Chapter 4 Unobserved Components Models

4.1 Introduction to Using Unobserved Components Models

Unobserved components models (UCMs) are substantially different from other models that are discussed in this course. They can accommodate and extrapolate more general features of the data (for example, seasonal patterns that change as a function of time). However, they are relatively easy to specify and refine. This chapter focuses on their ease of use and builds some intuition about the UCM framework.

Time Series Analysis: UCM Goals

Tasks:

- Forecast future Y values. (Interpolate past missing Ys.)
- Determine the nature of the relationship between Y and X1, X2, and so on.
- Decompose Y into some interpretable sub-components such as trend, cycles, seasons, and regression effects. Extrapolate these subcomponents into the future.

3

Time Series Analysis: UCM Goals

Answer questions such as the following:

- Did the behavior of Y qualitatively change at some past time *t*? Do some of the observed Y values look odd?
- Is the seasonal pattern changing over time?
- Did the nature of the relationship between Y and X1 remain stable through the life of the series?
- Is Y increasing or decreasing at a steady rate? If so, at what rate?

4

Unobserved Components Models

Response Time Series = Superposition of components such as Trend, Seasons, Cycles, and Regression effects

- Each component in the model captures some important feature of the series dynamics.
- Components in the model have their own probabilistic models.
- The probabilistic component models include meaningful deterministic patterns as special cases.

5

4.01 Multiple Choice Poll

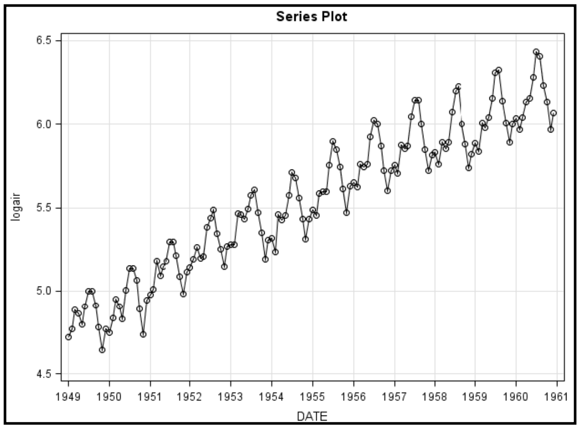
Which time series analysis tool do you use most often?

a. ARIMA modeling
b. exponential smoothing
c. X12 Census Bureau seasonal decomposition
d. spectral analysis
e. state space modeling
f. nonlinear time series modeling
g. other

6

Airline Data Example: *Hello, World!* of the Time Series Modeling

- Log-transformed monthly airline passenger series
- Source: Series G in Box and Jenkins (1976)



7

UCM Model for the Airline Series

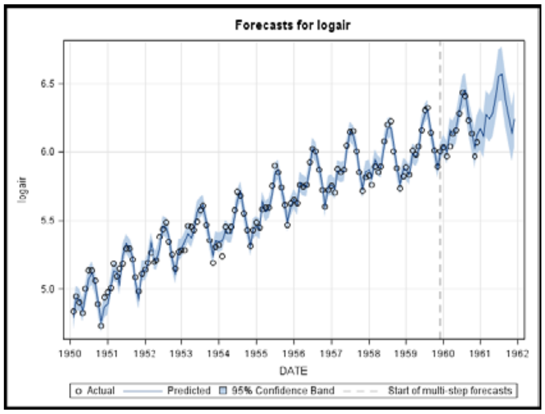
Basic UCM Model:

Logair ~ trend + season + noise

```
proc ucm data=airline;
   model logair;
   irregular;
   level;
   slope var=0 noest;
   season length=12 type=trig;
   estimate back=12;
   forecast back=12 lead=24;
run;
```

8

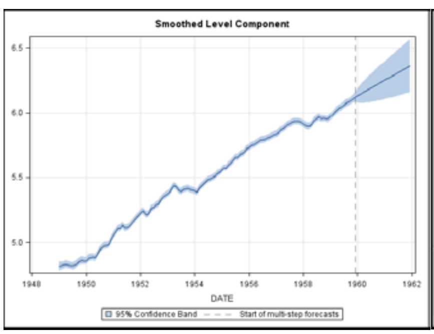
Airline Data: UCM Forecasts

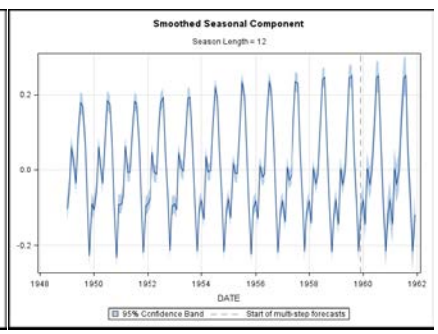


UCM

9

Airline Data: Component Estimates

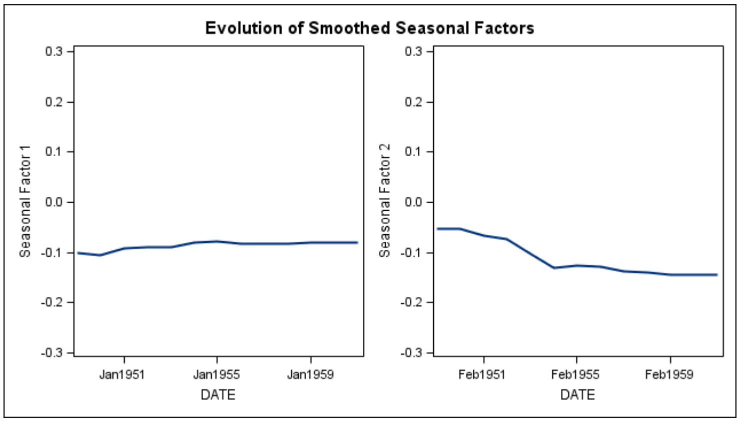




Estimated slope of the trend curve ~ 0.01/month
(in the log scale)

10

Annual Variation in the Seasonal Effects

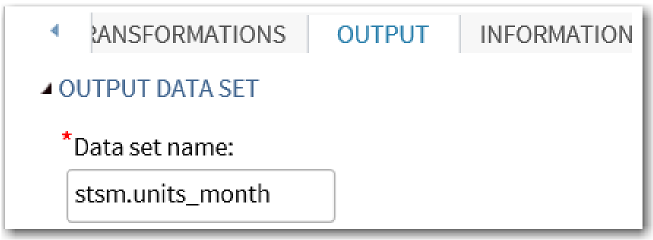


11

Creating the Unit Series on a Monthly Interval Using an Average Accumulation Method

Recall that in Chapter 1, you explored a monthly, intervaled time series that showed evidence of trend and seasonal component variation. This series was created from time-stamped observations on **units** in the **CH1_DEMODAT** table. The first part of this demonstration creates a new table, **stsm.units_month**, that aggregates the **units** time series to a monthly interval using the average aggregation method. A UCM model is then specified and fit to **units**.

1. Create a new Time Series Data Preparation task in SAS Studio.
2. On the DATA tab, set the data option to **STSM.CH1_DEMODAT**. Set the time series variable to **units**. Set the time ID to **date**, and set the interval of the time ID variable to **Month**.
3. Click the **TRANSFORMATIONS** tab, and set the Accumulation method to **Average**.
4. Click the **OUTPUT** tab, and change the output data set name to **units_month** in the **STSM** library.



Alternatively, write the SAS code directly as follows:

```
/* STSM04d01a.sas */
proc timeseries data=stsm.CH1_DEMODAT seasonality=12
                out=stsm.UNITS_MONTH;
   id date interval=month;
   var units / accumulate=average transform=none dif=0 sdif=0;
run;
```

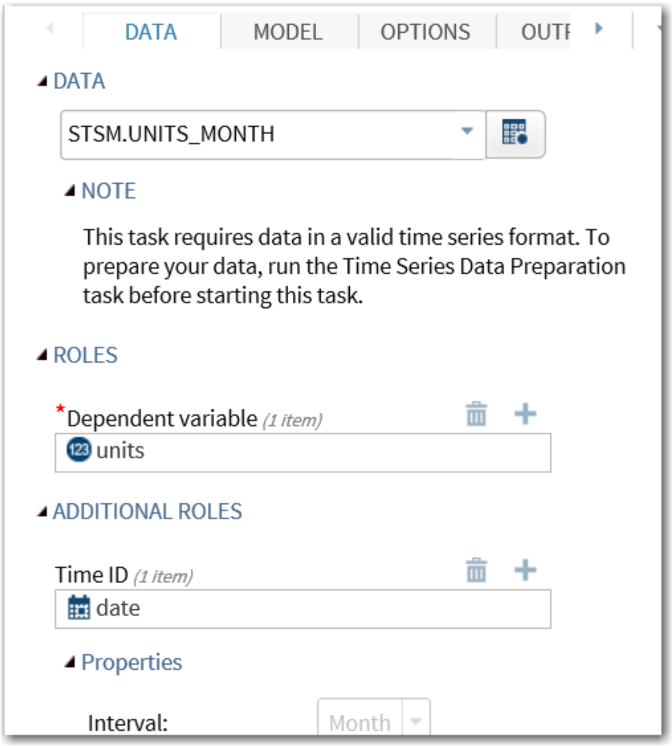
5. Run the task.

End of Demonstration

Specifying an Unobserved Components Model

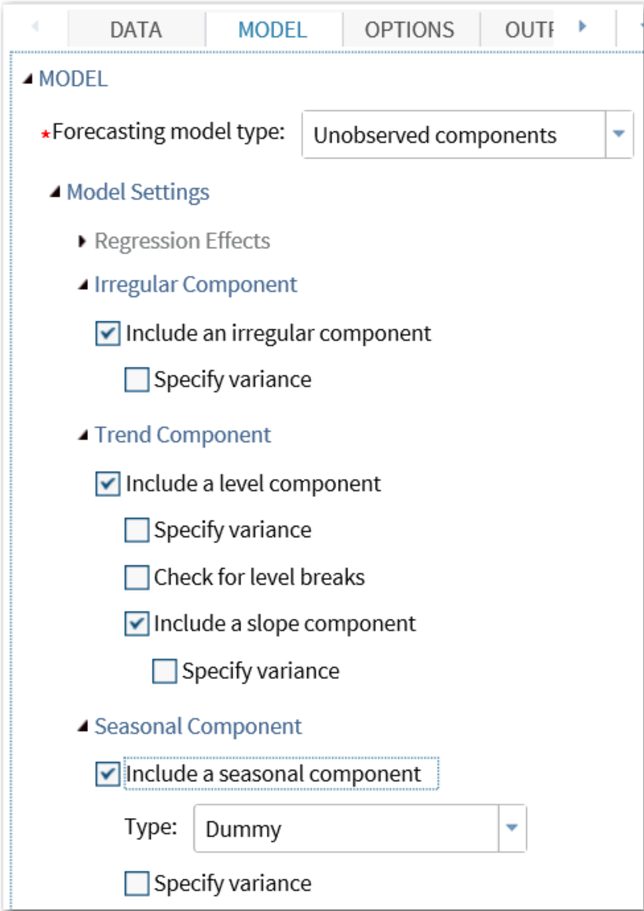
Build a UCM and explore the results.

1. Expand the **Forecasting** tasks, and create a new Modeling and Forecasting task.
2. On the DATA tab, assign table and variable roles as shown below.



3. On the MODEL tab, select **Unobserved Components** as the forecasting model type.
4. Recall that the component analysis from Chapter 1 suggested that the monthly, intervaled **units** data has trend, seasonal, and irregular (ARMA type) patterns. To accommodate these, modify the default model settings by selecting the **Include a slope component** check box. Also, expand the **Seasonal Component** dialog box, and select the **Include a seasonal component** check box. These settings are summarized below.

   The level and slope components combine to accommodate a trend in the model. More information about UCM components is provided later in this chapter.



5. Expand the **Plots** dialog box at the bottom of the MODEL tab, and select **Selected Plots** in the Selected Plots to Display box.

6. In addition to the default plots, select the following: **One Step Ahead Forecasts**, as well as smoothed **Irregular Component**, **Season Component**, **Level Component**, and **Slope Component**.

Alternatively, you can write the SAS code directly as follows:

```
/* STSM04d01b.sas */
/* Build and Explore a UCM */
proc ucm data=stsm.UNITS_MONTH;
   id date interval=month;
   model units;
   irregular plot=smooth;
   level plot=(smooth);
   slope plot=(smooth);
   season length=12 type=dummy plot=(smooth);
   estimate plot=(panel model loess);
   forecast lead=12 back=0 alpha=0.05 plot=(forecasts);
   outlier;
run;
```

Recall that the component analysis from Chapter 1 suggested that the monthly intervaled **units** data has trend, seasonal, and irregular (ARMA type) patterns. The UCM procedure syntax below accommodates these components.

The level and slope components combine to accommodate a trend in the model. More information about UCM components is provided later in this chapter.

Additional plot options were added to the component statements:

- The PLOT=(SMOOTH) options produce smoothed representations of the level, slope, and season components.
- The PLOT=(PANEL, MODEL, and LOESS) options in the ESTIMATE statement produce residuals diagnostics plots.
- The PLOT=(FORECASTS) option in the FORECAST statement produces a plot of historical and lead forecasted values.

7. Select **Run** to submit the generated UCM syntax.

   The Significance Analysis of Components table indicates that the data contain a significant level, slope, and seasonal component. However, the irregular component seems to be negligible in the presence of other components in the model. The slope estimate suggests that the data is increasing by approximately 14 units per month.

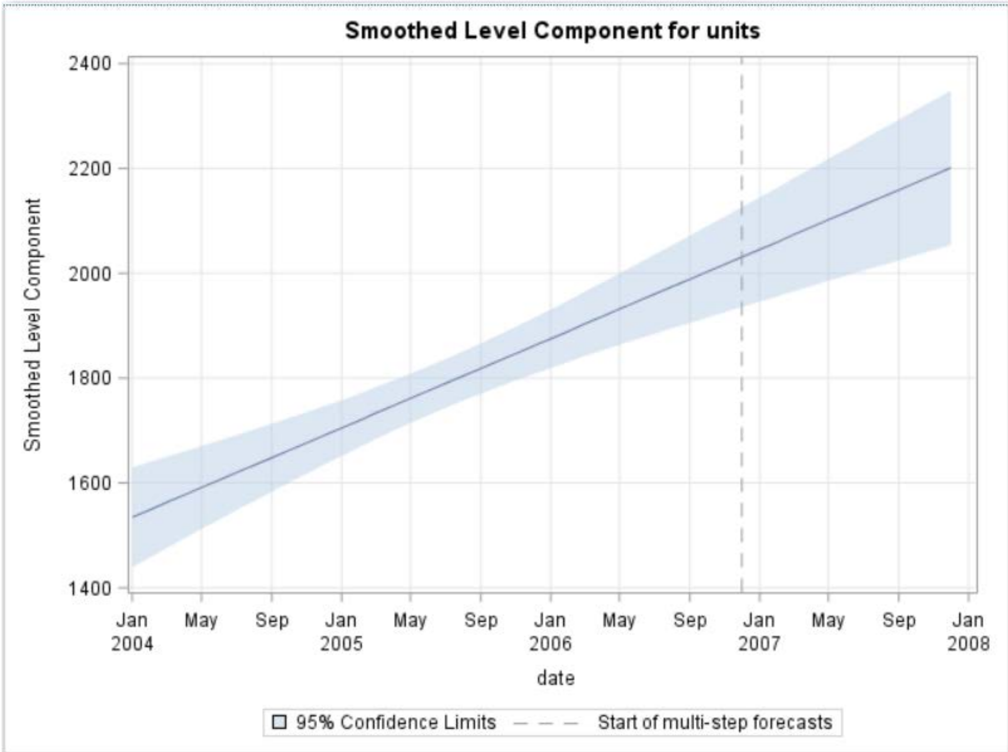
Significance Analysis of Components (Based on the Final State)

Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.00	0.9967
Level	1	1748.27	<.0001
Slope	1	34.28	<.0001
Season	11	26.55	0.0054

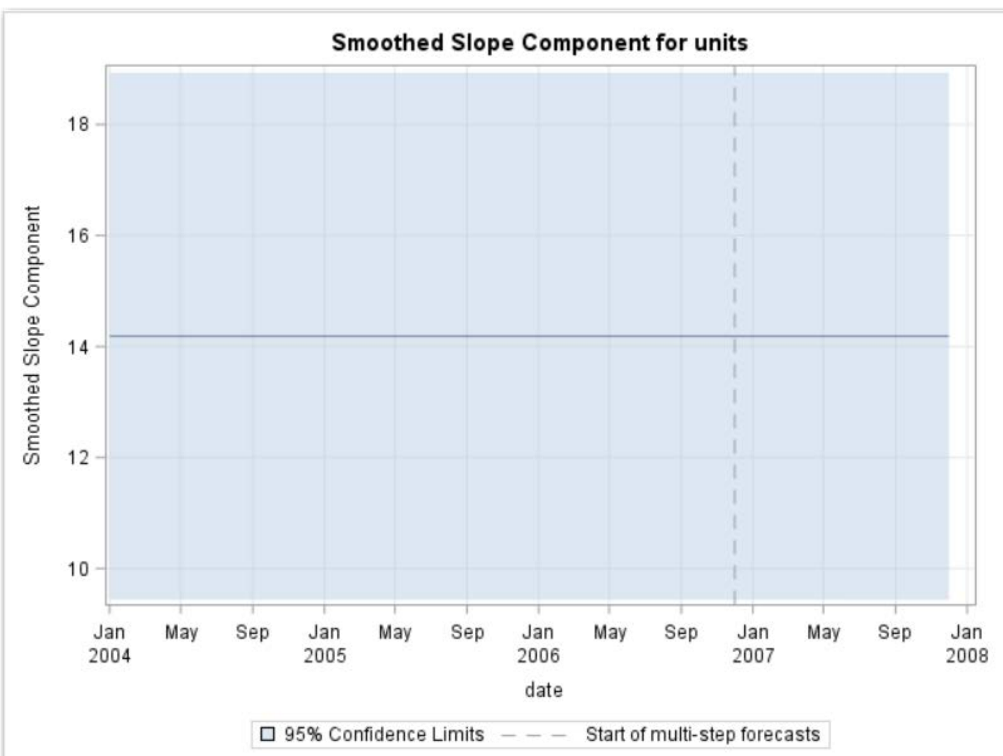
Trend Information (Based on the Final State)

Name	Estimate	Standard Error
Level	2031.166251	48.578223
Slope	14.18644584	2.4228224

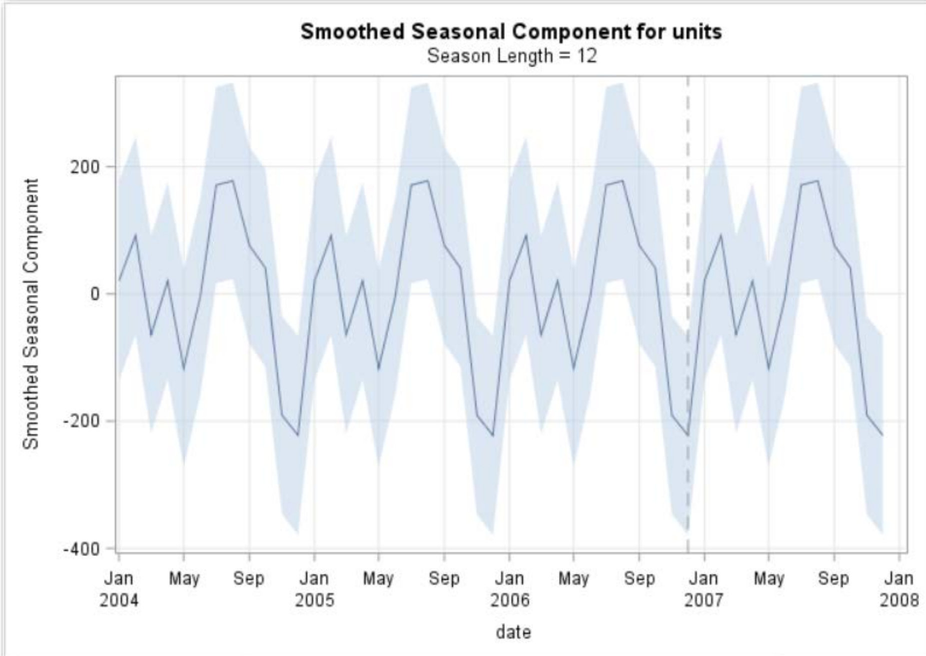
The Smoothed Level Component plot shows how the level of the data is estimated to evolve over the history and future forecast horizon.



The Smoothed Slope Component plot indicates that the slope in the data is a constant, and that the trend component is basically a deterministic linear trend.

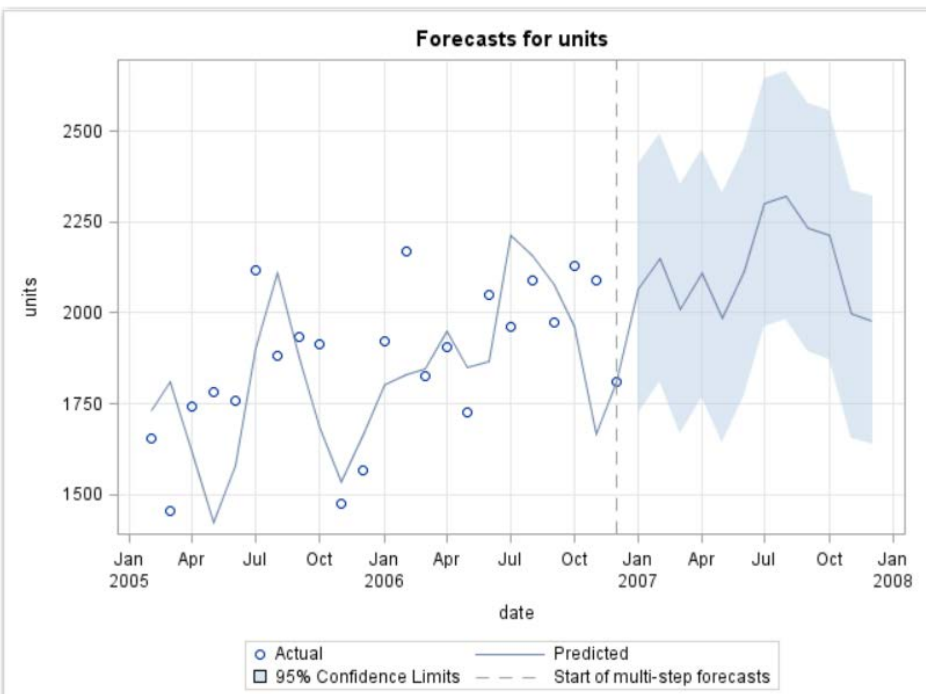


The Smoothed Seasonal Component plot shows in the sample and lead forecasts from the seasonal component model.



Smoothed component plots are based on parameter estimates derived from all of the observations in the data.

The forecast plot shows the extrapolated trend, level, and seasonal components in the lead forecast. The forecast is the sum of forecasts from the estimated components listed in the table above.



End of Demonstration

4.2 Unobserved Components Models

The Nature of Components

The examples shown previously share some commonly observed time series data qualities.

- If you call trend the long-term, slowly varying pattern of the series, it rarely has a definite shape, such as linear or quadratic or some other simple parametric curve.
- The periodic patterns exhibited by the series also rarely preserve their properties over the life of the series.
- Therefore, if the observed series is to be modeled as a sum of components, then these patterns must be flexible and adaptive.

15

Trend Component Example

Two models for trend:

- The *random walk trend* (RW) represents a slowly varying level without a drift in any particular direction.
- The *local linear trend* (LL) represents a locally linear pattern with slowly varying intercept and slope.

16

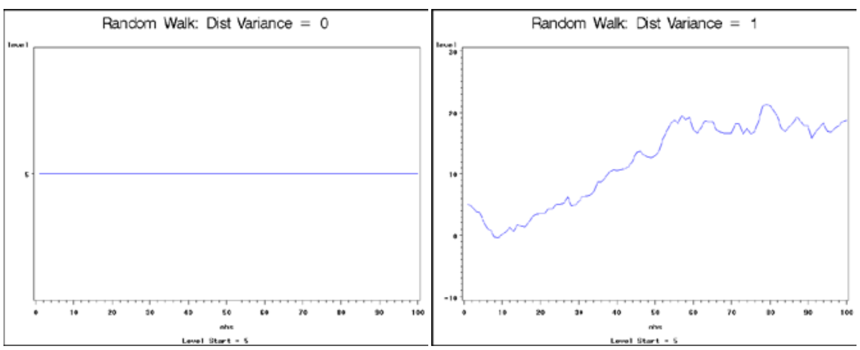
Additional trend specifications, such as trend specified using differencing, can also be considered.

Random Walk Trend

$$\mu_t = \mu_{t-1} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, \sigma^2_{\mu})$$

17

Random Walk Simulation



18

Local Linear Trend

Deterministic linear trend:

$$\mu_t = \mu_0 + \beta_0 * t$$

Recursive form:

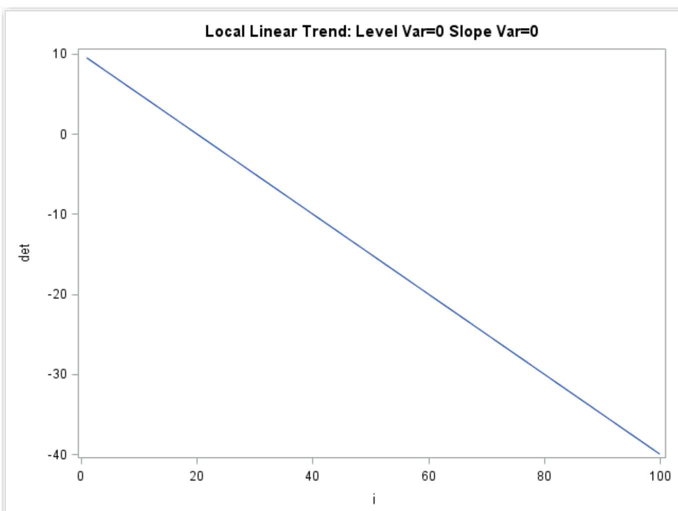
$$\mu_t = \mu_{t-1} + \beta_{t-1}$$
$$\beta_t = \beta_{t-1}$$

Local linear trend:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, \sigma^2_{\mu})$$
$$\beta_t = \beta_{t-1} + \xi_t \qquad \xi_t \sim \mathrm{N}(0, \sigma^2_{\beta})$$

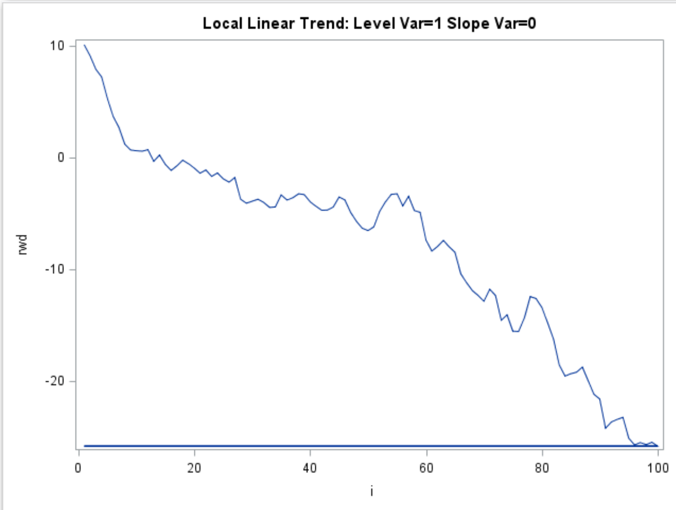
19

Local Linear Trend Simulation



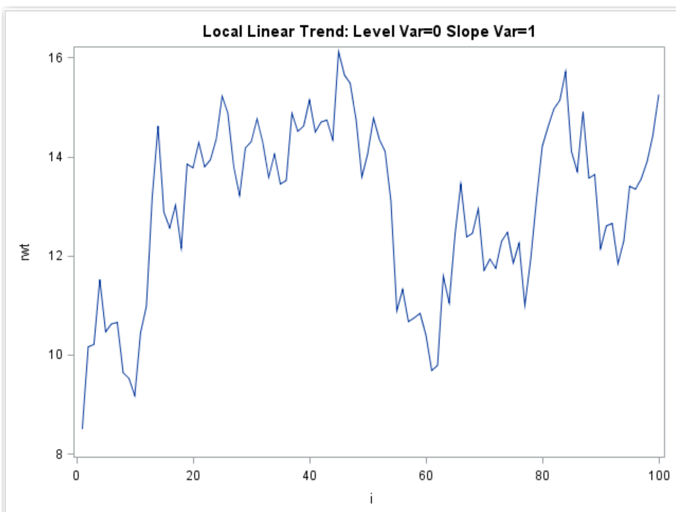
20

Local Linear Trend Simulation



21

Local Linear Trend Simulation



22

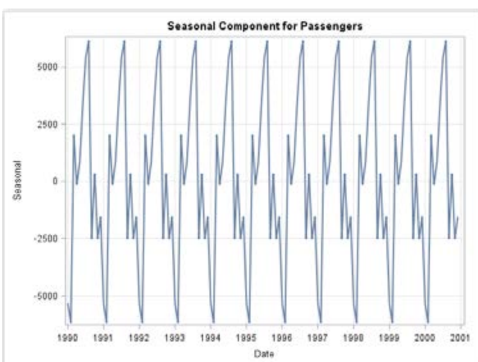
Season Component Example

1. The seasonal fluctuations are a common source of variation in the time series data.
2. The seasonal effects are regarded as corrections to the general trend of the series due to seasonal variations, and these effects sum to zero when summed over the full season cycle.
3. Therefore, a (deterministic) seasonal component γ_t is modeled as a periodic pattern of an integer period s so that the sum is as follows:

$$\sum_{i=0}^{s-1} \gamma_{t-i} = 0$$

23

Example of a (Deterministic) Seasonal Pattern (Period=12)



Seasonal Index	Seasonal Component
1	-5338.903472
2	-6177.449306
3	2037.0590278
4	-128.6951389
5	882.51736111
6	3367.1923611
7	5416.2756944
8	6168.7340278
9	-2525.282639
10	344.73819444
11	-2519.265972
12	-1526.920139

24

Stochastic Seasonal: Dummy Type

$$\sum_{i=0}^{s-1} \gamma_{t-i} = \omega_t, \qquad \omega_t \sim i.i.d.\, N\left(0, \sigma_{\varpi}^2\right)$$

1. The periodic pattern sums to zero **in the mean**.

2. The disturbance variance controls the variation in the seasons. If it is zero, the model reduces to a deterministic seasonal. This is equivalent to having (*s*-1) dummy regressors.

25

A General UCM

A general UCM can be described as follows:

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

$$\varepsilon_t \sim i.i.d.\, N\left(0, \sigma_{\varepsilon}^2\right)$$

- $\varepsilon_t, \mu_t, \gamma_t, \psi_t$, and r_t represent different stochastic components.
- The model can contain multiple seasons and cycles.
- The term $\sum_{j=1}^{m} \beta_j x_{jt}$ represents the effects of predictors.
- The term $\sum_{i=1}^{p} \phi_i y_{t-i}$ is a regression term involving the lags of the dependent variable.

26

Model Specification Syntax

A UCM is specified by describing the components in the model. For example, consider the following model:

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

It consists of the LL trend μ_t, monthly trigonometric season γ_t, and an irregular component ε_t. The corresponding syntax is as follows:

```
MODEL y;
    IRREGULAR;
    LEVEL;
    SLOPE;
    SEASON LENGTH=12 TYPE=TRIG;
```

27

continued...

A General Model Building Approach

A general modeling approach can be described as follows:

- Identify systematic components of variation in the data.
- Specify a general UCM that accommodates these components.
- Identify components that are non-stochastic. The variance of these components can be fixed at 0.
- Identify components that are not significant in explaining variation in the target. These components are candidates for removal from the model.

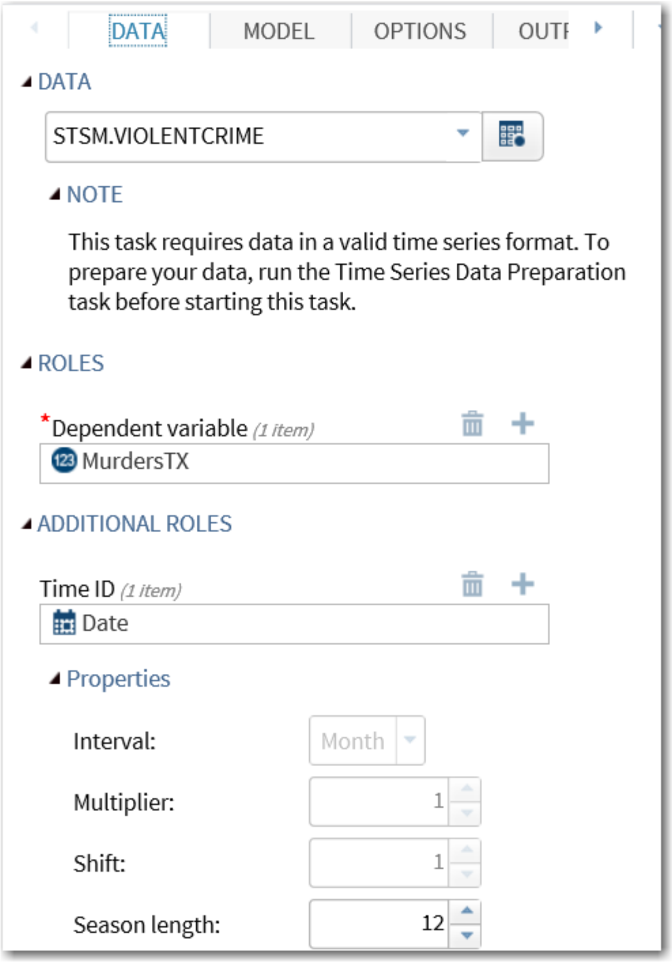
28

Refining an Unobserved Components Model

The time series **MurdersTX** in the **STSM.VIOLENTCRIME** data set was explored in an exercise in Chapter 1. A reasonable starting hypothesis is that the data contain trend (level + slope), seasonal, and irregular components.

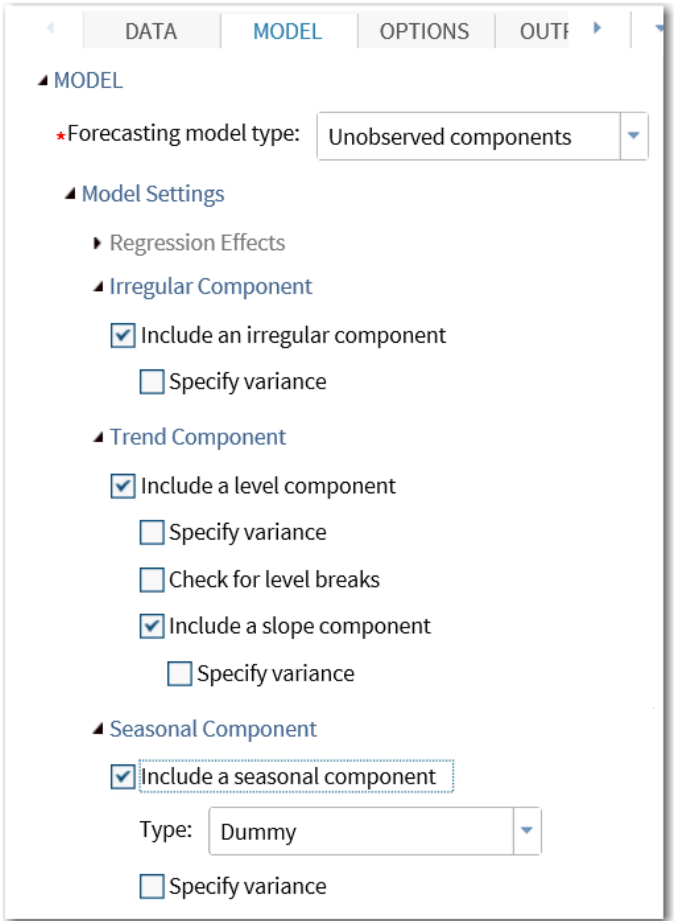
1. Specify a baseline UCM that accommodates the hypothesized components.

   a. Create a new Modeling and Forecasting task, and assign table and variable roles as shown below.



   b. Click the **MODEL** tab, and select **Unobserved components** as the forecasting model type.

   c. Specify that the UCM should contain *irregular*, *level*, *slope*, and *seasonal* components.



Alternatively, you can write the SAS code directly as follows:

```
/* STSM04d02.sas */
/* Specify the Baseline model */
proc ucm data=stsm.VIOLENTCRIME;
   id Date interval=month;
   model MurdersTX;
   irregular;
   level;
   slope;
   season length=12 type=dummy;
   forecast lead=12 back=0 alpha=0.05;
   outlier;
run;
```

d. Run the task to submit the generated code.

This model can be considered the baseline to judge model refinements. The initial fit statistics are shown below.

Likelihood Based Fit Statistics	
Statistic	**Value**
Full Log Likelihood	-431.5
Diffuse Part of Log Likelihood	-4.97
Non-Missing Observations Used	108
Estimated Parameters	4
Initialized Diffuse State Elements	13
Normalized Residual Sum of Squares	95
AIC (smaller is better)	870.96
BIC (smaller is better)	881.17
AICC (smaller is better)	871.4
HQIC (smaller is better)	875.08
CAIC (smaller is better)	885.17

The final estimates of the variances associated with each component indicate that only the Irregular component is stochastic.

Final Estimates of the Free Parameters					
Component	**Parameter**	**Estimate**	**Approx Std Error**	**t Value**	**Approx Pr > |t|**
Irregular	**Error Variance**	239.38162	48.55214	4.93	<.0001
Level	**Error Variance**	25.44934	18.72870	1.36	0.1742
Slope	**Error Variance**	0.08117	0.13315	0.61	0.5421
Season	**Error Variance**	1.87519	5.64980	0.33	0.7400

The Significance Analysis of Components table indicates that only the level and season components explain a substantial proportion of the variation in **MurdersTX**, in the presence of other components in the model.

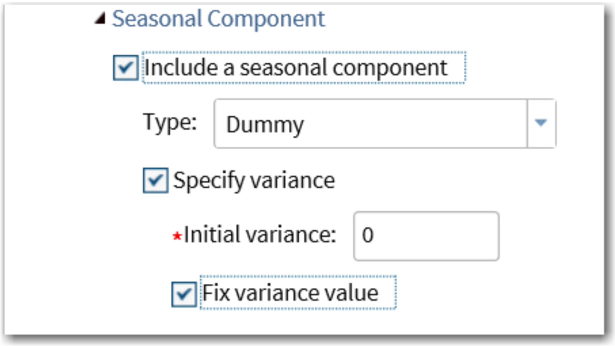
Significance Analysis of Components (Based on the Final State)			
Component	**DF**	**Chi-Square**	**Pr > ChiSq**
Irregular	1	0.55	0.4590
Level	1	133.11	<.0001
Slope	1	0.99	0.3202
Season	11	56.83	<.0001

2. Refine the baseline UCM. Fix the seasonal variance at 0.

   Based on the estimates of component variance, the season component is the most deterministic. Model refinement begins by fixing the variance of the component at 0.

   a. Select the **Specify variance** and **Fix variance value** check boxes in the Seasonal Component options area.

b. Verify that 0 is in the **Initial variance** field.



Alternatively, you can write the SAS code directly as follows:

```
/* Fix the Season component variance at 0 */
proc ucm data=stsm.VIOLENTCRIME;
   id Date interval=month;
   model MurdersTX;
   irregular;
   level;
   slope;
   season length=12 type=dummy variance=0 noest;
   forecast lead=12 back=0 alpha=0.05;
   outlier;
run;
```

Model refinement begins by fixing the variance of the season component at 0 using the VARIANCE and NOEST options as shown.

c. Select **Run** to fit the re-specified model.

The Fit Statistics table indicates that the penalized, overall fit of the model is better than the baseline.

Likelihood Based Fit Statistics	
Statistic	**Value**
Full Log Likelihood	-431.5
Diffuse Part of Log Likelihood	-4.97
Non-Missing Observations Used	108
Estimated Parameters	3
Initialized Diffuse State Elements	13
Normalized Residual Sum of Squares	95
AIC (smaller is better)	869.09
BIC (smaller is better)	876.75
AICC (smaller is better)	869.35
HQIC (smaller is better)	872.18
CAIC (smaller is better)	879.75

The slope and level components are still deterministic in the re-specified model.

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	247.85482	43.80894	5.66	<.0001
Level	Error Variance	24.50078	18.36962	1.33	0.1823
Slope	Error Variance	0.08448	0.13614	0.62	0.5349

The relative importance of the components in explaining variation in **MurdersTX** did not change from the baseline.

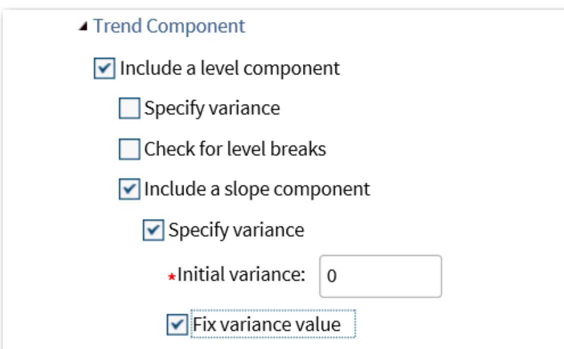
Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.62	0.4313
Level	1	132.54	<.0001
Slope	1	0.98	0.3227
Season	11	62.62	<.0001

3. Refine the UCM. Fix the slope variance at 0.

   Because the slope is indicated to be deterministic, the next step fixes the slope component variance at 0.

   a. In the Trend Component section, beneath the Include a slope component check box, select the **Specify variance** and **Fix variance value** check boxes in the Trend Component options area.

   b. Verify that 0 is in the **Initial variance** field.



Alternatively, you can write the code directly as follows:

```
/* Fix the Slope component variance at 0 */
proc ucm data=stsm.VIOLENTCRIME;
   id Date interval=month;
   model MurdersTX;
   irregular;
   level;
   slope variance=0 noest;
   season length=12 type=dummy variance=0 noest;
   forecast lead=12 back=0 alpha=0.05;
   outlier;
run;
```

c. Select **Run** to fit the re-specified model.

The Fit Statistics table indicates that the penalized, overall fit of the model is slightly better.

Likelihood Based Fit Statistics	
Statistic	**Value**
Full Log Likelihood	-432.1
Diffuse Part of Log Likelihood	-4.97
Non-Missing Observations Used	108
Estimated Parameters	2
Initialized Diffuse State Elements	13
Normalized Residual Sum of Squares	95
AIC (smaller is better)	868.26
BIC (smaller is better)	873.37
AICC (smaller is better)	868.39
HQIC (smaller is better)	870.32
CAIC (smaller is better)	875.37

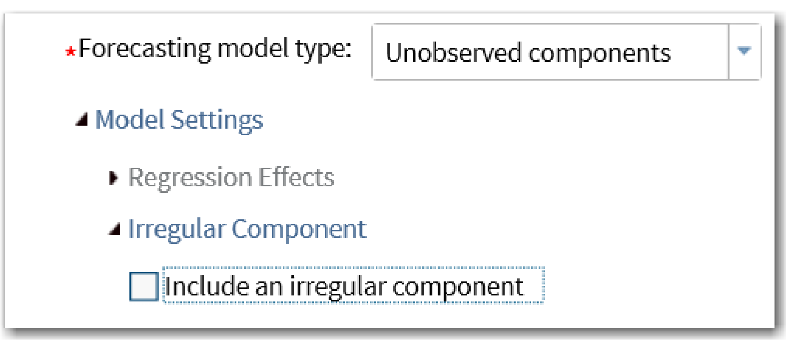
The Significance Analysis of Components table indicates that the irregular component is the least useful in terms of accounting for variation in **MurdersTX**.

Significance Analysis of Components (Based on the Final State)			
Component	**DF**	**Chi-Square**	**Pr > ChiSq**
Irregular	1	0.37	0.5442
Level	1	128.47	<.0001
Slope	1	1.01	0.3140
Season	11	59.03	<.0001

4. Refine the UCM. Remove the irregular component.

The irregular component is dropped from the model.

a. Clear the **Include an irregular component** check box.



Alternatively, you can write the code directly as follows:

```
/* Remove the Irregular component */
proc ucm data=stsm.VIOLENTCRIME;
   id Date interval=month;
   model MurdersTX;
   level;
   slope variance=0 noest;
   season length=12 type=dummy variance=0 noest;
   forecast lead=12 back=0 alpha=0.05;
   outlier;
run;
```

The IRREGULAR statement was removed.

b. Select **Run** to fit the re-specified model.

The Fit Statistics table indicates that the penalized, overall fit of the model is substantially worse after removing the irregular component.

Likelihood Based Fit Statistics	
Statistic	**Value**
Full Log Likelihood	-452.4
Diffuse Part of Log Likelihood	-4.97
Non-Missing Observations Used	108
Estimated Parameters	1
Initialized Diffuse State Elements	13
Normalized Residual Sum of Squares	95
AIC (smaller is better)	906.87
BIC (smaller is better)	909.42
AICC (smaller is better)	906.91
HQIC (smaller is better)	907.9
CAIC (smaller is better)	910.42

5. The final model, assessed below, is the one fit in the next-to-the-last step.

a. Select the **Irregular** component on the MODEL tab.

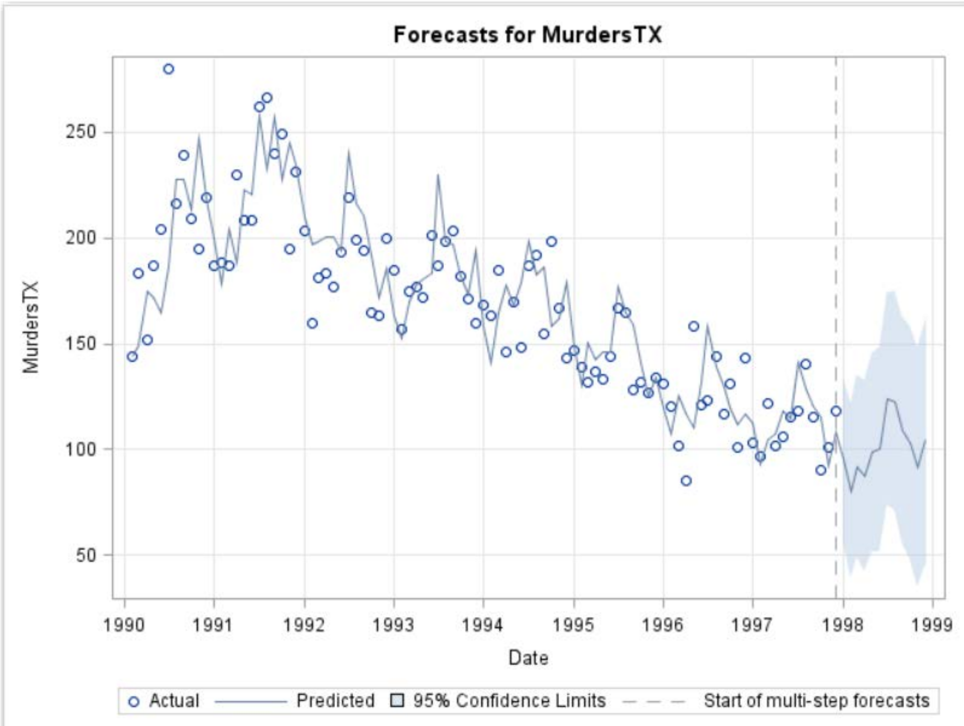
b. In the Plots options on the bottom of the MODEL tab, select **One-step-ahead Forecasts** and all of the **Smoothed Component Estimates** plots.

The final model syntax is shown below.

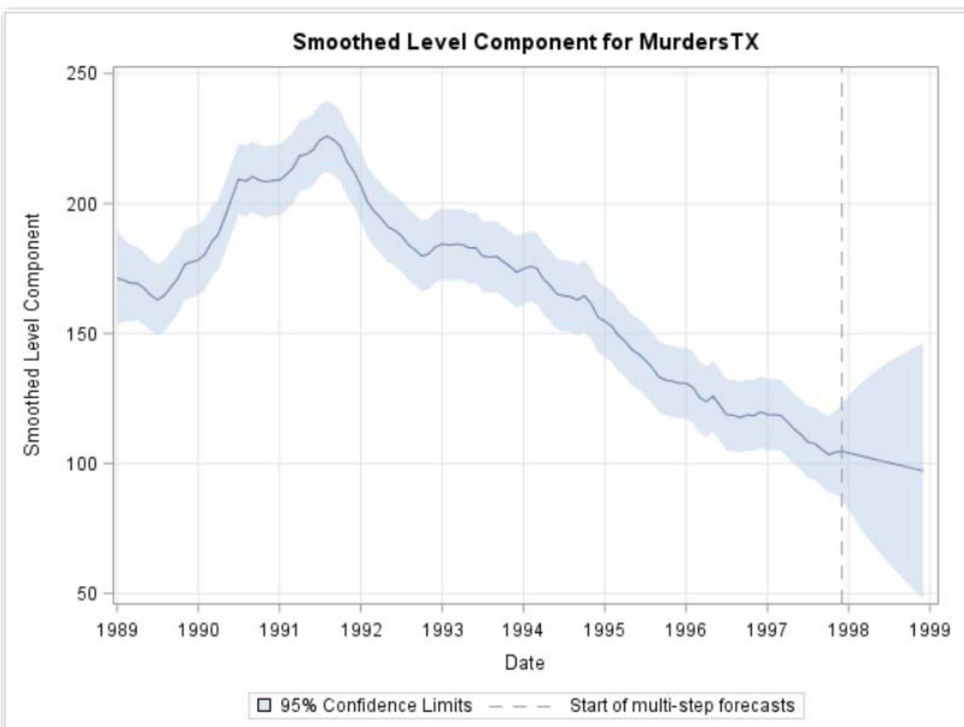
```
/* Add the Irregular component back in to get the Final model */
proc ucm data=stsm.VIOLENTCRIME;
   id Date interval=month;
   model MurdersTX;
   irregular plot=smooth;
   level plot=(smooth);
   slope variance=0 noest;
   season length=12 type=dummy variance=0 noest;
   estimate plot=(panel model loess);
   forecast lead=12 back=0 alpha=0.05 plot=(forecasts);
   outlier;
run;
```

c. Select **Run** to submit the final model.

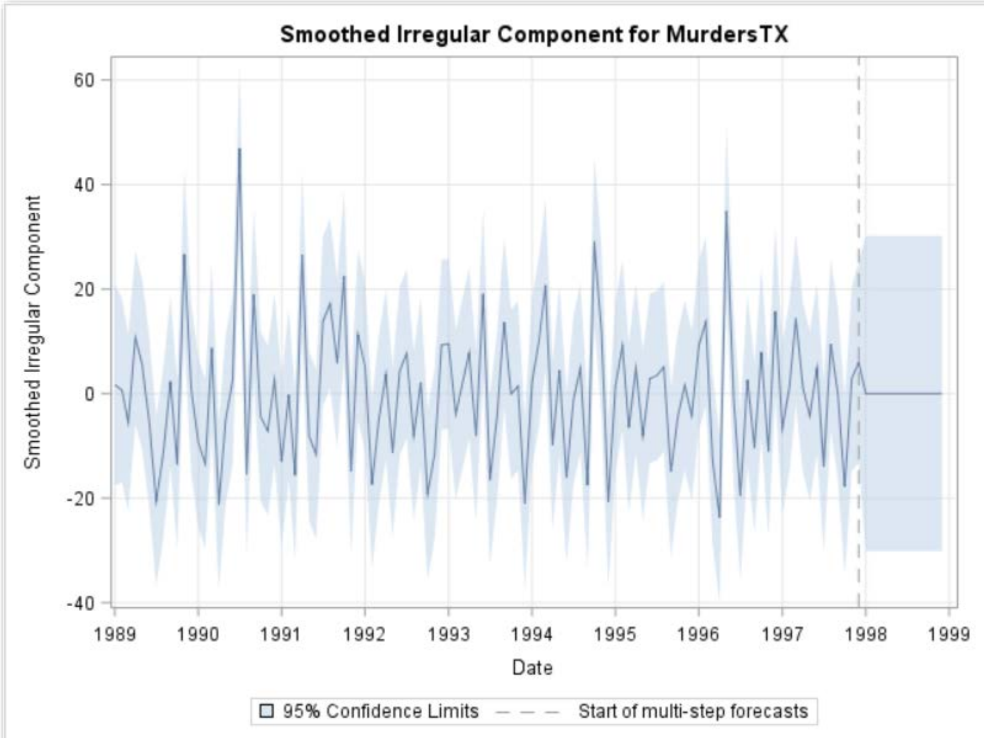
The overall model does a good job of accommodating and extrapolating the salient features of the data.



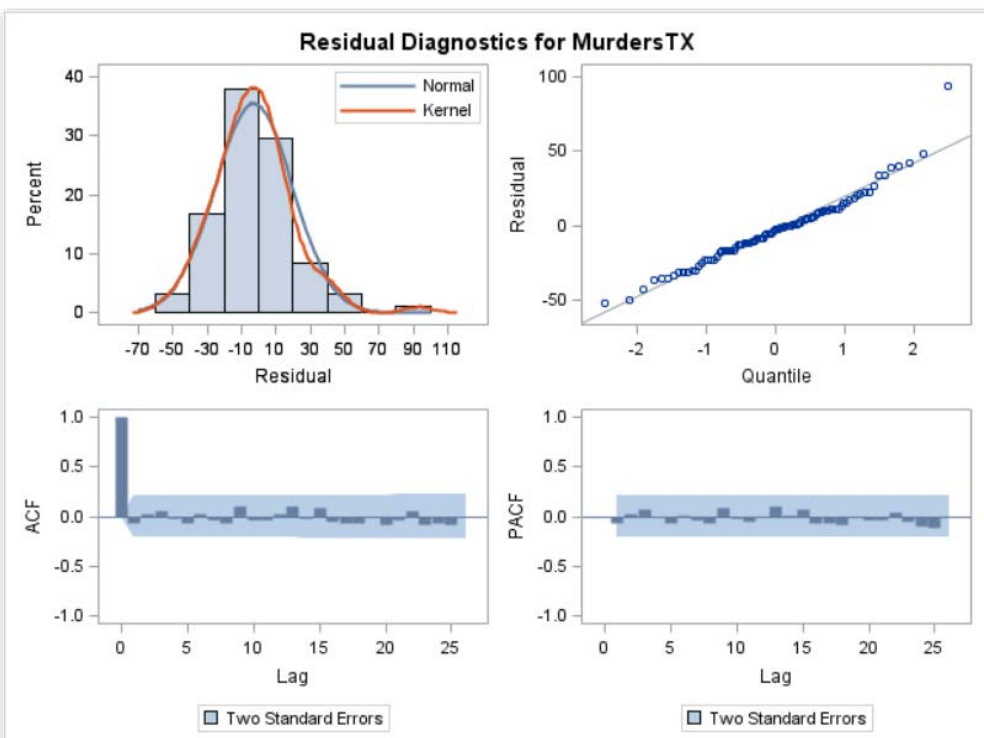
The Level Component plot illustrates how the level of the series changes as a function of time.



The Irregular Component plot represents the estimated, stationary variation in the series.



The Residual Diagnostics panel indicates that the model is adequately specified.



End of Demonstration