THE POWER TO KNOW.

SAS® Time Series Studio 12.1
User's Guide

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2012. *SAS® Time Series Studio 12.1: User's Guide*. Cary, NC: SAS Institute Inc.

SAS® Time Series Studio 12.1: User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

August 2012

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at **support.sas.com/publishing** or call 1-800-727-3228.

Contents

Using This Book

Audience

SAS Time Series Studio is designed for these users:

- any analyst who works with multiple time series
- an experienced user of SAS Forecast Server who relies on the user interface of SAS Forecast Studio for forecasting

Requirements

To use SAS Time Series Studio, you must have timestamped data. Your site administrator must have installed SAS Time Series Studio.

Recommended Reading

- *SAS Forecast Server: Administrator's Guide*
- *SAS Forecast Studio: User's Guide*
- *SAS/ETS User's Guide*
- *SAS/STAT User's Guide*

For a complete list of SAS publications, go to support.sas.com/bookstore. If you have questions about which titles you need, please contact a SAS Publishing Sales Representative:

SAS Publishing Sales
SAS Campus Drive
Cary, NC 27513-2414
Phone: 1-800-727-3228
Fax: 1-919-677-8166
E-mail: sasbook@sas.com
Web address: support.sas.com/bookstore

Part 1

Introduction to SAS Time Series Studio

2

1

About SAS Time Series Studio

What Is SAS Time Series Studio?

Many organizations collect large amounts of transactional and time series data, such as sales histories, inventory histories, customer transactions, insurance claim histories, and Internet data. When reviewing this data, here are some key tasks for analysts:

- understand the patterns within each time series and the relationships between time series
- subset or segment the data by using graphical, hierarchical, and parameter queries

SAS Time Series Studio is experimental for the 12.1 release. It enables you to interactively explore time series data. You can interactively structure the time series data from several hierarchical and frequency perspectives. You can interactively query (or subset) the time series data using hierarchical queries, graphical queries, parametric queries, or manual selection. Given a target series, you can interactively search for similar series. After acquiring the time series data of interest, you can analyze the time series data using common statistical time series analysis techniques. After you

understand the patterns in the selected time series, the data can be exported for subsequent analysis such as forecasting, econometric analysis, pricing analysis, risk analysis, time series mining, and other analyses related to time series data.

Benefits of Using SAS Time Series Studio

SAS Time Series Studio provides users with the following benefits:

the ability to explore multiple time series simultaneously to better understand your data
: You can quickly identify anomalies (such as outliers or missing values) and determine which time series do not look like the others. You can explore the effect of transformations of variables and the effect of different time series.

the ability to identify series that require specialized methods for analysis
: Using SAS Time Series Studio, you can identify short series (new products or a short lifecycle product) and intermittent series.

the ability to identify and group similar time series
: You can quickly identify similar time series and group them.

How SAS Time Series Studio Works

When you first open SAS Time Series Studio, you need to create a project. When you create a project, you specify the following information:

- the input data set.
- at least one variable to the role of dependent variable or other variable. “Other” means that the role of the variable is unclear. You can have as many dependent variables or other variables as you need.

After a project is created, you can perform a variety of analyses on the time series in your project. You can use a variety of options to add structure to your data. You can

also subset your data. After you prepare your time series for further analysis, you can export the data to a SAS data set, which then can be used in SAS Forecast Studio.

How SAS Time Series Studio Relates to Other SAS Software

SAS Time Series Studio is a component of SAS Forecast Server. Use SAS Time Series Studio to turn your semi-structured (timestamped) data into structured (time series) data for forecasting purposes. When you use SAS Time Series Studio to create your time series data, your data will meet all of the requirements for analyzing data in SAS Forecast Studio.

Here are the requirements for analyzing time series data in SAS Forecast Studio:

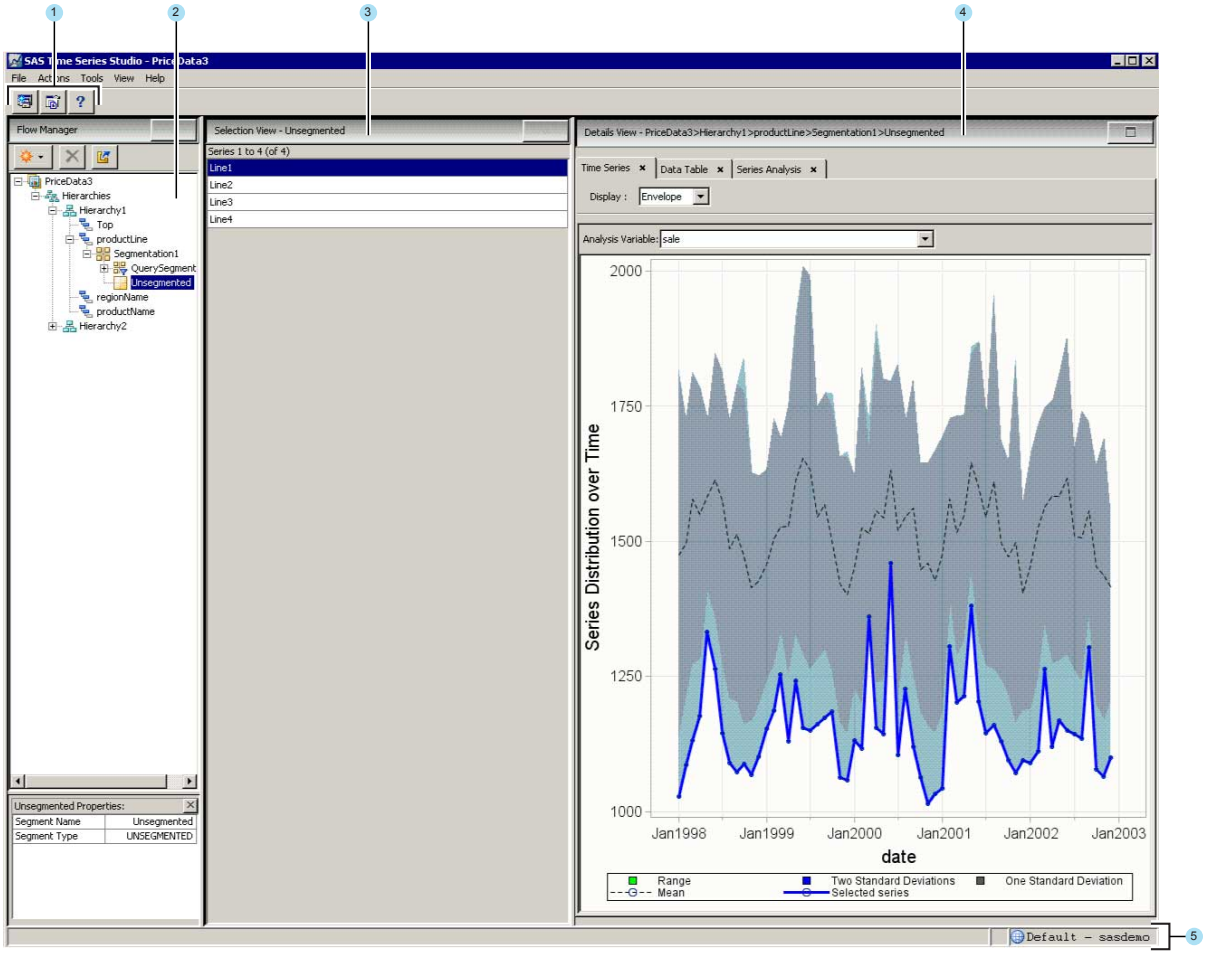
- The data set contains one variable for each analysis variable.
- The data set contains a time ID variable that identifies the time period for each observation. When SAS Forecast Server creates a project, the data is sorted by the time ID variable so that the observations are in order according to time.
- The data is equally spaced in time. This means that successive observations are a fixed time interval apart, and the data can be described by a single interval, such as hourly, daily, or monthly.

2

Using the Workspace

Overview of the Workspace

The SAS Time Series Studio workspace consists of the following components:



1 The toolbar displays some of the most commonly used SAS Time Series Studio options, so that you can quickly and easily manage your project.

2 The Flow Manager shows the elements (hierarchies, levels, queries, subsets, and segmentations) in your project. The contents of the Selection Manager depend on what you have selected in the Flow Manager.

3. The Selection Manager lists the series that make up the element (hierarchy, level, query, subset, or segmentation) selected in the Flow Manager. The selected series determines the content of the graphs and tables in the View Manager.
4. The Details View includes several analysis views. The content of each view depends on what is selected in the Selection Manager.
5. The status bar can contain the following information:
   - the status of the current action in SAS Time Series Studio. For example, the status bar displays the name of the SAS procedure that is currently executing.
   - the environment and your user ID.

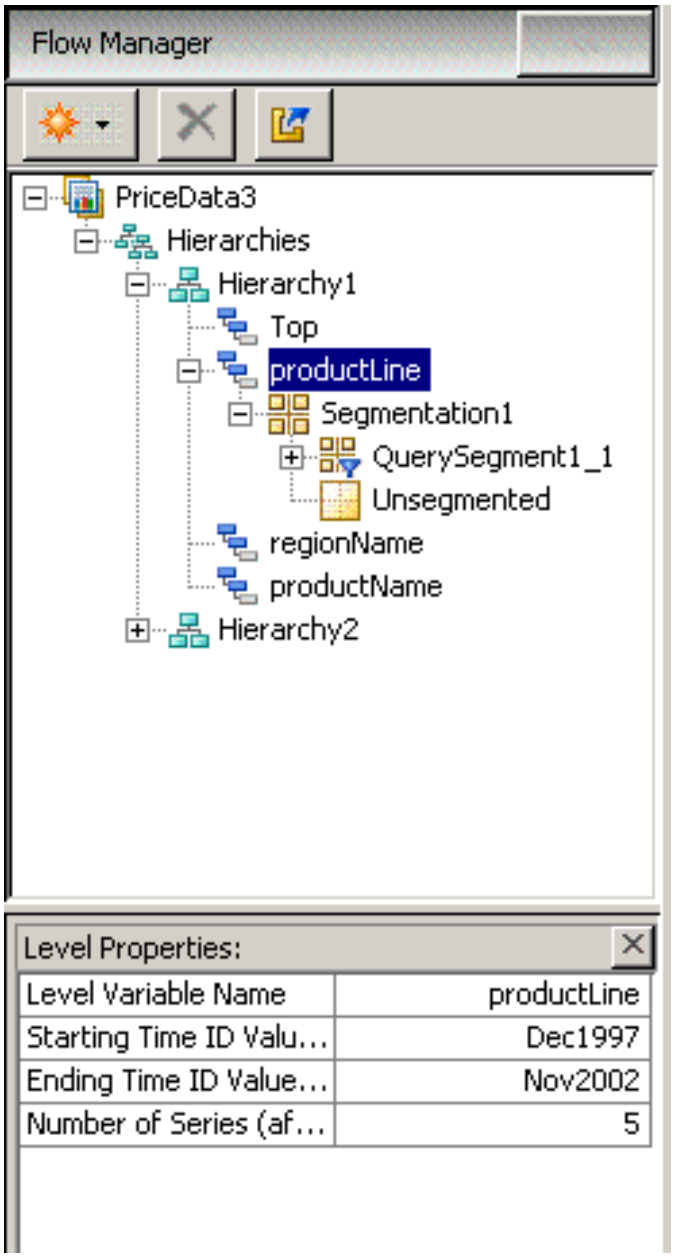
The content of the Selection Manager and Details View depends on which node is selected in the navigation tree in the Flow Manager.

Flow Manager	Selection Manager	Details View
project node	List of variables in the data source	**Distribution** and **Data Set** tabs
hierarchies node	faceted search of hierarchies	**Time Series** and **Series Analysis** tabs
hierarchy node or level	list of units in the hierarchy	**Time Series** and **Series Analysis** tabs
level, query node, or segmentation	list of units in the level, query, or segmentation	**Time Series**, **Series Analysis**, and **Multi-Variable** tabs

If you close a tab, you can reopen it by selecting **View** ▶ **Show View** ▶ ***name-of-tab***.

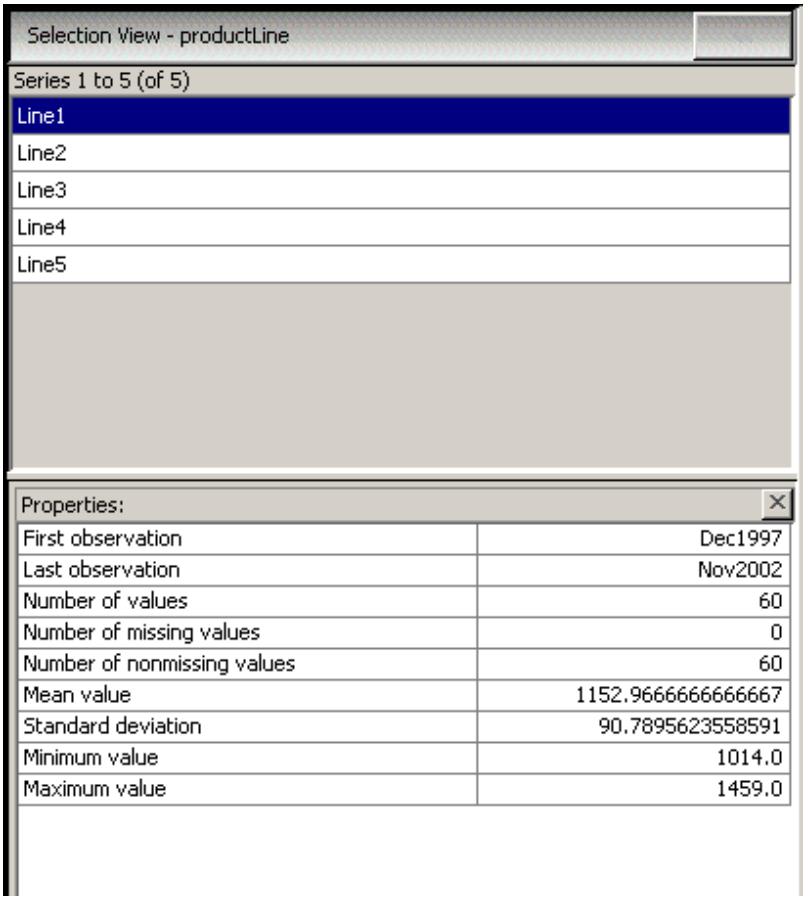
Flow Manager

The Flow Manager contains a project tree that visually depicts the organization and relationship between elements in your project. At the bottom of the Flow Manager, properties are displayed for the selected hierarchy or level.


Flow Manager
PriceData3
Hierarchies
Hierarchy1
Top
productLine
Segmentation1
QuerySegment1_1
Unsegmented
regionName
productName
Hierarchy2
Level Properties:
Level Variable Name
productLine
Starting Time ID Valu...
Dec1997
Ending Time ID Value...
Nov2002
Number of Series (af...
5

Selection Manager

The Selection Manager displays the series that are associated with the selected element in the Flow Manager. To view the properties for the selected series, select **View** ▶ **Show Properties** ▶ **Show series properties**. The properties appear at the bottom of the Selection Manager.

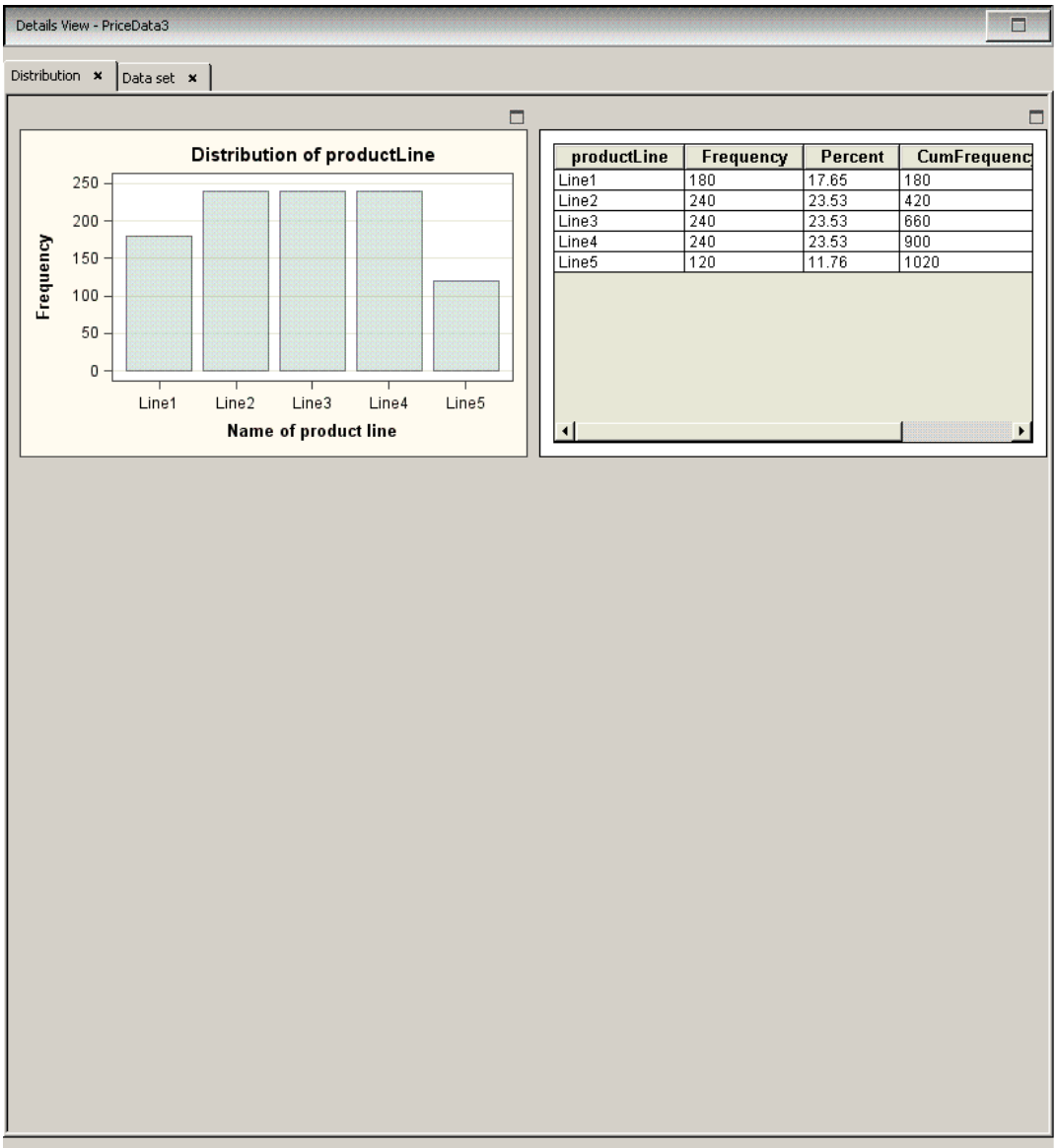


Details View

The Distribution View

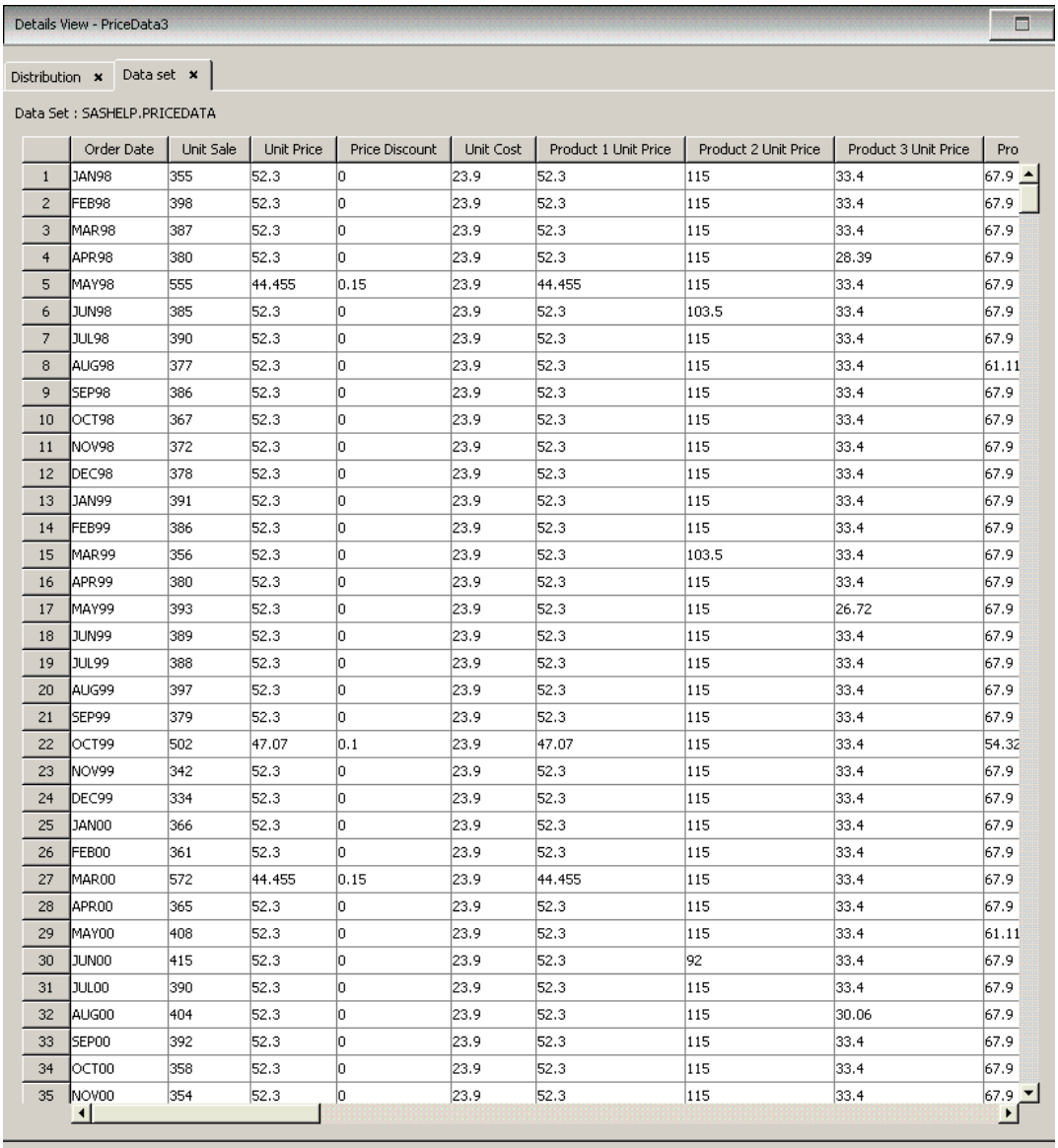
The Distribution View appears when you select the project node in the Flow Manager. This view shows the distribution for each variable in the project. The information is presented in a histogram. For discrete data, you can also see the following statistics: frequency, percent, cumulative frequency, and cumulative percentage. For continuous data, you see the following statistics: sum, mean, and standard deviation.

Note: If the number of series exceeds 1,000, the information displayed is based on a sample of the data. If a sample is used, a note is displayed in the workspace.



The Data Set View

The Data Set View is available when you select the project node in the Flow Manager. This view displays the input data set.



Details View - PriceData3

Distribution | Data set

Data Set : SASHELP.PRICEDATA

	Order Date	Unit Sale	Unit Price	Price Discount	Unit Cost	Product 1 Unit Price	Product 2 Unit Price	Product 3 Unit Price	Pro
1	JAN98	355	52.3	0	23.9	52.3	115	33.4	67.9
2	FEB98	398	52.3	0	23.9	52.3	115	33.4	67.9
3	MAR98	387	52.3	0	23.9	52.3	115	33.4	67.9
4	APR98	380	52.3	0	23.9	52.3	115	28.39	67.9
5	MAY98	555	44.455	0.15	23.9	44.455	115	33.4	67.9
6	JUN98	385	52.3	0	23.9	52.3	103.5	33.4	67.9
7	JUL98	390	52.3	0	23.9	52.3	115	33.4	67.9
8	AUG98	377	52.3	0	23.9	52.3	115	33.4	61.11
9	SEP98	386	52.3	0	23.9	52.3	115	33.4	67.9
10	OCT98	367	52.3	0	23.9	52.3	115	33.4	67.9
11	NOV98	372	52.3	0	23.9	52.3	115	33.4	67.9
12	DEC98	378	52.3	0	23.9	52.3	115	33.4	67.9
13	JAN99	391	52.3	0	23.9	52.3	115	33.4	67.9
14	FEB99	386	52.3	0	23.9	52.3	115	33.4	67.9
15	MAR99	356	52.3	0	23.9	52.3	103.5	33.4	67.9
16	APR99	380	52.3	0	23.9	52.3	115	33.4	67.9
17	MAY99	393	52.3	0	23.9	52.3	115	26.72	67.9
18	JUN99	389	52.3	0	23.9	52.3	115	33.4	67.9
19	JUL99	388	52.3	0	23.9	52.3	115	33.4	67.9
20	AUG99	397	52.3	0	23.9	52.3	115	33.4	67.9
21	SEP99	379	52.3	0	23.9	52.3	115	33.4	67.9
22	OCT99	502	47.07	0.1	23.9	47.07	115	33.4	54.32
23	NOV99	342	52.3	0	23.9	52.3	115	33.4	67.9
24	DEC99	334	52.3	0	23.9	52.3	115	33.4	67.9
25	JAN00	366	52.3	0	23.9	52.3	115	33.4	67.9
26	FEB00	361	52.3	0	23.9	52.3	115	33.4	67.9
27	MAR00	572	44.455	0.15	23.9	44.455	115	33.4	67.9
28	APR00	365	52.3	0	23.9	52.3	115	33.4	67.9
29	MAY00	408	52.3	0	23.9	52.3	115	33.4	61.11
30	JUN00	415	52.3	0	23.9	52.3	92	33.4	67.9
31	JUL00	390	52.3	0	23.9	52.3	115	33.4	67.9
32	AUG00	404	52.3	0	23.9	52.3	115	30.06	67.9
33	SEP00	392	52.3	0	23.9	52.3	115	33.4	67.9
34	OCT00	358	52.3	0	23.9	52.3	115	33.4	67.9
35	NOV00	354	52.3	0	23.9	52.3	115	33.4	67.9

The Time Series View

About the Time Series View

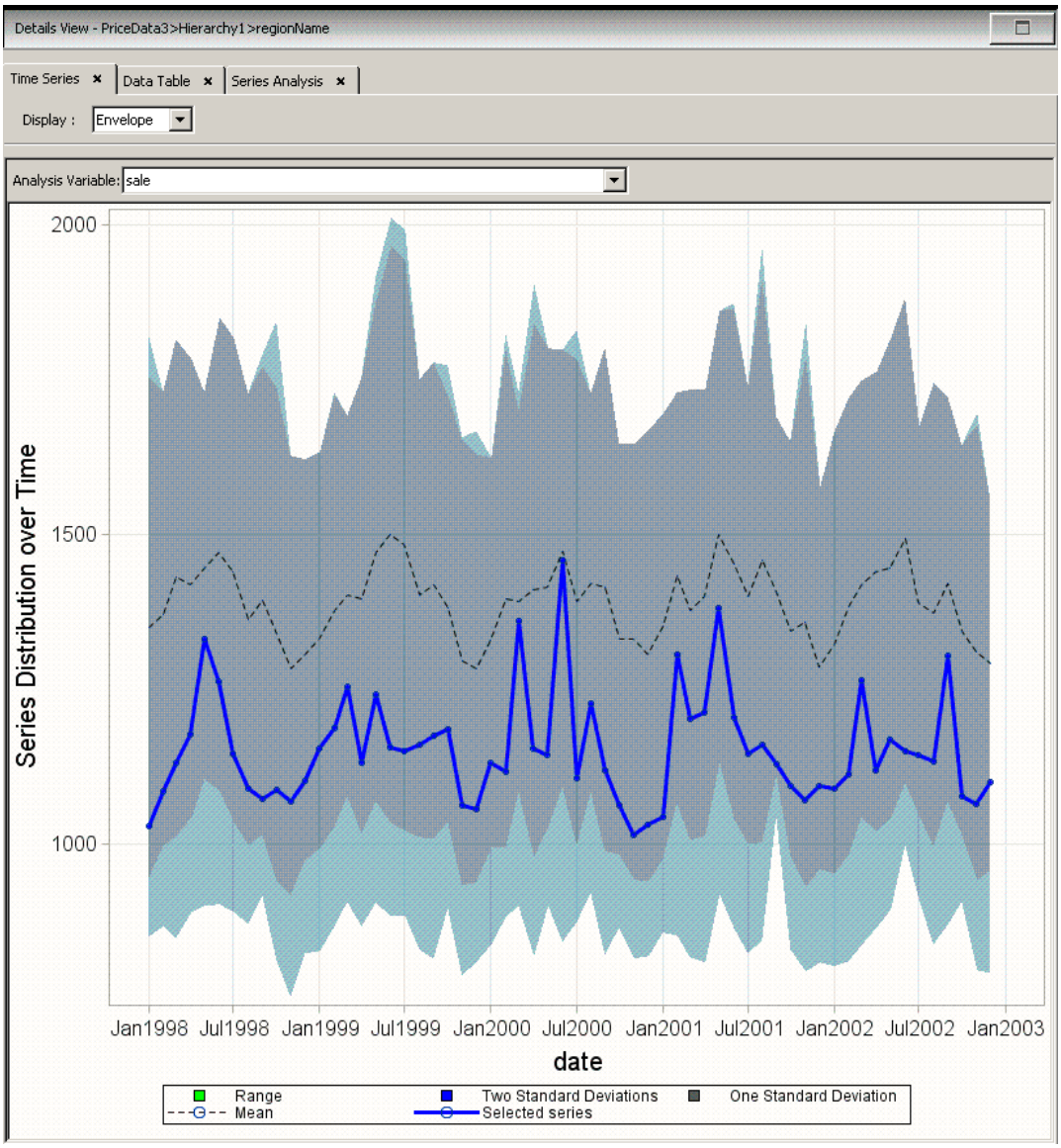
By default, when you open the Time Series View and select a level in the hierarchy, you see the envelope plot, which is an overview of all the data. (The envelope plot is not available if you select the top level in the hierarchy because there is only one series to

display.) The time series plot provides an interactive way for you to analyze the time series. When you select a point in a series, the graph displays the value at that point. You can also zoom in and out of the graph.

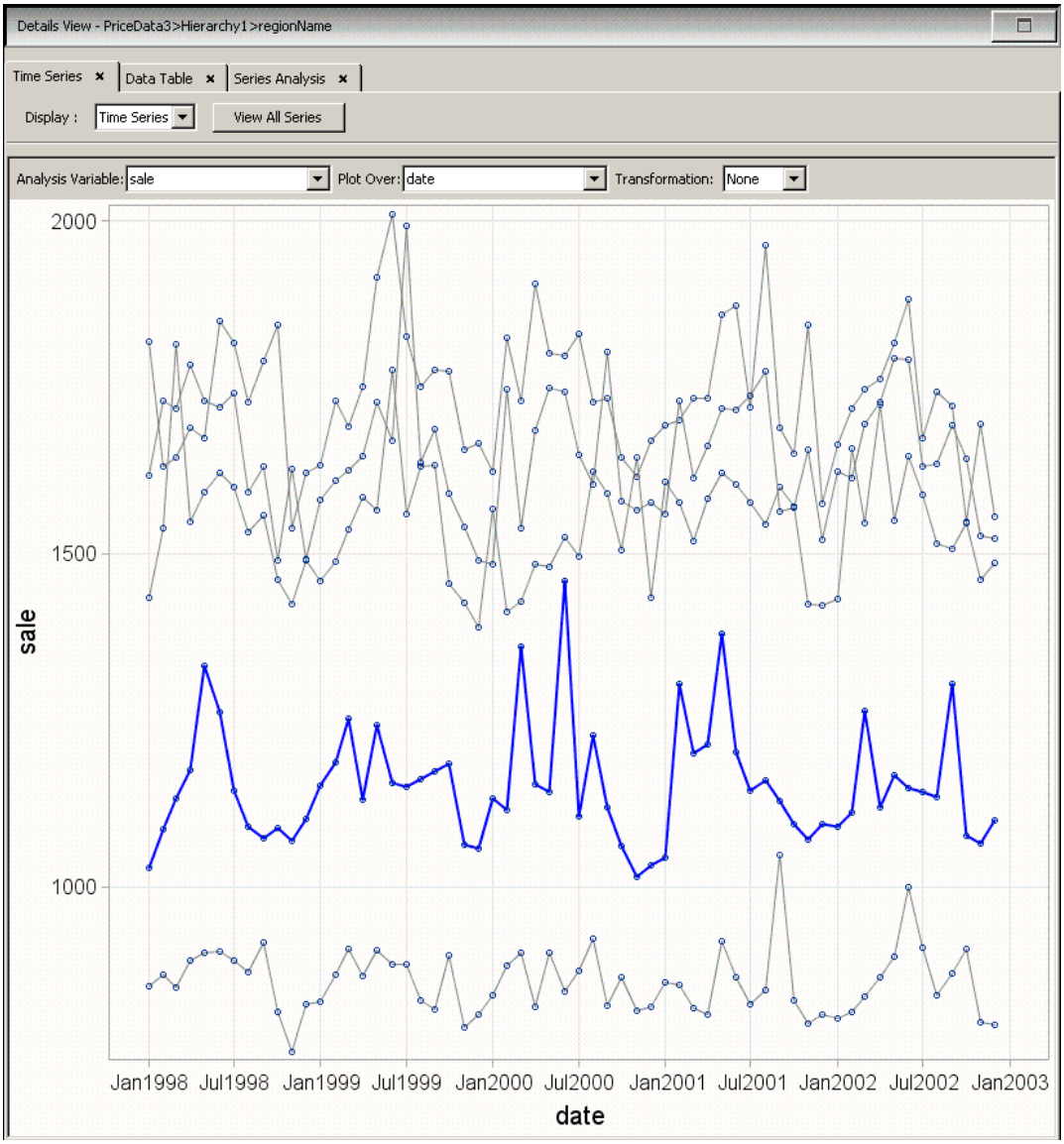
In the Time Series View, the graph shows all of the units for the node that is currently selected in the navigation tree. All of the units for this node are listed in the Selection Manager. The Time Series View shows the plot of the analysis variable over time or the cycle ID. You can also choose to transform the data and see the result of the transformation in the plot.

If you select an individual series in the Selection Manager, the plot for that series is highlighted in the Time Series View. By default, the Time Series View plots a subset of the data to optimize performance. Additional plots are shown if you page through the list of series in the Selection Manager.

To view the envelope plot for the selected series, select **Envelope** from the **Display** drop-down list. Here is an example of an envelope plot:

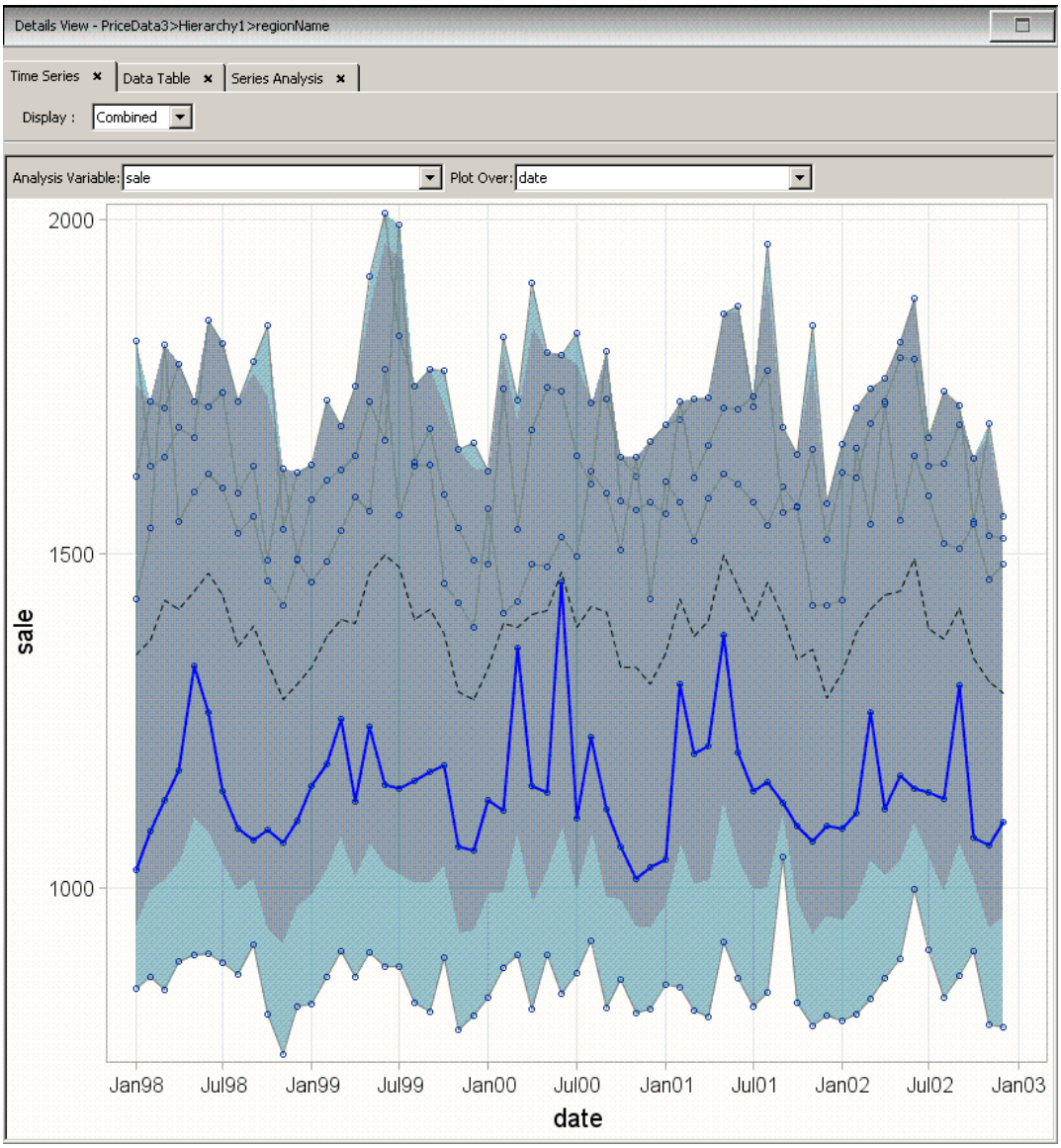


To display the time series plot for the selected series, select **Time Series** from the **Display** drop-down list. Here is an example of a time series plot:



If there are too many series to plot, the **View All Series** button is available. If you click **View All Series** in the Time Series View, the Multiple Series View appears and displays a random sample (200 by default) of all the series. You can increase the sample size by clicking **Increase Sample Size**.

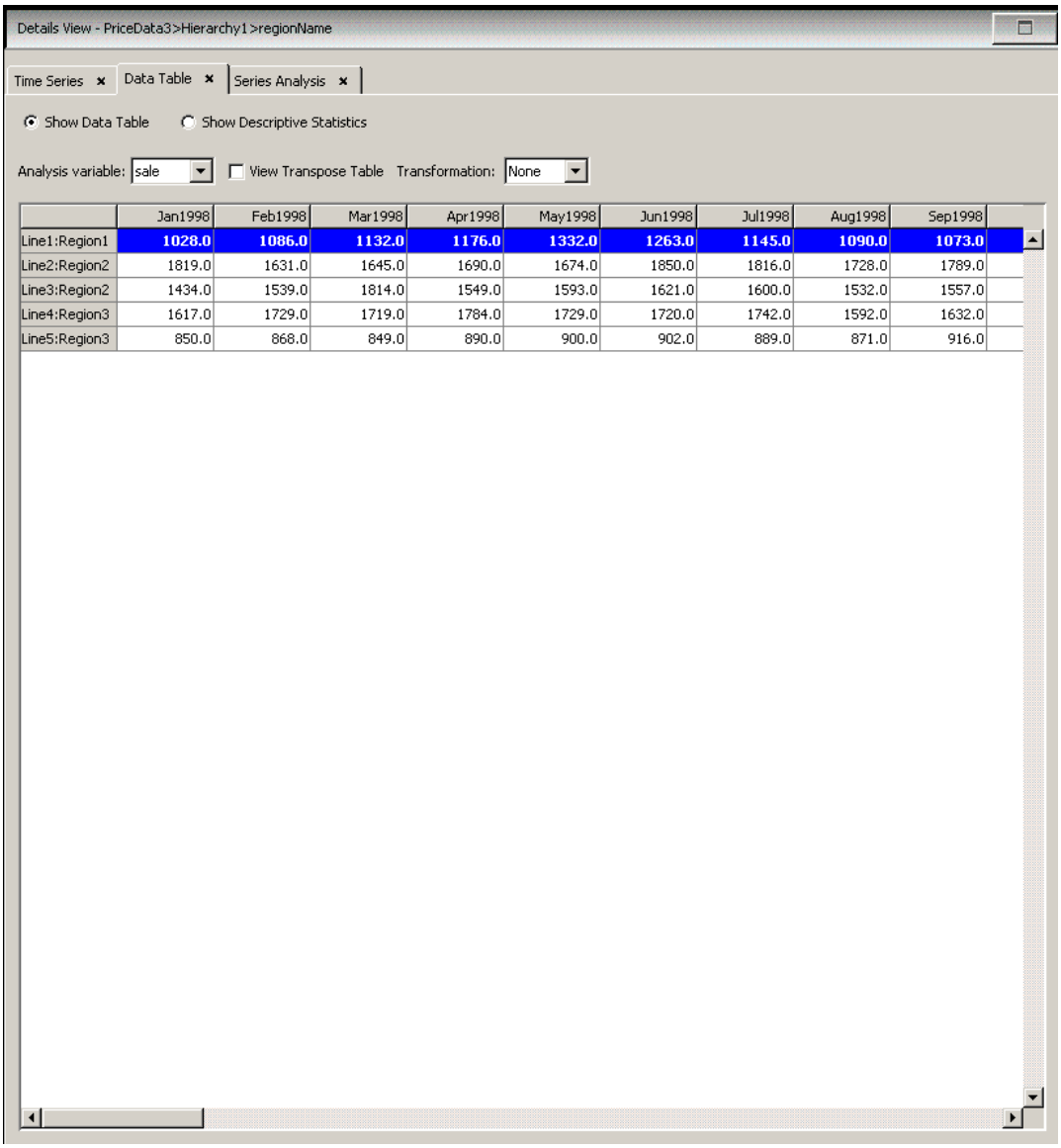
To view both the envelope plot and the time series plot, select **Combined** from the **Display** drop-down list. Here is an example of a combined plot:



The Data Table View

The Data Table View shows you the values of the selected variable for each time period in the data. Alternatively, you can choose to show the descriptive statistics in the data

table. Descriptive statistics include items such as the start and end date for the series, the number of observations, the number of missing values, mean, sum, and standard deviation.



	Jan1998	Feb1998	Mar1998	Apr1998	May1998	Jun1998	Jul1998	Aug1998	Sep1998
Line1:Region1	**1028.0**	**1086.0**	**1132.0**	**1176.0**	**1332.0**	**1263.0**	**1145.0**	**1090.0**	**1073.0**
Line2:Region2	1819.0	1631.0	1645.0	1690.0	1674.0	1850.0	1816.0	1728.0	1789.0
Line3:Region2	1434.0	1539.0	1814.0	1549.0	1593.0	1621.0	1600.0	1532.0	1557.0
Line4:Region3	1617.0	1729.0	1719.0	1784.0	1729.0	1720.0	1742.0	1592.0	1632.0
Line5:Region3	850.0	868.0	849.0	890.0	900.0	902.0	889.0	871.0	916.0

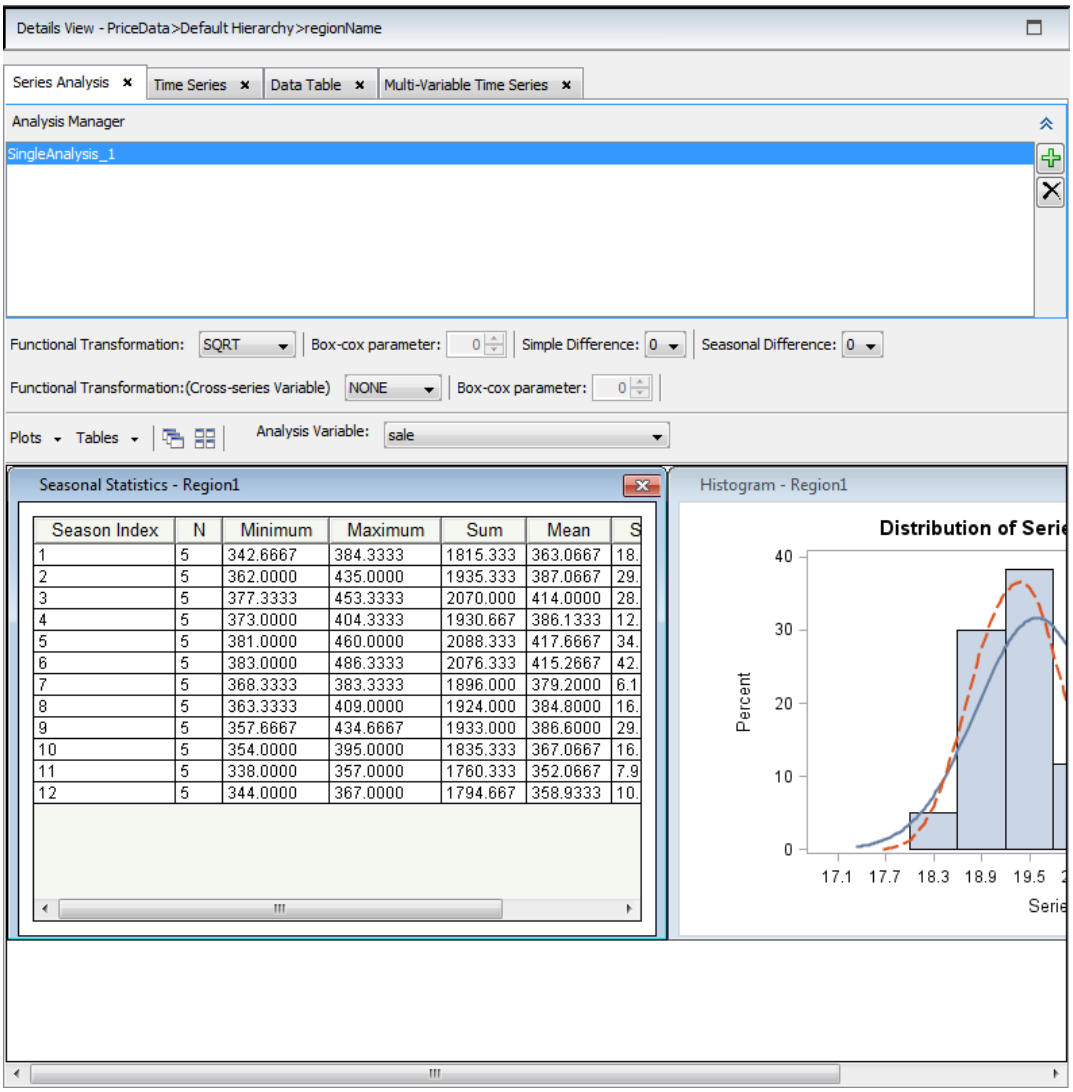
In the Data Table view, you can perform the following tasks:

- Choose to view the data table or the descriptive statistics.
- Select the analysis variable that you want to use.

- Select the **View Transpose Table** option to view the table with the variable names as columns and the dates as rows.
- Specify how you want to transform the data. For more information, see “Using Transformations” on page 23.

The Series Analysis View

In the Series Analysis View, you can perform an in-depth analysis on the series that is selected in the Selection Manager.



In the Series Analysis View, you can perform the following tasks:

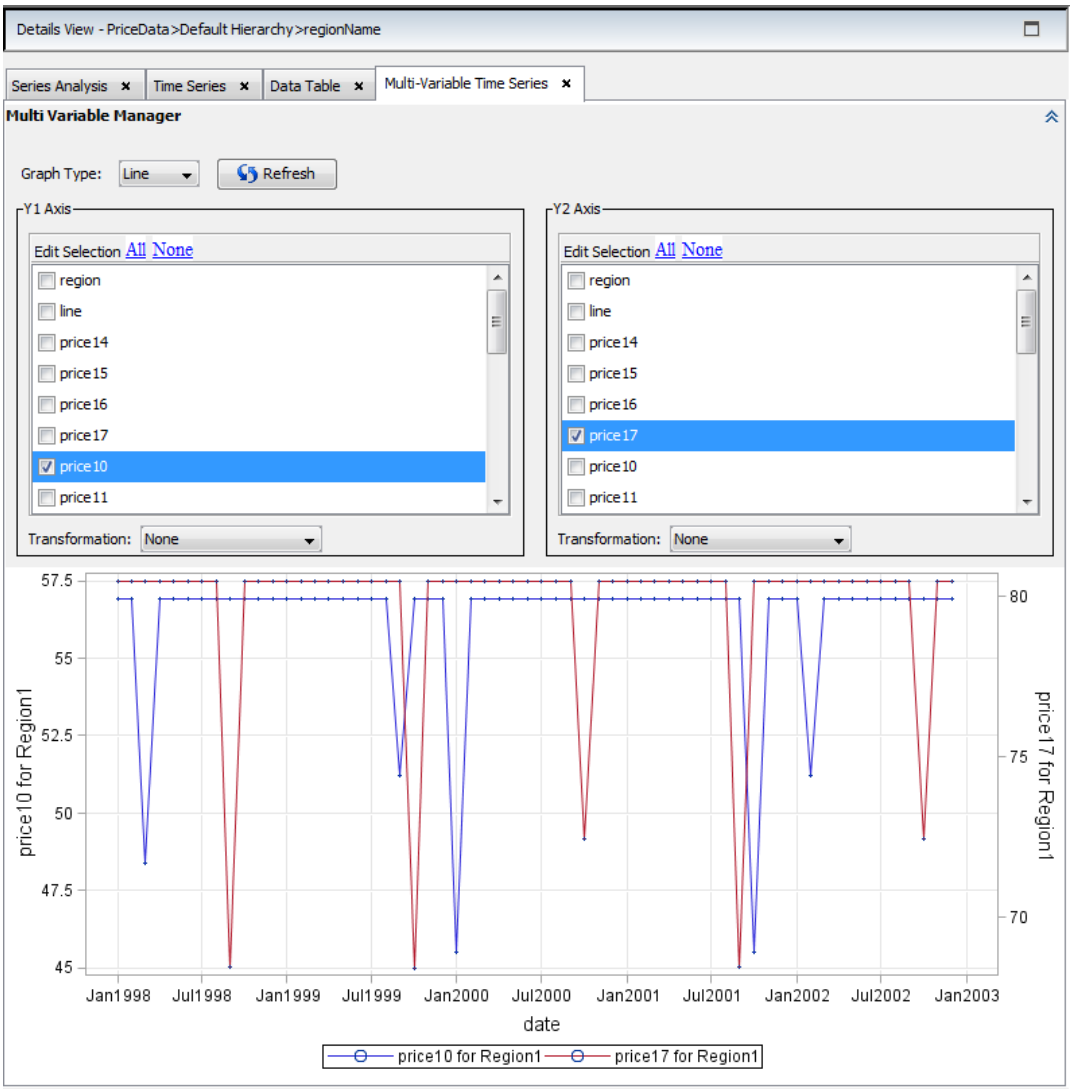
- Use the Analysis Manager to create and compare analyses.
- Specify the following options for your analysis:
  - **Functional Transformation** — specifies the functional transformation to apply to the dependent series. By default, no transformation is applied. If you select the Box-Cox transformation, then you must specify a parameter value between -5 and 5 in the **Box-Cox parameter** box.

- **Simple Difference** — specifies the simple differencing order. You can specify values from 0 to 9. The default value is 0.
- **Seasonal Difference** — specifies the seasonal differencing order. You can specify values from 0 to 9. The default value is 0.
- **Functional Transformation (Cross-series variable)** specifies the functional transformation to apply to the cross-series variable. By default, no transformation is applied. If you select the Box-Cox transformation, then you must specify a parameter value between -5 and 5 in the **Box-Cox parameter** box.

- Select the analysis variable.
- Select the plots and tables to display. You can display these plots and tables as tiled or cascade. When you close a plot or table, the plot or table is removed from the current analysis.

The Multi-variable Time Series View

You can compare the time series plots for multiple variables by using the Multi-Variable Time Series View.



Using Transformations

Overview of Transformations

SAS Time Series Studio is shipped with these transformations:

- cumulative sum transformation
- logistic transformation

- log transformation
- square root transformation

For more information about these transformations, see the EXPAND and TIMESERIES procedures in the *SAS/ETS: User's Guide*.

Create a Transformation

In addition to the default transformations, you can create Box-Cox, scale, simple difference, and seasonal difference transformations.

To create a transformation:

1. Select **Tools** ▶ **Manage Transformations**. The Manage Transformations dialog box appears.
2. Click  . The Add Composite Transformations dialog box appears.
3. Specify the name of the transformation. You can also specify a description.
4. Click  to add an individual transformation to the table. The Add Individual Transformation dialog box appears.
5. Select the type of transformation that you want to create, any options that are associated with that transformation, and click **OK**. The new transformation appears in the **Individual Transformations** table in the Add Composite Transformations dialog box.
6. (Optional) Repeat steps 4 and 5 until your transformation is completely defined.
7. Click **OK**.

The transformation now appears in the Manage Transformations dialog box. Click **Close** to return to the main workspace. Your new transformation is also available in any **Transformation** drop-down list that is in SAS Time Series Studio (for example, the **Transformation** drop-down list in the Time Series View).

Example: Create a Box-Cox Transformation with a Seasonal Difference

To create this transformation:

1 Select **Tools** ▶ **Manage Transformations**. The Manage Transformations dialog box appears.

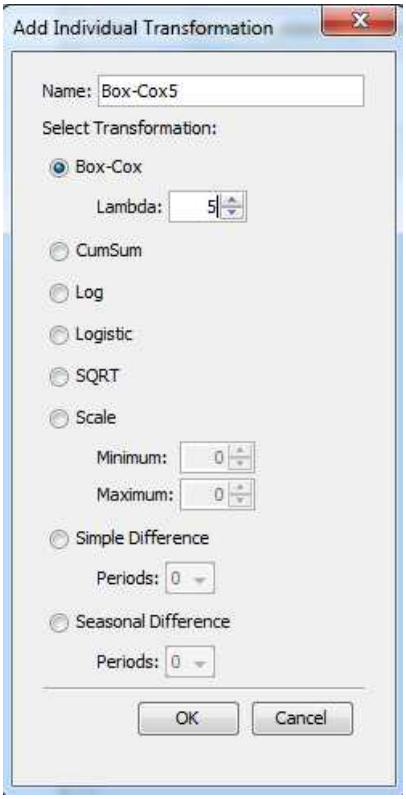
2 Click  . The Add Composite Transformations dialog box appears.

3 For the name of the transformation, type **`Box-Cox5_seasonal_difference`**.

4 Click  next to the table. The Add Individual Transformation dialog box appears.

5 For the name of this individual transformation, type **`Box-Cox5`**.

6 For the transformation, select the **Box-Cox** option and type `5` in the **Lambda** box.



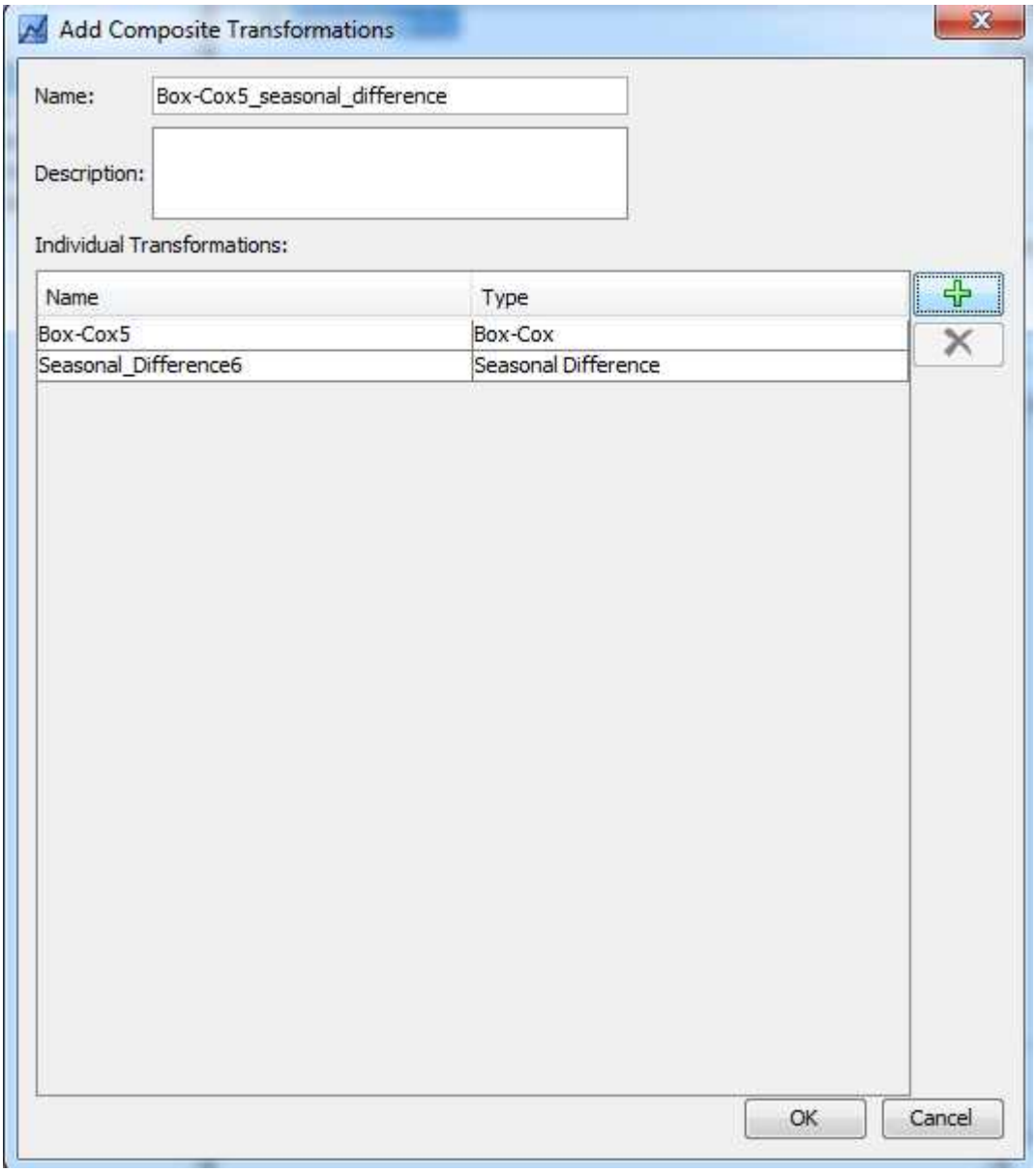
Click **OK**. The Box-Cox5 transformation appears in the list of individual transformations in the Add Composite Transformation dialog box.

7 In the Add Composite Transformation dialog box, click  . The Add Individual Transformations dialog box appears.

8 For the name, type `Seasonal_Difference6`.

9 Select the **Seasonal Difference** option and type `6` in the **Periods** box. Click **OK**.

The Seasonal_Difference6 transformation now appears in the Add Composite Transformations dialog box.



10 In the Add Composite Transformations dialog box, click **OK**.

The new transformation now appears in the Manage Transformations dialog box. It is available from the **Functional transformation** drop-down list in the Time Series View and the **Transformation** drop-down list in the Multi-variable Time Series View.

Part 2

Getting Started in SAS Time Series Studio

3

Understanding the Data Preparation Options

Understanding Aggregation and Accumulation

Difference between Aggregation and Accumulation

Aggregation is the process of combining more than one time series to form a single series. Aggregation combines data within the same time interval. For example, you can aggregate data into a total or average. In the New Project wizard, you can specify the aggregation and accumulation options for the dependent, independent, adjustment, and reporting variables in your project.

The following examples explain when you might want to use an aggregation method:

- Your data set contains the sales for a group of products. If you want to know the total sales for a category, you would choose **Sum of values** as the aggregation method.
- Your data contains the price of each product. If you want to know the average price for a product line, you would choose **Average of values** as the aggregation method.

Accumulation can be either of the following:

- the process of converting a time series that has no fixed interval into a time series that has a fixed interval (such as hourly or monthly)
- the process of converting a time series that has a fixed interval into a time series with a lower frequency time interval (such as hourly into daily)

Accumulation combines data within the same time interval into a summary value for that time period.

Because this process is not dependent on whether you have a hierarchy, you might need to accumulate data regardless of whether you forecast your data hierarchically.

The Accumulation and Aggregation Methods

Note: The difference between accumulation and aggregation is the dimension along which each method is applied. The following equations focus on accumulation, but the same equations also apply to the aggregation methods that you specify when you configure the hierarchy.

Let $R = \{r_q\}_{q=1}^{Q}$ be the data vector that is ordered by the time series occurrence in the data set with respect to the observation index. Let $q = 1, \ \ldots \ , Q$ be the index that represents this ordering. Let Q_N be the number of nonmissing values, and let $Q_{NMISS} = Q - Q_N$ be the number of nonmissing values in the data vector. Let $\bar{r} = \frac{1}{Q_N} \sum_{q=1}^{Q} r_q$ be the average value of the data vector with the missing values ignored.

The following example accumulates the observation series $Z^{(N)} = \{z_i\}_{i=1}^{N}$ to the time series $Y^{(T)} = \{y_t\}_{t=1}^{T}$, $y_t = Accumulate(Z_t^{(T)})$, for $t = 1, \ldots, T$. In this situation, $R = Z_t^{(T)}$ and $Q = N_t^{(T)}$ for $t = 1, \ldots, T$.

Let $a = Accumulate(R)$ be this accumulated value for this data vector when the following accumulation methods are applied:

None
: does not accumulate the vector values.

Sum
: accumulates the vector values based on the summation of their values.

$$a = \sum_{q=1}^{Q} r_q$$

Missing values are ignored in the summation. If $Q_N = 0$, then a is set to missing.

Average
: accumulates the vector values based on the average of their values.

$$a = \bar{r} = \frac{1}{Q_N} \sum_{q=1}^{Q} r_q$$

Missing values are ignored in the summation. If $Q_N = 0$, then a is set to missing.

Minimum
: accumulates the vector values based on the minimum of their values.

$$a = \min\left(\{r_q\}_{q=1}^{Q}\right)$$

Missing values are ignored in the summation. If $Q_N = 0$, then a is set to missing.

Maximum
: accumulates the vector values based on the maximum of their values.

$$a = \max\left(\{r_q\}_{q=1}^{Q}\right)$$

Missing values are ignored in the summation. If $Q_N = 0$, then a is set to missing.

Median
accumulates the vector values based on the median of their values.

$$a = median(\{r_q\}_{q=1}^{q})$$

Missing values are ignored in the summation. If $Q_N = 0$, then a is set to missing.

Number of Nonmissing Observations
accumulates the vector values based on the number of nonmissing values.

$$a = Q_N$$

Number of Observations
accumulates the vector values based on the number of values.

$$a = Q$$

Number of Missing Observations
accumulates the vector values based on the number of missing values.

$$a = Q_{NMISS}$$

First Occurrence
accumulates the vector values based on the first observation in the data.

$$a = r_1$$

Last Occurrence
accumulates the vector values based on the last observation in the data.

$$a = r_Q$$

Standard Deviation
accumulates the vector values based on their standard deviation.

$$a = \sqrt{\frac{1}{Q_N - 1} \sum_{q=1}^{Q} (r_q - \bar{r})^2}$$

Missing values are ignored in the summation. If $Q_N \leq 1$, then a is set to missing.

Uncorrected Sum of Squares
accumulates the vector values based on their uncorrected sum of squares.

$$a = \sum_{q=1}^{Q} (r_q)^2$$

Missing values are ignored in the summation. If $Q_N = 0$, then a is set to missing.

Corrected Sum of Squares
accumulates the vector values based on their corrected sum of squares.

$$a = \sum_{q=1}^{Q} (r_q)^2$$

Missing values are ignored in the summation. If $Q_N = 0$, then a is set to missing.

Working with Missing Values

How SAS Time Series Studio Interprets Missing Values

If your data contains missing values in variables other than the time ID variable (such as the dependent, independent, and reporting variables), you can specify how to interpret missing values (regardless of the variable role).

- You can specify how to replace missing values in the data. You can specify that SAS Time Series Studio should set missing values to one of the following values:
  - 0.
  - the accumulated average value.
  - the accumulated first nonmissing value.
  - the accumulated last nonmissing value.
  - the accumulated maximum value.
  - the accumulated median value.
  - the accumulated minimum value.
  - missing.
  - the next accumulated nonmissing value. Missing values at the end of the accumulated series remain missing.
  - the previous accumulated nonmissing value. Missing values at the beginning of the accumulated series remain missing.
- You can specify how missing values are removed from the accumulated time series. You can choose to keep all of the missing values, remove the beginning missing values, remove the ending missing values, or remove both the beginning and ending missing values.
- You can also specify how beginning and ending zero values are interpreted in the accumulated time series. You can choose to keep the beginning and ending zeros, set the beginning zero values to missing, set the ending zero values to missing, or set both the beginning and ending zero values to missing.

Specify How to Interpret Missing Values

When you create a project, you can specify how to treat missing values from the data preparation step in the New Project wizard.

4

Understanding Variable Roles

Overview of the Types of Roles

When you create a project in SAS Time Series Studio, you can assign the following roles to the variables in the data set:

- time ID variable
- classification (BY) variable
- dependent variable

- independent variable
- rejected variables
- other variables

Time ID Variables

What Is the Time ID Variable?

You specify the time ID variable when you create the project using the New Project wizard. After the project has been created, you cannot change the time ID variable. The time ID variable is a variable in the input data set that contains the SAS date or datetime value for each observation. This variable is used to determine the frequency and ordering of the data and to extrapolate the time ID values for the forecasts. You can assign only one variable to this role. That variable must be either a date variable, a datetime variable, or a numeric variable that contains date or datetime values.

Basics on Time Intervals

All time intervals (whether they are shipped with SAS or custom intervals that you create) must meet the following criteria:

- A discrete time interval has a beginning and an ending SAS date or SAS datetime.
- For SAS date intervals, the ending date is defined as 1 day before the beginning of the next interval.
- For SAS datetime intervals, the ending time is 1 second before the beginning of the next interval.
- All observations with an identifying SAS date or SAS datetime that is between the beginning and the end of the interval t_i correspond to the interval t_i.
- Alignment refers to the identifying date of the interval and does not affect the definition of the interval.

Supported Time Intervals

SAS Time Series Studio supports the following time intervals:

Day
: specifies daily intervals.

Hour
: specifies hourly intervals.

ISO 8601 year
: specifies ISO 8601 yearly intervals. The ISO 8601 year starts on the Monday on or immediately preceding January 4. Note that it is possible for the ISO 8601 year to start in December of the preceding year. Also, some ISO 8601 years contain a leap week.

ISO 8601 week
: specifies ISO 8601 weekly intervals of seven days. Each week starts on Monday. The starting subperiod (or subperiods) is in days (DAY). Note that WEEKV differs from WEEK in that WEEKV.1 starts on Monday, WEEKV.2 starts on Tuesday, and so on.

Minute
: specifies minute intervals.

Month
: specifies monthly intervals.

Quarter
: specifies quarterly intervals (every three months). The starting subperiod is in months.

Retail 4-4-5 Year
: specifies ISO 8601 weekly interval, except that the starting subperiod (or subperiods) is in retail 4-4-5 months.

Retail 4-5-4 Year
: specifies ISO 8601 weekly interval, except that the starting subperiod (or subperiods) is in retail 4-5-4 months.

Retail 5-4-4 Year
: specifies ISO 8601 weekly interval, except that the starting subperiod (or subperiods) is in retail 5-4-4 months.

Retail 4-4-5 Month
: specifies retail 4-4-5 monthly intervals. The 3rd, 6th, 9th, and 12th months are five ISO 8601 weeks long with the exception that some 12th months contain leap weeks. All other months are four ISO 8601 weeks long. R445MON intervals begin with the 1st, 5th, 9th, 14th, 18th, 22nd, 27th, 31st, 35th, 40th, 44th, and 48th weeks of the ISO year.

Retail 4-5-4 Month
: specifies retail 4-5-4 monthly intervals. The 2nd, 5th, 8th, and 11th months are five ISO 8601 weeks long. All other months are four ISO 8601 weeks long with the exception that some 12th months contain leap weeks. R454MON intervals begin with the 1st, 5th, 10th, 14th, 18th, 23rd, 27th, 31st, 36th, 40th, 44th, and 49th weeks of the ISO year.

Retail 5-4-4 Month
: specifies retail 5-4-4 monthly intervals. The 1st, 4th, 7th, and 10th months are five ISO 8601 weeks long. All other months are four ISO 8601 weeks long with the exception that some 12th months contain leap weeks. R544MON intervals begin with the 1st, 6th, 10th, 14th, 19th, 23rd, 27th, 32nd, 36th, 40th, 45th, and 49th weeks of the ISO year.

Retail 4-4-5 Quarter
: specifies retail 4-4-5 quarterly intervals (every 13 ISO 8601 weeks). Some fourth quarters contain a leap week. The starting subperiod (or subperiods) is in retail 4-4-5 months.

Retail 4-5-4 Quarter
: specifies retail 4-5-4 quarterly intervals (every 13 ISO 8601 weeks). Some fourth quarters contain a leap week. The starting subperiod (or subperiods) is in retail 4-5-4 months.

Retail 5-4-4 Quarter
: specifies retail 5-4-4 quarterly intervals (every 13 ISO 8601 weeks). Some fourth quarters contain a leap week.

Second
specifies second intervals.

Semimonth
specifies semimonthly intervals. Each month consists of two periods. The first period starts on the first, and the second period starts on the 16th.

Semiyear
specifies intervals every six months. The starting subperiod is in months.

Ten-day
specifies 10-day intervals. Each month consists of three periods. The first period is the 1st through the 10th day of the month. The second period is the 11th through the 20th day of the month. The third period is the 21st through the end of the month.

Week
specifies weekly intervals of seven days.

The days of the week are numbered as follows:

Value of the Shift	Day of the Week
1	Sunday
2	Monday
3	Tuesday
4	Wednesday
5	Thursday
6	Friday
7	Saturday

Weekday
specifies daily intervals with weekend days included in the preceding weekday. The weekday interval is the same as the day interval, except that the weekend days are absorbed into the preceding weekday. The default weekend days are Saturday and Sunday, but you can specify the days to include in the weekend. If you use the

default weekend, then there are five weekday intervals in a calendar week: Monday, Tuesday, Wednesday, Thursday, and the three-day period Friday, Saturday, and Sunday.

Year
: specifies yearly intervals. The starting subperiod is in months.

Understanding SAS Time Intervals

SAS Time Series Studio analyzes the variable that is assigned to the time ID role to detect the time interval of the data. SAS assumes that all of the values in the time ID variable are either date or datetime values and distinguishes between the values by their magnitude. This assumption fails if you have dates that extend beyond July 21, 2196, or datetimes before January 1, 1960.

For many businesses, their time series data is equally spaced, or any two consecutive indices have the same difference between the time intervals. The following table shows an equally spaced time series with a one-year interval.

Year	Number of Sales
2005	42,100
2006	45,000
2007	47,000
2008	50,000

If the time interval cannot be detected from the variable that you assign, then you need to specify the interval and seasonal cycle length. For example, the following table shows an unequally spaced time series.

Year	Number of Sales
2003	32,100
2004	45,000

Year	Number of Sales
2007	47,000
2008	50,000

Often the time interval cannot be detected with transactional data (timestamped data that is recorded at no particular frequency). If this is the case, then SAS Time Series Studio accumulates the data into observations that correspond to the interval that you specify. For nontransactional data, you might need to specify the interval and seasonal cycle length if there are numerous gaps (missing values) in the data. In this case, SAS Time Series Studio supplies the missing values. A validation routine checks the values of the time ID to determine whether they are spaced according to the interval that you specified.

In SAS Time Series Studio, the interval determines the frequency of the output. You can modify the time interval. You can change the interval from a higher frequency to a lower frequency or from a lower frequency to a higher frequency. Time intervals are specified in SAS by using character strings. Each of these strings is formed according to a set of rules that enables you to create an almost infinite set of attributes. For each time interval, you can specify the type (such as monthly or weekly), a multiplier, and a shift (the offset for the interval). You can specify a greater time interval than that found in the input data. A smaller interval should not be used, because a small interval generates a large number of observations.

Seasonal cycle length specifies the length of a season. This value is populated automatically if SAS Time Series Studio can determine the seasonal cycle length from the time ID variable. However, you can specify a seasonal cycle length other than the default if you want to model a cycle in the data. For example, your data might contain a 13-week cycle, so you need to specify a 13-week seasonal cycle length in SAS Time Series Studio.

Here is the syntax for an interval:

name<multipler><.starting-point>

Here is an explanation of each of the user-supplied values:

name
is the name of the interval.

multipler
specifies the multiplier of the interval. This value can be any positive number. By default, the multiplier is 1. For example, YEAR2 indicates a two-year interval.

.starting-point
specifies the starting point for the interval. By default, this value is one. A value greater than 1 shifts the start to a later point within the interval. The unit for the shift depends on the interval. For example, YEAR.4 specifies a shift of three months, so the year is from April 1 through March 31 of the following year.

The examples in the following table show how the values that you specify for the interval, seasonal cycle length, multiplier, and shift work together.

Interval Name (in SAS code format)	Default	Shift Period	Example
YEARm.s	January 1	Months	YEAR2.7 specifies an interval of every two years. Because the value for the shift is 7, the first month in the year is July.
SEMIYEARm.s	January 1 July 1	Months	SEMIYEAR.3 - six-month intervals, March-August and September-February.
QTRm.s	January 1 April 1 July 1 October 1	Months	QTR3.2 - three-month intervals that start on April 1, July 1, October 1, and January 1.
SEMIMONTHm.s	First and 16th of each month	Semimonthly periods	SEMIMONTH2.2 - intervals from the 16th of one month through the 15th of the next month.

Interval Name (in SAS code format)	Default	Shift Period	Example
MONTHm.s	First of each month	Months	MONTH2.2 - February-March, April-May, June-July, August-September, October-November, and December-January of the following year.
TENDAYm.s	First, 11th, and 21st of each month	Ten-day periods	TENDAY4.2 - Four ten-day periods that start at the second ten-day period.
WEEKm.s	Each Sunday	Days (1=Sunday . . . 7=Saturday)	WEEK6.3 specifies six-week intervals that start on Wednesdays.
DAYm.s	Each day	Days	DAY3 - three-day intervals that start on Sunday.
HOURm.s	Start of the day (midnight)	Hours	HOUR8.7 specifies eight-hour intervals that start at 6:00 a.m., 2:00 p.m., and 10:00 p.m.

Classification (BY) Variables

SAS Time Series Studio groups together observations that have the same value for the BY variable. Assigning a BY variable enables you to obtain separate analyses for groups of observations. You can assign character and numeric variables to this role. The order of the BY variables describes the structure of the hierarchy.

One of the primary goals of SAS Time Series Studio is to figure out a useful hierarchy for modeling. When creating a project, you should assign all potential variables to the BY variables role in the New Project wizard. You cannot assign additional BY variables after you have created the project. However, you can specify the order of the BY variables and thus, change the order of the hierarchy after the project is created.

Dependent Variables

Dependent variables are the variables that you want to model and forecast. These dependent variables are used by SAS Forecast Studio when you import a data set that was created in SAS Time Series Studio.

When you create a project in SAS Time Series Studio, you can assign multiple variables as dependent variables in the New Project wizard. You must assign at least one dependent variable or one other variable when you create a project.

Independent Variables

Independent variables are any explanatory, input, predictor, or causal factor variables. These independent variables are used by SAS Forecast Studio when you import a data set that was created in SAS Time Series Studio.

Rejected Variables

Rejected variables are not included in the project and are not included in any resulting data sets. However, these variables are still part of the original data set.

Other Variables

Variables that are assigned to the Other role are included in the project but are not assigned to a specific purpose. You use this role to learn more about these variables. After a project is created, you might choose to assign a role to these variables.

5

Managing Environments

SAS Environments and Environments for Projects in SAS Time Series Studio

A SAS environment represents a deployment. An environment in SAS Time Series Studio represents a product workspace for product sessions. These environments are created by an administrator by using the functionality in SAS Time Series Studio.

SAS environments are available across products. Typically, your site administrator decides which SAS environments are available to you. The administrator also configures these environments. After an environment is configured, the Log On dialog box identifies the SAS environments that are available, applies any product-specific filtering to the list of environments, and then presents the list of available environments in the Log On dialog box. If your site administrator provided a description of a SAS environment, then this description is available as well. If you need to access an

environment that does not appear in the drop-down list or you receive an error message instead of the Log On dialog box, contact your site administrator.

Managing Environments

About the Default Environment

When your site administrator installs SAS Time Series Studio, your administrator probably created a default environment for you. You cannot delete this default environment.

However, you can create new environments, rename existing environments, or delete an environment. (You should not delete the Default environment.)

Create a New Environment

To create a new environment:

1 Open the Projects dialog box.

- When you start SAS Time Series Studio, the Projects dialog box appears automatically.
- After you close a project in SAS Time Series Studio, you can open the Projects dialog box by selecting **File** ▶ **Open Project**.

2 Next to the **Environment** drop-down list, click **Manage**. The Environments dialog box appears.

3 Click **New**.

Delete an Environment

To delete an environment:

1 Open the Projects dialog box.

- When you start SAS Time Series Studio, the Projects dialog box appears automatically.
- After you close a project in SAS Time Series Studio, you can open the Projects dialog box by selecting **File** ▶ **Open Project**.

2 Next to the **Environment** drop-down list, click **Manage**. The Environments dialog box appears.

3 Select the environment that you want to delete and click **Delete**.

Rename an Environment

To rename an environment:

1 Open the Projects dialog box.

- When you start SAS Time Series Studio, the Projects dialog box appears automatically.
- After you close a project in SAS Time Series Studio, you can open the Projects dialog box by selecting **File** ▶ **Open Project**.

2 Next to the **Environment** drop-down list, click **Manage**. The Environments dialog box appears.

3 Select the environment that you want to rename and click **Rename**.

View the Properties of an Environment

To view the properties of an environment:

1 Open the Projects dialog box.

- When you start SAS Time Series Studio, the Projects dialog box appears automatically.
- After you close a project in SAS Time Series Studio, you can open the Projects dialog box by selecting **File** ▶ **Open Project**.

2 Next to the **Environment** drop-down list, click **Manage**. The Environments dialog box appears.

3 Select the environment and click **Properties**. The Environment Properties dialog box appears.

6

Working with Projects

What Is a Project?

SAS Time Series Studio organizes the files that are generated with an analysis into a project. In a Windows environment, each project is saved by default to `C:\SAS\TSS\Lev1\AppData\SASTimeSeriesStudioMidTier12.1`.

Creating a New Project

Overview of the New Project Wizard

You can have only one project open in SAS Time Series Studio at a time. You can create a new project either when you first open SAS Time Series Studio or after you close an existing project. You create a new project by using the New Project wizard.

How you open the wizard depends on where you are in the application.

- To create a new project when you open SAS Time Series Studio, click **New** in the Projects dialog box.
- To create a new project when SAS Time Series Studio is already running, select **File** ▶ **New Project**.

Step 1: Name Your Project

Specify a name and description for your project. The project name must be a valid SAS name and cannot exceed 32 characters. For more information about SAS naming conventions, see *SAS Language Reference: Concepts*. You can also specify whether other users at your site can open and edit this project. By default, only the creator of the project has permission to open and edit the project.

Click **Next**.

Step 2: Select from a SAS Library the Data Set That You Want to Analyze

Select the data set that you want to analyze. You can preview the contents of the data set by clicking **View**. There is no limit on the size of the data set or the number of series that you want to analyze.

Step 3: Specify the Classification Variables and Whether to Create a Default Hierarchy

Assign one or more variables to the classification (BY) variables role. Classification variables are used to obtain separate analyses for groups of observations, or they can be used to create hierarchies. To view the distribution graph for a selected variable, click **Distribution**. To have SAS Time Series Studio select the classification (BY) variables automatically, click **Recommend**. By default, SAS Time Series Studio assigns all character variables as BY variables.

In the **Selected variables** panel, select the check boxes for the BY variables to include in the hierarchy. The order of the BY variables in this panel determines the order of the hierarchy. To preview the hierarchy, click **Preview**.

Click **Next**.

Step 4: Specify the Properties of the Time Dimensions of Your Data

1. From the **Time ID variable** drop-down list, select the variable in the data that contains the timestamped values.
2. (Optional) Specify the interval, multiplier, shift, seasonal cycle length, and format of the time ID variable. For more information, see “Time ID Variables” on page 38.
3. Click **Next**.

Step 5: Assign Roles to Variables in Your Data

1 Specify the roles of the remaining variables in your data set. By default, all variables are **Rejected**, which means that they are not included in the project. For more information about each type of role, see “Understanding Variable Roles” on page 37.

2 (Optional) To view the distribution graph for a selected variable, click **Distribution**.

3 Select the accumulation and aggregation methods for each variable. You can choose to use the same method for accumulation and aggregation. For more information, see “Understanding Aggregation and Accumulation” on page 31.

4 Specify how SAS Time Series Studio should interpret missing values in the data. For more information, see “Working with Missing Values” on page 35.

5 Click **Finish**.

Open an Existing Project

You can open a project in the following ways:

- When you first invoke SAS Time Series Studio, you can open a project from the Projects dialog box.
- After SAS Time Series Studio is running, select **File** ▶ **Projects** to open the Projects dialog box.

Note: You can have only one project open at a time in SAS Time Series Studio.

Import New Data

After creating a project, other users at your site might add, remove, or change the values in the input data set.

Note: SAS Time Series Studio does not support structural changes (such as the removal of columns) to the input data set.

To update the input data set for your project:

1 Select **File** ▶ **Import Data**. The Import New Data dialog box appears.

2 Select the **Update project** check box and click **OK**.

SAS Time Series Studio re-creates the project using the updated data source.

Update Project

After creating a project, you might need to update it.

To update a project:

1 Select **File** ▶ **Update Project**. The Update Project dialog box appears.

2 Click **OK**.

Delete a Project

When you delete a project, you remove all of the content in that project. You cannot retrieve a project after it has been deleted. Before you can delete a project, you must close any projects that are currently open in SAS Time Series Studio.

To delete a project:

1 Open the Projects dialog box.
   - When you start SAS Time Series Studio, the Projects dialog box appears automatically.
   - After you close a project in SAS Time Series Studio, you can open the Projects dialog box by selecting **File** ▶ **Open Project**.

2 Select the project that you want to delete and click **Delete**.

3 In the confirmation dialog box that appears, click **Yes**.

Copy a Project

Before you can copy a project, you must close any open projects in SAS Time Series Studio.

1 Open the Projects dialog box.
   - When you start SAS Time Series Studio, the Projects dialog box appears automatically.
   - After you close a project in SAS Time Series Studio, you can open the Projects dialog box by selecting **File** ▶ **Open Project**.

2 Select the project that you want to copy and click **Copy**. The Copy Project dialog box appears.

3 Specify a name for the copied project. The project name must be a valid SAS name. By default, the name is *original-project-name*_Copy_`n`, where `n` is an integer value.

4 Click **OK**. The copied project now appears in the Projects dialog box.

Rename a Project

Before you can rename a project, you must close any open projects in SAS Time Series Studio.

1 Open the Projects dialog box.

- When you start SAS Time Series Studio, the Projects dialog box appears automatically.
- After you close a project in SAS Time Series Studio, you can open the Projects dialog box by selecting **File** ▶ **Open Project**.

2 Select the project that you want to rename and click **Rename**. The name must be a valid SAS name.

3 In the confirmation dialog box that appears, click **Rename**.

View Project Properties

1 Open the Projects dialog box.

- When you start SAS Time Series Studio, the Projects dialog box appears automatically.
- After you close a project in SAS Time Series Studio, you can open the Projects dialog box by selecting **File** ▶ **Open Project**.

2 Select the project and click **Properties**. The Project Properties dialog box appears.

3 (Optional) To view the input data source for the project, click **View**.

4 (Optional) To change the project description, enter the description in the **Description** field.

5 (Optional) To share this project with other users at your site, select the **Allow other users to view and edit this project** check box. By default, this option is not selected, and only the project creator can view and edit a project.

6 Click **OK**.

Export Data

You can export the data from SAS Time Series Studio as a data set and then import it into another SAS application, such as SAS Forecast Studio. You can also export data as a JMP table and then open it in a local installation of JMP.

To export data:

1 In the Flow Manager, right-click the node that contains the data that you want to export, and from the pop-up menu, select **Export**. You can export data from a level in the hierarchy, a segment, or a subset.

2 Select the content that you want to export.

3 Specify how you want to export the data. If you export the data as a SAS data set, you must specify the SAS library where to save the data set. If you export the data as a JMP table, you must specify a local file system path.

4 Click **OK**.

Part 3

Exploring Your Data

7

Understanding the Distribution of Your Data

Starting to Explore Your Data

When you are first asked to explore a data source, you might not know anything about the data set (other than where it came from). In SAS Time Series Studio, you can create a simple project so that you can view the distribution of each variable. After you see how the data is distributed, you can decide how to structure your data.

Example: Examine the Distribution of the PriceData Data Set

This example assumes that you are not familiar with the SASHelp:PriceData data set. If you already know how you want to structure your data, you can create this hierarchy in the New Project wizard. This example assumes that you do not know how you want to structure your data, so you will not be creating the hierarchy in the New Project wizard. Instead, you will create the hierarchy after the project is created and you have had the chance to study the distribution of the data.

To examine the distribution of the variables in the PriceData data set:

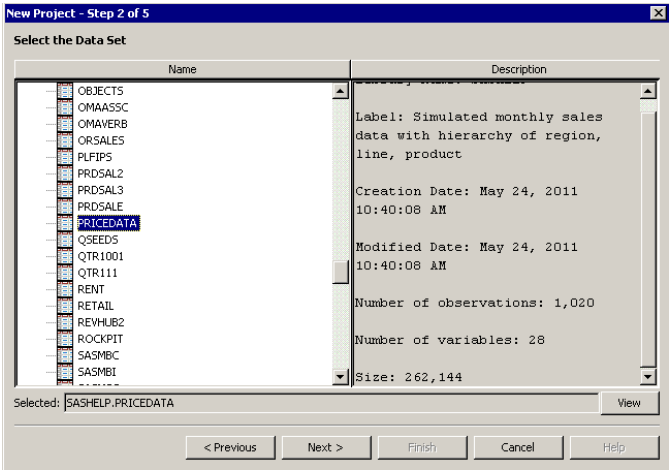
1 In SAS Time Series Studio, select **File** ▶ **New Project**. The New Project wizard appears.

Note: When you start SAS Time Series Studio, the Project dialog box appears by default. Click **New** to open the New Project wizard.

2 In the New Project wizard, specify a valid SAS name for the project. You can also provide a description.

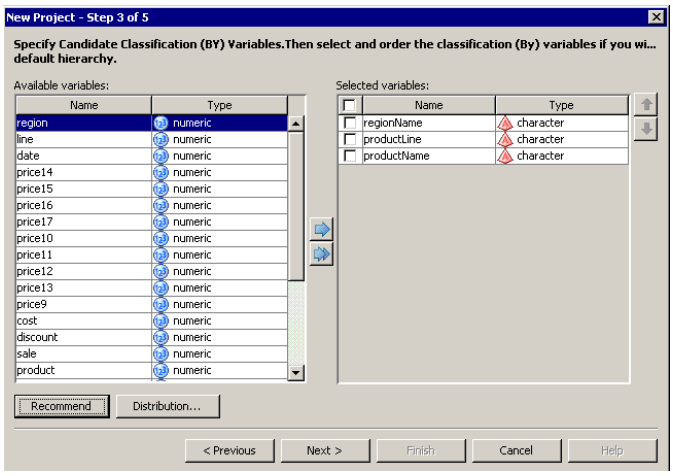
Click **Next**.

3 In the SASHelp library, select the PriceData data set. Click **Next**.

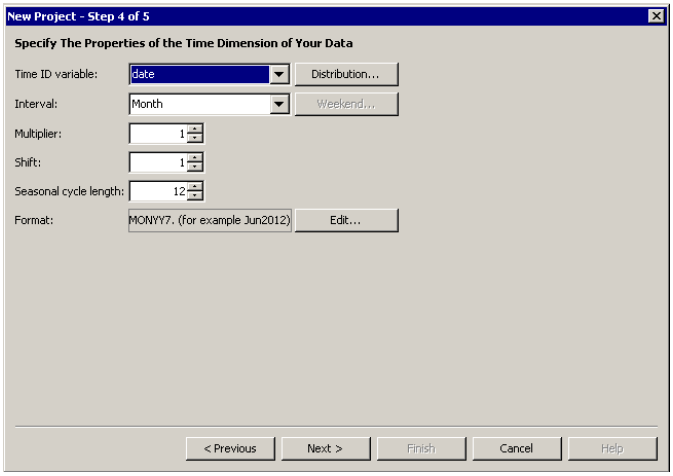


4 To select the BY variables for the project, click **Recommend**. SAS Time Series Studio selects all of the character variables: regionName, productLine, and productName.

Click **Next**.



5 For the Time ID variable, select `date`. Click **Next**.

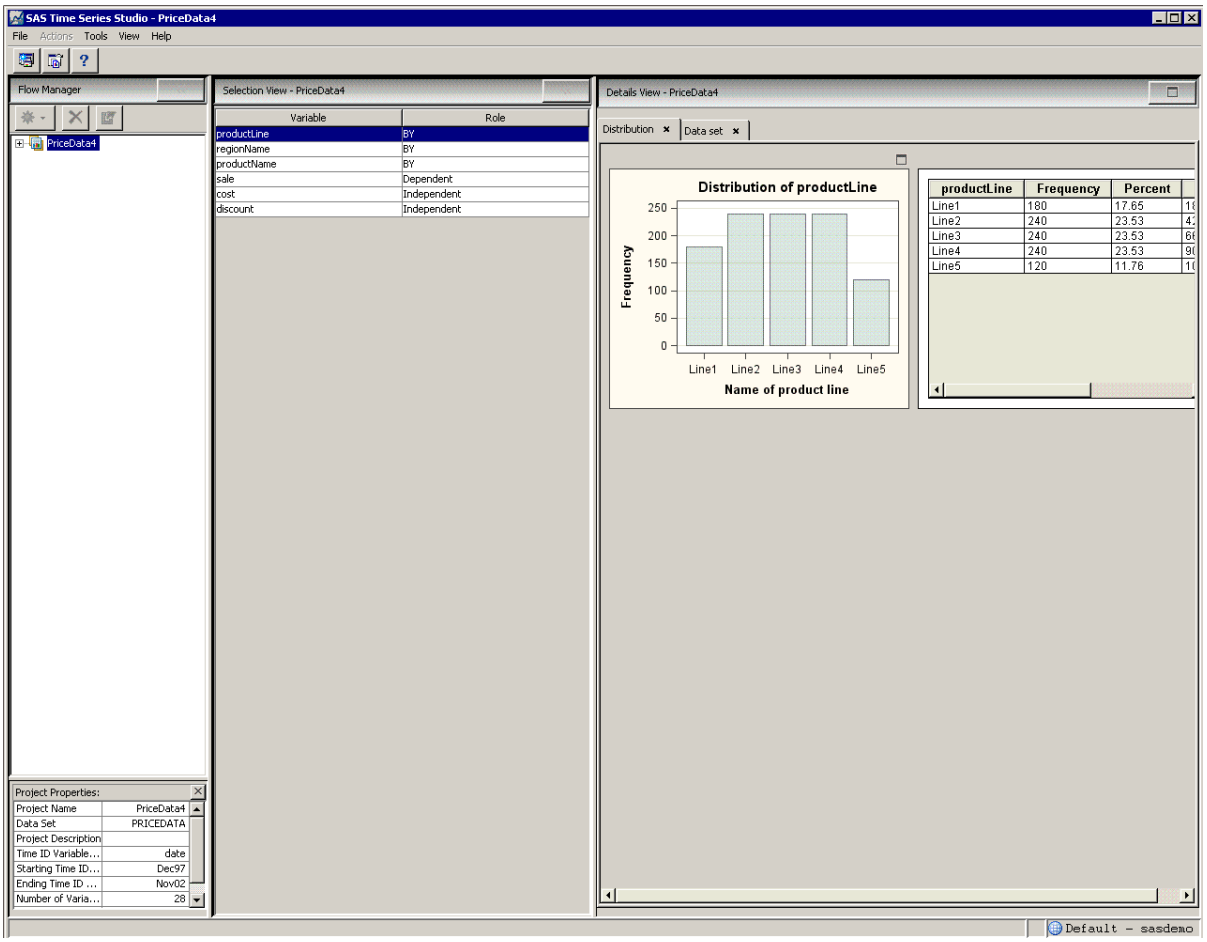


6 In the Specify Analysis Variable Roles and How to Prepare the Data step, assign roles to the following variables:

Variable Name	Role to Assign
cost	Independent
discount	Independent
sales	Dependent

Click **Finish**.

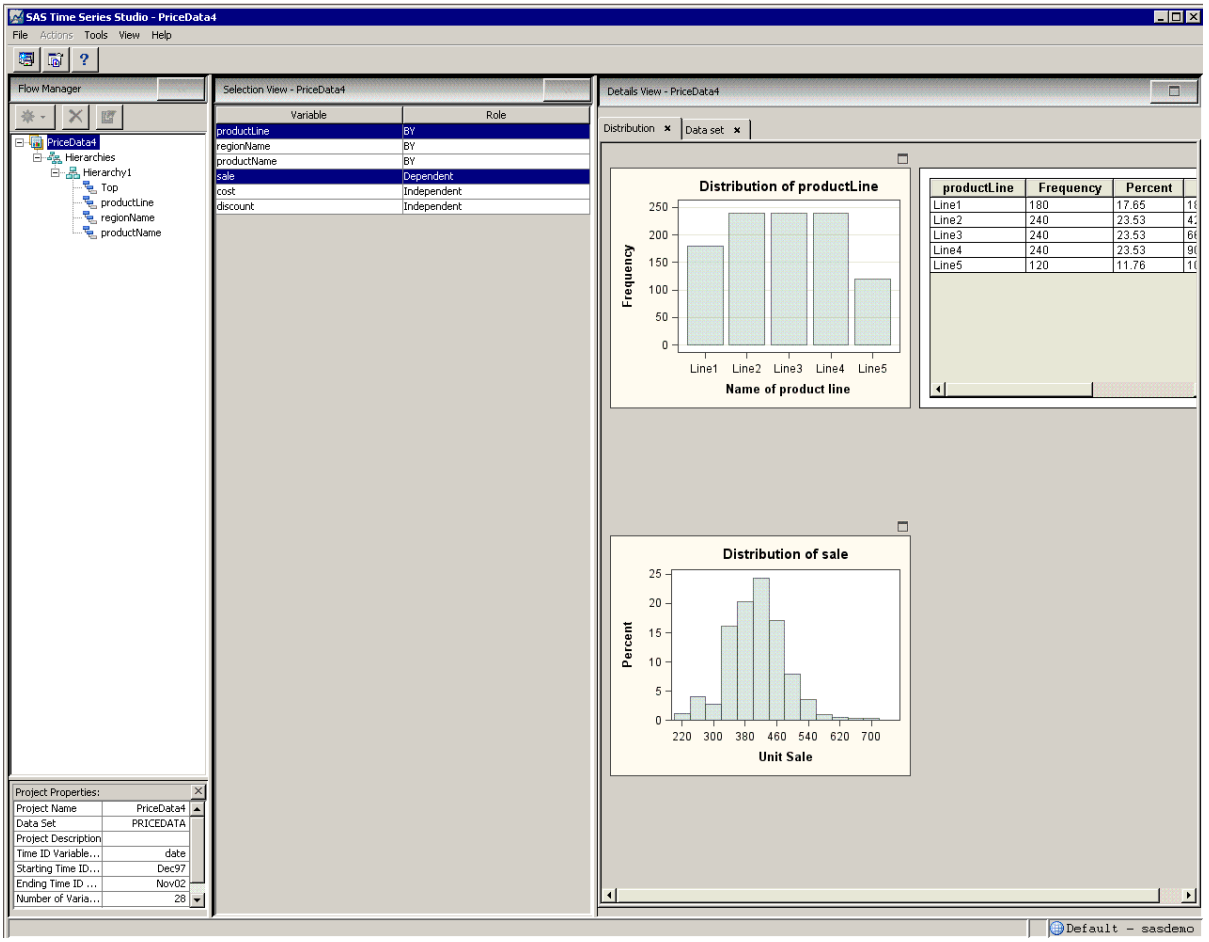
Here are the results in the SAS Time Series Studio workspace:



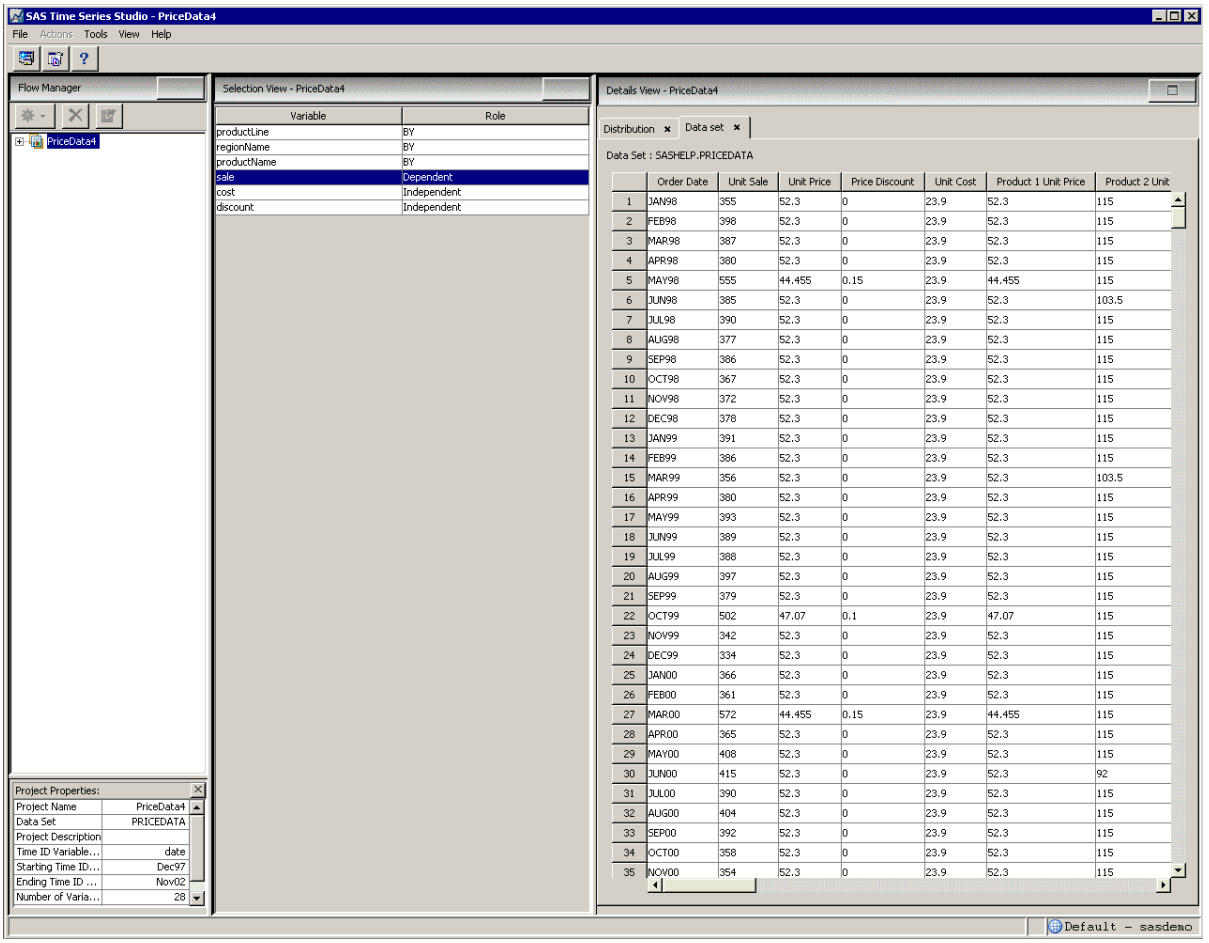
The Selection Manager displays the name and role of each variable in the project. By default, the first variable, productLine, is selected. So in the Distribution tab of the View Manager, you can see a graph of and statistics for the discrete distribution for the productLine variable.

You can view the distributions for multiple variables at the same time. This enables you to compare the distributions visually and to make more informed decisions about how to specify the order of the BY variables in the hierarchy.

To view the distribution for a different variable, press CTRL and select that variable from the Selection Manager. In the following example, the distribution for the sales variable has been added to the Distribution View. This is a continuous distribution.



To view the data for the project, click the **Data set** tab in the View Manager.



Note: When you have the project node selected in the Flow Manager, the Data Set view displays the entire input data set. However, not all of this data might be used in the project. For example, any variables that were assigned a role of Rejected are not included in the project data set. However, these variables are still part of the original data set. If you select a hierarchy node in the Flow Manager, the Data Set view shows only those variables that are included in the project.

Now that you understand the distribution of your data, you might want to structure the data using hierarchies. For more information, see “Creating Hierarchies to Structure Your Data” on page 69.

8

Creating Hierarchies to Structure Your Data

About Hierarchies

You can use hierarchies to structure your data. To create a hierarchy, you must assign at least one classification (BY) variable. These hierarchies can be created when you create the project (by using the options in the New Project wizard). Or you can create a hierarchy by using the options in the Flow Manager.

Create a Hierarchy

To create a hierarchy:

1. In the Flow Manager, expand the project node. Select the **Hierarchies** node and select **Actions** ▶ **Add Hierarchy**. The New Hierarchy dialog box appears.
2. Select the BY variables to include in the hierarchy.
3. (Optional) To specify the time interval and seasonality for the hierarchy, click **Hierarchy Options**. By default, the hierarchy options that you specified when you created the project using the New Project wizard are used.
4. Click **OK**.

Examples: Using Hierarchies to Explore Your Data

About These Examples

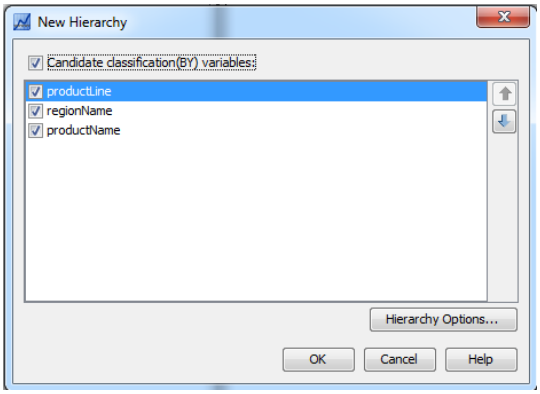
The following steps build on the example for viewing the distribution of your data. (For more information, see “Example: Examine the Distribution of the PriceData Data Set” on page 63.)

In these examples, you create hierarchies of different frequencies and compare the differences in SAS Time Series Studio.

Example 1: Create a Hierarchy with No Additional Options

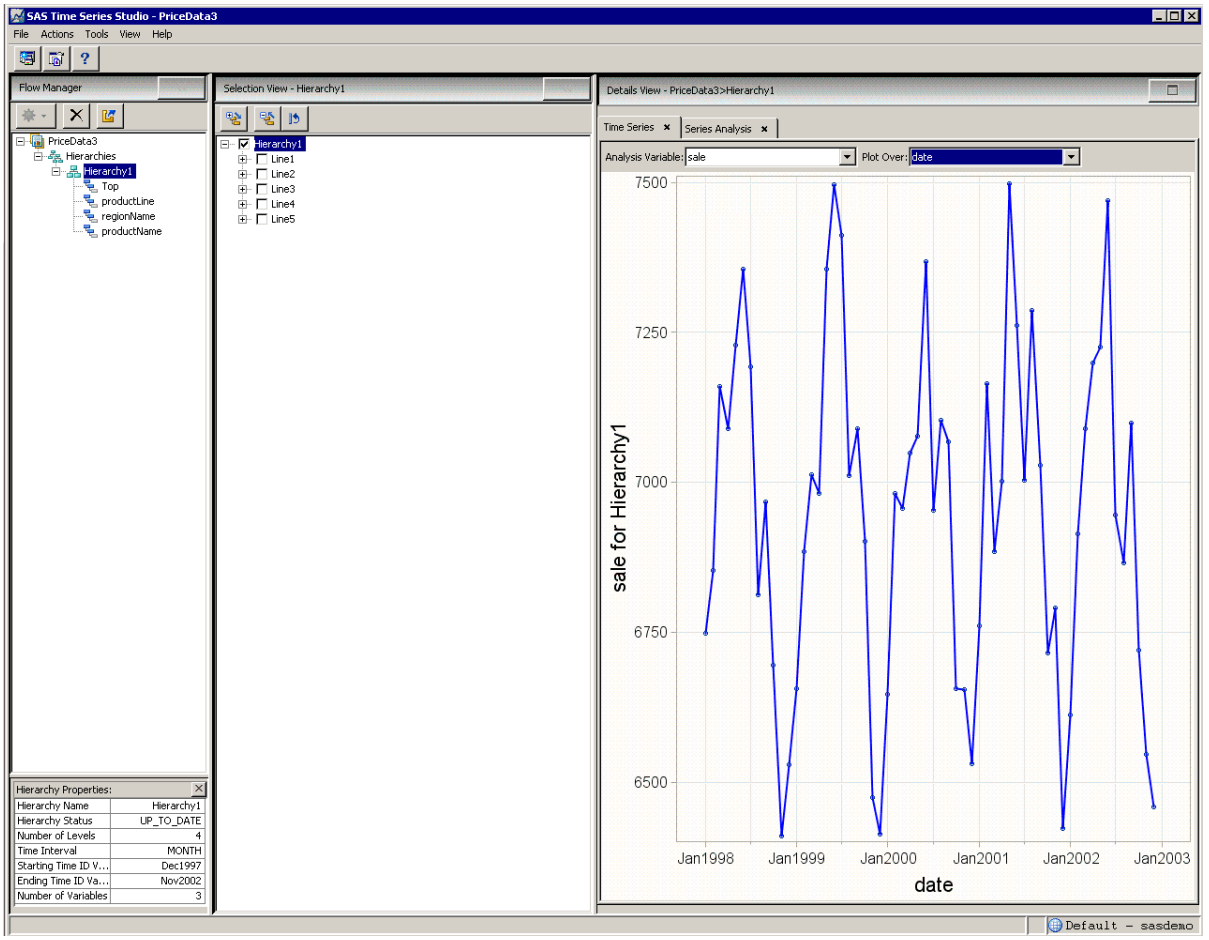
1 In the Flow Manager, select the **Hierarchies** node and select **Tools** ▶ **Add Hierarchy**. The New Hierarchy dialog box appears.

2 Select the **Candidate classification (BY) variables** check box.



3 Click **OK** to add the productLine > regionName > productName hierarchy to the project.

In the workspace, this hierarchy is labeled Hierarchy1.



Example 2: Create a Hierarchy and Specify a Quarterly Time Interval

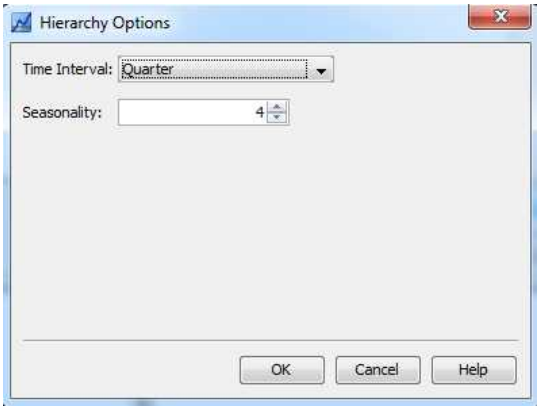
In this example, you create the same hierarchy (productLine > regionName > productName) but change the frequency.

1 In the Flow Manager, select the **Hierarchies** node and select **Tools** ▶ **Add Hierarchy**. The New Hierarchy dialog box appears.

2 Select the **Candidate classification (BY) variables** check box.

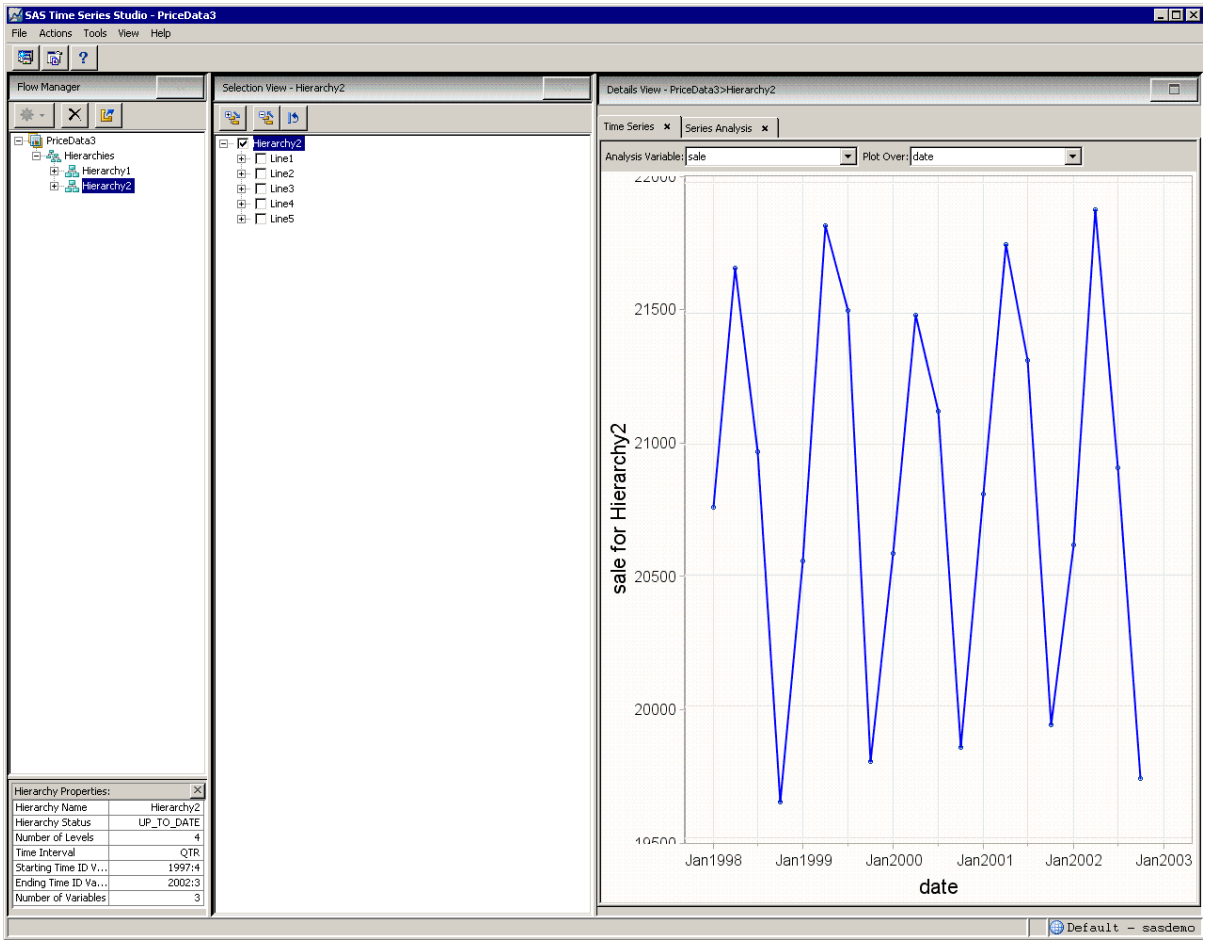
3 Click **Hierarchy Options**. The Hierarchy Options dialog box appears.

4 From the **Time Interval** drop-down list, select **Quarter**. Click **OK**.



5 In the New Hierarchy dialog box, click **OK**.

This hierarchy is called Hierarchy2 in the workspace.



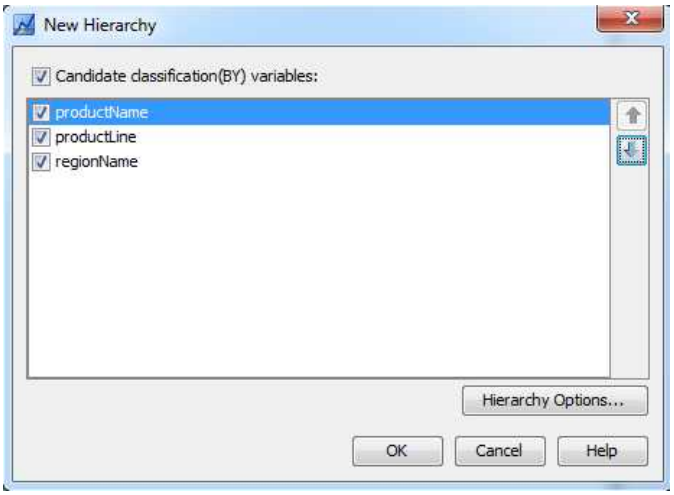
Example 3: Change the Order of the Levels in the Hierarchy

In this example, you want to change the order of the levels in the hierarchy. To do this, you must create a new hierarchy. You cannot change the order of the levels in an existing hierarchy.

1 In the Flow Manager, select the **Hierarchies** node and select **Tools** ▶ **Add Hierarchy**. The New Hierarchy dialog box appears.

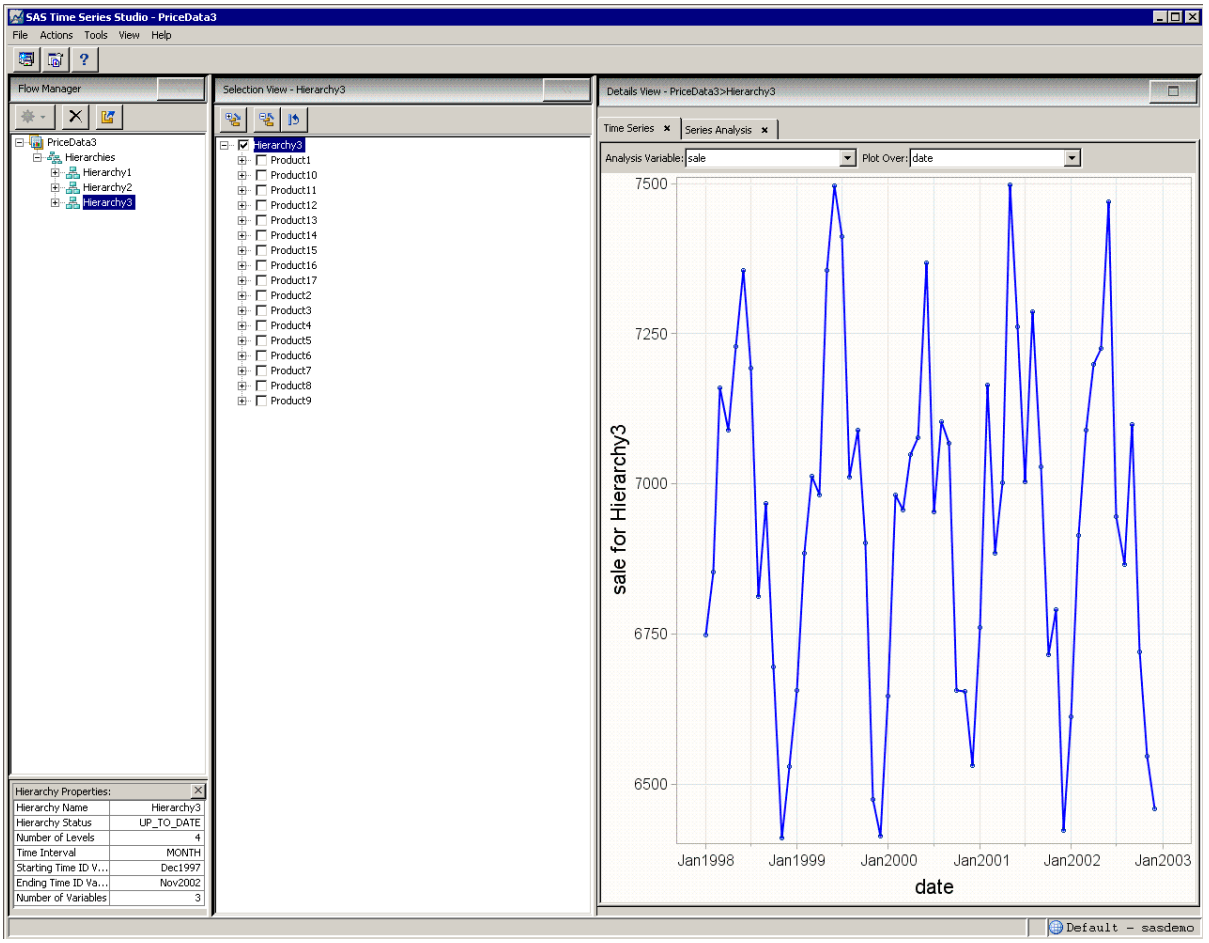
2 Select the **Candidate classification (BY) variables** check box.

3 Select the **productName** level. Click  twice to move the productName level to the top of the list.



4 Click **OK** to add the productName > productLine > regionName hierarchy to the project.

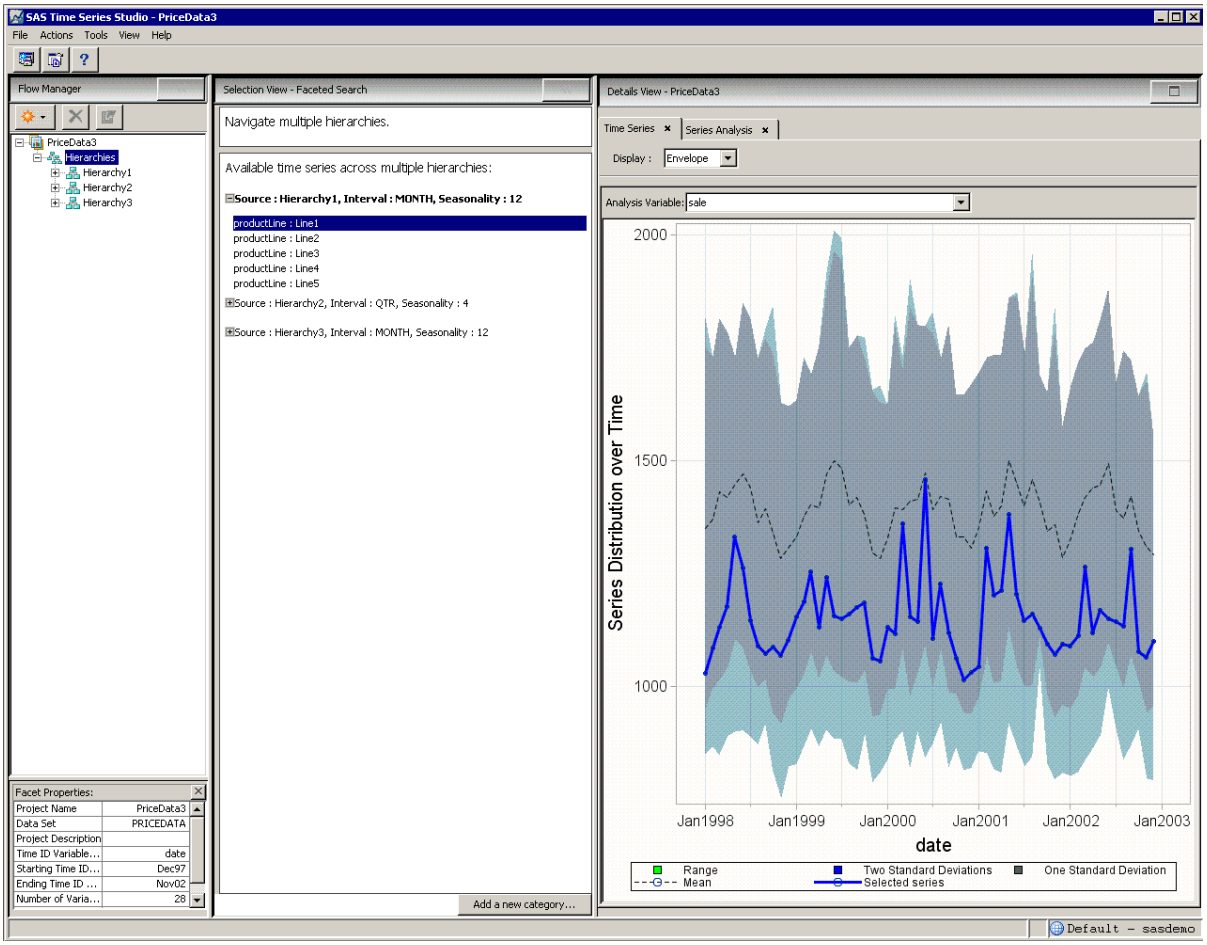
This hierarchy is called Hierarchy3 in the workspace.



Viewing the Results of Several Hierarchies

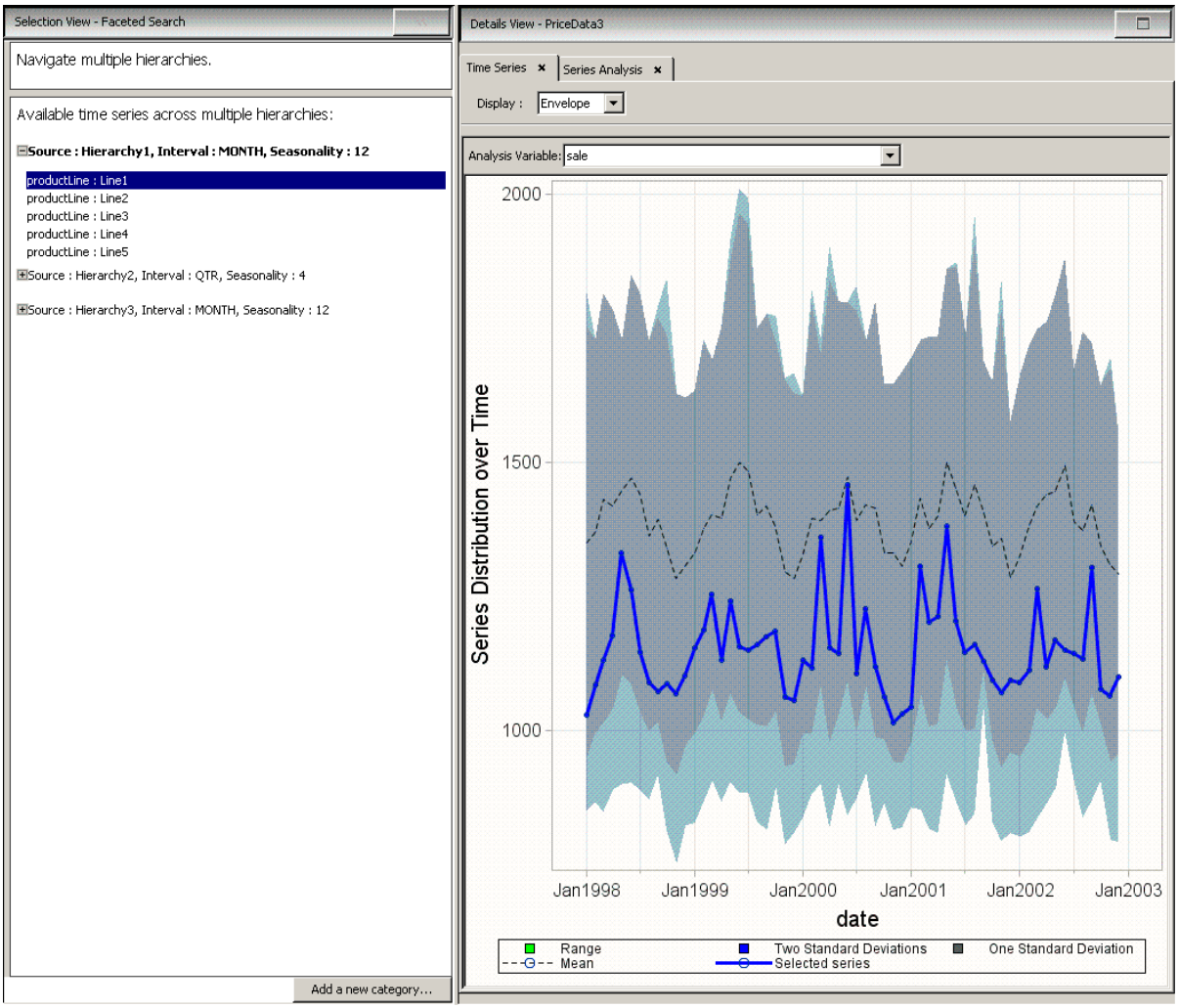
The Top Level Hierarchy

Now, you have added three hierarchies to your project. When you select the **Hierarchies** node in the Flow Manager, you get an overview of all the hierarchies in the Selection View – Faceted Search panel.

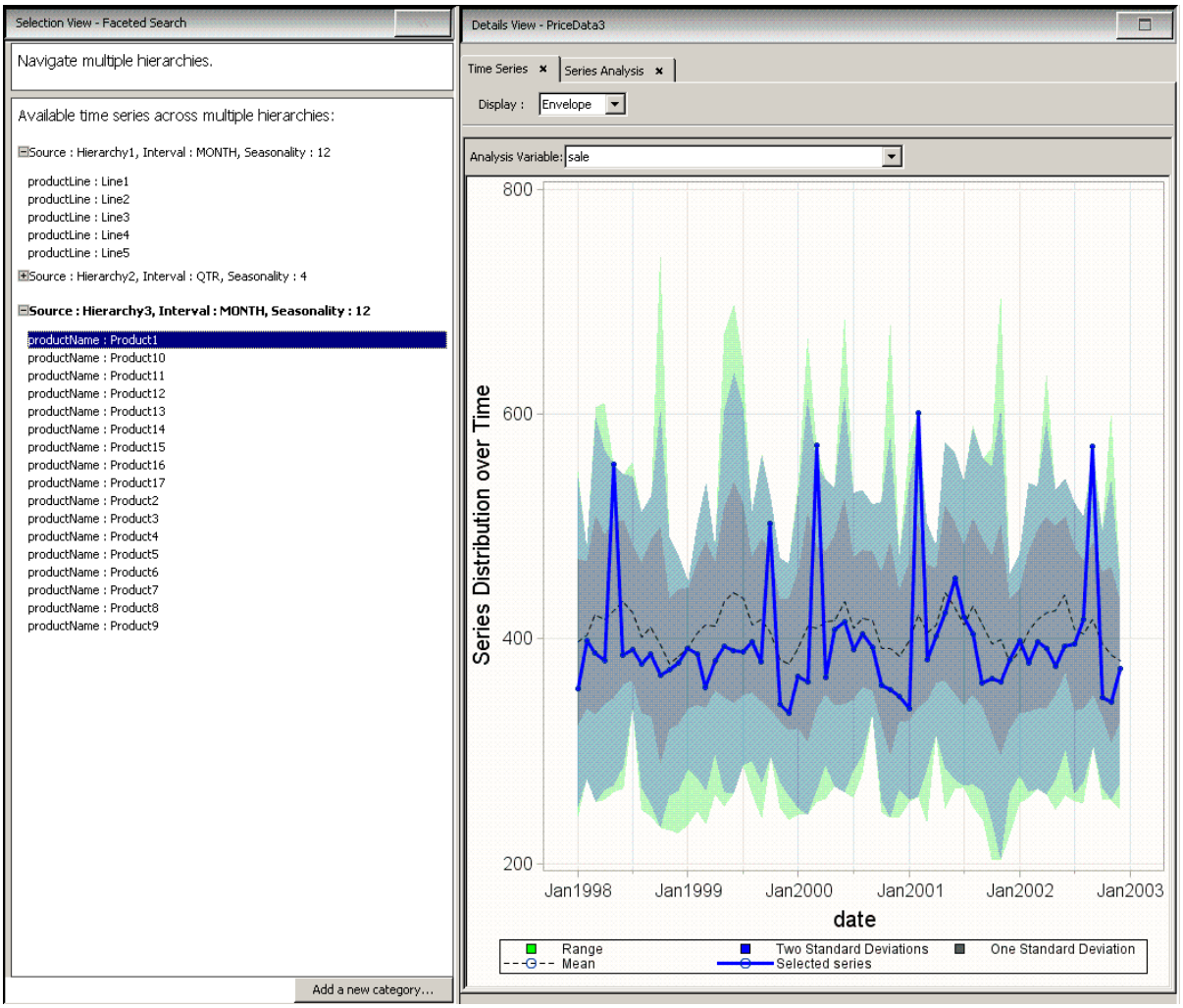


When you select a series in a hierarchy, the graphs in the Details View update to reflect the analyses for the selected series.

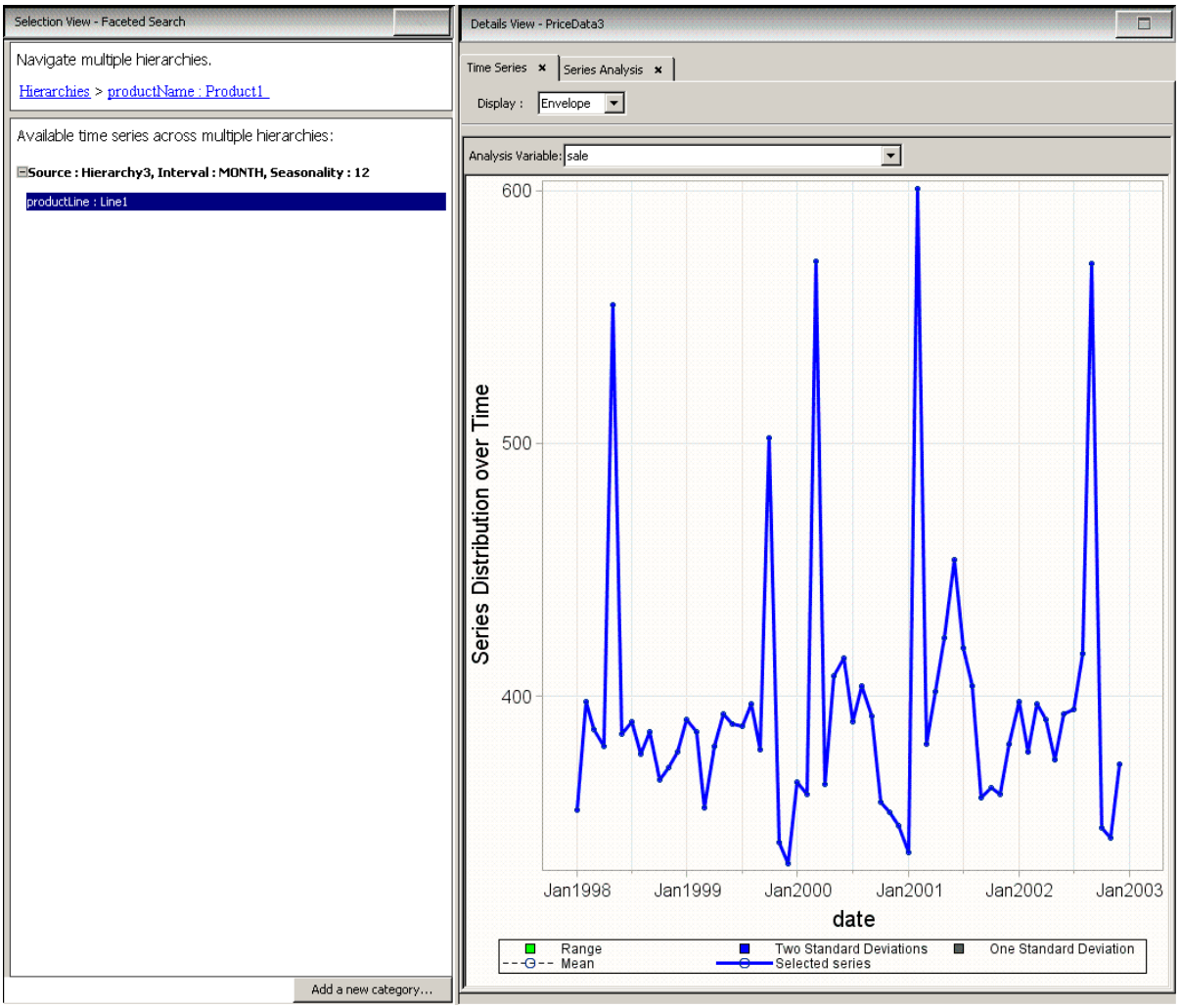
For example, here are the results when you select **productLine: Line1** in Hierarchy1. In the graph for the time series analysis, the selected series appears in blue.



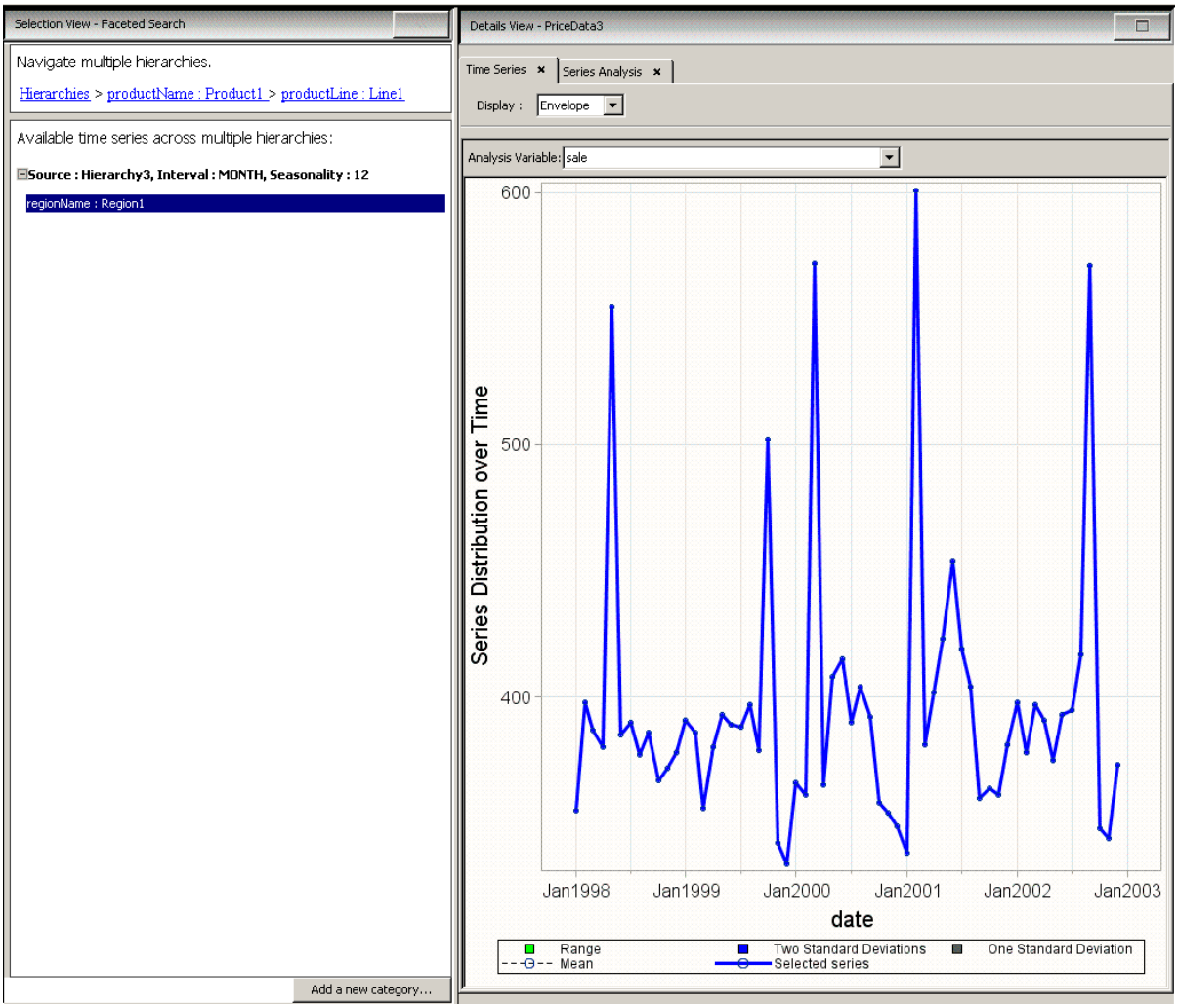
This example shows the productName: Product1 series in Hierarchy3.



If you double-click the name of a hierarchy in the faceted search, you are drilling down through the hierarchy. In this example, you double-click the **productName: Product1** series in Hierarchy3 and get the following results:



You can continue drilling down through the hierarchy until you reach the bottom level.

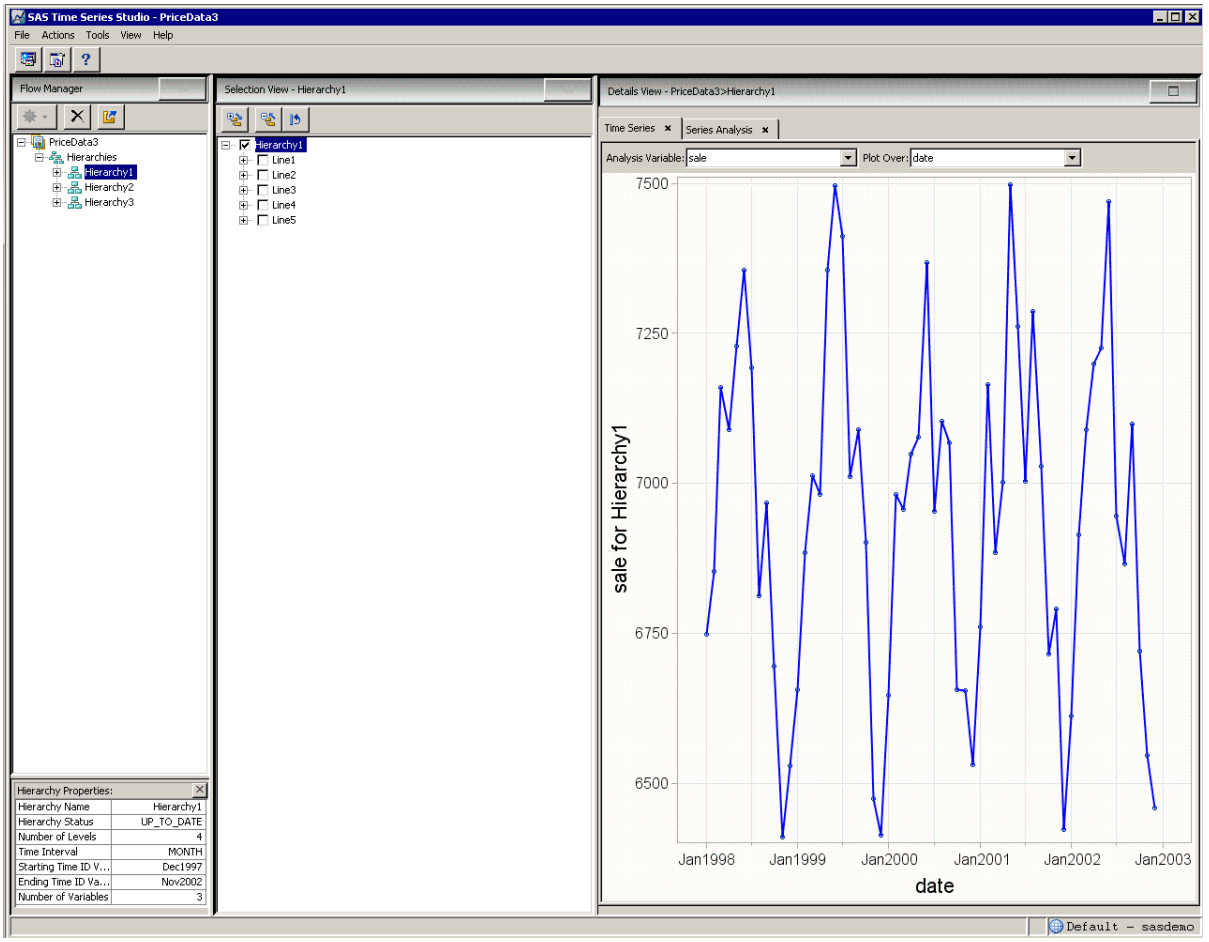


Note: Drilling down through multiple hierarchies is possible only if these hierarchies have the same organization of at least one BY variable.

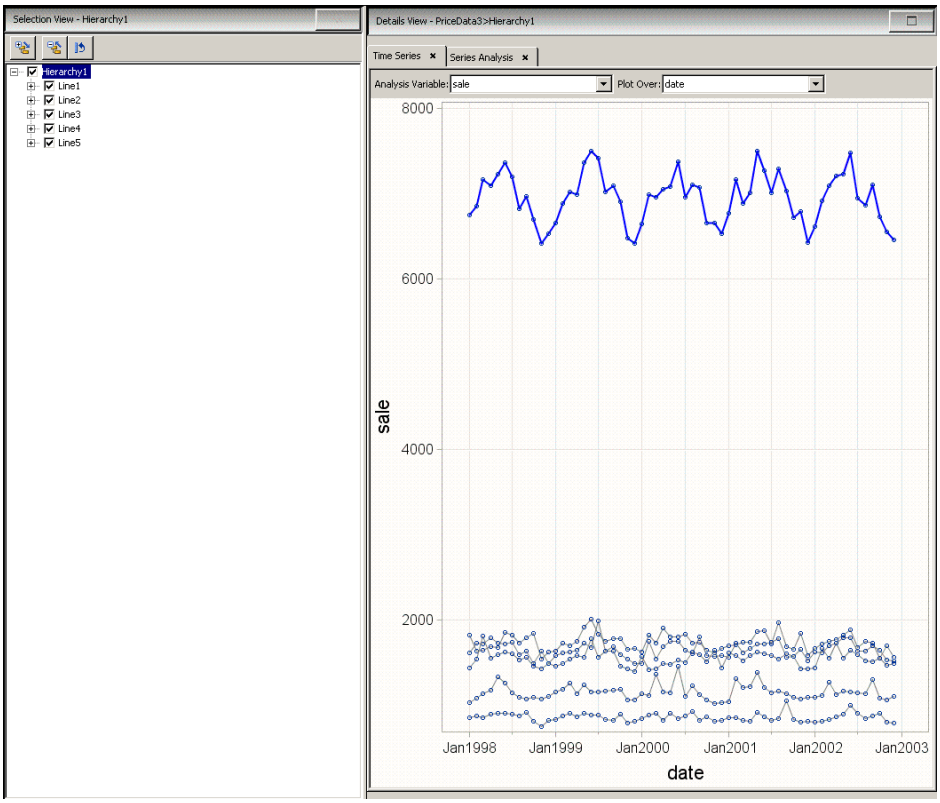
Individual Hierarchies

Using the Flow Manager, you can also view individual hierarchies. For example, if you click the **Hierarchy1**, **Hierarchy2**, or **Hierarchy3** node, then you are seeing the accumulated hierarchy for that node.

Here are the results when you select **Hierarchy1** in the Flow Manager:

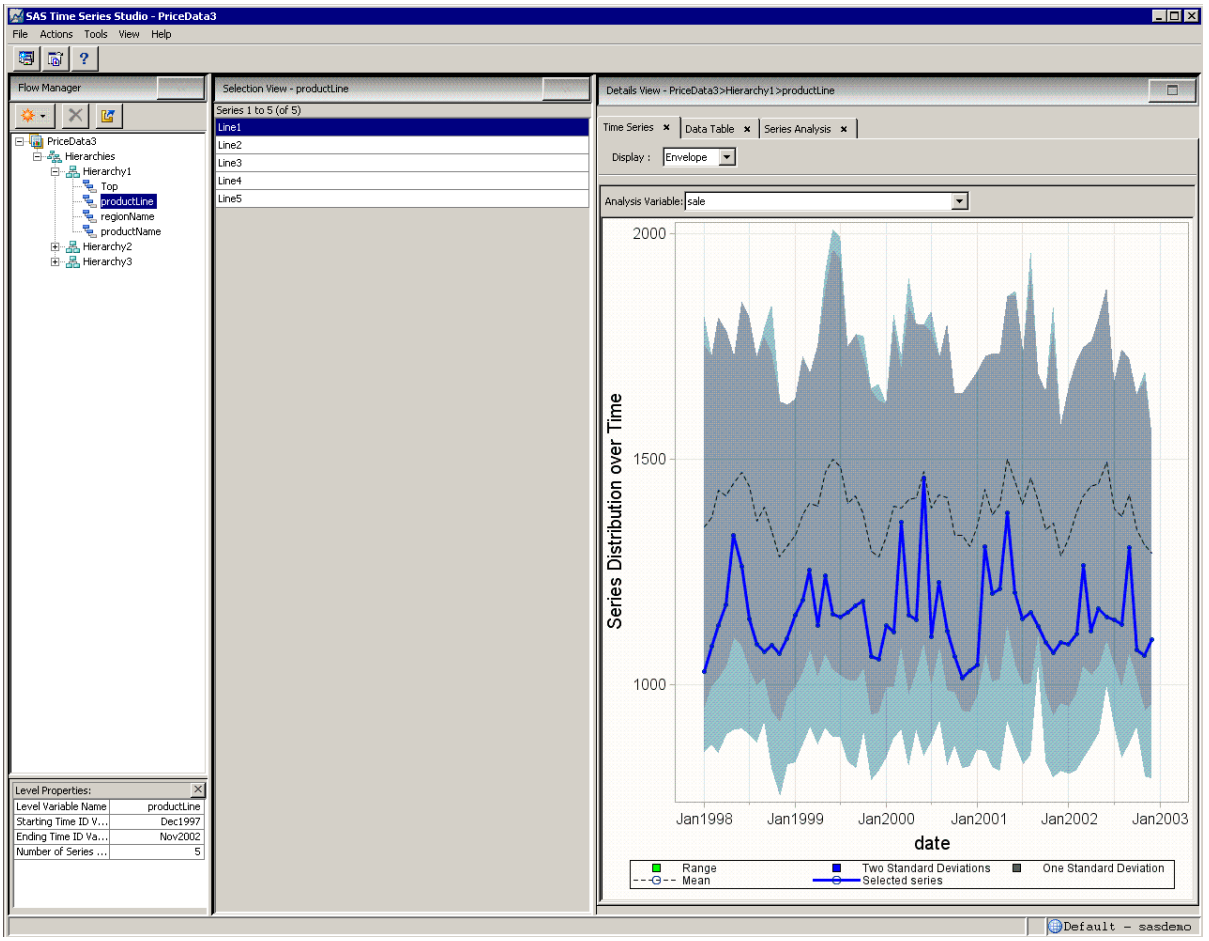


Using the Selection Manager, you can view the series for all of the children in the hierarchy. For example, right-click **Hierarchy1** and select **Select Children**. In the time series graph, the line for Hierarchy1 appears in blue because Hierarchy1 is selected in the Selection Manager.

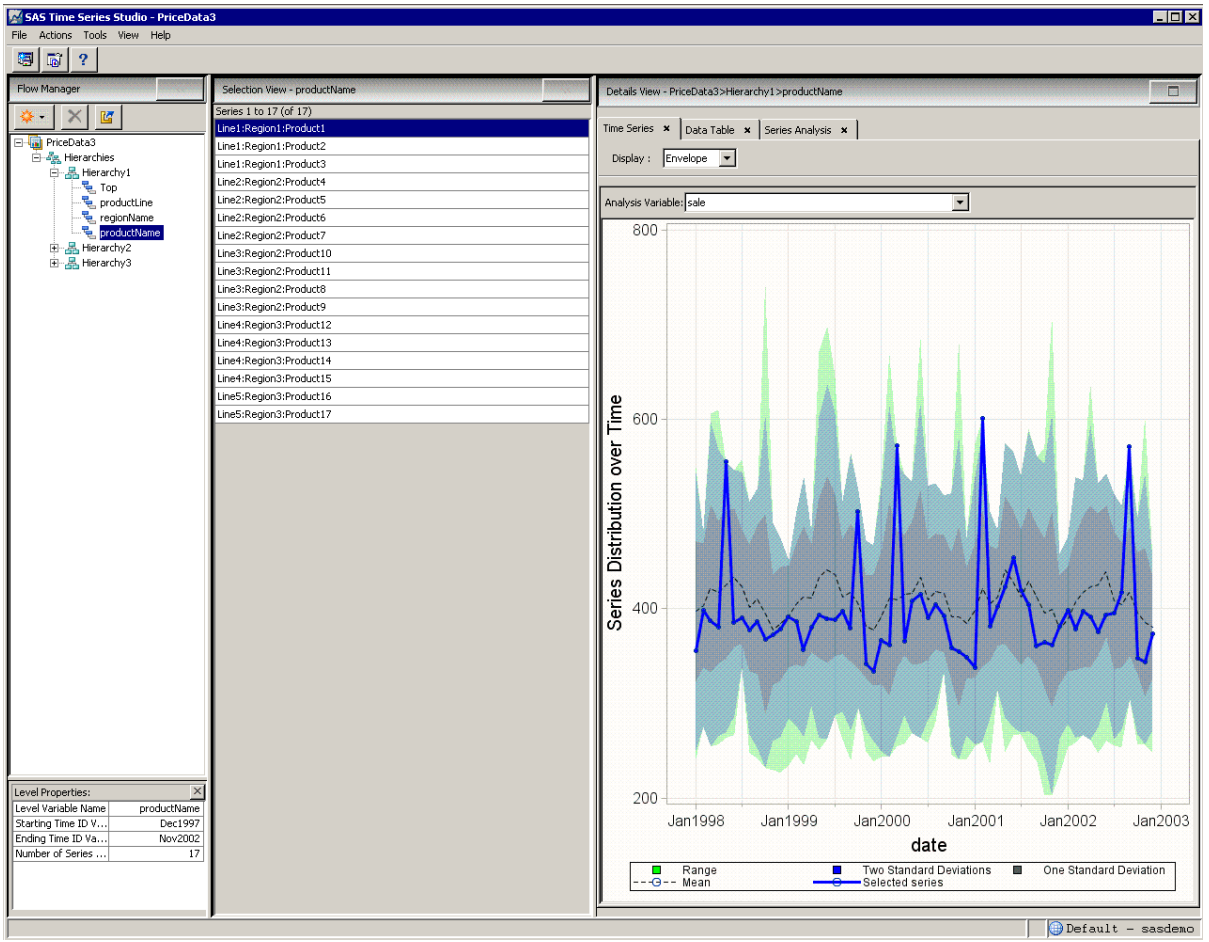


You can also view the data from the levels in the hierarchy. For example, in the Flow Manager, expand the Hierarchy1 node, and you see four levels: Top, productLine, regionName, and productName. The lower that you go in the hierarchy, the more data that appears in the time series graph. By comparing the data from different levels, you can determine how the various “children” impact the aggregation of data at the higher level.

For example, if you select productLine for Hierarchy1 in the Flow Manager, the results show the data for all of the product lines. The blue line in the time series graph is the series that is selected in the HierarchyLevel(productLine) panel.

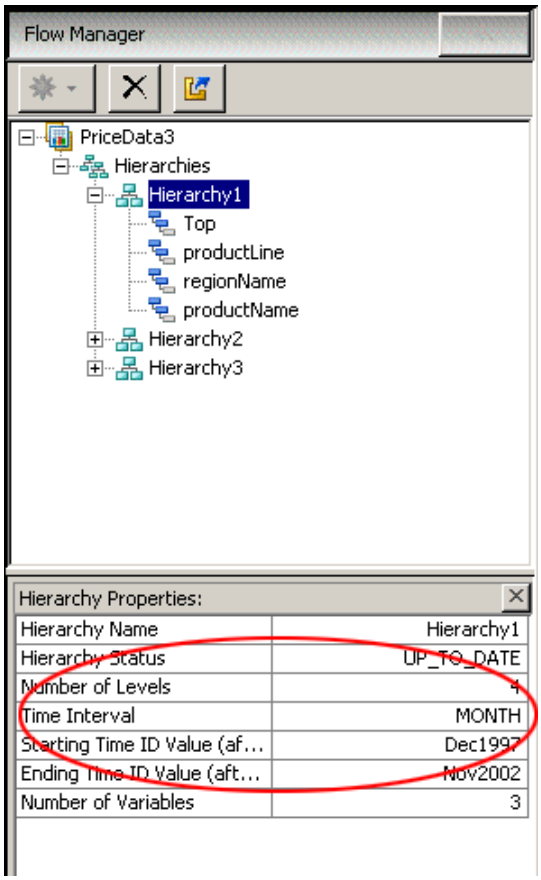


If you select the productName level, then all of the data is displayed in the results.

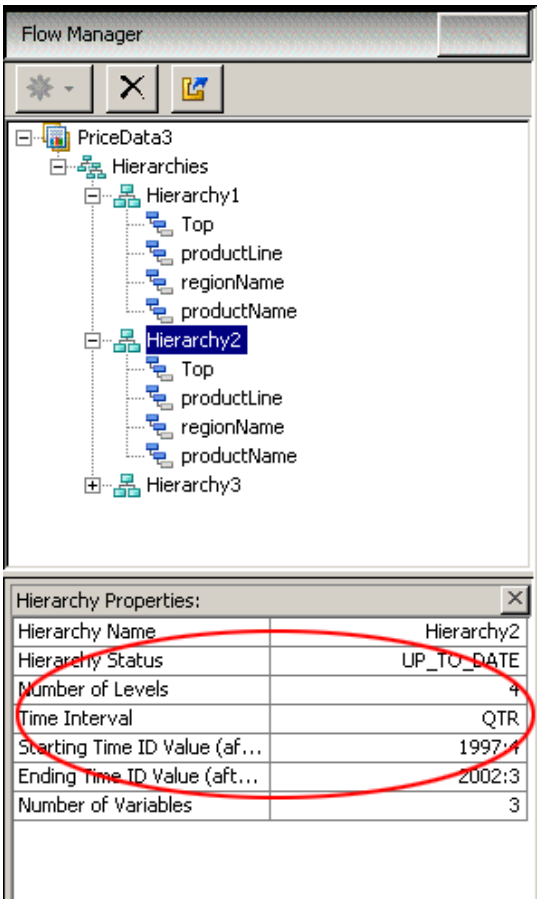


Comparing the Results of Several Hierarchies

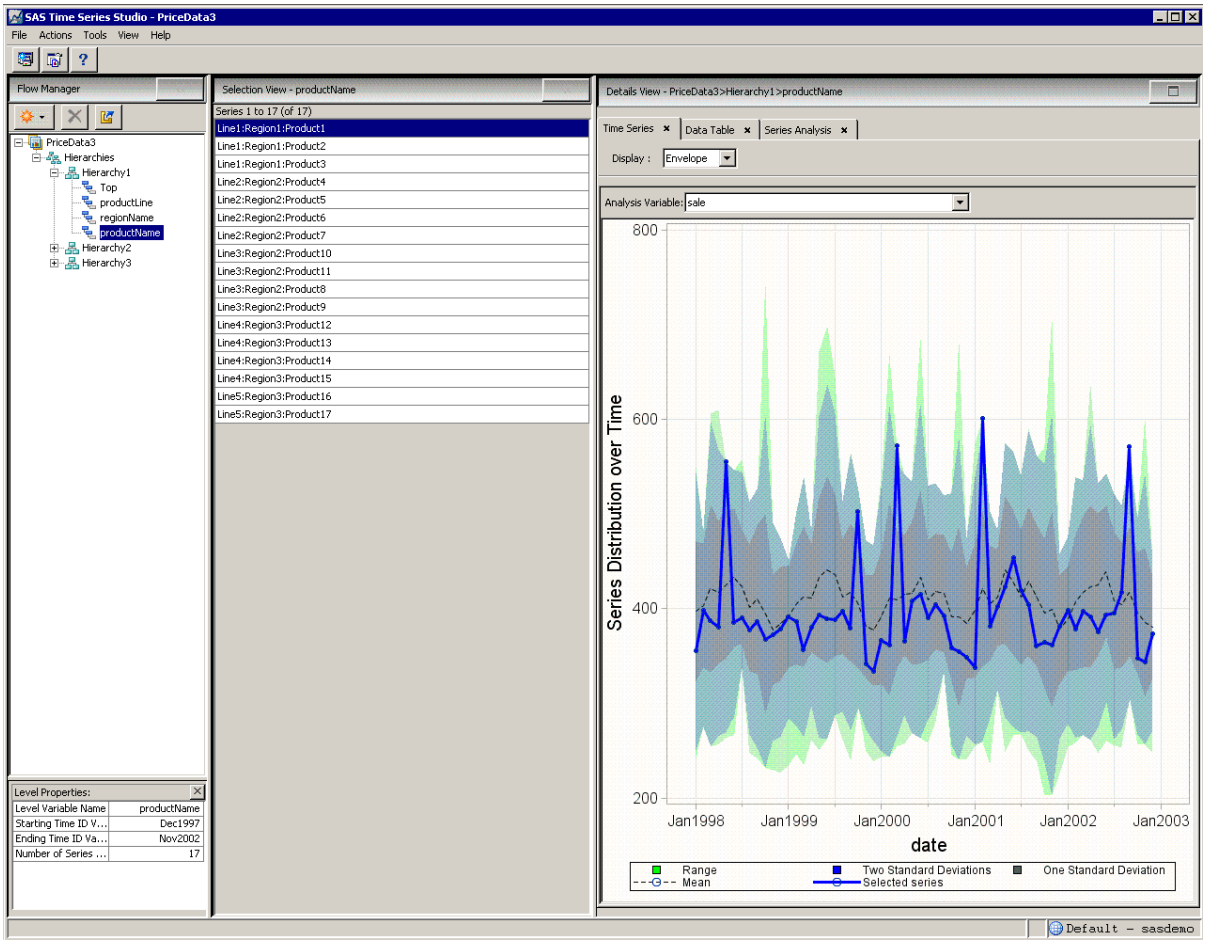
In the example, Hierarchy1 has three classification (BY) variables and is ordered productLine > regionName > productName. The hierarchy uses a monthly time interval with a start date of Dec1997 and an end date of Nov2002.



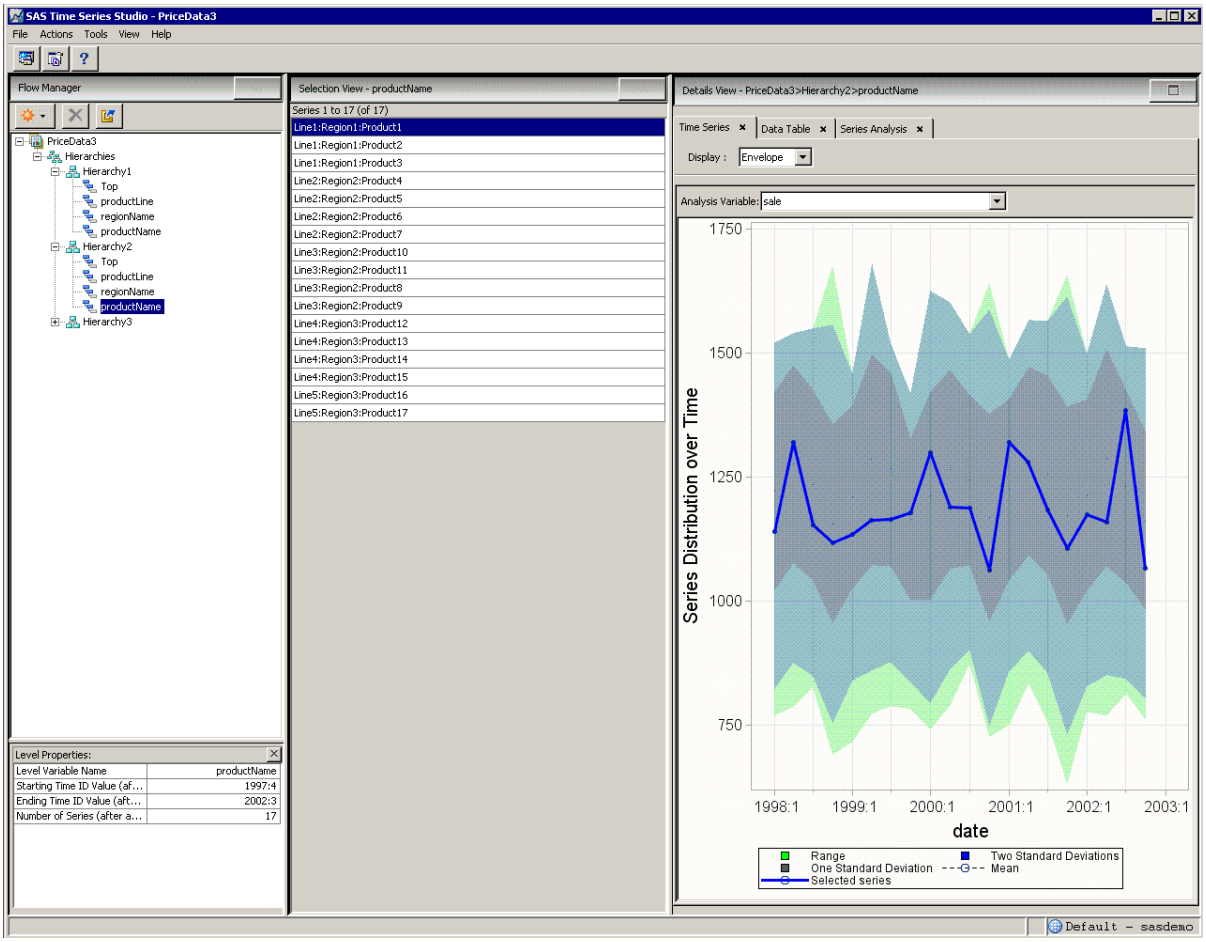
Hierarchy2 has three classification (BY) variables and is ordered by productLine > regionName > productName. This hierarchy uses a quarterly time interval with a start date of 1997:4 (the fourth quarter in 1997) and an end date of 2002:3 (the third quarter in 2002).



You want to see how the change in time interval has affected the results in the productName level. In the Flow Manager, you click productName under Hierarchy1.



Then in the Flow Manager, you select productName under Hierarchy2.



As expected, the time series plots in Hierarchy2 are smoother because the data is recorded at a lower frequency.

9

Understanding Segmentation and Subsets

Understanding Segments

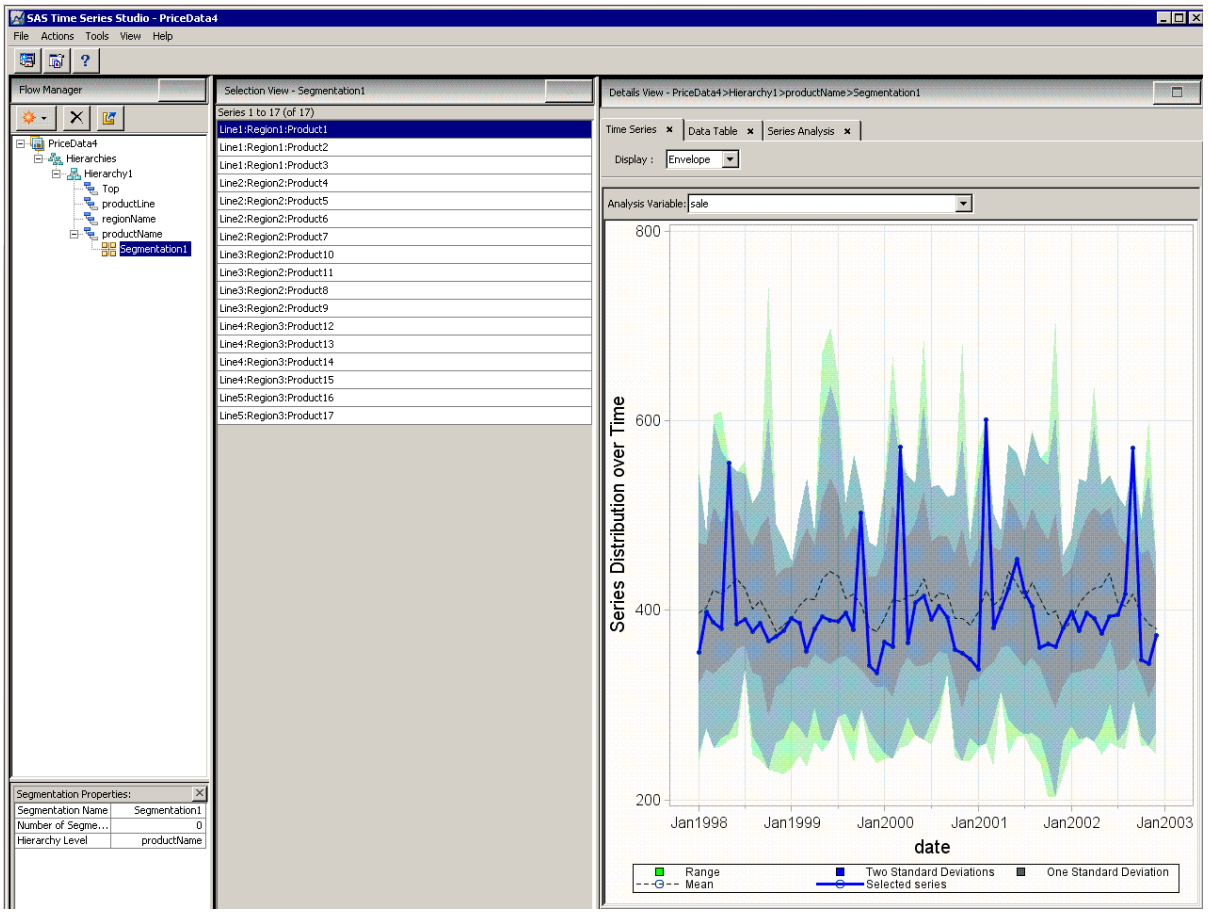
What Is a Segment?

To create a segment, you start with all of the data. When you create queries, you are splitting the data into buckets, but the size of the overall data remains 100%. In SAS Time Series Studio, the data that is not part of the query result is included in the unsegmented bucket.

Example: Creating a Segment in SAS Time Series Studio

When you first create the segment, the segment contains the same number of series as the parent node. In this project, productName contains 17 series. When you add Segmentation1, the result is the same 17 series. To create Segmentation1 node in the Flow Manager, select **productName**, and from the main menu, select **Actions** ▶ **Add Segmentation**.

Here are the 17 series in Segmentation1:



Now, you add a hierarchical query to Segmentation1.

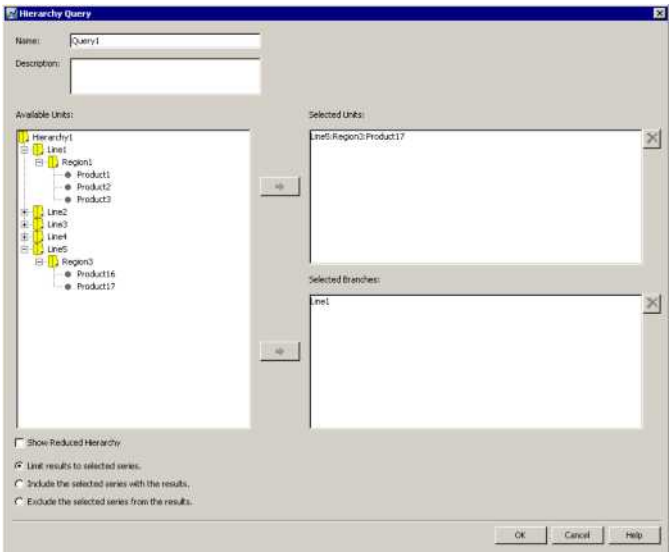
To create this hierarchical query:

1 In the Flow Manager, select the **Segmentation1** node, and from the main menu, select **Actions** ▶ **New Segment** ▶ **Add Hierarchical Query** from the pop-up menu. The Hierarchical Query dialog box appears.

2 Specify the following information for the hierarchical query:

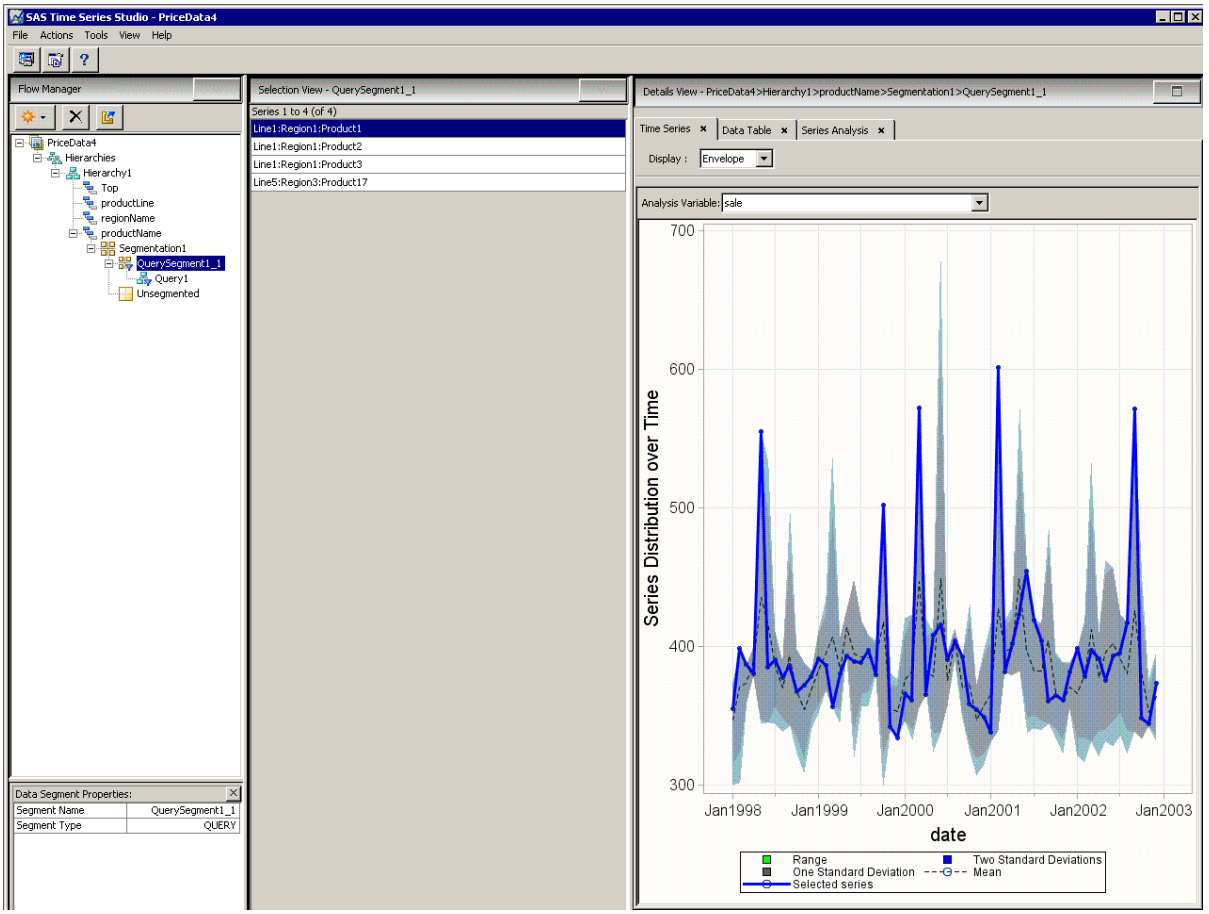
- For the name of the query, enter **`Query1`**.
- From the list of available units, select **Line1** and **Line5:Region3:Product17**.
- Select **Limit results to the selected series**.

When you are finished, the Hierarchy Query dialog box should appear similar to the following:

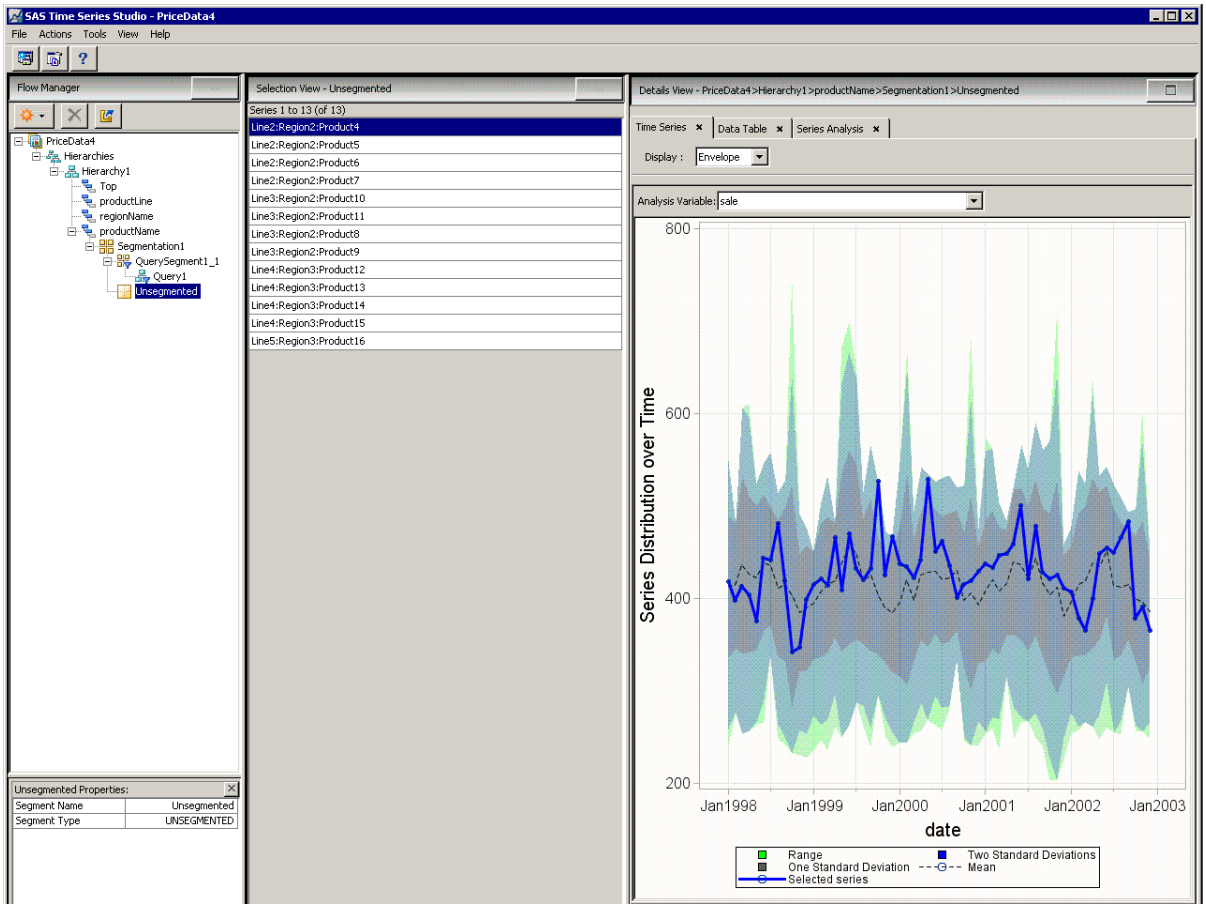


Click **OK**.

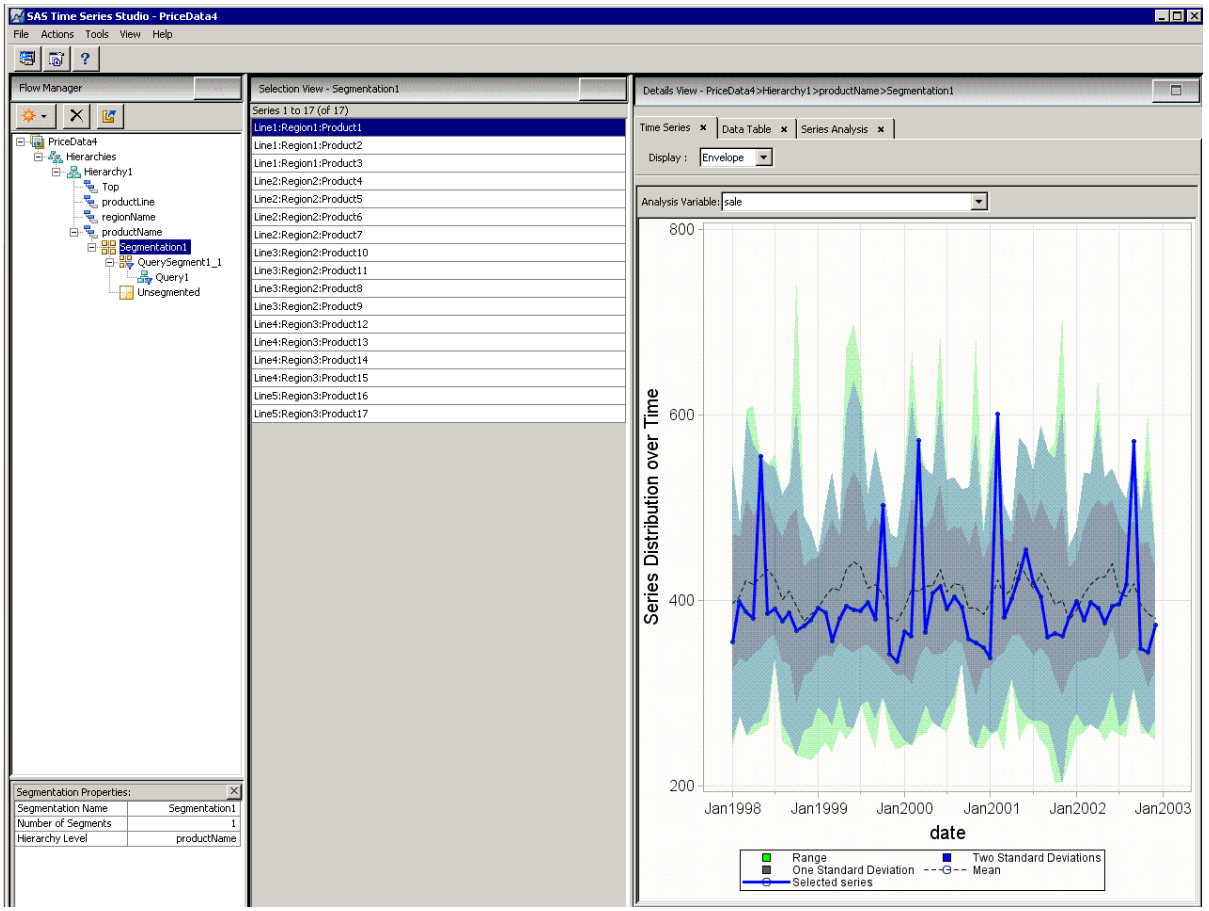
In the Flow Manager, a node for the query and an unsegmented node appear. QuerySegment1_1 contains four series. When you created the query, you specified that the results contain only the results of the query. So the four series in the QuerySegment1_1 node are the only series that met the criteria for the query.


SAS Time Series Studio - PriceData4
File Actions Tools View Help
Flow Manager
PriceData4
Hierarchies
Hierarchy1
Top
productLine
regionName
productName
Segmentation1
QuerySegment1_1
Query1
Unsegmented
Data Segment Properties:
Segment Name QuerySegment1_1
Segment Type QUERY
Selection View - QuerySegment1_1
Series 1 to 4 (of 4)
Line1:Region1:Product1
Line1:Region1:Product2
Line1:Region1:Product3
Line5:Region3:Product17
Details View - PriceData4>Hierarchy1>productName>Segmentation1>QuerySegment1_1
Time Series
Data Table
Series Analysis
Display : Envelope
Analysis Variable: sale
Series Distribution over Time
700
600
500
400
300
Jan1998
Jan1999
Jan2000
Jan2001
Jan2002
Jan2003
date
Range
One Standard Deviation
Selected series
Two Standard Deviations
Mean

The Unsegmented node contains the remaining 13 series, the series that did not meet the criteria for the query.



Segmentation1 contains all of the series in the QuerySegment1_1 node and the Unsegmented node, which in this case is all 17 series.



You decide to add another query to the segmentation.

To create this hierarchical query:

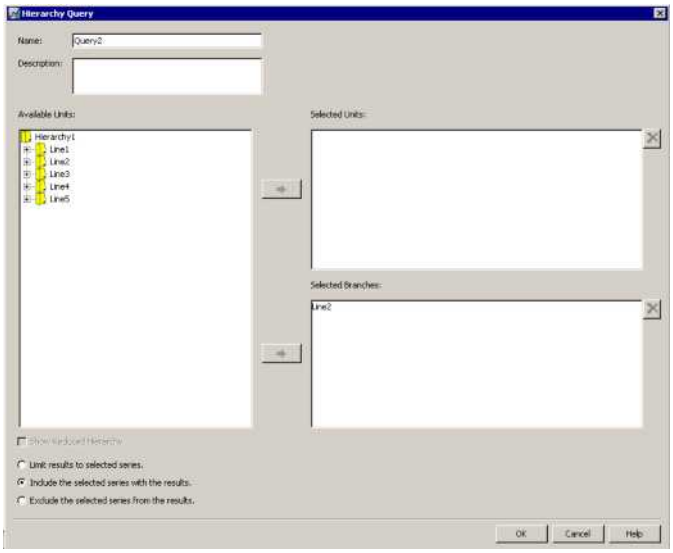
1 In the Flow Manager, select the **Segmentation1** node, and from the main menu, select **Actions** ▶ **New Segment** ▶ **Add Hierarchical Query** from the pop-up menu. The Hierarchical Query dialog box appears.

2 Specify the following information for the hierarchical query:

- For the name of the query, enter `Query2`.
- From the list of available units, select **Line2**.

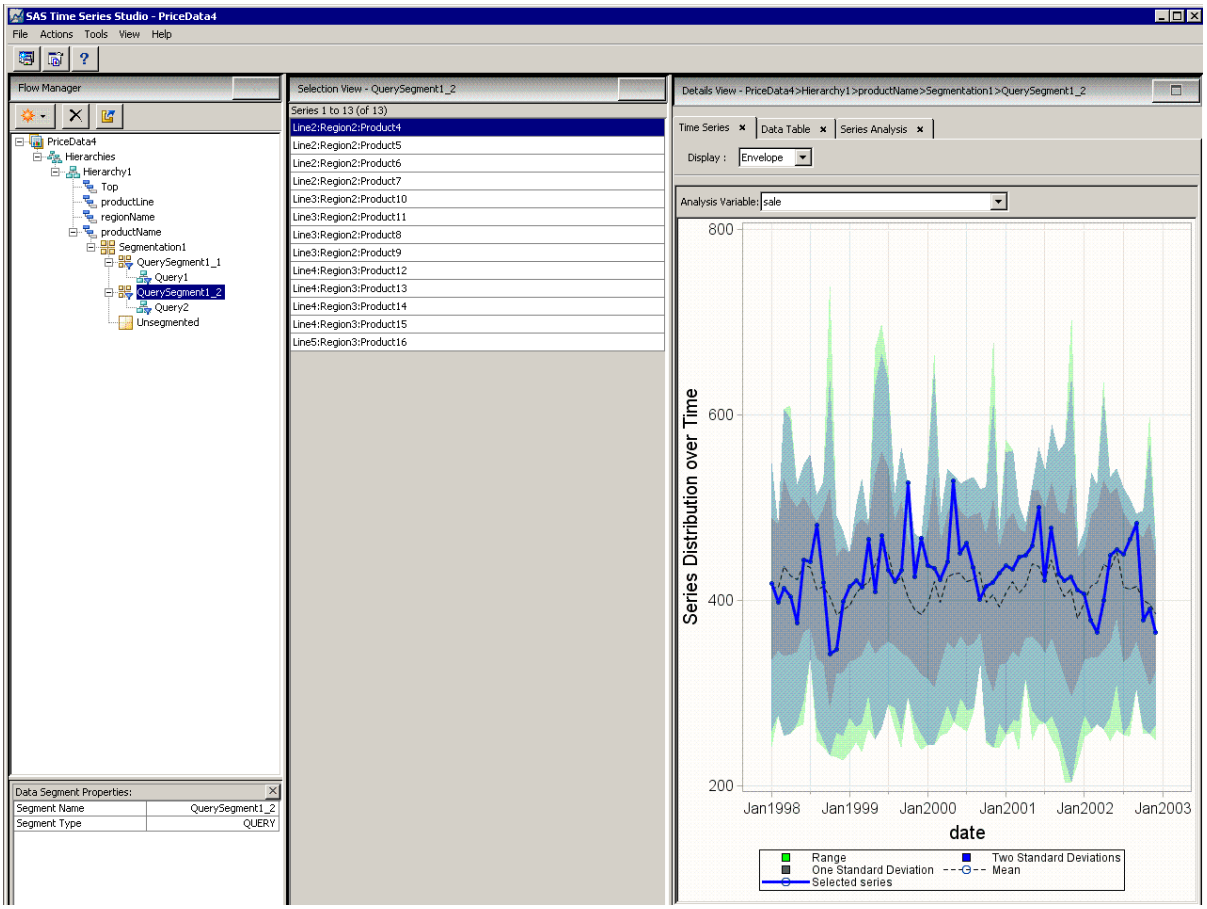
- Select **Include the selected series with the results**.

When you are finished, the Hierarchy Query dialog box should appear similar to the following:



Click **OK**.

In the Flow Manager, a node for the query and an unsegmented node appear. QuerySegment1_2 contains 13 series. When you created the query, you specified that the selected series (the series that met the criteria) should be included with the results, so 13 series are shown in total.

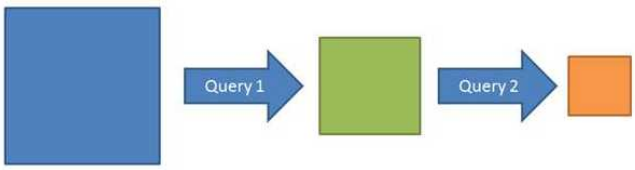


The Unsegmented node still contains 13 series, the series that did not meet the criteria for query 1, and the Segmentation1 node contains all 17 series.

Understanding Subsets

What Is a Subset?

To create a subset, you start with all of the data. When you define a query, you are filtering out the data that you are not interested in. The result of the query will contain only a portion of the original data.

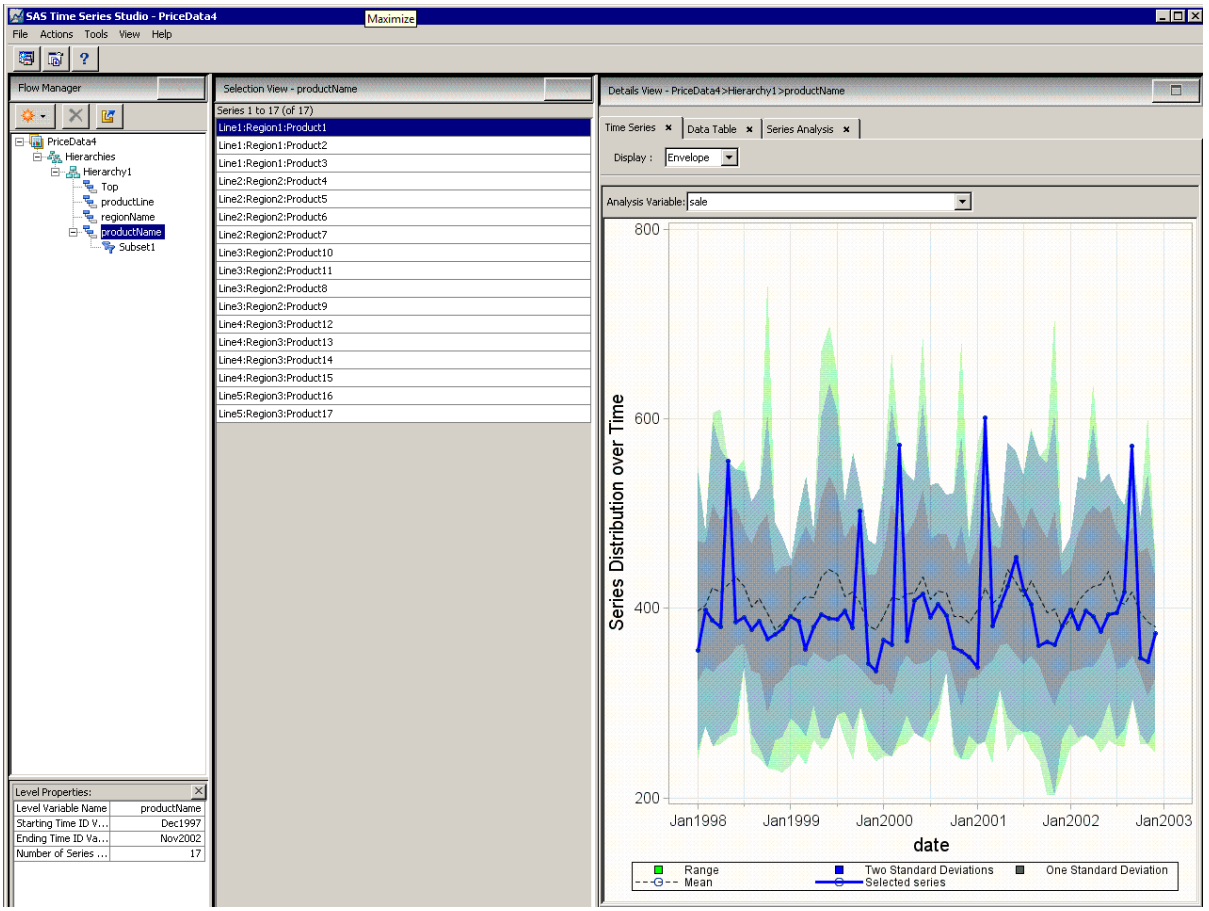


Example: Creating Subsets in SAS Time Series Studio

In SAS Time Series Studio, a subset could consist of individual queries. How those queries are combined depends on whether you specify to limit the results to the series in the query, exclude the selected series from the results, or include the selected series in the results.

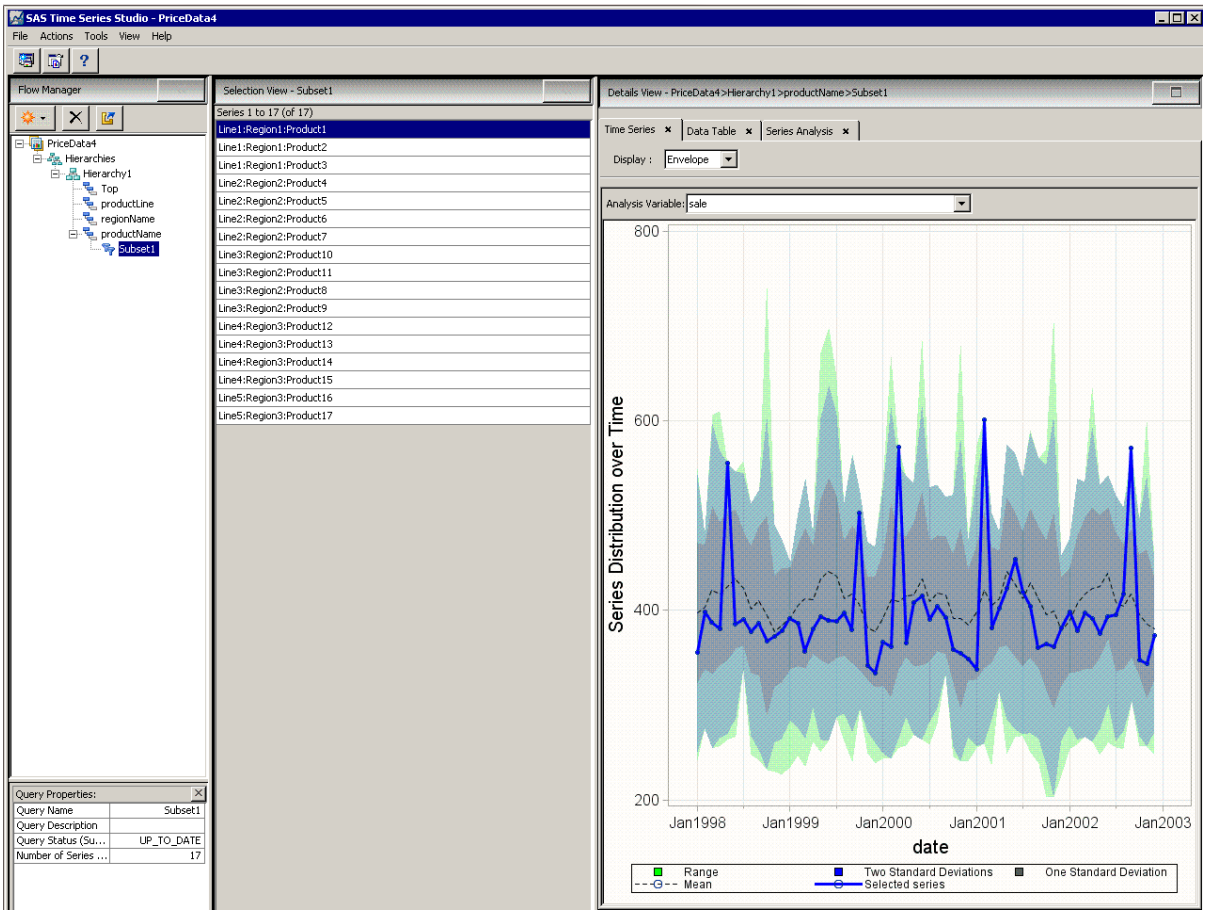
When you first create a subset, the subset contains the same number of series as the parent node. In the following example, productName and Subset1 contain the same number of series (17). You can view the number of series from the Selection Manager.

Here are the 17 series for productName:



To create the subset, select the **productName** node, and from the main menu, select **Actions** ▶ **Add Subset**. Subset1 is added to the Flow Manager.

Here are the 17 series for Subset1:



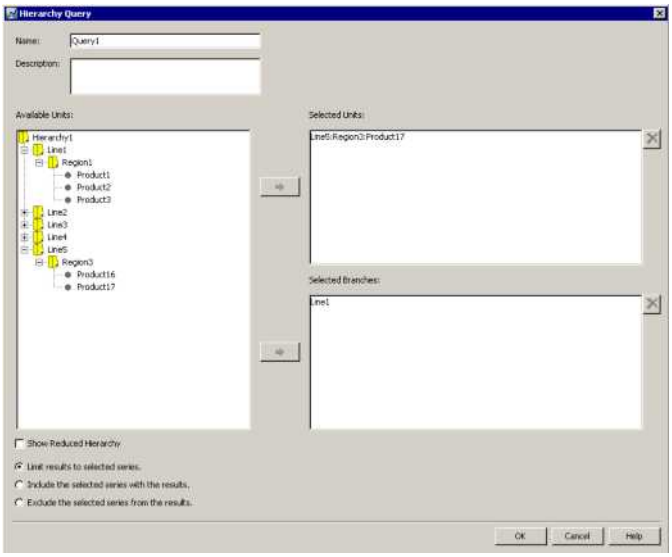
To create this hierarchical query:

1 In the Flow Manager, select the **Subset1** node, and from the main menu, select **Actions** ▶ **Add Hierarchical Query** from the pop-up menu. The Hierarchical Query dialog box appears.

2 Specify the following information for the hierarchical query:

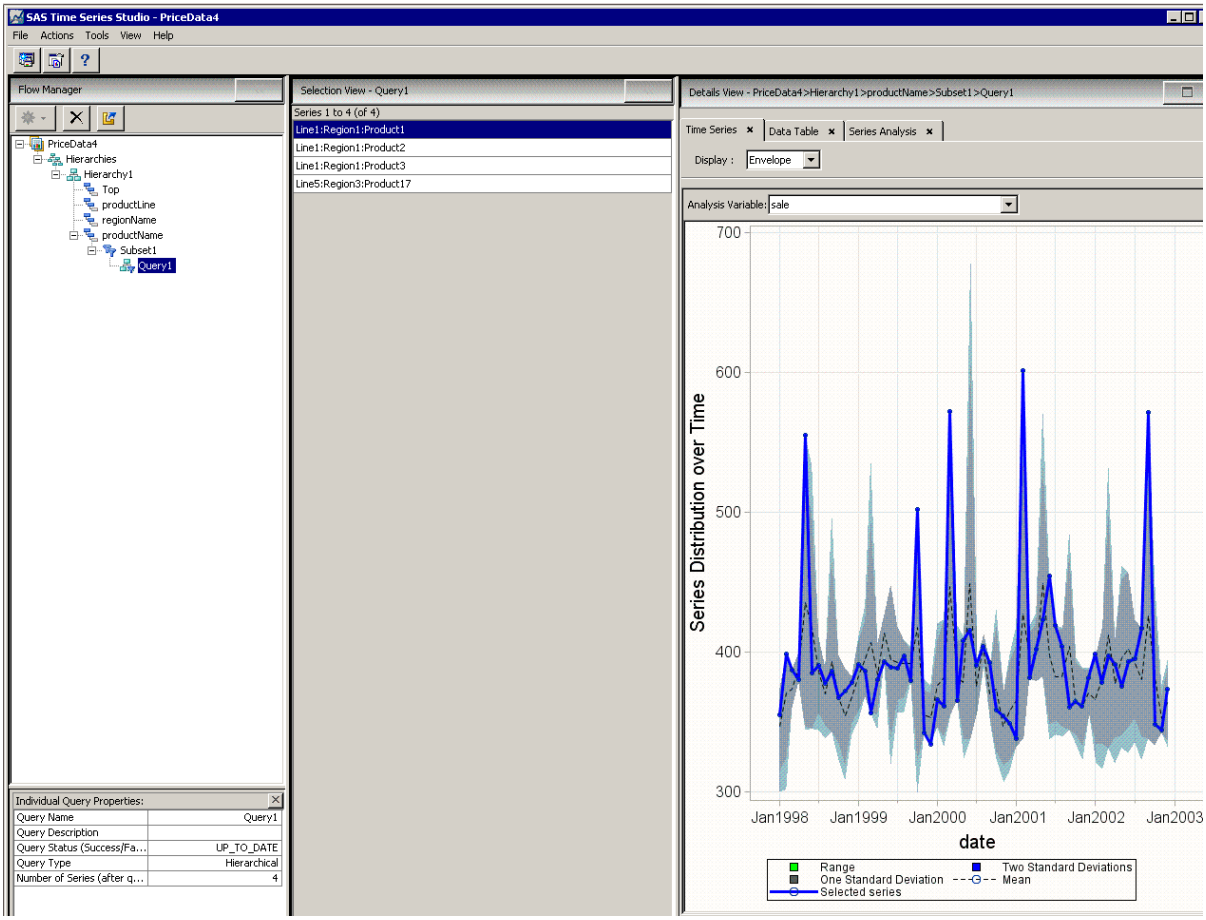
- For the name of the query, enter `Query1`.
- From the list of available units, select **Line1** and **Line5:Region3:Product17**.
- Select **Limit results to the selected series**.

When you are finished, the Hierarchy Query dialog box should appear similar to the following:



Click **OK**.

When you created the query, you specified that the results should contain only the series from that query, so Query1 contains only the four series that met the criteria.



The Subset1 node still contains 17 series.

You decide to add another query to the subset.

To create this hierarchical query:

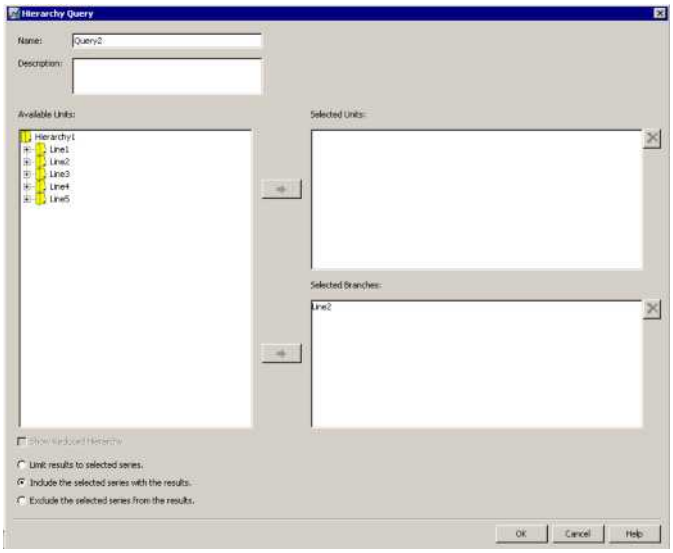
1 In the Flow Manager, select the **Subset1** node, and from the main menu, select **Actions** ▶ **Add Hierarchical Query** from the pop-up menu. The Hierarchical Query dialog box appears.

2 Specify the following information for the hierarchical query:

- For the name of the query, enter `Query2`.
- From the list of available units, select **Line2**.

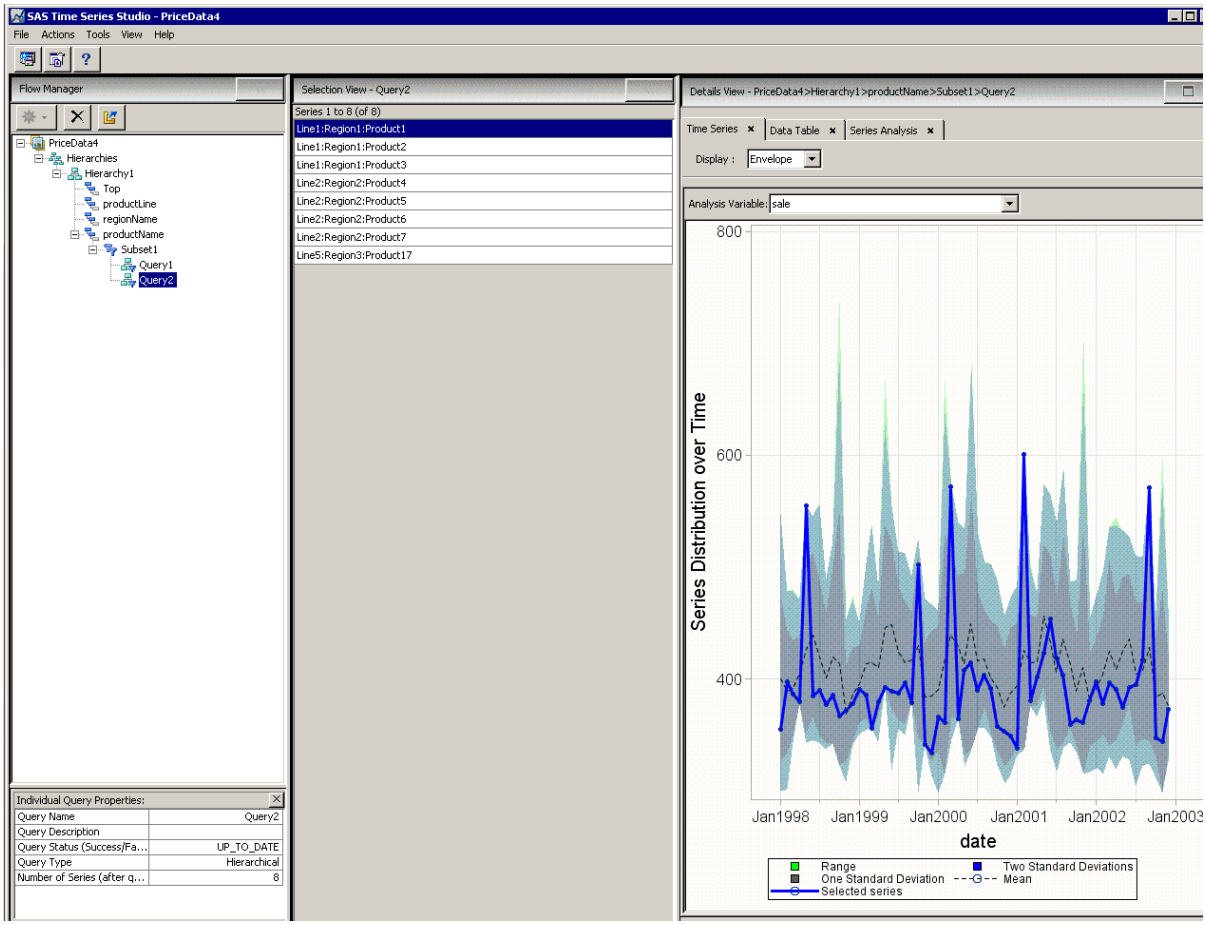
- Select **Include the selected series with the results**.

When you are finished, the Hierarchy Query dialog box should appear similar to the following:

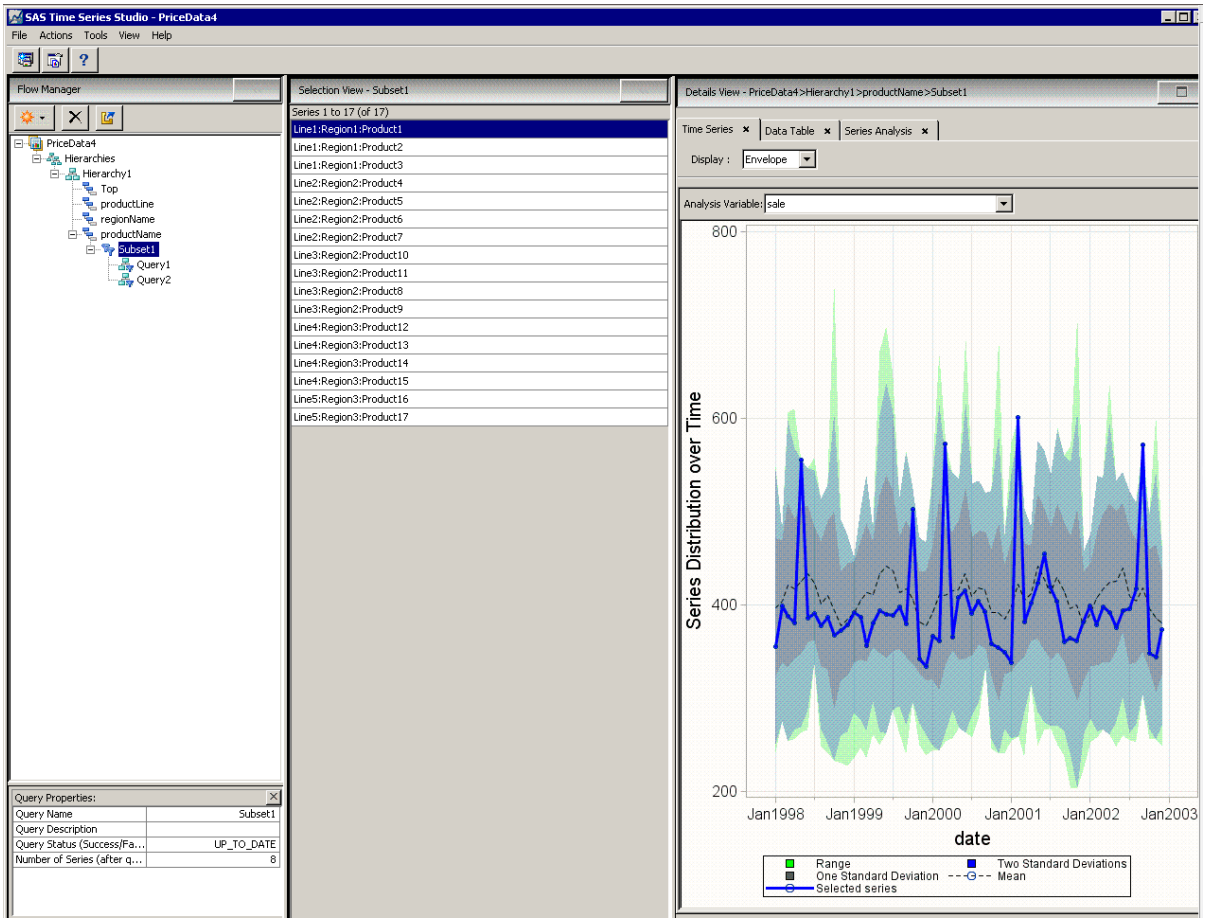


Click **OK**.

Because you specified that the results from second query should be included with the results, Query2 contains eight series: the four series for Line2 and the four series that met the criteria for the first query.


SAS Time Series Studio - PriceData4
File Actions Tools View Help
Flow Manager
PriceData4
Hierarchies
Hierarchy1
Top
productLine
regionName
productName
Subset1
Query1
Query2
Individual Query Properties:
Query Name Query2
Query Description
Query Status (Success/Fa... UP_TO_DATE
Query Type Hierarchical
Number of Series (after q... 8
Selection View - Query2
Series 1 to 8 (of 8)
Line1:Region1:Product1
Line1:Region1:Product2
Line1:Region1:Product3
Line2:Region2:Product4
Line2:Region2:Product5
Line2:Region2:Product6
Line2:Region2:Product7
Line5:Region3:Product17
Details View - PriceData4>Hierarchy1>productName>Subset1>Query2
Time Series
Data Table
Series Analysis
Display : Envelope
Analysis Variable: sale
800
600
400
Series Distribution over Time
Jan1998
Jan1999
Jan2000
Jan2001
Jan2002
Jan2003
date
Range
Two Standard Deviations
One Standard Deviation
Mean
Selected series

The Subset1 node still contains 17 series.



10

Using Queries to Segment and Subset Your Data

About Queries

To segment or subset your data, you use queries. These different types of queries help you filter the series of interest with a variety of techniques.

In SAS Time Series Studio, you can use the following queries:

- graphical queries
- hierarchical queries
- parameter queries

Graphical Query

What Is a Graphical Query?

A graphical query enables you to create a query by selecting the desired time series from a graph. To select the series, you use the mouse pointer to draw a rectangular box around the area of interest. The selected series are highlighted in blue, but this selection is an approximation of the query results.

Note: A series is selected only if a data point in the series is included in the rectangle. If the series passes through the rectangle but the data point is outside the rectangle, the series is not selected.

Create a Graphical Query

To create a graphical query:

1. In the Flow Manager, right-click the subset where you want to add the query and select **Add Graphical Query**. Or right-click the segment where you want to add the query and select **New Segment** ▸ **Add Graphical Query**. The Graphical Query dialog box appears.
2. Specify the name of the query. By default, the name is Query`n` where `n` is an integer value. You can also provide a description.
3. Select the analysis variable to display in the graph.

4 In the graph, select the series of interest. You can select an entire series (which is highlighted in red). Or you can draw a rectangle around the data points that you want to select (any series that contain these data points are highlighted in blue).

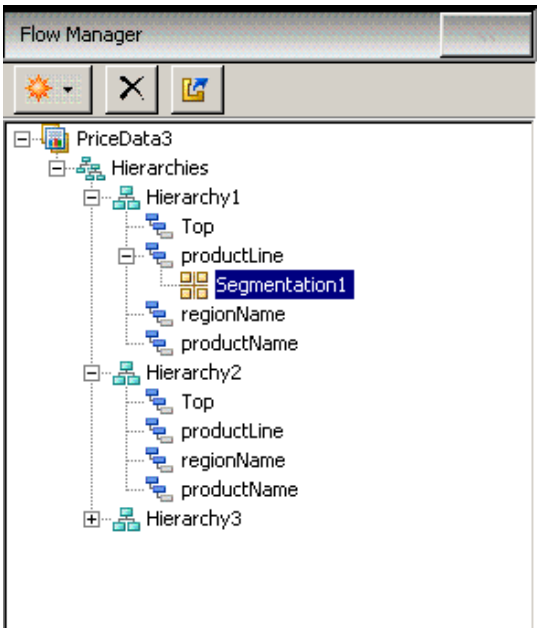
5 Specify whether to include only the selected series in the results, exclude the series from the results, or add the series to the results.

6 Click **OK**.

Example: Segmenting productLine by Using a Graphical Query

To create a segment for productLine in Hierarchy1:

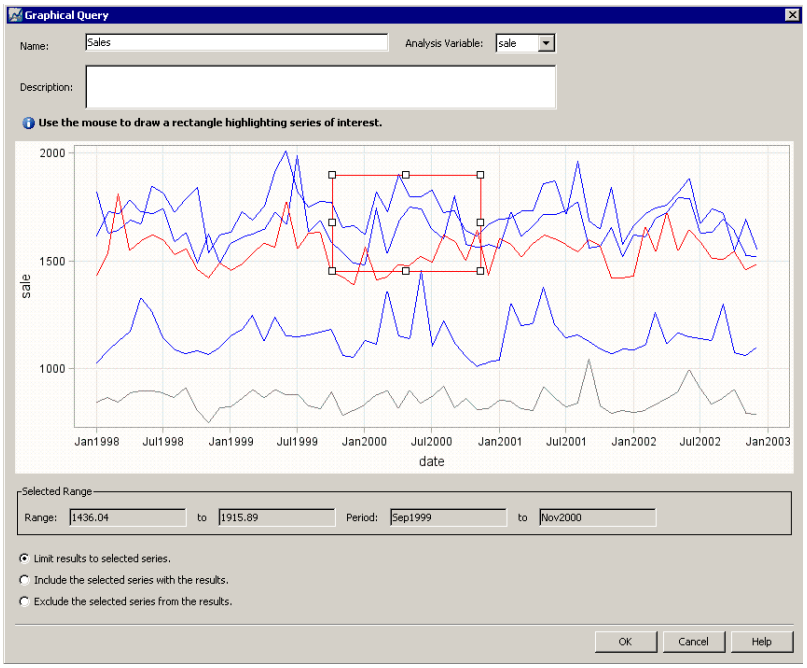
1 Select **productLine** in Hierarchy1 and select **Actions** ▶ **Add Segmentation**. Segmentation1 is added to the Flow Manager. For now, this segment contains the same data as productLine1.



2 Right-click **Segmentation1** and select **New Segment** ▶ **Add Graphical Query**. The Graphical Query dialog box appears.

3 In the Graphical Query dialog box, type `Sales` as the name of the query.

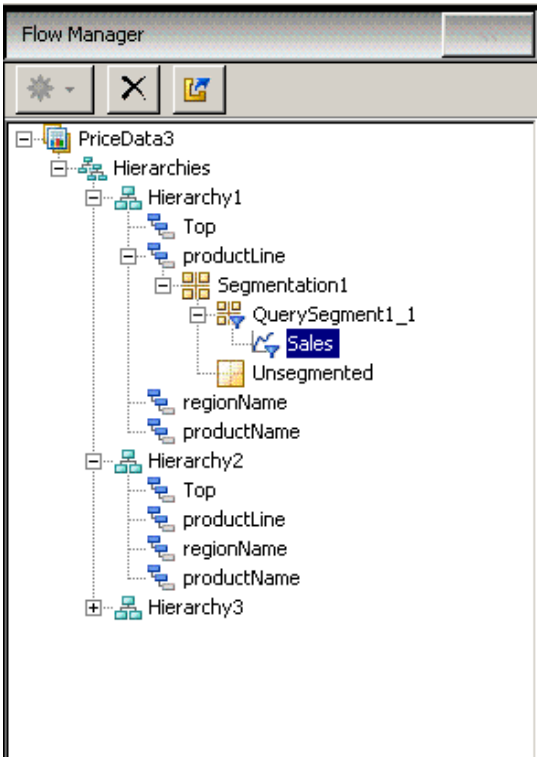
4 In the graph, select the series of interest.



5 Select **Exclude the selected series from the results**.

6 Click **OK**.

The new Sales segment appears in the Flow Manager.



Hierarchical Query

What Is a Hierarchical Query?

A hierarchical query enables you to select a leaf or level in a hierarchy. If you select a branch in the hierarchy, then all the nodes in that level are selected. The result of the query depends on the selected operator.

Create a Hierarchical Query

To create a hierarchical query:

1 In the Flow Manager, select the node where you want to add the query and select **Actions** ▸ **Add Hierarchial Query**. The Hierarchical Query dialog box appears.

2 Specify the name of the query. By default, the name is Query`n` where `n` is an integer value. You can also provide a description.

3 Select the units and branches in the hierarchy that you want to include in the query. Click  to move them to the list of selected units and branches.

4 Specify whether to include only the selected series in the results, exclude the series from the results, or add the series to the results.

5 Click **OK**.

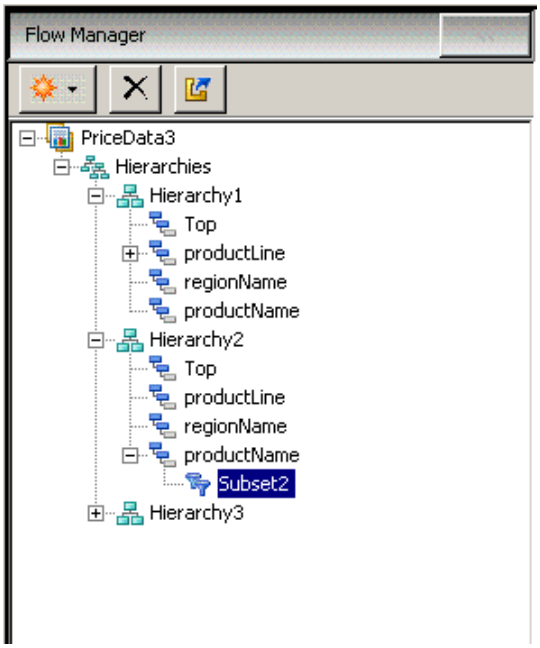
Example: Subsetting Your Data Using a Hierarchical Query

After studying the results for the productName level in Hierarchy1 and Hierarchy2, you realize that the monthly time interval is inappropriate for several of the products and that you should use the quarter interval instead. You decide to subset the data in the productName level, so you can easily view the analyses for those products that are better represented by the quarter interval.

To subset the series in the productName level of Hierarchy2:

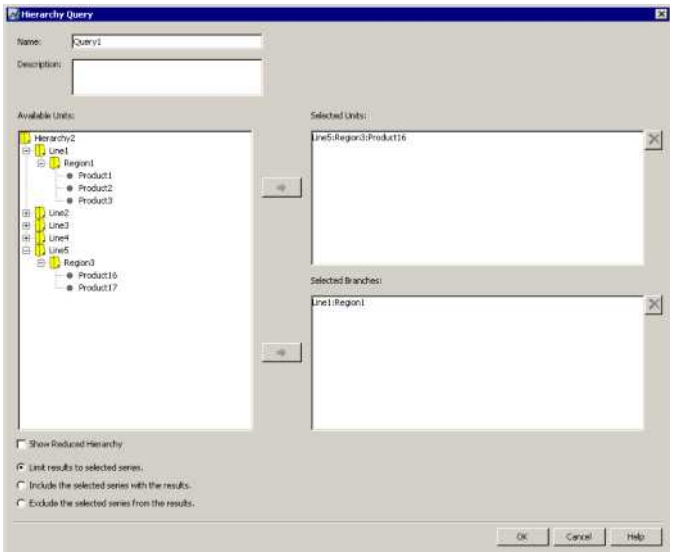
1 In the Flow Manager, select **productName** under Hierarchy3 and select **Actions** ▶ **Add Subset**.

Subset1 now appears in the Flow Manager. In the Selection Manager and Details View, the results of Subset1 are the same as the results for productName.



2 Right-click **Subset1** and select **Add Hierarchical Query**. The Hierarchy Query dialog box appears.

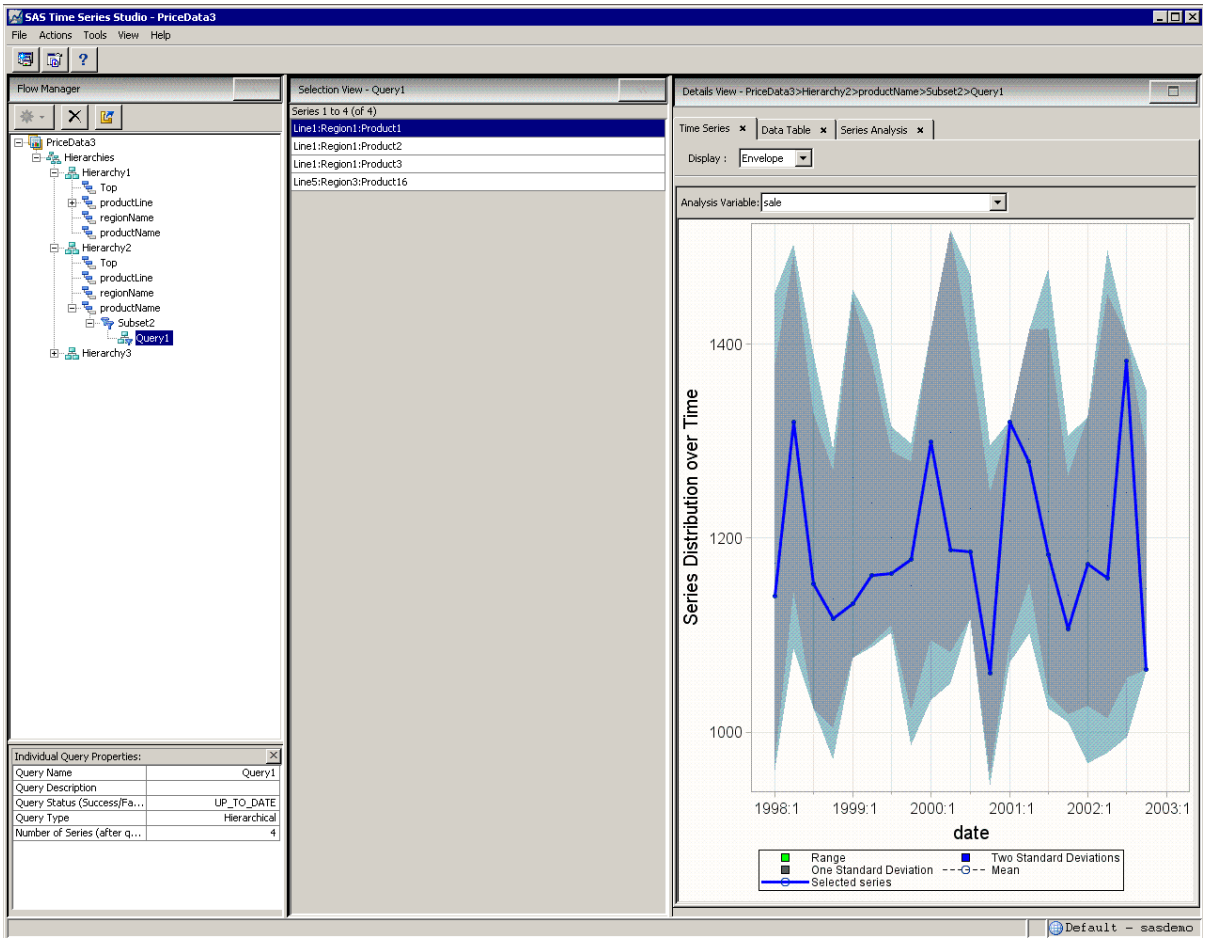
3 From the list of available units, select Line1 and Product16 in Line 5.



Click  to move them to the list of selected units and branches.

4 Select **Limit results to the selected series**. Click **OK**.

The results from the query now appear in the SAS Time Series Studio workspace.



Parameter Queries

What Is a Parameter Query?

A parameter query is the most flexible way of performing a query operation. You can specify the parameters based on either the input level data set or descriptive statistics. With this type of query, you build WHERE clauses using the variables in the input data set and descriptive statistics (such as start, end, maximum value, and number of missing values). You can specify multiple WHERE clauses and combine these clauses by using an AND or OR operator.

Create a Parameter Query

To create a parameter query:

1. In the Flow Manager, select the node where you want to add the query and select **Tools** ▶ **Add Parameter Query**. The Parameter Query dialog box appears.
2. Specify the name of the query. By default, the name is Query`n` where `n` is an integer value. You can also provide a description.
3. Specify whether to include only the selected series in the results, exclude the series from the results, or add the series to the results.
4. Specify whether to create the query based on the variables in the project data set or descriptive statistics.
5. Create the query and click **OK**.

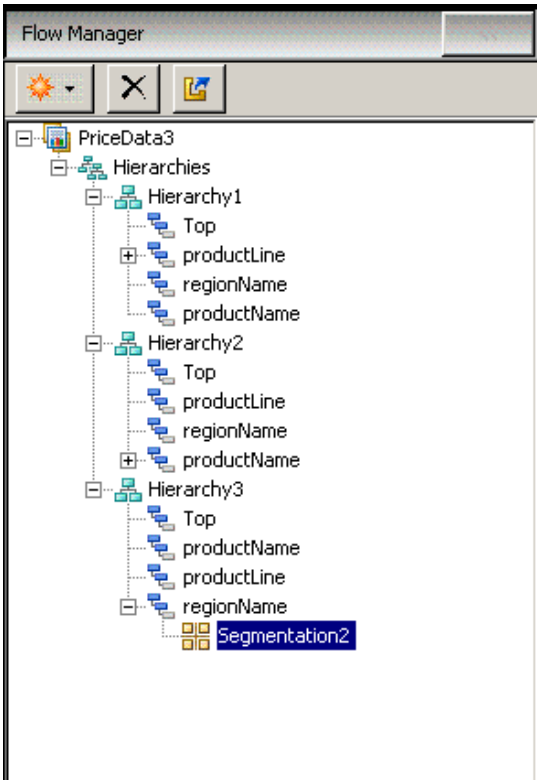
Example: Segmenting the Data by Using a Parameter Query

In Hierarchy3, you want to segment the data in the regionName level. You want to create a separate segment for all of the series in Region2.

To create a segment that contains the series in Region2:

1 In the Flow Manager, select **regionName** in Hierarchy3 and select **Actions** ▶ **Add Segmentation**.

Segmentation2 is added to the Flow Manager. In the Selection Manager and Details View, the results of Segmentation2 are the same as the results for regionName.



2 Right–click **Segmentation2** and select **New Segment** ▶ **Add Parameter Query** from the pop-up menu. The Parameter Query dialog box appears.

3 In the Parameter Query dialog box, type `Region2` as the name of the query.

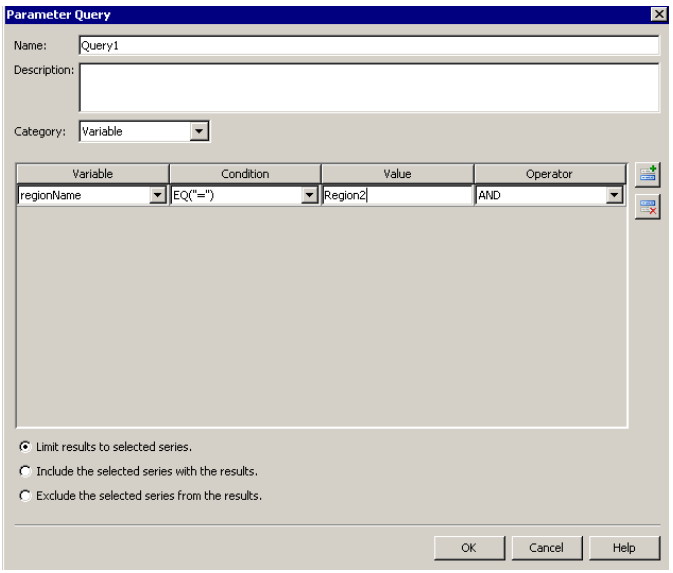
4 Select **Limit results to selected series** include only the selected series in the results.

5 For the category, select **Variable**.

6 Click  to add a row to the table.

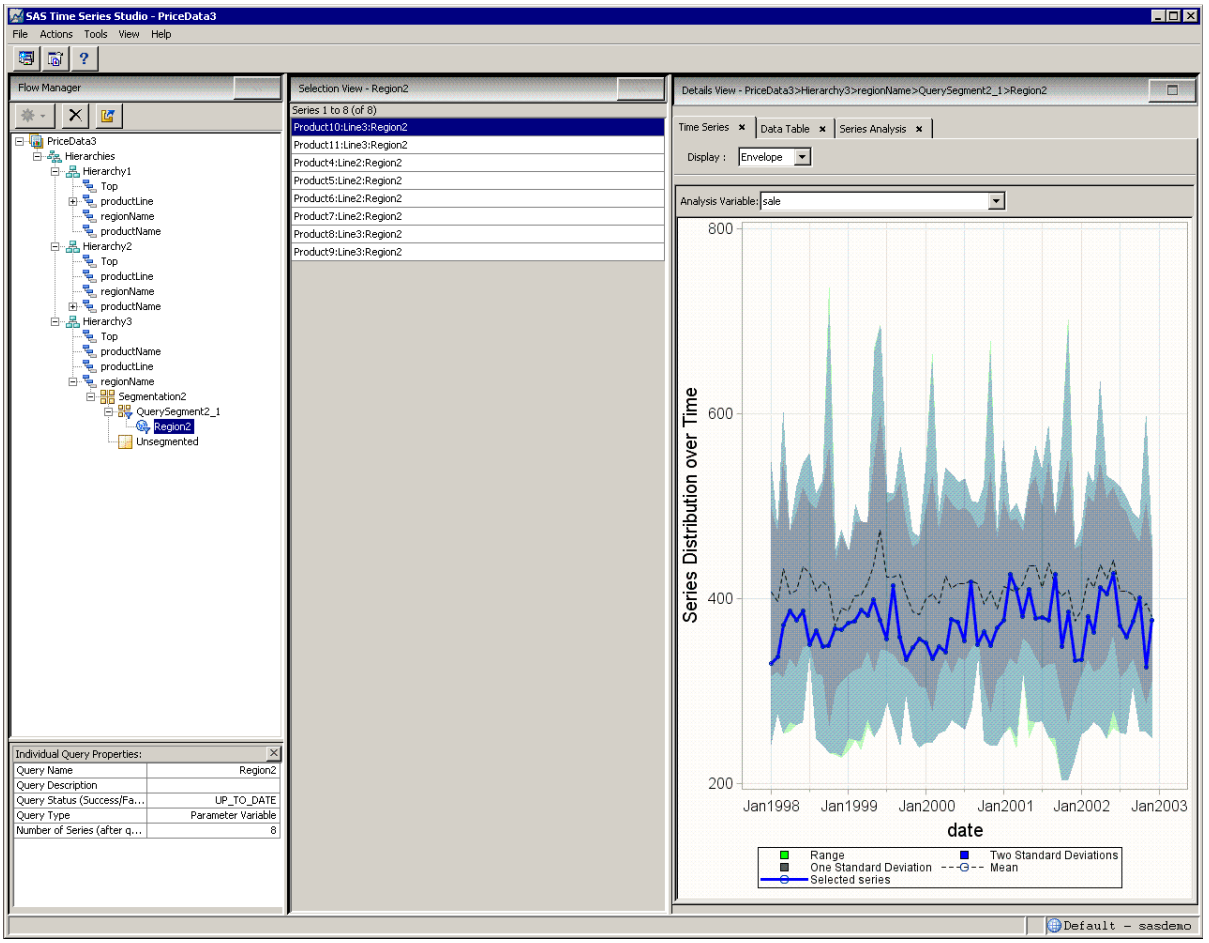
7 In the table, select the following options:

- In the **Variable** column, select **regionName**.
- In the **Condition** column, select **EQ("=")**.
- In the **Value** column, type `Region2`.



8 Click **OK**.

The Region2 segment is added to the Flow Manager. When this node is selected, the series for Region2 appear in the Selection Manager and Details View.



The Unsegmented node is a segment that contains the remaining series. These series are not used by previously defined segments, such as Region2.

Part 4

Appendixes

Appendix 1

Troubleshooting Tips

View the SAS Code

To view the SAS code that was used to generate different parts of the hierarchy (such as a subset or segment), right-click the node in the hierarchy, and select **View SAS Code** from the pop-up menu.

View the SAS Log

From the SAS log, you can view the code that was run to generate your results. You can also use the log to troubleshoot any errors that occurred when the code was run.

To view the SAS log that SAS Time Series Studio creates when exploring the time series data, select **Tools** ▶ **SAS Log**. You can save the SAS log to a file for future reference. Saving the log is helpful, especially if you need to contact SAS Technical Support for assistance.