

Chapter 469

Decomposition Forecasting

Introduction

Classical time series decomposition separates a time series into five components: mean, long-range trend, seasonality, cycle, and randomness. The decomposition model is

$$\text{Value} = (\text{Mean}) \times (\text{Trend}) \times (\text{Seasonality}) \times (\text{Cycle}) \times (\text{Random}).$$

Note that this model is multiplicative rather than additive. Although additive models are more popular in other areas of statistics, forecasters have found that the multiplicative model fits a wider range of forecasting situations.

Decomposition is popular among forecasters because it is easy to understand (and explain to others). While complex ARIMA models are often popular among statisticians, they are not as well accepted among forecasting practitioners. For seasonal (monthly, weekly, or quarterly) data, decomposition methods are often as accurate as the ARIMA methods and they provide additional information about the trend and cycle which may not be available in ARIMA methods.

Decomposition has one disadvantage: the cycle component must be input by the forecaster since it is not estimated by the algorithm. You can get around this by ignoring the cycle, or by assuming a constant value. Some forecasters consider this a strength because it allows the forecaster to enter information about the current business cycle into the forecast.

Decomposition Method

The basic decomposition method consists of estimating the five components of the model

$$X_t = UT_t C_t S_t R_t$$

where

X_t denotes the series or, optionally, log of series.

U denotes the mean of the series.

T_t denotes the linear trend.

C_t denotes cycle.

S_t denotes season.

R_t denotes random error.

t denotes the time period.

We will now take you through the steps used by the program to perform a decomposition of a time series. Most of this information is from Makridakis (1978), chapter 15.

Step 1 – Remove the Mean

The first step is to remove the mean by dividing each individual value by the series mean. This creates a new series that has values near one. This step is represented symbolically as

$$Y_t = X_t / U$$

If the absolute value of the mean of the series is less than 0.0000001, no division takes place.

Step 2 – Calculate a Moving Average

The next step calculates an L -step moving average centered at the time period, t , where L is the length of the seasonality (e.g., L would be 12 for a monthly series or 4 for quarterly series). Since the moving average gives the mean of a year's data, the seasonality factor is removed. Usually, the averaging removes the randomness component as well. Symbolically, this step is represented as

$$M_t = \sum Y_t$$

where for odd L , the summation runs from $t-[L/2]$ to $t+[L/2]$. The symbols $[x]$ mean take the integer part of x . Hence $[6.43]=6$ and $[11/2]=5$. Notice that this summation range centers the moving average at t .

For even L , the values usually found in practice (2, 4 and 12), it is a little more difficult to center the moving average on the time period t . For example, the average of the first 12 terms of a series would be centered at 6.5 rather than 6. To center the average right on 7, we must compute the moving average centered at 6.5 and at 7.5 and then average these. The resulting double moving average is centered at the desired value of 7.

Another complexity that must be dealt with is what to do at the ends of the series. Because the average is centered, the first and last $L/2$ averages cannot be computed (because of the lack of data). Many different end-effect techniques have been proposed.

Our end-effect strategy can best be explained by considering an example. Suppose we have a monthly series that runs from January of 1980 to December of 1988. To compute the moving average centered at January, 1980, we will need estimated data back through July, 1979. The estimate of July 1979 is obtained by subtracting the difference of July, 1980 and July, 1981 from July, 1980.

At the other end of the series we will need estimated values through June, 1989. To compute the estimated value for June, 1989, we add the difference of June, 1987 and June, 1988 to June, 1988.

This method of estimating end-effects preserves local trends in the series. However, it is especially sensitive to outliers. You should remember that strange patterns in the last $L/2$ time periods may be from the end-effect calculation and not from a pattern in the series itself.

Step 3 – Calculate the Trend

The next step is to calculate and remove the trend component of the series. This calculation is made on the moving averages, M_t , rather than on the Y_t series. A least squares fit is made of the of the model

$$M_t = a + bt + e_t$$

where

- a is the intercept.
- b is the slope.
- e_t is the residual or lack-of-linear-fit.

Decomposition Forecasting

The linear portion of the above model is used to define the trend. That is, we use

$$T_t = a + bt$$

Note that because of the problems of end-effects, the first and last $L/2$ terms are omitted in the trend calculation.

Step 4 – Calculate the Cycle

The cycle term is found by dividing the moving average by the computed trend. Symbolically, this is

$$C_t = \frac{M_t}{T_t}$$

Step 5 – Calculate the Seasonality

The seasonality is computed by dividing the Y series by the moving averages. Symbolically, this is

$$K_t = \frac{Y_t}{M_t}$$

Note that the K series is composed of both the seasonality and the randomness. To calculate the seasonal component for each season, we simple average all like seasons. That is, the average of all Januarys is computed, the average of all Februarys gives the seasonal value for February, and so on. Mathematically, this is stated as

$$S_g = \sum K_t$$

where the summation is over all t in which the season is g .

Step 6 – Calculate the Randomness

The final step is to calculate the randomness component. This is accomplished by dividing the K series by S_i where the values of S_1, S_2, \dots, S_g are repeated as needed. This is represented mathematically as follows

$$R_t = \frac{K_t}{S_t}$$

Creating Forecasts

Once the series decomposition is complete, forecasts may be generated fairly easily. The trend component is calculated using

$$T_t = a + bt ,$$

the seasonal factor is read from

$$S_g = \sum K_t ,$$

the cycle factor is input by hand, and the random factor is assumed to be one. If the series was transformed using the log transformation, the forecasts are transformed back using the appropriate inverse function.

Assumptions and Limitations

These algorithms are useful for forecasting seasonal time series with (local or global) trend.

Data Structure

The data are entered in a single column.

This section describes the options available in this procedure.

Missing Values

When missing values are found in the series, they are either replaced or omitted. The replacement value is the average of the nearest observation in the future and in the past or the nearest non-missing value in the past.

If you do not feel that this is a valid estimate of the missing value, you should manually enter a more reasonable estimate before using the algorithm. These missing value replacement methods are particularly poor for seasonal data. We recommend that you replace missing values manually before using the algorithm.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

Specify the variable(s) on which to run the analysis.

Time Series Variable

Time Series Variable

Specify the variable(s) on which to run the analysis. A separate analysis will be conducted for each variable listed.

Use Logarithms

Specifies that the log (base 10) transformation should be applied to the values of the variable. The forecasts are converted back to the original metric before display.

Missing Values

Choose how missing (blank) values are processed.

The algorithm used in this procedure cannot tolerate missing values since each row is assumed to represent the next point in a time sequence. Hence, when missing values are found, they must be removed either by imputation (filling in with a reasonable value) or by skipping the row and pretending it does not exist.

Whenever possible, we recommend that you replace missing values manually.

Here are the available options.

Average the Adjacent Values

Replace the missing value with the average of the nearest values in the future (below) and in the past (above).

Carry the Previous Value Forward

Replace the missing value with the first non-missing value immediately above (previous) this value.

Decomposition Forecasting

Omit Row from Calculations

Ignore the row in all calculations. Analyze the data as if the row was not on the database.

Cycle-Input Variable

Cycle-Input Variable

Specifies the name of a variable containing estimated values of the cycle component. If this option is left blank, a value of one is used for all future cycle components. This ignores the cycle in the forecasts.

The program ignores the first n rows (where n is the number of rows in the original series) and begins reading the cycle ratios in the rows that match the forecast rows. For example, suppose you have a series with 48 rows over data and you want to forecast the next 12 rows. The first 48 rows of this variable are ignored. The 49th through 60th rows are used to provide the cyclical component.

Note that since the forecasting model is multiplicative, if you enter a '1' for the cycle, the forecast will not be changed. If you feel the cycle influence will be a 5% increase, you would enter '1.05' for this time period.

Use Cycle

This option specifies whether a cycle component is found or ignored.

Forecasting Options

Number of Forecasts

This option specifies the number of forecasts to be generated.

Seasonality Options

Number of Seasons

Specify the number of seasons in the series. Use '4' for quarterly data or '12' for monthly data.

First Season

Specify the first season of the series. This value is used to format the reports and plots. For example, if you have monthly data beginning with March, you would enter a '3' here.

First Year

Specify the first year of the series. This value is used to format the reports and plots.

Reports Tab

The following options control which reports are displayed.

Select Reports

Summary Report

This option specifies whether the indicated report is displayed.

Forecast Report

This option specifies which parts of the series are listed on the numeric reports: the original data and forecasts, just the forecasts, or neither.

Decomposition Forecasting

Report Options

Precision

Specify the precision of numbers in the report. Single precision will display seven-place accuracy, while the double precision will display thirteen-place accuracy. Note that all reports are formatted for single precision only.

Variable Names

Specify whether to use variable names or (the longer) variable labels in report headings.

Page Title

Specify a title to be shown at the top of the reports.

Plots Tab

This section controls the inclusion and the settings of the plots.

Select Plots

Forecast Plot- Decomposition Ratio Plots

These options control whether each plot is displayed. Click the plot format button to change the plot settings.

Horizontal Axis Variable

Horizontal Variable

This option controls the spacing on the horizontal axis when missing or filtered values occur.

Your choices are

Actual Row Number

Use the actual row number of each row from the dataset along the horizontal axis.

Constructed Date

Construct a date value from the sequence (relative row) number and the *Seasonality Options* settings. Any missing or filtered values are skipped when forming the sequence number.

Storage Tab

Several statistics, including the forecasts and residuals, may be stored on the current dataset for further analysis. This group of options lets you designate which statistics (if any) should be stored and which columns should receive these statistics.

Note that existing data is replaced. Be careful that you do not specify columns that contain important data.

Data Storage Columns

Forecasts

The forecasts are stored in this column.

Residuals

The residuals are stored in this column.

Decomposition Forecasting

Trend Ratios

The trend ratios, the T_t series, are stored in this column.

Cycle Ratios

The cycle ratios, the C_t series, are stored in this column.

Season Ratios

The season ratios, the S_t series, are stored in this column.

Error Ratios

The error, or random, ratios, the R_t series, are stored in this column.

Example 1 – Decompositions Forecasting

This section presents an example of how to generate forecasts of a series using the time series decomposition forecasting method. The data in the Sales dataset will be used. We will forecast the values of the Sales variable for the next twelve months.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Decomposition Forecasting window.

1 Open the Sales dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Sales.NCSS**.
- Click **Open**.

2 Open the Decomposition Forecasting window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Decomposition Forecasting** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- On the Decomposition Forecasting window, select the **Variables tab**.
- Double-click in the **Time Series Variable** box. This will bring up the variable selection window.
- Select **Sales** from the list of variables and then click **Ok**.
- Enter **1970** in the **First Year** box.

4 Specify the Reports.

- On the Decomposition Forecasting window, select the **Reports tab**.
- Set **Forecast Report** to **Data and Forecasts**.

5 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Forecast Summary Section

Forecast Summary Section

Forecast	(Mean) x (Trend) x (Cycle) x (Season)
Variable	Sales
Number of Rows	144
Missing Values	None
Mean	174.2847
Pseudo R-Squared	0.9872795
Forecast Std. Error	3.211923
Trend Equation	Trend = (0.762387) + (0.003224) * (Time Period Number)
Number of Seasons	12
First Year	1970
First Season	1

Seasonal Component Ratios

No.	Ratio	No.	Ratio	No.	Ratio	No.	Ratio
1	0.901933	2	0.855207	3	0.971591	4	0.998289
5	1.030011	6	1.028496	7	0.997262	8	1.005337
9	0.975528	10	1.024248	11	1.007887	12	1.205982

This report summarizes the forecast equation.

Variable

The name of the variable for which the forecasts are generated.

Number of Rows

The number of rows that were in the series. This is provided to allow you to double-check that the correct series was used.

Missing Values

If missing values were found, this option lists the method used to estimate them.

Mean

The mean of the variable across all time periods.

Pseudo R-Squared

This value generates a statistic that acts like the R-Squared value in multiple regression. A value near zero indicates a poorly fitting model, while a value near one indicates a well-fitting model. The statistic is calculated as follows:

$$R^2 = 100 \left(1 - \frac{SSE}{SST} \right)$$

where *SSE* is the sum of square residuals and *SST* is the total sum of squares after correcting for the mean.

Forecast Std. Error

The estimated standard deviation of the forecast errors (the difference between the actual and predicted). This value is calculated by squaring and summing all of the forecast errors, dividing by the number of observations, and taking the square root.

Trend Equation

The equation used to predict the trend. The equation is

$$\text{Trend} = a + bt$$

Decomposition Forecasting

where

a is the intercept.

b is the slope.

t is the time period.

Note that the trend value obtained from this equation will be a ratio type value that will be multiplied by the mean to obtain the actual forecast.

Number of Seasons

The number of rows per year. For example, monthly data would have a value of 12.

First Year

The value of the first year of the series.

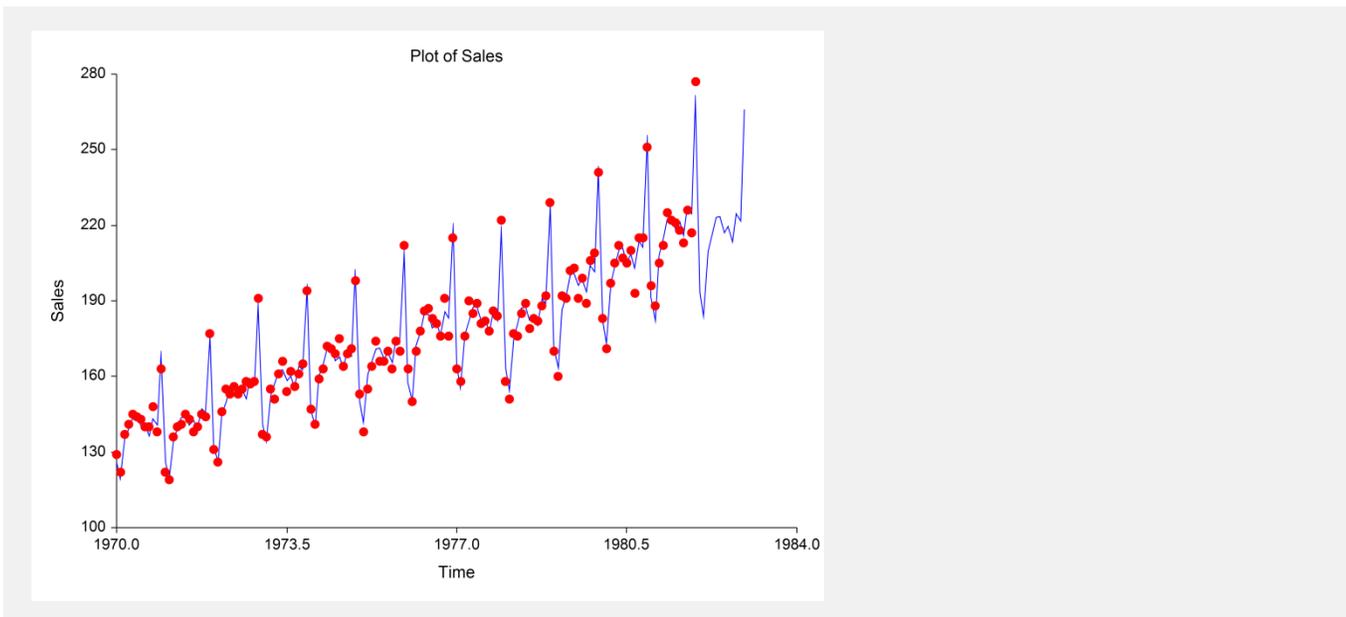
First Season

The value of the season of the first observation.

Season Component Ratios

The ratios used to adjust for each season (month or quarter). For example, the last ratio in this example is 1.205982. This indicates that the December correction factor is a 20.5982% increase in the forecast.

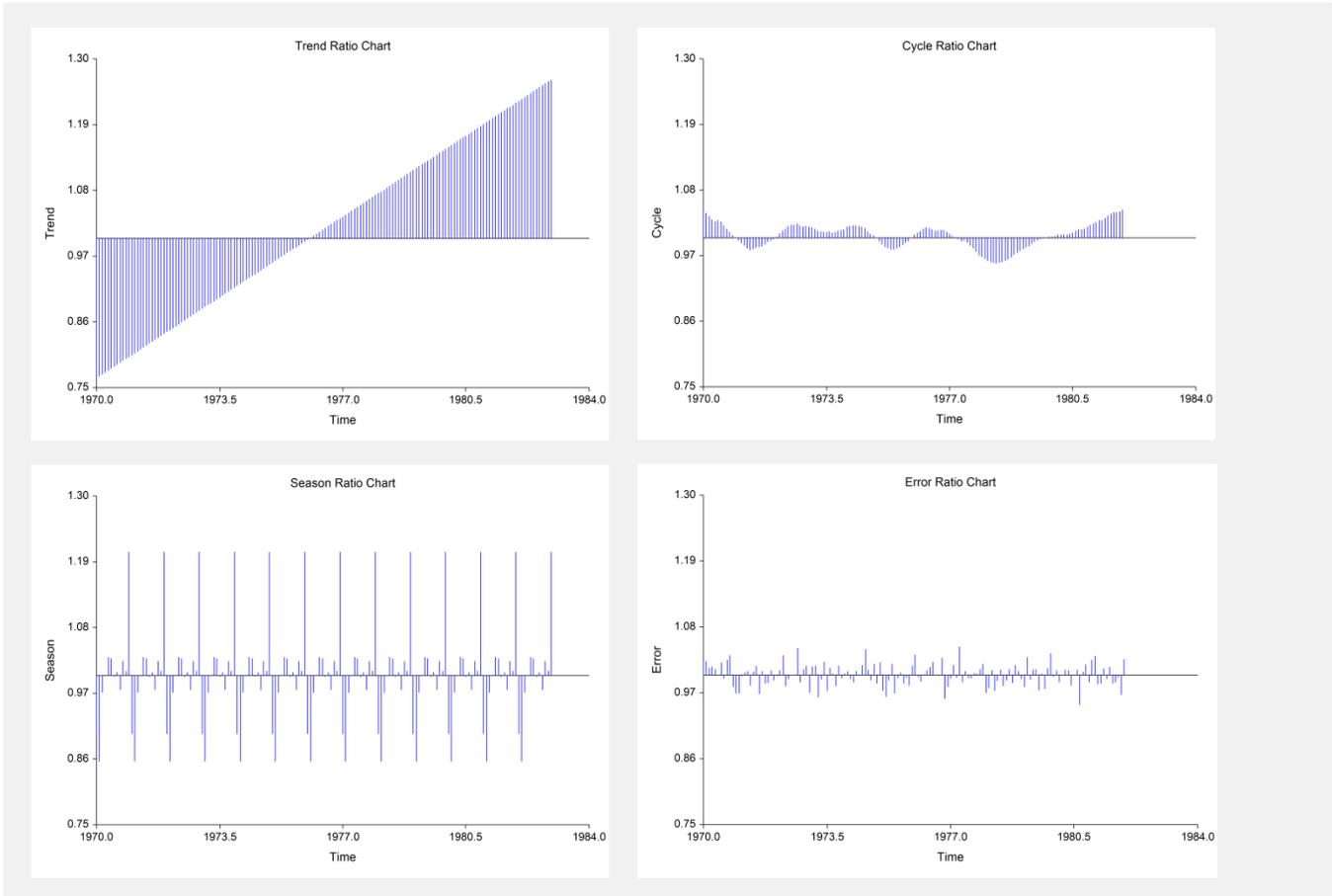
Data and Forecast Plot



The data plot lets you analyze how closely the forecasts track the data. The plot also shows the forecasts at the end of the data series.

Decomposition Forecasting

Ratio Plots



Ratio Plots

These plots let you see the various components of the forecast. Each of these plots is centered at one since this is the value that will leave the forecast unchanged. By studying these plots, you can see which factors influence the forecasts the most.

Forecasts Section

Forecasts Section									
Row No.	Year Season	Forecast Sales	Actual Sales	Residual	Trend Factor	Cycle Factor	Season Factor	Error Factor	
1	1970 1	125.8949	129	3.105124	0.7656	1.0461	0.9019	1.0247	
2	1970 2	119.3013	122	2.698687	0.7688	1.0411	0.8552	1.0226	
3	1970 3	135.456	137	1.544002	0.7721	1.0361	0.9716	1.0114	
.
144	1981 12	269.8886	277	7.111356	1.2266	1.0468	1.2060	1.0263	
145	1982 1	193.3271			1.2299	1.0000	0.9019	1.0000	
146	1982 2	183.7919			1.2331	1.0000	0.8552	1.0000	
147	1982 3	209.3499			1.2363	1.0000	0.9716	1.0000	
.
.

This section shows the values of the forecasts, the dates, the actual values, the residuals, and the forecast ratios. Note that the forecasted cycle ratios are all equal to one. This is because we did not supply cycle values to be used. If we had, they would have shown up here.