecast that is to be transformed, and "std" is the standard deviation of the log forecast. In (14) "passlcl" denotes the desired lower confidence limit for passengers forecasted and "lcl" is the lower confidence limit for the log forecast. In (15) "passucl" denotes the desired upper confidence limit for passengers forecasted and "ucl" denotes the upper confidence limit for the log forecast.

Forecasts produced by BSM4

Obs	Obs	FORECAST	STD	LCL	UCL	passf	passlcl	passucl
1	145	6.09666	0.035833	6.02643	6.16689	444.658	414.235	476.703
2	146	6.04763	0.039551	5.97011	6.12515	423.439	391.548	457.211
3	147	6.17027	0.043828	6.08437	6.25617	478.775	438.943	521.221
4	148	6.20533	0.047668	6.11191	6.29876	495.948	451.198	543.899
5	149	6.24241	0.051027	6.14240	6.34242	514.766	465.169	568.170
6	150	6.38805	0.053942	6.28233	6.49378	595.564	535.034	661.017
7	151	6.53967	0.056508	6.42892	6.65042	693.164	619.503	773.111
8	152	6.54427	0.058839	6.42894	6.65959	696.450	619.518	780.229
9	153	6.36839	0.061048	6.24874	6.48804	584.206	517.360	657.236
10	154	6.24796	0.063205	6.12408	6.37184	517.990	456.723	585.132
11	155	6.08873	0.065249	5.96084	6.21661	441.799	387.936	501.004
12	156	6.18190	0.066620	6.05133	6.31247	484.986	424.677	551.408

IX. Conclusion

We have seen that the UCM model is a convenient way of additively decomposing a time series into components for trend, season, cycle, and irregular movements. Proc UCM in SAS uses the Kalman filter to provide this decomposition and some useful diagnostics for determining just what kind of UCM model should be specified for the data at hand. In particular, careful use of tests of error component variances and component effects can provide us with a time series model that is not only meaningfully interpretable but also competitive as far as forecasting accuracy is concerned.

Of course the UCM model is not the only time series model that we might consider for modeling and forecasting time series data like the Box-Jenkins Airline data that we have just analyzed. We will consider the Deterministic Trend / Deterministic Seasonal model next and then go on to investigate exponential smoothing models and Box-Jenkins models in turn.

APPENDIX

BSM.sas Program
That Fits Various UCMs
To the Box-Jenkins Airline Data

```
/* This is the Airline Passenger data originally analyzed by Box and
   Jenkins in their classic textbook (1970).  */

/*  Here we apply the Basic Structural Model (BSM) of Harvey (1989)
    and some extensions thereof to the Airline data.  */

data airline;
   input date:monyy5. pass @@;
   datalines;
jan49 112 feb49 118 mar49 132 apr49 129 may49 121 jun49 135
jul49 148 aug49 148 sep49 136 oct49 119 nov49 104 dec49 118
jan50 115 feb50 126 mar50 141 apr50 135 may50 125 jun50 149
jul50 170 aug50 170 sep50 158 oct50 133 nov50 114 dec50 140
jan51 145 feb51 150 mar51 178 apr51 163 may51 172 jun51 178
jul51 199 aug51 199 sep51 184 oct51 162 nov51 146 dec51 166
jan52 171 feb52 180 mar52 193 apr52 181 may52 183 jun52 218
jul52 230 aug52 242 sep52 209 oct52 191 nov52 172 dec52 194
jan53 196 feb53 196 mar53 236 apr53 235 may53 229 jun53 243
jul53 264 aug53 272 sep53 237 oct53 211 nov53 180 dec53 201
jan54 204 feb54 188 mar54 235 apr54 227 may54 234 jun54 264
jul54 302 aug54 293 sep54 259 oct54 229 nov54 203 dec54 229
jan55 242 feb55 233 mar55 267 apr55 269 may55 270 jun55 315
jul55 364 aug55 347 sep55 312 oct55 274 nov55 237 dec55 278
jan56 284 feb56 277 mar56 317 apr56 313 may56 318 jun56 374
jul56 413 aug56 405 sep56 355 oct56 306 nov56 271 dec56 306
jan57 315 feb57 301 mar57 356 apr57 348 may57 355 jun57 422
jul57 465 aug57 467 sep57 404 oct57 347 nov57 305 dec57 336
jan58 340 feb58 318 mar58 362 apr58 348 may58 363 jun58 435
jul58 491 aug58 505 sep58 404 oct58 359 nov58 310 dec58 337
jan59 360 feb59 342 mar59 406 apr59 396 may59 420 jun59 472
jul59 548 aug59 559 sep59 463 oct59 407 nov59 362 dec59 405
jan60 417 feb60 391 mar60 419 apr60 461 may60 472 jun60 535
jul60 622 aug60 606 sep60 508 oct60 461 nov60 390 dec60 432
;

data airline;
  set airline;
  logpass=log(pass);

title1 'Airline Passengers Jan. 1949 - Dec. 1960';
title2 '(thousands of passengers)';

axis1 label=('Year');

axis2 order=(100 to 700 by 50)
      label=(angle=90 'Passengers');

proc gplot data=airline;
```

```
   plot pass*date / haxis=axis1 vaxis=axis2;
   symbol1 i=join;
   format date year4.;

run;

title1 'Log of Airline Passengers Jan. 1949 - Dec. 1960';
title2 '(Log of thousands of passengers)';

axis1 label=('Year');
axis2 order=(4 to 7 by .25)
      label=(angle=90 'Log of Passengers');

proc gplot data=airline;
   plot logpass*date / haxis=axis1 vaxis=axis2;
   symbol1 i=join;
   format date year4.;

run;

ods html;
   ods graphics on;

 /* Here we try out the BSM and Several Extensions of BSM */

 /* BSM: stochastic level, stochastic slope, stochastic dummy
    seasonal */

title 'BSM: stochastic level, stochastic slope, stochastic dummy
seasonal';

 proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope plot=smooth;
        season length = 12 type=dummy plot=smooth;
        estimate plot=(residual normal acf);
        run;

  /* BSM2: stochastic level, fixed slope, stochastic dummy seasonal */

title 'BSM2: stochastic level, fixed slope, stochastic dummy seasonal';

       proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy plot=smooth;
        estimate plot=(residual normal acf);
        run;

 /* BSM3: stochastic level, fixed slope, deterministic dummy
          seasonal  */
```

```
title 'BSM3: stochastic level, fixed slope, non-stochastic dummy
seasonal';

       proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy var = 0 noest plot=smooth;
        estimate plot=(residual normal acf);
        run;

 /* BSM4: stochastic level, fixed slope, stochastic dummy seasonal,
              one stochastic cycle */

title1 'BSM4: stochastic level, fixed slope, stochastic dummy
        seasonal,';
title2 'one stochastic cycle';

       proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy plot=smooth;
        cycle plot=smooth;
        estimate plot=(residual normal acf);
        run;

/* BSM5: stochastic level, fixed slope, stochastic dummy seasonal,
              two stochastic cycles */

title1 'BSM5: stochastic level, fixed slope, stochastic dummy
seasonal,';
title2 'two stochastic cycles';

       proc ucm data = airline;
        id date interval = month;
        model logpass;
        irregular;
        level plot=smooth;
        slope var = 0 noest plot=smooth;
        season length = 12 type=dummy plot=smooth;
            cycle plot=smooth;
            cycle plot=smooth;
            estimate plot=(residual normal acf);
        run;

/* BSM6: stochastic level, fixed slope, stochastic dummy seasonal,
             non-stochastic cycle  */

title1 'BSM6: stochastic level, fixed slope, stochastic dummy
seasonal,';
title2 'non-stochastic cycle';
```

```
       proc ucm data = airline;
         id date interval = month;
         model logpass;
         irregular;
         level plot=smooth;
         slope var = 0 noest plot=smooth;
         season length = 12 type=dummy plot=smooth;
             cycle noest=variance var = 0 plot=smooth;
             estimate plot=(residual normal acf);
         run;

/* BSM7: stochastic level, fixed slope, stochastic dummy seasonal,
              autoregressive component for cycle */

title1 'BSM7: stochastic level, fixed slope, stochastic dummy
seasonal,';
title2 'autoregressive component for cycle';

        proc ucm data = airline;
         id date interval = month;
         model logpass;
         irregular;
         level plot=smooth;
         slope var = 0 noest plot=smooth;
         season length = 12 type=dummy plot=smooth;
         autoreg plot=smooth;
         estimate plot=(residual normal acf);
         run;

        /*  According to AIC and BIC and significance of components,
            BSM4 is apparently the best model.  */

        /* Now we forecast with BSM4.  */

title 'Forecasts produced by BSM4';

        proc ucm data = airline;
         id date interval = month;
         model logpass;
         irregular;
         level plot=smooth;
         slope var = 0 noest plot=smooth;
         season length = 12 type=dummy plot=smooth;
         cycle plot=smooth;
         forecast lead=12 plot=(forecasts decomp) outfor = results;
         run;

/* Here we are translating the log forecasts into forecasts of
passengers (in thousands) and then printing them out.  Also the lower
and upper confidence intervals have been translated as well.  */

            data results;
              set results;
              passf = exp(forecast + 0.5*std**2);
              passlcl = exp(lcl);
              passucl = exp(ucl);
```

```
  if _n_ > 144;
  keep obs forecast std lcl ucl passf passlcl passucl;
  run;

proc print data = results;
run;
```