SAS® Time Series Studio 13.1

User's Guide

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2013. *SAS® Time Series Studio 13.1: User's Guide*. Cary, NC: SAS Institute Inc.

SAS® Time Series Studio 13.1: User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

December 2013

SAS provides a complete selection of books and electronic products to help customers use SAS® software to its fullest potential. For more information about our offerings, visit **support.sas.com/bookstore** or call 1-800-727-3228.

Contents

Using This Book

Audience

SAS Time Series Studio is designed for any analyst who needs to analyze and structure their time-stamped data.

Requirements

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

What's New

What's New in SAS Time Series Studio 13.1

Overview

SAS Time Series Studio 13.1 now supports the following:

- custom time intervals
- Integrated Windows Authentication (IWA)
- clustered metadata servers
- clustered middle-tier servers

For more information, see *SAS Time Series Studio: Administrator's Guide*.

You can also now edit composite transformations in the Manage Transformations dialog box. For more information about transformations, see "Using Transformations" on page 26.

Accessibility

Accessibility Features of SAS Time Series Studio

Overview

SAS Time Series Studio has been tested with assistive technology tools. It includes accessibility and compatibility features that improve the usability of the product for users with disabilities. (Some accessibility issues remain and are noted below.) These features are related to accessibility standards for electronic information technology that were adopted by the U.S. Government under Section 508 of the U.S. Rehabilitation Act of 1973 (2008 draft proposal initiative update). For detailed information about the accessibility of this product, send e-mail to **accessibility@sas.com** or call SAS Technical Support.

Documentation Format

Please contact **accessibility@sas.com** if you need this document in an alternative digital format.

Exceptions to Accessibility Standards

Exceptions to accessibility standards are documented in the following table.

Accessibility Issue	Workaround
The high contrast settings for some operating environments are not fully supported. As a result, in some operating environments, you might have difficulty following the focus in the user interface.	None
Some items are not properly labeled for a screen reader. As a result, the screen reader might read a file path instead of announcing the item properly. For example, in the Flow Manager, the first TAB stop is the collapse button. JAWS reads this button as a long string of text that sounds like a file path.	None
Some actions might result in a loss of focus. For example, select a project in the Flow Manager. Now, press TAB to move to the Details View and move focus to the **Time Series** tab. If you press the RIGHT ARROW key, the focus is moved to the **Series Analysis** tab. After focus is removed, tab navigation is no longer possible.	Try using one of the mnemonic key combinations to return focus.
In step 3 of the New Project wizard, the TAB order is incorrect.	In order to move a variable from the **Available variables** pane to the **Selected variables** pane, keyboard users must navigate to the **Selected variables** pane and then press CTRL+TAB to reach the arrow buttons. To remove a variable from the **Selected variables** pane, press SHIFT+TAB to move focus to the **Selected variables** pane. Use the arrow keys to select the variable to remove and press CTRL+TAB to reach the arrow buttons.

Accessibility Issue	Workaround
In step 4 of the New Project wizard, there is no visual indication of focus on the **Interval** drop-down list.	None
In step 5 of the New Project wizard, the focus is on the **Next** button.	Use a mnemonic key combination to move focus to another item in the wizard.
Screen readers (such as JAWS) cannot read items in the Multivariable Time Series View.	None
Items in context menus are not accessible to keyboard-only users.	None

Recommended Reading

- *SAS Time Series Studio: Administrator's Guide*
- *SAS/ETS User's Guide*
- *SAS/STAT User's Guide*
- *SAS/GRAPH: Reference*

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

SAS Books
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. Given the transactional or time series data, analysts often need to structure the time series data into hierarchal time series at particular frequencies to enhance their understanding of the data and to improve the accuracy of their analyses. Given large amounts of data, analysts often need to subset their time series data to support different analysis on different subsets of their data. Using SAS Time Series Studio, you can subset the data using hierarchical queries, graphical queries, parametric queries, or manual selection.

By using SAS Time Series Studio, you can help you achieve these tasks:

- time series data preparation
- time series forecasting

- forecasting by similarity analysis
- econometric analysis
- risk analysis
- time series data mining

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 does not look like the others. You can explore the effect of transformations of variables and the effect of different time series.

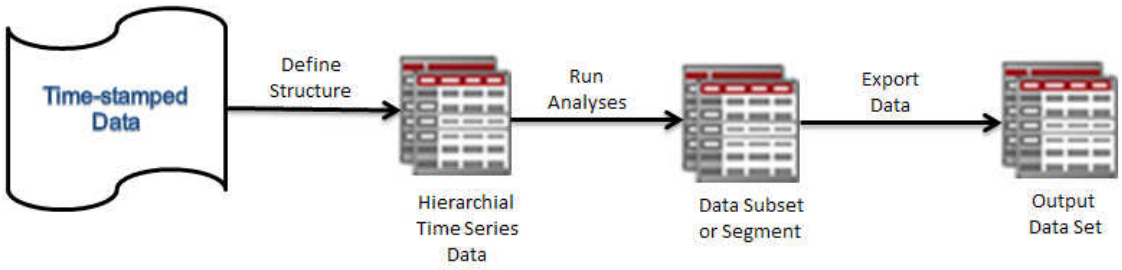
the ability to identify series that require specialized methods for analysis
: Using SAS Time Series Studio, you can identify short series (new products or a short lifecycle product) and intermittent time series (for example, a series that contains a large number of zero values).

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

data for use by various analytical tools
: Using SAS Time Series Studio, you can export a data source that then can be used by SAS High-Performance Forecasting, SAS Forecast Studio, SAS Enterprise Miner, SAS Revenue Optimization, and SAS Risk Dimensions.

How SAS Time Series Studio Works

Here is a basic use case for SAS Time Series Studio:



1. You select a raw data source that contains timestamped data.
2. Using the New Project wizard in SAS Time Series Studio, you specify how to structure the data. For example, you assign roles to variables in the data (such as BY variables) and specify whether to structure the data hierarchically. SAS Time Series Studio accumulates the data based on the options that you specify.
3. SAS Time Series Studio creates the time series data set. If you created a default hierarchy, SAS Time Series Studio accumulates and aggregates the data across the levels in the hierarchy.
4. Now, you can subset or segment the data to get only the data of interest. You create these subsets and segments by defining queries. After viewing the results from your queries, you might decide to perform additional analyses or transform the data.
5. After you have finished your analyses, SAS Time Series Studio can export the data as an output data set. This output data set can then be imported into SAS Forecast Studio or other SAS applications that support data sets created in SAS Time Series Studio.

How SAS Time Series Studio Relates to Other SAS Software

SAS Time Series Studio uses many of the procedures and options from other SAS products (such as SAS/ETS, Base SAS, and SAS/GRAPH). Many features not found in this solution software are available in other SAS solutions or in SAS products that are used with this SAS solution. If you do not find a feature that you need in this software, you might find it in one of the following SAS solutions or products.

SAS/ETS
: SAS/ETS software provides SAS procedures that perform econometric and time series analysis and forecasting, as well as financial analysis and reporting. The software also provides an interactive environment for time series forecast and investment analysis. For more information, see *SAS/ETS Users's Guide*.

Base SAS
: Base SAS delivers a highly flexible and extensible fourth-generation programming language that is specially designed for data access, transformation, and reporting. It includes a rich library of encapsulated programming procedures for data manipulation, information storage and retrieval, descriptive statistics, and report writing. The output for SAS Time Series Studio is generated by the Output Delivery System that is part of Base SAS.

SAS/GRAPH
: SAS/GRAPH software provides high-impact visuals for all levels of your organization, enabling customers to readily understand complex information and empowering them to make informed, timely decisions. SAS/GRAPH software extends the power of SAS data management, business intelligence, and analytic tools, enabling customers to turn data into full-color graphs and charts.

Data exported from SAS Time Series Studio can be imported into SAS Forecast Studio. Any metadata that was defined in SAS Time Series Studio (for example, the role assignments for each variable) is also imported into SAS Forecast Studio.

2

Starting SAS Time Series Studio

SAS Environments and Product Environments

A SAS environment represents a SAS deployment. A product environment in SAS Time Series Studio represents a product workspace for product sessions. Product 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.

Starting SAS Time Series Studio

How You Can Run SAS Time Series Studio

SAS Time Series Studio runs on Windows. Depending on your site, you can run SAS Time Series Studio in either of the following ways:

- from a local installation of SAS Time Series Studio on your computer. If you installed SAS Time Series Studio on a computer in the Windows operating environment, select **Start** ▶ **Programs** ▶ **SAS** ▶ **SAS Time Series Studio 12.3**.
- by using Java Web Start. Contact your site administrator for the URL for your site.

Log On to SAS Time Series Studio

Whether you run a local installation of SAS Time Series Studio or you use Java Web Start, you need to log on to SAS Time Series Studio.

To log on to SAS Time Series Studio:

1. Select your SAS environment. The list of environments is managed by your site administrator. For more information, see *SAS Time Series Studio: Administrator's Guide*.
2. Specify your user name and password for SAS Time Series Studio.
3. (Optional) If security policies at your site allow it, select the **Remember password** check box, so SAS Time Series Studio remembers your password.

   This option saves the password for the selected SAS environment and automatically uses that password for subsequent logon. Each SAS deployment has an independent policy setting to control whether passwords can be stored. This policy setting is decided by your organization. For more information about client-side storage of passwords, see the *SAS Intelligence Platform: Security Administration Guide*.

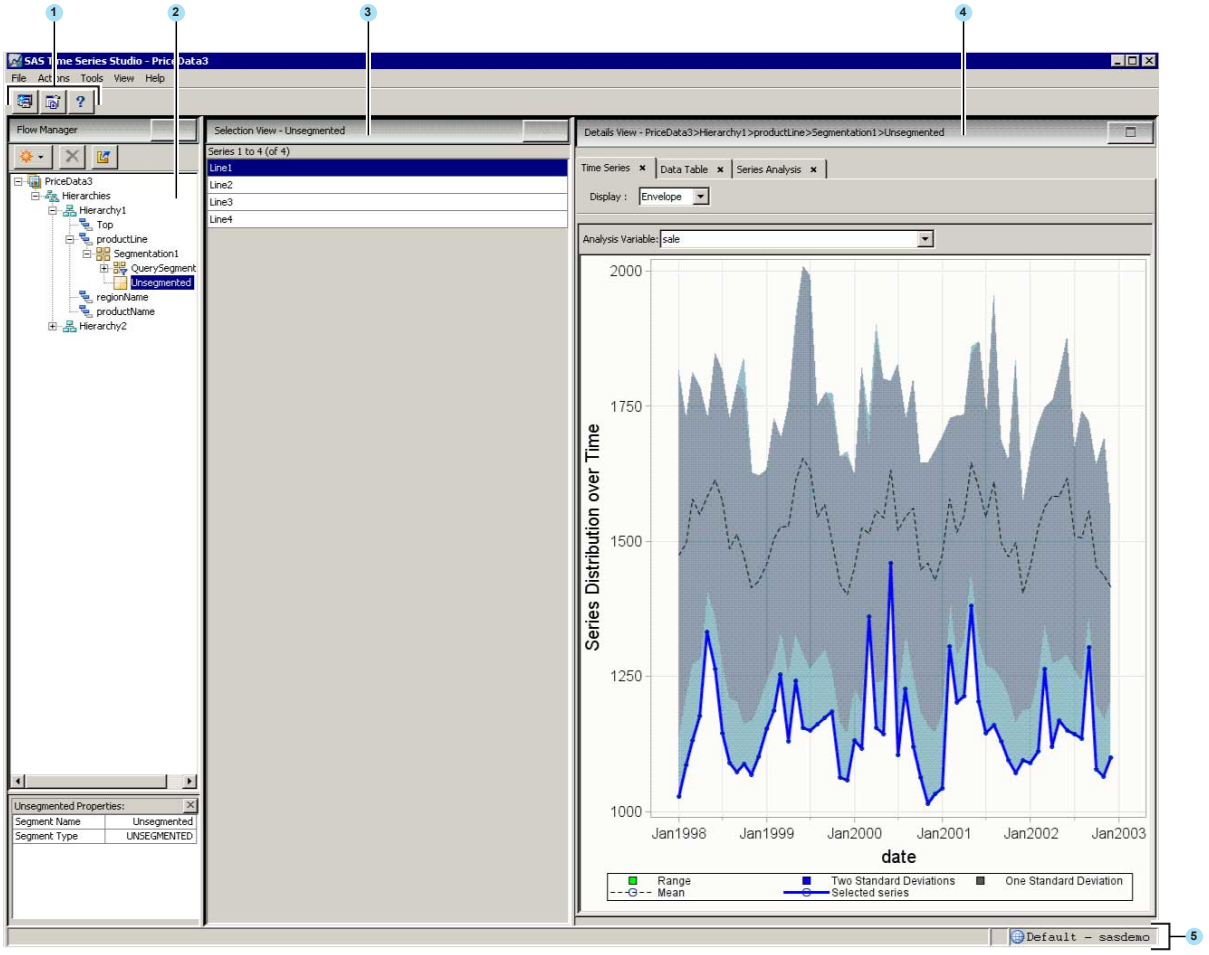
4 Click **Log On** to start SAS Time Series Studio.

3

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 steps that were used to structure your data. The contents of the Selection View depend on what you have selected in the Flow Manager.

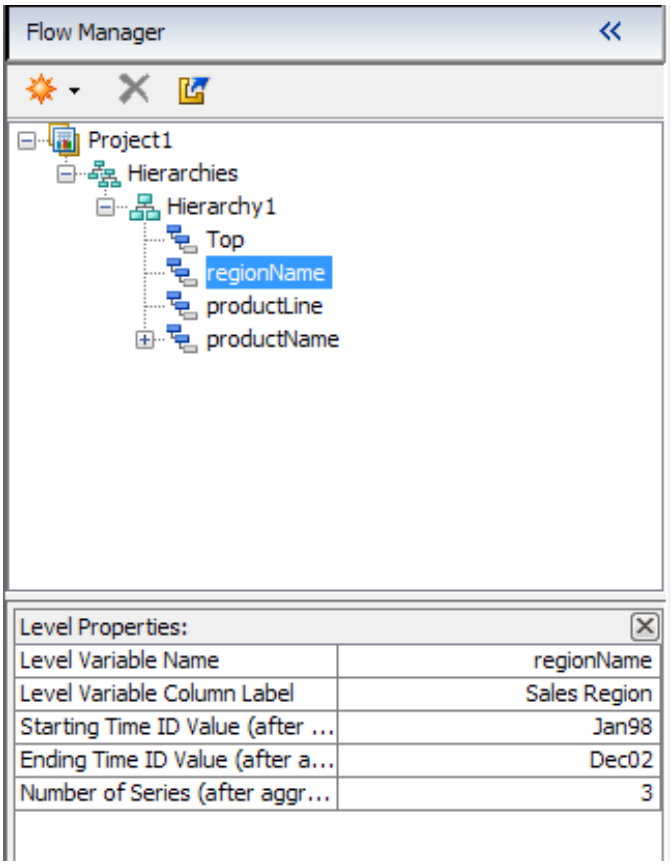
3. The Selection View lists the series that make up the element (hierarchy, level, query, subset, or segmentation) selected in the Flow Manager. The selected series determines the content of the graphs and tables in the Details View.
4. The Details View includes several analysis views. The content of each view depends on what is selected in the Selection View. If you close a tab in the **Details View**, you can reopen it by selecting **View** ▶ **Details** ▶ ***name-of-tab***.
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.

When using the SAS Time Series Studio workspace, the progression is from left to right. First you select a node in the Flow Manager, which determines the series that are displayed in the Selection View. You select a series in the Selection View to specify what is displayed in the Details View.

Flow Manager

The Flow Manager represents a program flow that turns your unstructured time-stamped data into time series data. The highest node in the project represents the unstructured data. At the bottom of the Flow Manager, properties are displayed for the selected node.

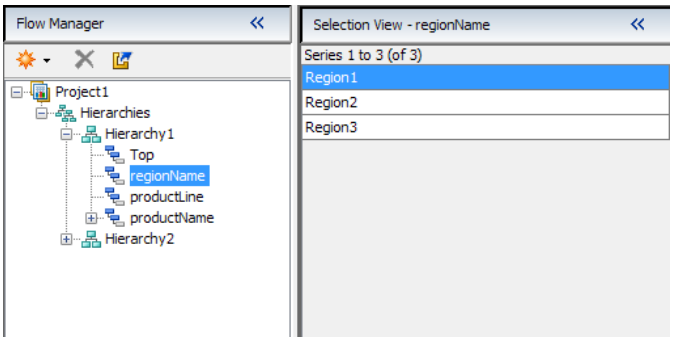
In the following display, the properties are shown for the regionName level.



To view the node properties, select **View** ▶ **Properties** ▶ **Node properties**.

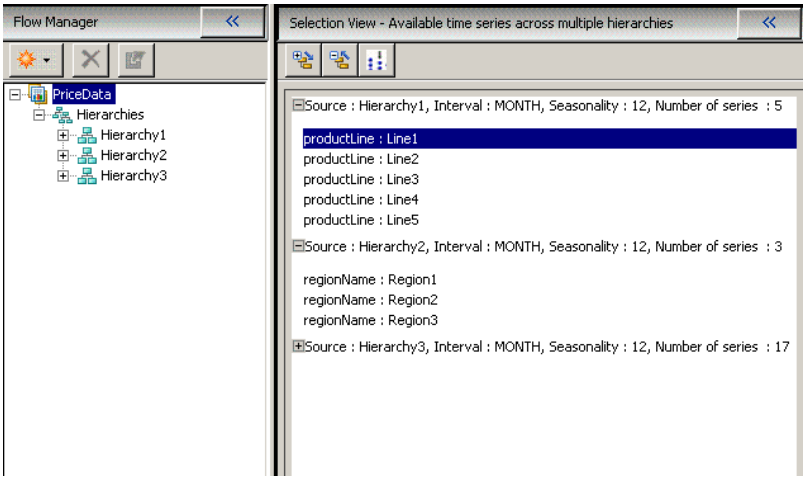
Selection View

When you select a level of a hierarchy in the Flow Manager, the Selection View displays the series that are associated with that level. For example, the following display lists the three series for the regionName level.

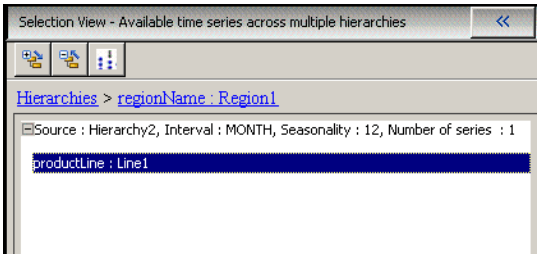


You can also use the Selection View to explore your hierarchies in these ways:

- If you select the **Hierarchies** node in the Flow Manager, you can simultaneously explore all of the hierarchies in your project.

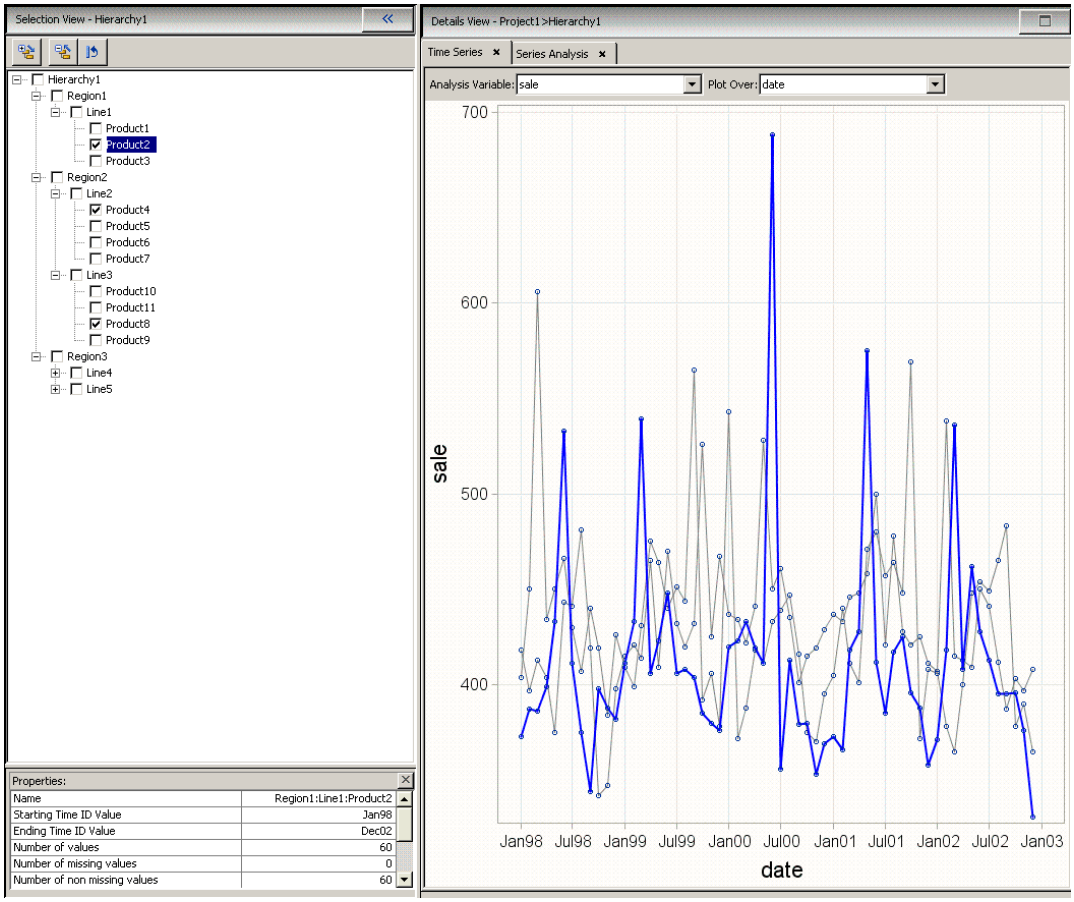


You can drill-down through the data by double-clicking a facet in the hierarchy. Use the breadcrumbs to drill back up the hierarchy. The following displays shows what happens if you drill down on regionName: Region1 in Hierarchy1.

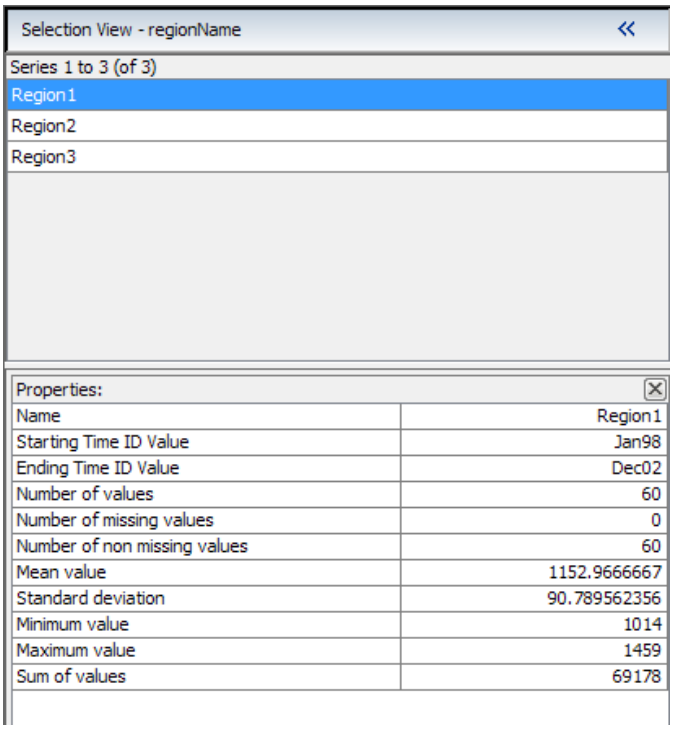


- If you select a single hierarchy node in the Flow Manager, you can explore the hierarchy by using a tree view in the Selection View. You can expand the tree and select the nodes that you want to display in the Details View. The highlighted node is the highlighted time series in the Details View.

  In the following display, the check boxes for Product2, Product4, and Product8 are selected, so all of these time series appear in the graph. In the Selection View, Product2 is also highlighted, so the time series for Product2 appears in blue in the Details View.



To view the properties for the selected series in the Selection View, select **View** ▶ **Properties** ▶ **Series properties**. The properties appear at the bottom of the Selection View.



Name	Region1
Starting Time ID Value	Jan98
Ending Time ID Value	Dec02
Number of values	60
Number of missing values	0
Number of non missing values	60
Mean value	1152.9666667
Standard deviation	90.789562356
Minimum value	1014
Maximum value	1459
Sum of values	69178

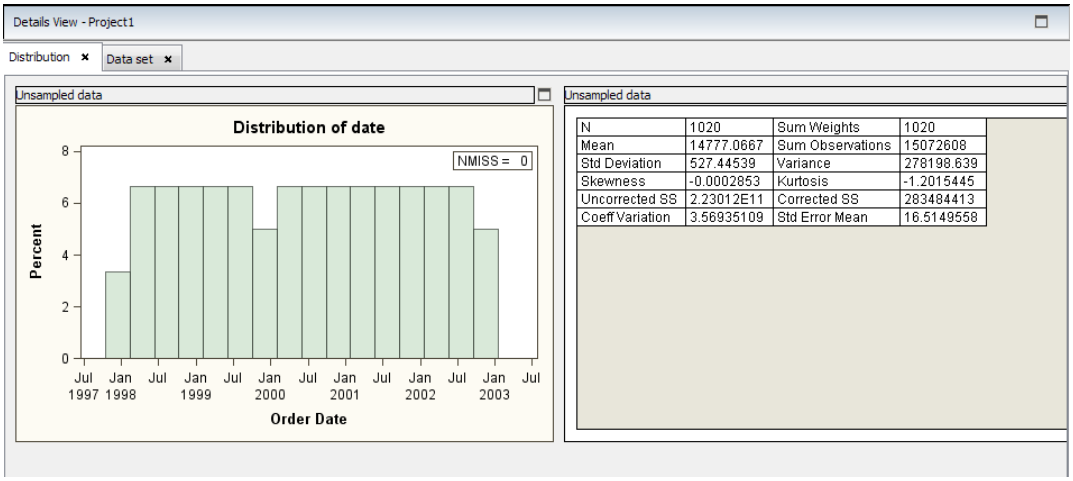
Details View

Several tabs are available from the Details View. The content in the Details View depends on the node that you selected in the Selection View.

The Distribution View

The Distribution View is available from the Details View when you select a variable in the Selection View or 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 and tables. For classification (BY) variables, the distribution is created by the FREQ procedure, and you can see these statistics: frequency, percent, cumulative frequency, and cumulative percentage. For all other variables, the distribution is created by the UNIVARIATE procedure, and you see these statistics: sum, mean, and standard deviation.

Note: If the input data source contains more than 100,000 observations, SAS Time Series Studio uses a random sample of the data. The default sample size is 10,000 observations. For information about how to change the default sample size, see the *SAS Forecast Server: Administrator's Guide*.



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 for the project.

Details View - Project1

Distribution | Data set

Data Set : SASHELP.PRICEDATA

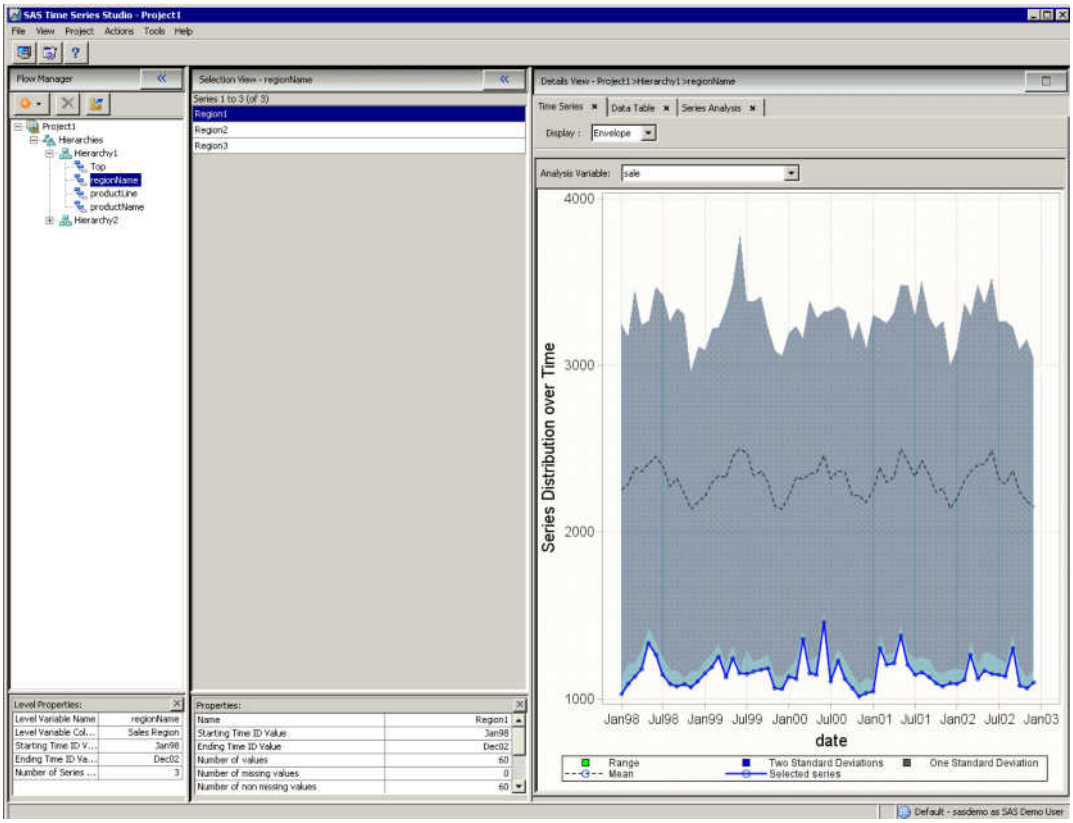
	Order Date	Unit Sale	Unit Price	Price Discount	Unit Cost	Product 1 Unit Price	Product 2 Unit Price	Product 3 Unit Price	Product 4 Unit Price	Produc
1	JAN98	355	52.3	0	23.9	52.3	115	33.4	67.9	36
2	FEB98	398	52.3	0	23.9	52.3	115	33.4	67.9	36
3	MAR98	387	52.3	0	23.9	52.3	115	33.4	67.9	36
4	APR98	380	52.3	0	23.9	52.3	115	28.39	67.9	36
5	MAY98	555	44.455	0.15	23.9	44.455	115	33.4	67.9	36
6	JUN98	385	52.3	0	23.9	52.3	103.5	33.4	67.9	36
7	JUL98	390	52.3	0	23.9	52.3	115	33.4	67.9	36
8	AUG98	377	52.3	0	23.9	52.3	115	33.4	61.11	36
9	SEP98	386	52.3	0	23.9	52.3	115	33.4	67.9	30.6
10	OCT98	367	52.3	0	23.9	52.3	115	33.4	67.9	36
11	NOV98	372	52.3	0	23.9	52.3	115	33.4	67.9	36
12	DEC98	378	52.3	0	23.9	52.3	115	33.4	67.9	36
13	JAN99	391	52.3	0	23.9	52.3	115	33.4	67.9	36
14	FEB99	386	52.3	0	23.9	52.3	115	33.4	67.9	36
15	MAR99	356	52.3	0	23.9	52.3	103.5	33.4	67.9	36
16	APR99	380	52.3	0	23.9	52.3	115	33.4	67.9	36
17	MAY99	393	52.3	0	23.9	52.3	115	26.72	67.9	36
18	JUN99	389	52.3	0	23.9	52.3	115	33.4	67.9	36
19	JUL99	388	52.3	0	23.9	52.3	115	33.4	67.9	32.4
20	AUG99	397	52.3	0	23.9	52.3	115	33.4	67.9	36
21	SEP99	379	52.3	0	23.9	52.3	115	33.4	67.9	36
22	OCT99	502	47.07	0.1	23.9	47.07	115	33.4	54.32	36
23	NOV99	342	52.3	0	23.9	52.3	115	33.4	67.9	36
24	DEC99	334	52.3	0	23.9	52.3	115	33.4	67.9	36
25	JAN00	366	52.3	0	23.9	52.3	115	33.4	67.9	36
26	FEB00	361	52.3	0	23.9	52.3	115	33.4	67.9	36
27	MAR00	572	44.455	0.15	23.9	44.455	115	33.4	67.9	36
28	APR00	365	52.3	0	23.9	52.3	115	33.4	67.9	30.6
29	MAY00	408	52.3	0	23.9	52.3	115	33.4	61.11	36
30	JUN00	415	52.3	0	23.9	52.3	92	33.4	67.9	36
31	JUL00	390	52.3	0	23.9	52.3	115	33.4	67.9	36
32	AUG00	404	52.3	0	23.9	52.3	115	30.06	67.9	36
33	SEP00	392	52.3	0	23.9	52.3	115	33.4	67.9	36
34	OCT00	358	52.3	0	23.9	52.3	115	33.4	67.9	36
35	NOV00	354	52.3	0	23.9	52.3	115	33.4	67.9	36

The Time Series View

The Time Series View is available when time series appear in the Selection View.

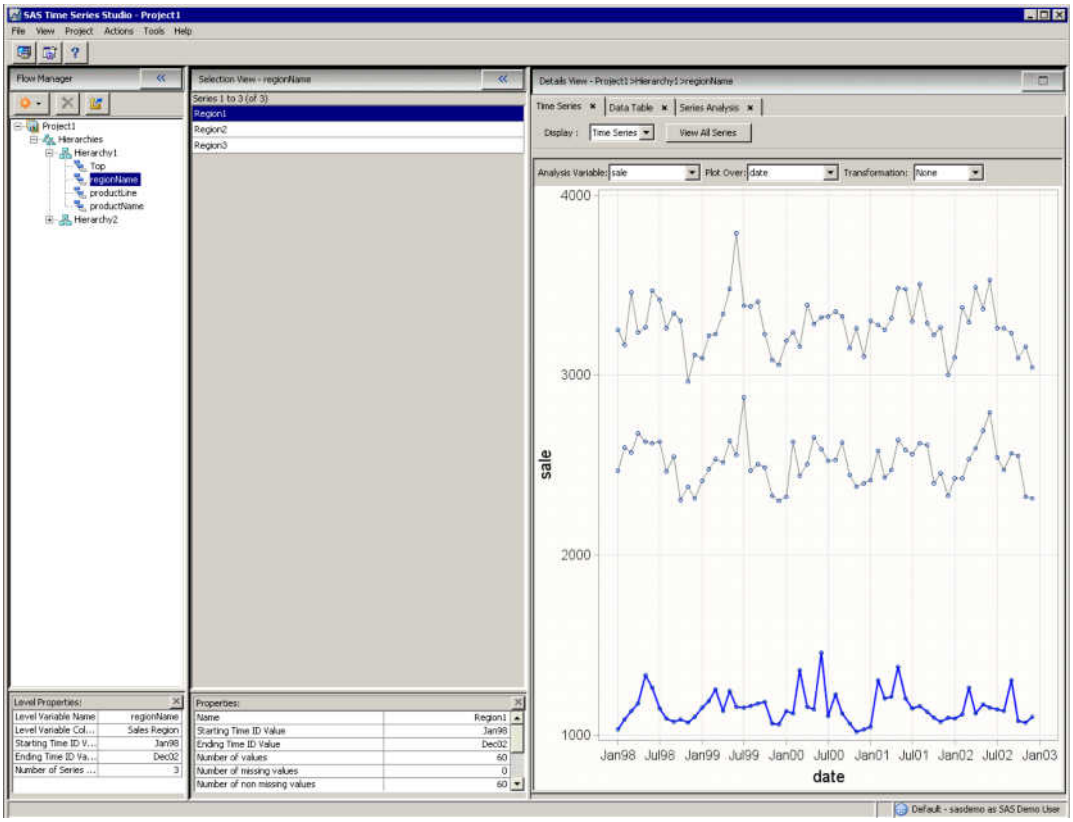
When you select a level in the hierarchy in the Selection View, the Time Series View displays the envelope plot. The envelope plot displays the descriptive statistics (mean, minimum, maximum, and so on) for each time period in a time series plot. (The envelope plot is not available if you select the top level in the hierarchy because there is only one series to display.)

In the following display, the regionName level is selected in the Flow Manager. As a result, the Selection View displays the BY groups for the regionName level. You can select a unit in the Selection View to see the time series of that specific unit in the envelope plot. In this example, the time series for Region1 is highlighted in the envelope plot.



In the Time Series View, you can also view the time series plot. The time series plot displays all the units in the Selection View. The unit highlighted in the Selection View appears in blue in the time series plot. To view the time series plot, select **Time Series** from the **Display** drop-down list.

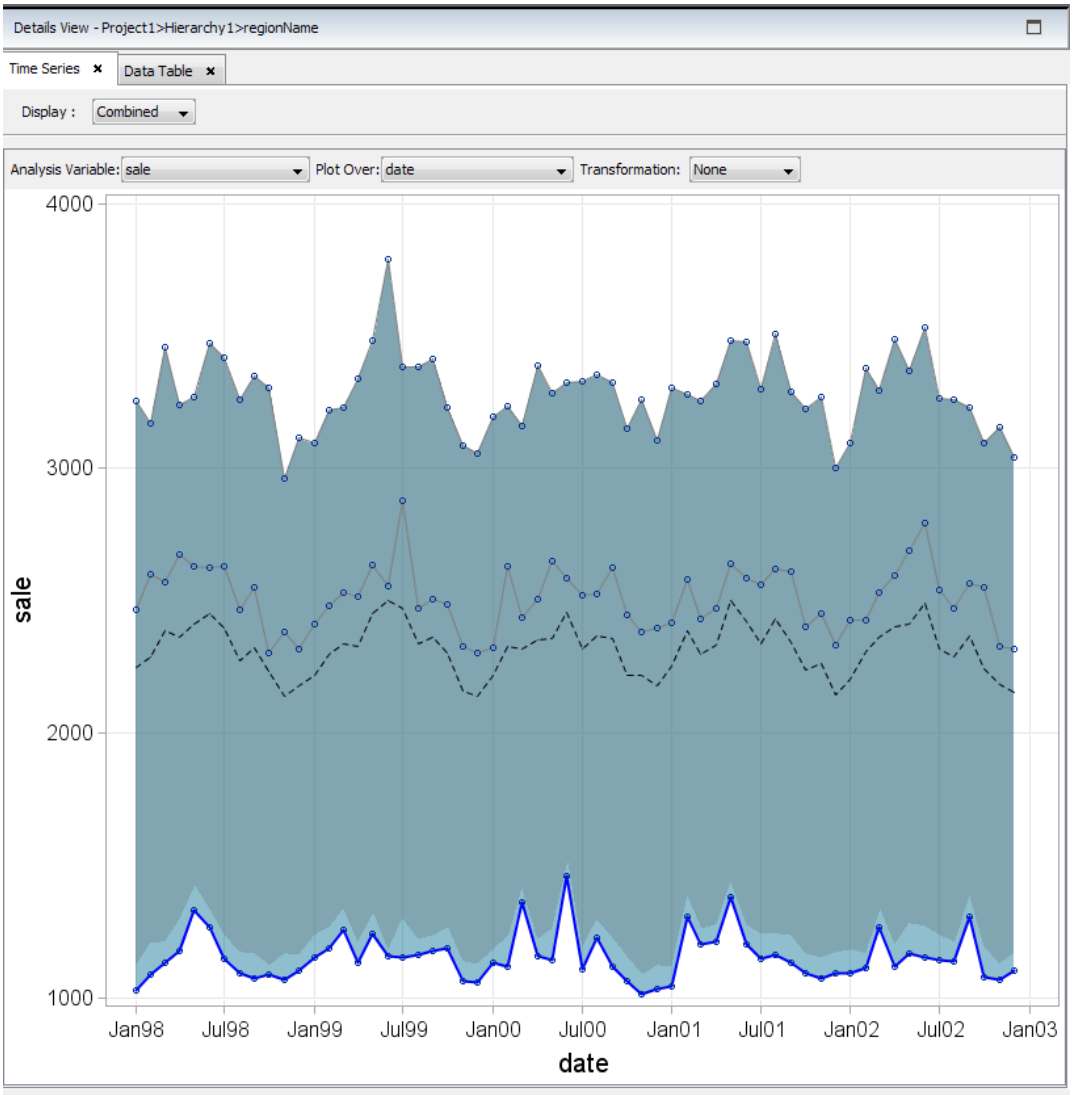
Here is an example of a time series plot:



For time series plots, you can choose to plot 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. For more information, see “Using Transformations” on page 26.

If you have a large amount of data, 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 View. To view all the series, click **View All Series**. 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**.

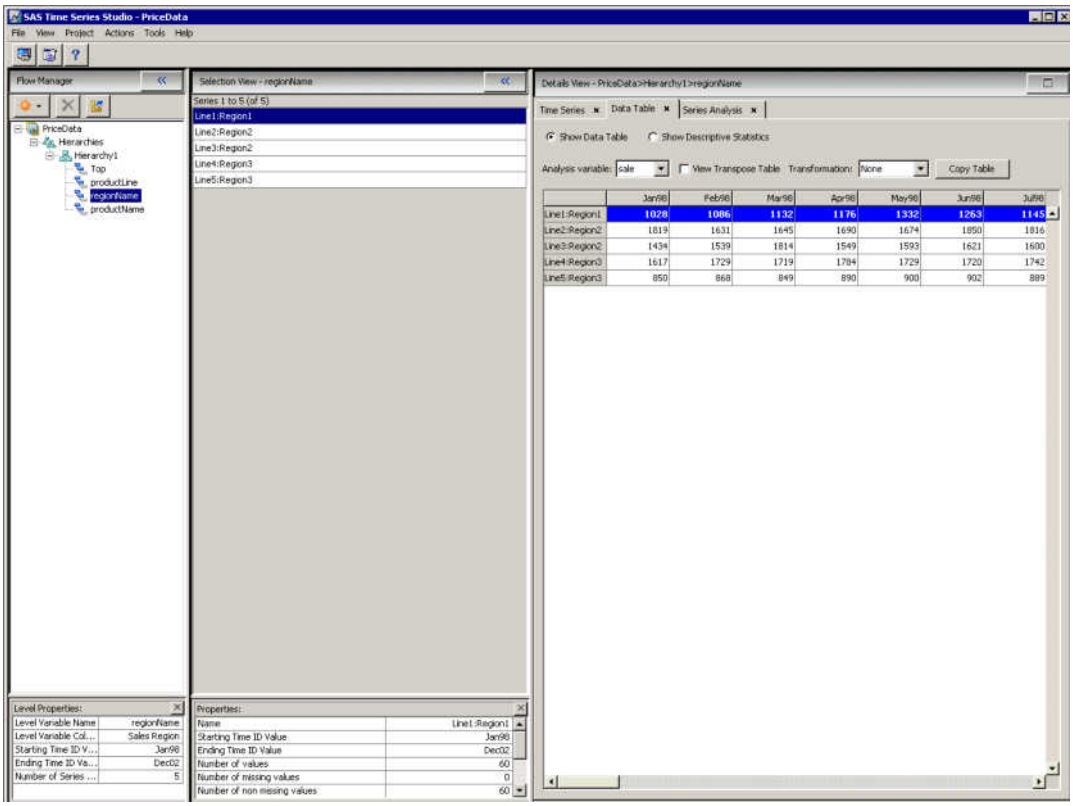
To view both the envelope plot and the time series plot, select **Combined** from the **Display** drop-down list. Here is an example of a combined plot:



The Data Table View

The Data Table View shows you the values of the series in the Selection View. 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.

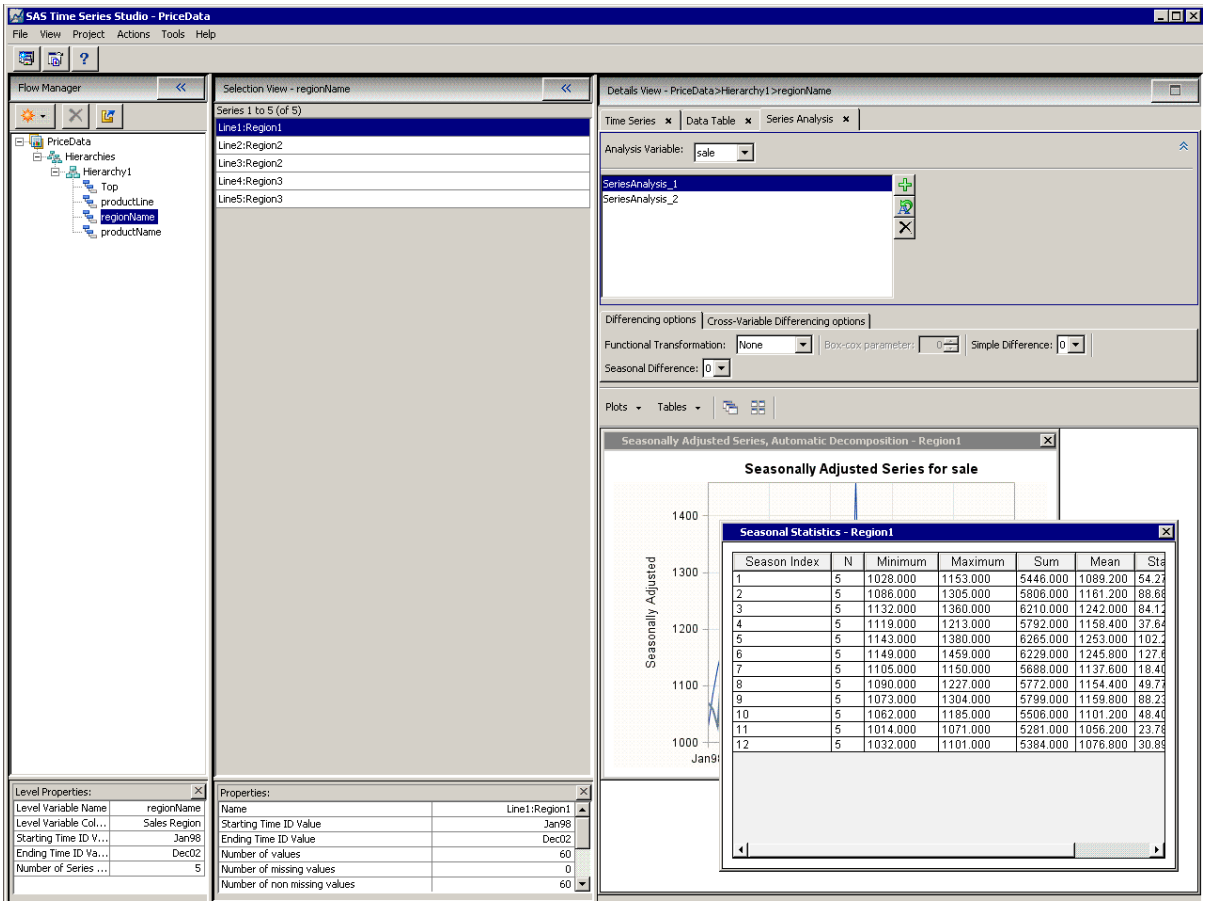


In the Data Table view, you can perform the following tasks:

- 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 values for the time ID variable as rows.
- Specify how you want to transform the data. For more information, see “Using Transformations” on page 26.
- Copy the table to your Windows clipboard. Then you can paste the table in another application, such as Microsoft Excel.

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 View.



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

- Create and compare analyses.
- Specify the following options for your analysis:
  - **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. The default value is 0.
- **Seasonal Difference** — specifies the seasonal differencing order. The default value is 0.

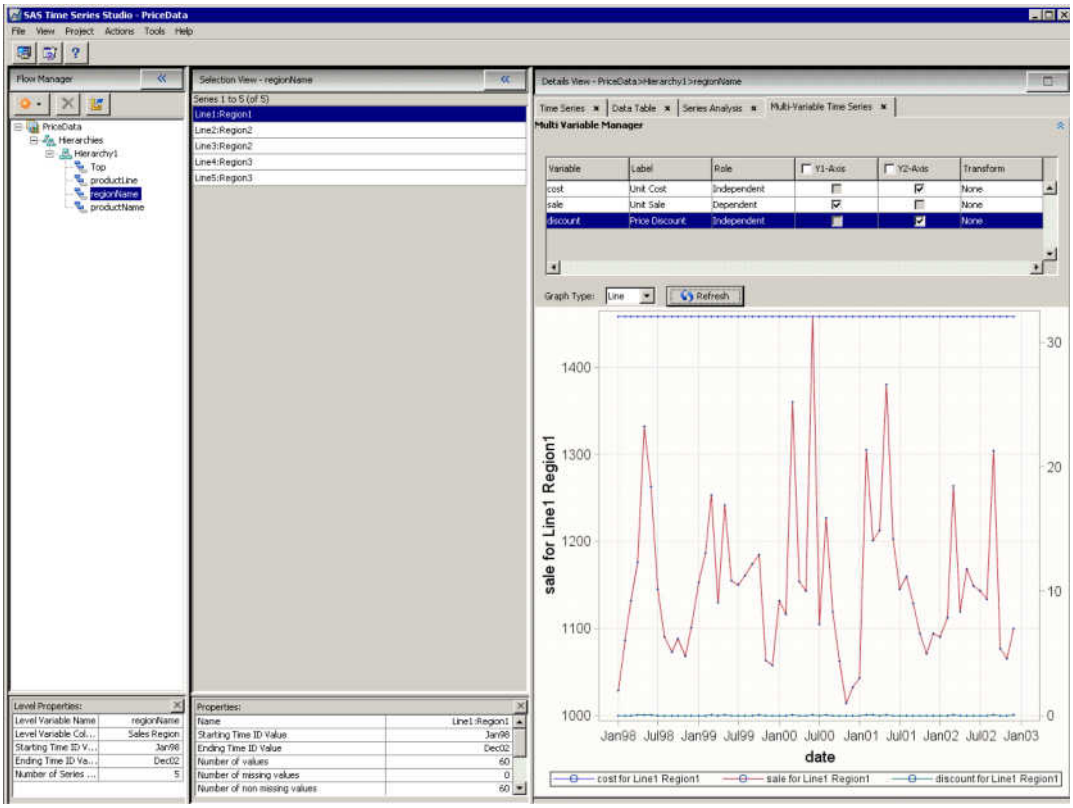
- Select the analysis variable.
- Select the plots and tables to display. You can display these plots and tables as tiled or cascade. When you close a plot or table, the plot or table is removed from the current analysis.

The Multi-variable Time Series View

Using the Multi-Variable Time Series View, you can view the effect of one variable on another. To open the Multi-Variable Time Series View, select **View** ▶ **Details** ▶ **Multi-Variable Time Series**.

In the table, SAS Time Series Studio displays the variables that you assigned to the dependent, independent, or other roles. In the graph, the X axis represents time, and there are two Y axes. You can assign several variables to each Y-axis. You can also apply transformations to the variables.

To create the graph, select the variables for the Y axes and click **Refresh**.



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 Composite Transformation

In addition to the default transformations, you can create composite transformations by combining individual transformations.

To create a composite 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). A transformation is not applied until you select it from the transformation option in one of the Detail Views.

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.

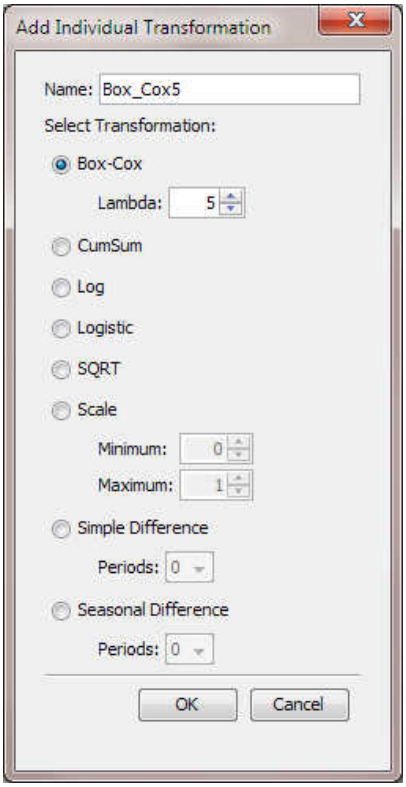
2 Click . The Add Composite Transformations dialog box appears.

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.



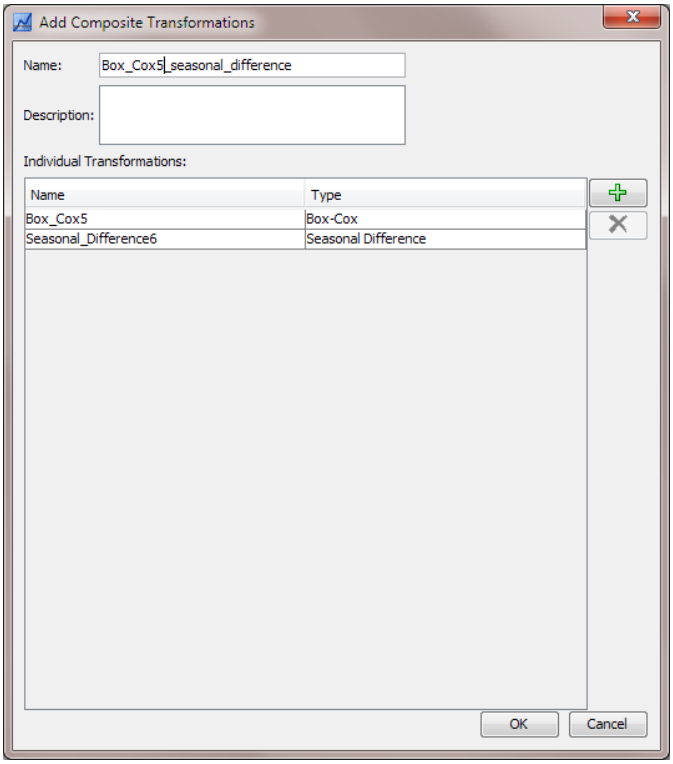
Click **OK**. The Box-Cox5 transformation appears in the list of individual transformations in the Add Composite Transformation dialog box.

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.

Part 2

Getting Started in SAS Time Series Studio

4

Understanding the Data Preparation Options

Understanding Aggregation and Accumulation

Differences between Aggregation and Accumulation

Aggregation is the process of combining more than one time series to form a single series. Accumulation combines data within the same time interval. For example, you can aggregate all of the series of each product in the Electronic group into a single series that represents the group as a whole. 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 refers to data aggregation in the time domain. It 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 Aggregation and Accumulation Methods

Note: The difference between accumulation and aggregation is the dimension along which each method is applied. The following equations focus on accumulation. Aggregation is restricted to the total and average methods, but the same equations apply.

Let $R = \{r_q\}_{q=1}^{Q}$ be the data vector 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 missing 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.

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

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.

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

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

Working with Missing Values

About Missing Values

Missing data can occur within a series. Missing values that appear after the beginning of a time series and before the end of the time series are called *embedded missing values*. SAS Time Series Studio expects the input data sets to contain observations for a contiguous time sequence. Omitted observations will cause errors.

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 and independent variables), you can specify how to interpret missing values (regardless of the variable role). You can specify these data preparation options when you create the project or when you create a hierarchy.

- Using the **Missing Interpretation** option, 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:
  - zero.
  - 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.
- Using the **Missing Trim** option, you can specify how beginning and ending 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.
- Using the **Zero Miss Interpretation**, 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.

5

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. 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
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.

Interval Name (in SAS code format)	Default	Shift Period	Example
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.
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.

Custom Time Intervals

What Is a Custom Time Interval?

Depending on your data, the standard time intervals that are available in SAS Time Series Studio might not be appropriate. Here are some examples of when you might need to create a custom interval:

- In SAS, the MONTH interval begins on the 1st of every month, but your fiscal months begin on a different day (such as the 10th) of every month.
- Using a standard interval results in gaps in your data. For example, you have retail data that is collected hourly. However, the business is closed at night. You do not want to include the hours when the business is closed.
- You have irregular gaps in the data, such as holidays.
- You need to transform a function to improve stationarity.

A custom time interval is a base time interval that you define by using a SAS data set. For more information about custom intervals, see *SAS Language Reference: Concepts* and the *SAS/ETS User's Guide*.

Creating a Custom Time Interval

In order to create a custom time interval, you must be familiar with SAS programming and have access to the autoexec.sas or sasv9_usermods.cfg file for the SAS server. Therefore, it is recommended that you ask your site administrator to create the custom interval that you need. For more information about creating a custom time interval, see *SAS Time Series Studio: Administrator's Guide*.

How to Select a Custom Time Interval

Custom intervals should appear in the same drop-down lists as the time intervals that are shipped with SAS Time Series Studio. For example, you can select a custom interval when creating a new project in the New Project wizard. If you do not see the interval that you need or you need more information about an interval, contact your site administrator.

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 of concern in the analysis. When you create a project in SAS Time Series Studio, you can assign multiple, numeric 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. For more information about "other" variables, see “Other Variables” on page 52.

Independent Variables

Independent variables (the explanatory, input, predictor, or causal factor variables) are the variables that potentially influence the dependent variables. You can assign multiple, numeric variables to this role. These independent variables are considered by SAS Forecast Studio when you import a data set that was created in SAS Time Series Studio.

Rejected Variables

If your input data source contains a large number of variables, you might want to remove the unnecessary variables. Rejected variables are not included in the project and are not included in any resulting data sets. These variables are also excluded from any exported 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. SAS Time Series Studio considers these numeric variables in the analysis, but their role is not defined. 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.

6

Working with Projects

What Is a Project?

SAS Time Series Studio organizes the files that are generated with an analysis into a project. The default location for the project files depends on your SAS configuration directory. An example of a default directory path in a Windows operating environment is `C:\SAS\TSS\Lev1\AppData\SASTimeSeriesStudioMidTier13.1`.

The name of the project directory is the name that you specify when you create a project. For a complete list of the directories that are part of an individual project, see *SAS Time Series Studio: Administrator's Guide*.

Note: For a project to work correctly, the input data set must be available. If you move, migrate, or copy a project to a different environment, the input data set must also be available in that environment.

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 the Data Set

Select the data set that you want to analyze. You can preview the contents of the data set by clicking **View**.

Click **Next**.

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

Assign one or more variables to the classification (BY) variables role. Classification variables are used to create hierarchies.

- To view the distribution graph for a selected variable, click **Distribution**.

  Note: If the variable contains a large number of observations, SAS Time Series Studio plots a sample of the data.

- For SAS Time Series Studio to select the classification (BY) variables automatically, click **Recommend**. By default, SAS Time Series Studio assigns all character variables as BY variables.

To create a hierarchy, select the check boxes for the BY variables in the **Selected variables** panel. 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 Dimension of Your Data

1 From the **Time ID variable** drop-down list, select the variable in the data that contains the timestamped values. By default, SAS Time Series Studio assigns the first date or datetime variable to the time ID role.

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 42.

Note: When you select a time ID variable, SAS Time Series Studio tries to detect the time interval (or frequency) of the data. If the time interval cannot be detected or it is not valid, you must specify the time interval.

3 Click **Next**.

Step 5: Assign Roles to Variables in Your Data and Specify How to Prepare the Data

In this step, you assign roles to the variables in your input data set, and you specify how to prepare the 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 41.

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

3 Select the accumulation and aggregation methods for each variable. By selecting the **Set accumulation to the value used for aggregation** check box, you can choose to use the same method for accumulation and aggregation. For more information, see "Understanding Aggregation and Accumulation" on page 33.

4 Specify how SAS Time Series Studio should interpret missing values in the data. For more information, see "Working with Missing Values" on page 37.

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.

Updating the Input Data

Because of the dynamic nature of time series data, new values might be added to your input data source. By default, SAS Time Series Studio detects when the data is updated and prompts you to update the data in your project. You can choose to continue working with the old data or incorporate the new data in your project. However, if you choose to incorporate the new data in your project, the project structure is not updated until you update the project. For more information, see “Updating a Project” on page 58.

Note: SAS Time Series Studio does not support structural changes (such as the removal of columns) to the input data set. Also, the attributes for the variables in the input data should not have changed. An example of a variable attribute is length.

If you previously decided not to include the updated data, you can check for new project data. To update the input data set for your project:

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

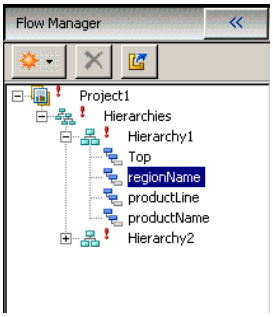
2 Click **OK**.

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

Updating a Project

About Updating a Project

When you update the input data for the project, the project structure does not change. Exclamation points next to the node in the Flow Manager can indicate that you need to update your project.



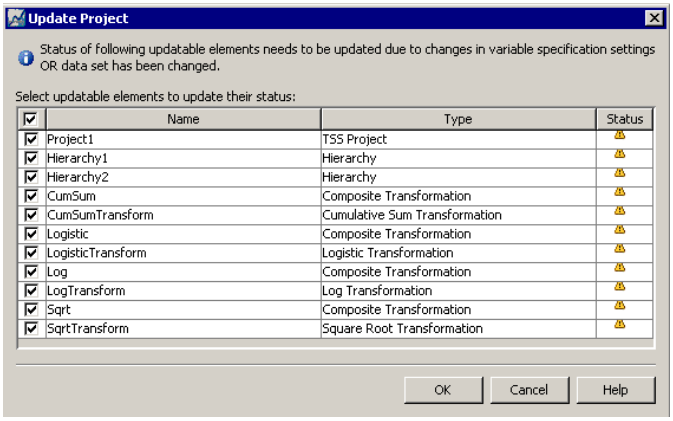
Note: SAS Time Series Studio does not support structural changes to the input data set. For example, if you assign a variable to a role and that variable is removed from the input data set, SAS Time Series Studio cannot update the project.

Update a Project

To update a project:

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

2 (Optional) Select the check boxes for any out-of-date elements.



3 Click **OK**.

Delete a Project

When you delete a project, you remove all of the content and metadata for that project. You cannot retrieve a project after it has been deleted.

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** ▶ **Projects**.

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

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** ▶ **Projects**.

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

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** ▶ **Projects**.

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 Select **File** ▶ **Project Properties**. The Project Properties dialog box appears.

Note: You can also view the project properties from the Projects dialog box.

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

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

4 (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.

5 Click **OK**.

Save the Project Code

You might want to save the project code, so that you can share the code with someone else or run the project in batch mode.

To save the project code:

1 Select **Project** ▶ **SAS Code**. The SAS Code dialog box appears.

2 (Optional) Specify whether to include library declarations in the SAS code. By default, the library declarations are saved in a separate file from the code.

3 Click **Save**. The Save As dialog box appears.

4 Specify where you want to save the SAS code, and click **Save**.

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 view the distribution of each variable from the New Project wizard or by creating a simple project. 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

To create a project in SAS Time Series Studio, you use the New Project wizard. If you already know how you want to structure your data, you can create a default hierarchy when you create the project. However, you can also choose to create this hierarchy later.

This example assumes that you are not familiar with the SASHelp:PriceData data set. 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:

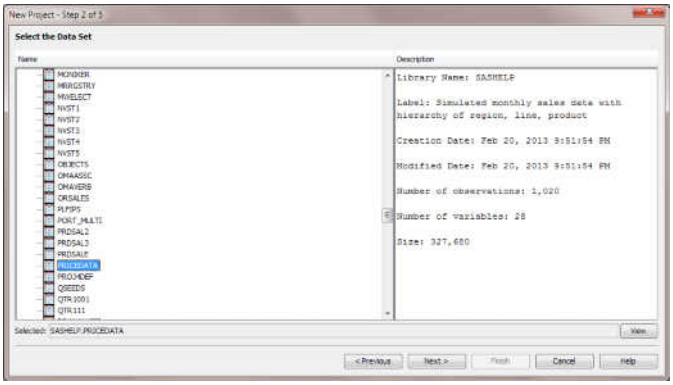
1 In SAS Time Series Studio, select **File** ▶ **New Project**. The New Project wizard appears.

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 `PriceData` as the name of the project. You can also provide a description.

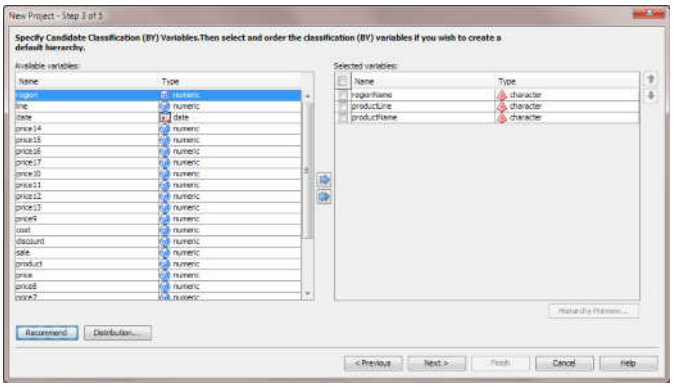
Click **Next**.

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



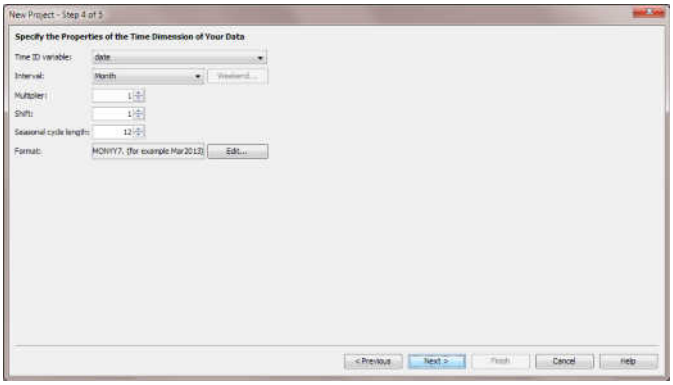
4 To automatically select the classification (BY) variables for the project, click **Recommend**. SAS Time Series Studio selects all of the character variables: regionName, productLine, and productName.

Click **Next**.



Note: In this example, you are not creating a default hierarchy at this time, so do not select the check boxes in the **Selected variables** pane.

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

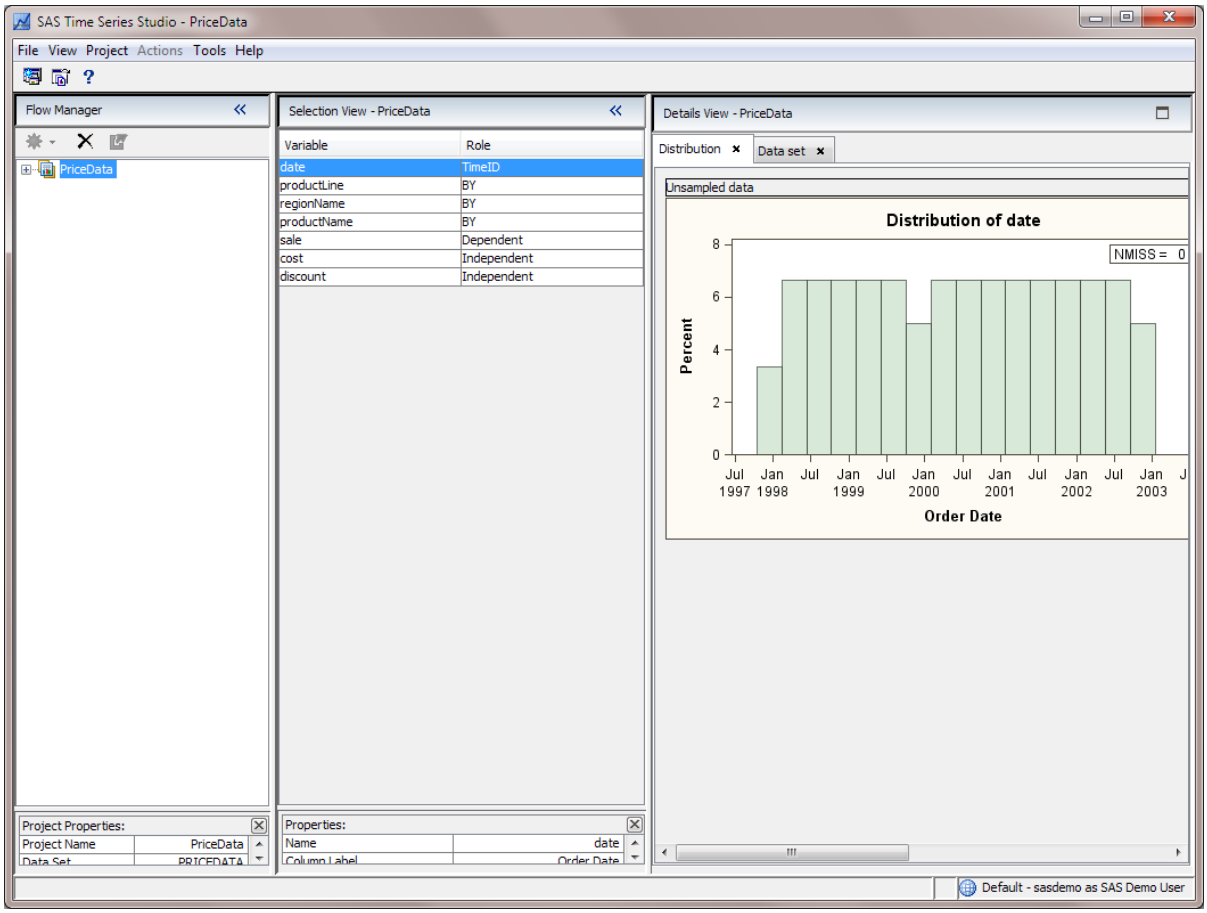


6 In the Specify Analysis Variable Roles and How to Prepare the Data step, assign roles to the following variables:

Variable Name	Role to Assign
cost	Independent
discount	Independent
sales	Dependent

Click **Finish**.

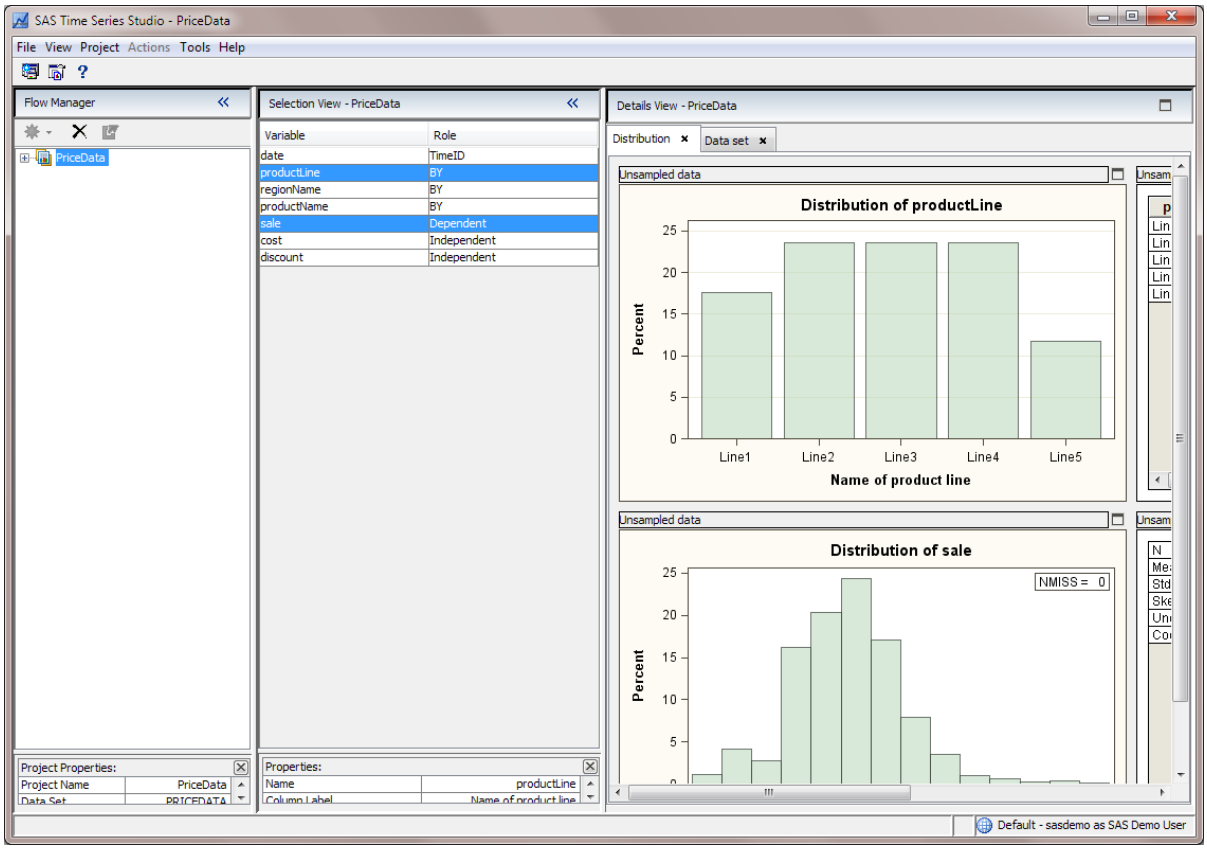
Here are the results in the SAS Time Series Studio workspace:



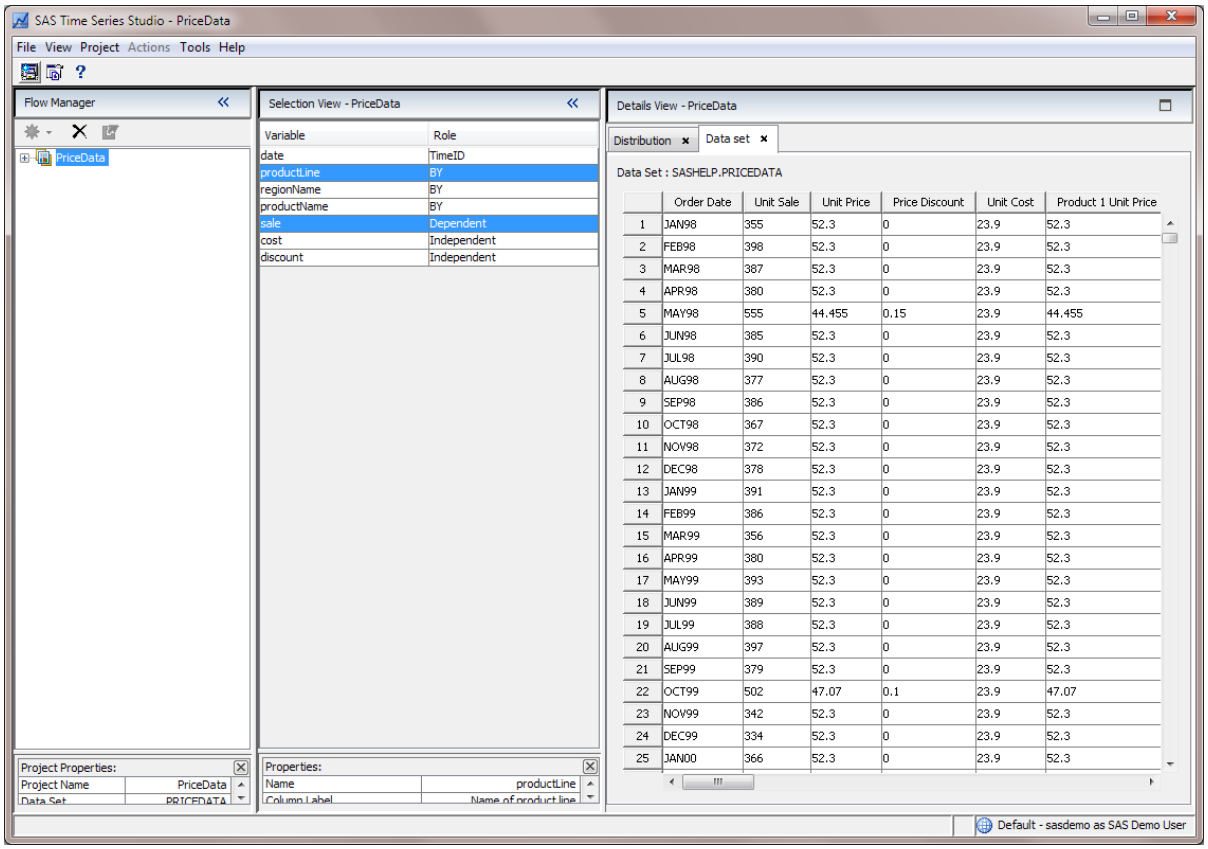
The Selection View displays the name and role of each variable in the project. By default, the first variable, date, is selected. So in the **Distribution** tab of the Details View, you can see a graph of and statistics for the discrete distribution for the date variable.

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 View. In the following example, the distributions for the productLine and sales variables are displayed in the Distribution View. productLine is a discrete distribution, and sale is a continuous distribution.



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



Note: When you have selected the project node 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.

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 71.

8

Creating Hierarchies to Structure Your Data

About Hierarchies

You can use hierarchies to structure your data. Hierarchies can be created with different hierarchical aggregations and for different frequencies (time intervals).

To create a hierarchy, you must assign at least one classification (BY) variable when you create the project. 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.

Note: For ideal performance, do not create a hierarchy with more than 10 BY variables. The number of BY variables determines the number of levels in your hierarchy, and more levels can result in slower performance. In general, you specify only those hierarchy levels that are beneficial from a modeling point of view. If you need to use complex hierarchies for reporting purposes, you can do that as a post-processing step.

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 classification (BY) variables to include in the hierarchy.
3. (Optional) To specify the time interval attributes and the data preparation options, click **Hierarchy Options**. For more information about these options, see “Time ID Variables” on page 42 and “Understanding the Data Preparation Options” on page 33.
4. Click **OK**.

Rename a Hierarchy

To rename a hierarchy:

1 Select the hierarchy name in the Flow Manager, and select **Actions** ▶ **Rename**. The Edit Hierarchy dialog box appears.

2 Specify the new name of the hierarchy.

3 (Optional) Specify a description.

4 Click **OK**.

Display Formatted Values for BY Variables

By default, the values for the BY variables in the hierarchy are displayed as unformatted values. You can choose to display the formatted values. However, before any formatting changes are applied, SAS Time Series Studio must close and reopen the project.

To display the formatted values for the BY variables in the hierarchy:

1 Select **Projects** ▶ **Preferences** ▶ **Display Formatted BY Variable Values**.

2 In the warning dialog box, click **Apply Now**. SAS Time Series Studio must close and reopen the project in order to display the formatted values. If you click **Do Not Apply**, the unformatted values are still displayed.

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 65.)

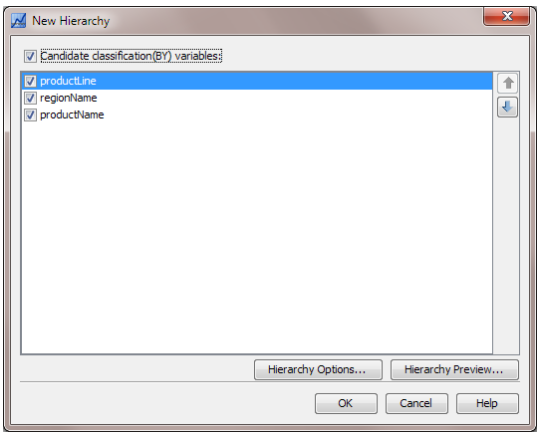
In these examples, you use the same classification (BY) variables to create three different hierarchies in SAS Time Series Studio.

Example 1: Create the productLine > regionName > productName Hierarchy

After looking at the distribution of the data, you decide to structure the data by using the productLine > regionName > productName hierarchy.

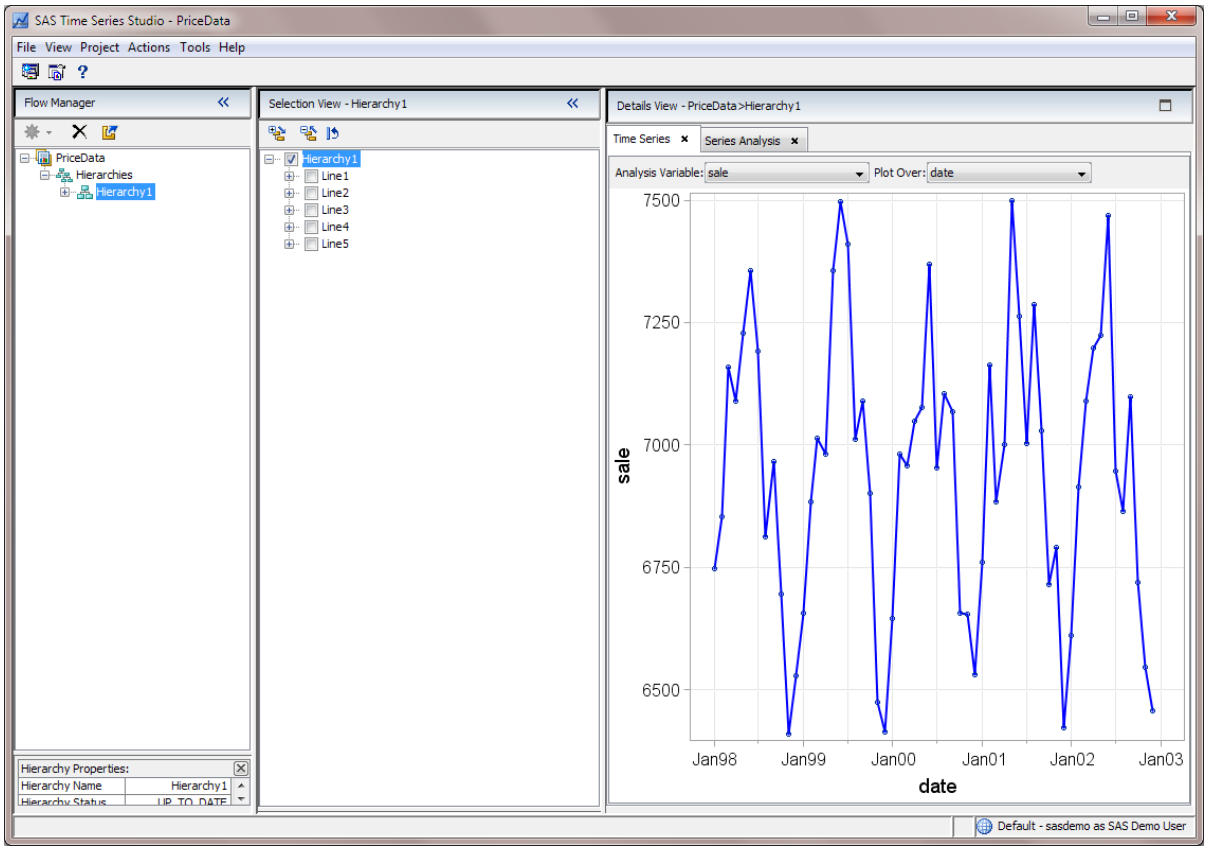
1 In the Flow Manager, select the **Hierarchies** node and select **Actions** ▶ **Add Hierarchy**. The New Hierarchy dialog box appears.

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 the productLine > regionName > productName Hierarchy with Quarterly Time Interval

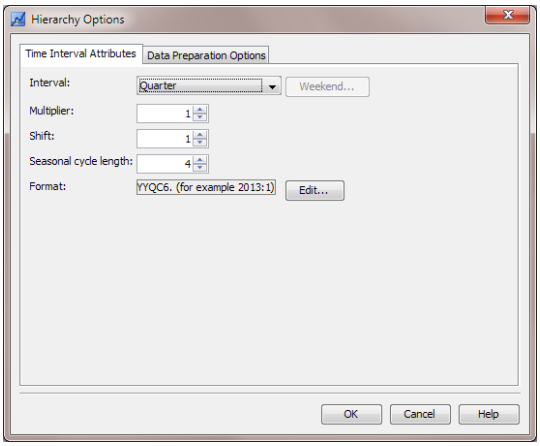
In this example, you create the same hierarchy (productLine > regionName > productName) but change the time interval to quarterly.

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

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

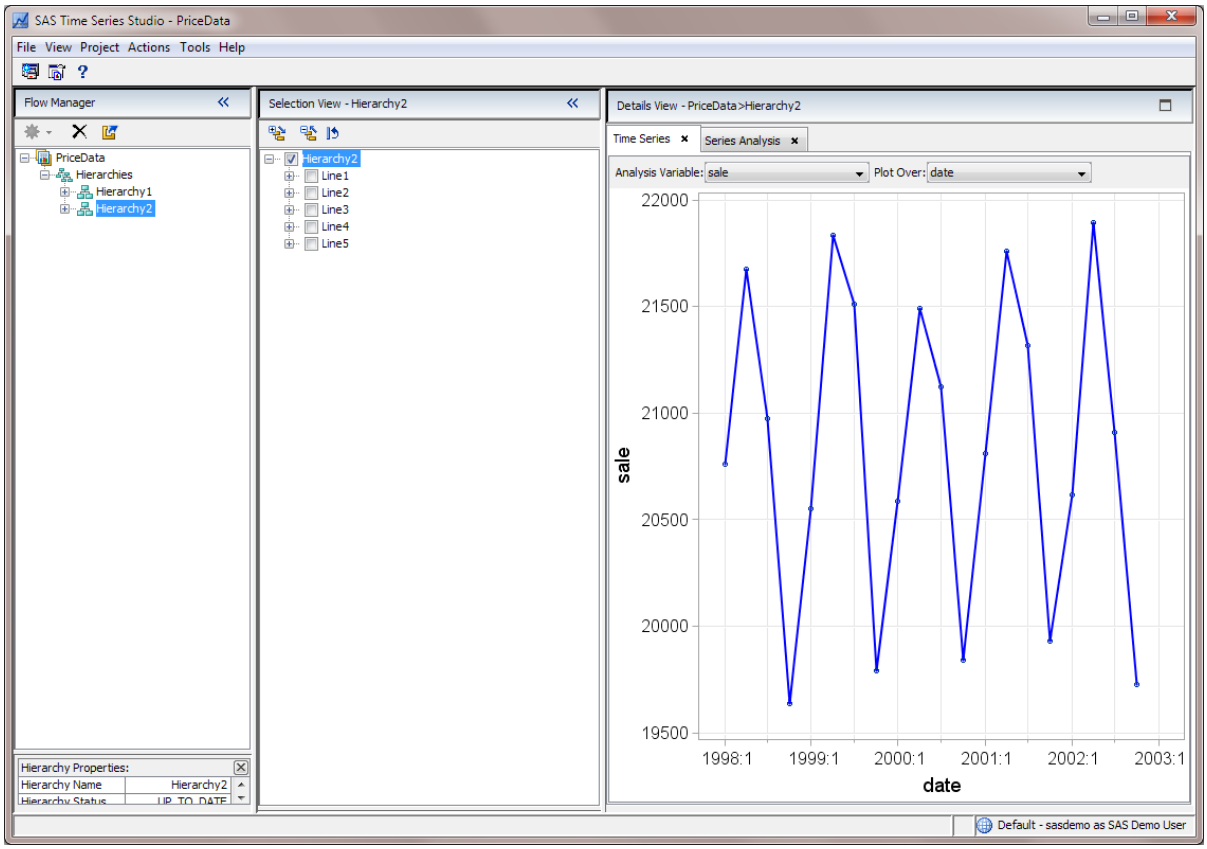
3 Click **Hierarchy Options**. The Hierarchy Options dialog box appears.

4 From the **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.



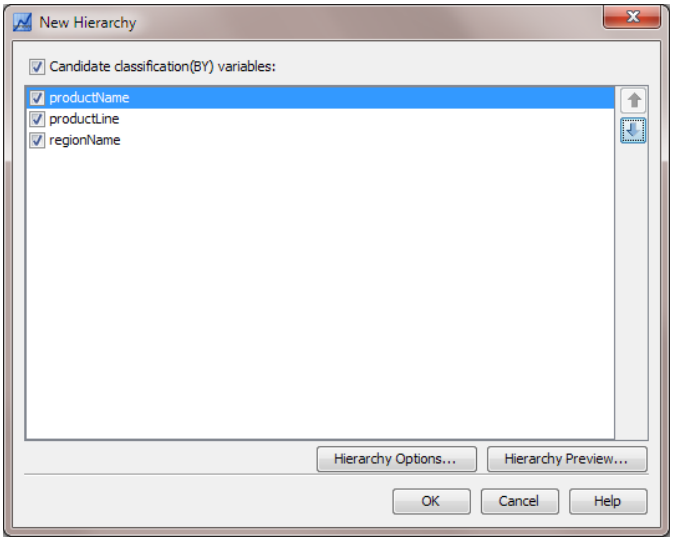
Example 3: Create the productName > productLine > regionName Hierarchy

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 **Actions** ▶ **Add Hierarchy**. The New Hierarchy dialog box appears.

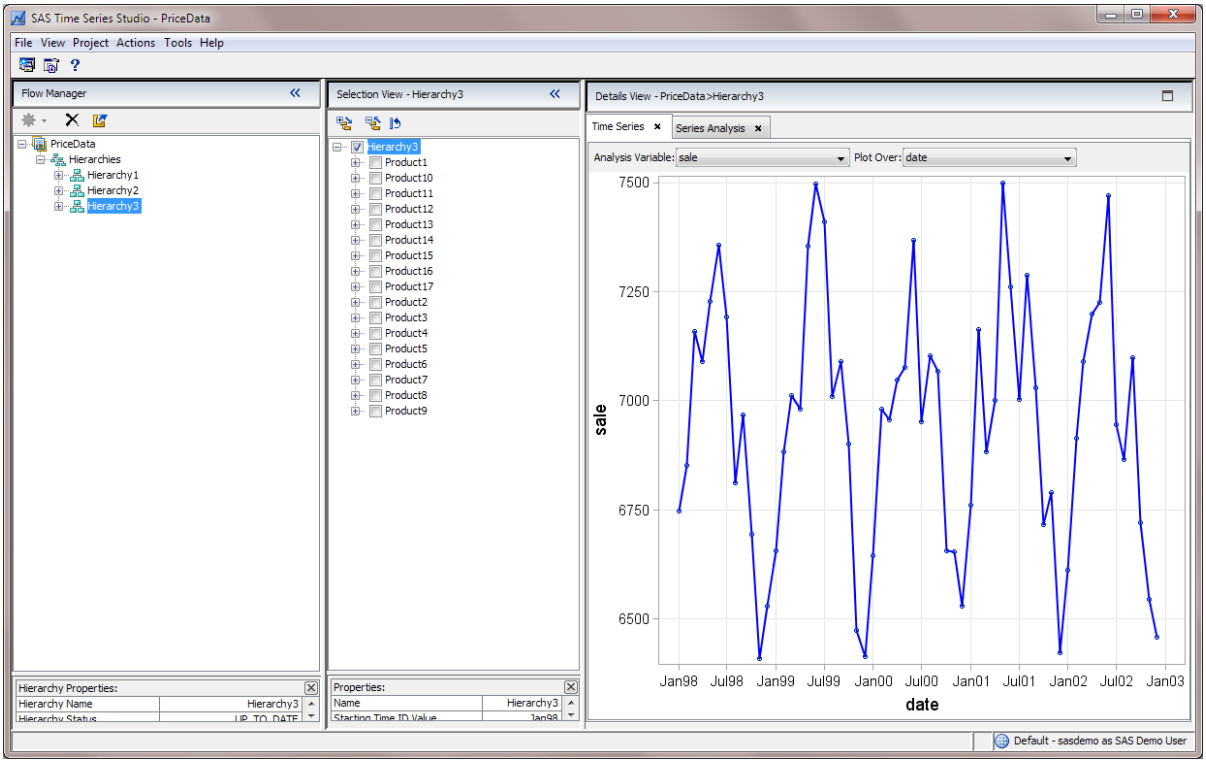
2 Select the **Candidate classification (BY) variables** check box.

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.

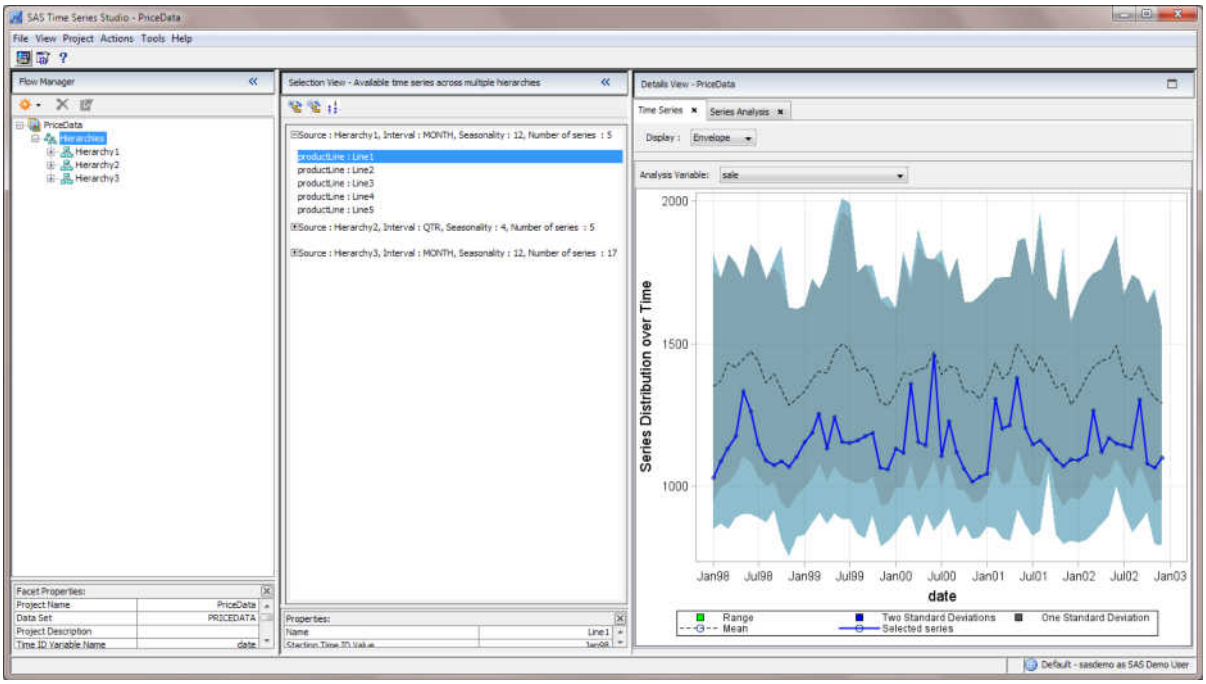
This hierarchy is called Hierarchy3 in the workspace.



Comparing Hierarchies in the Selection View

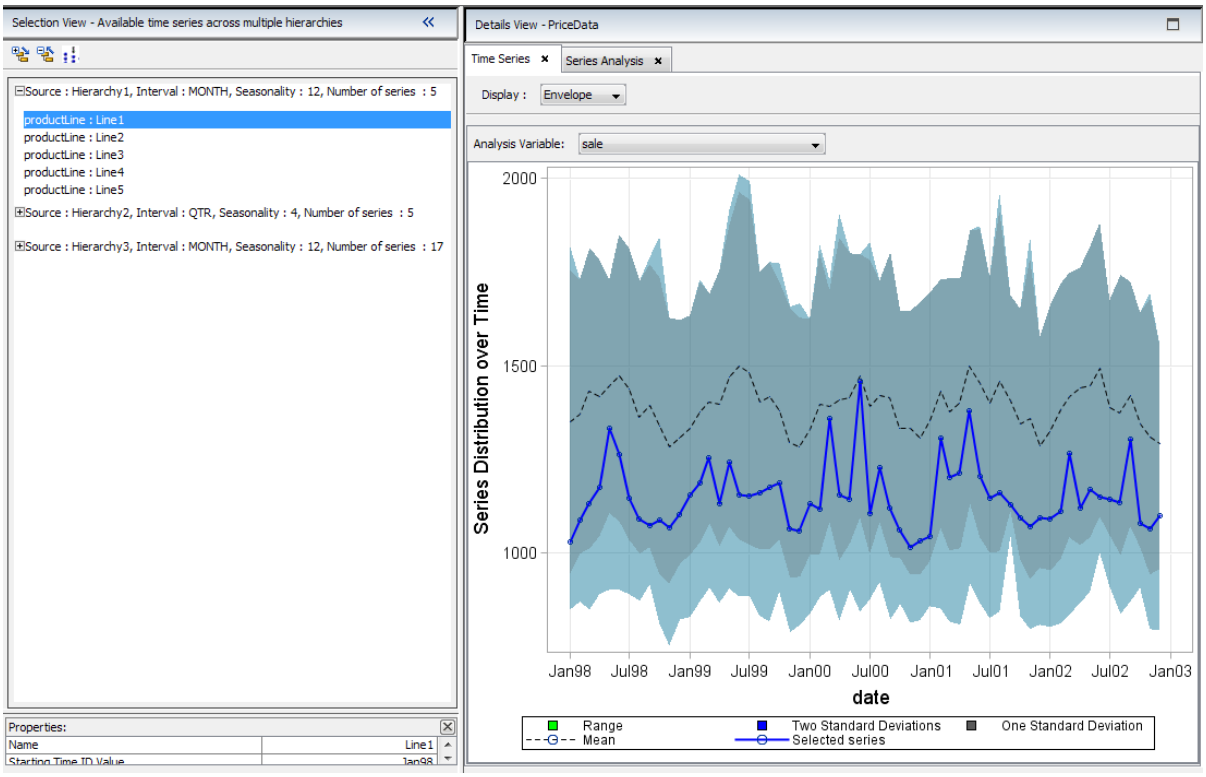
The Hierarchies Node

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.

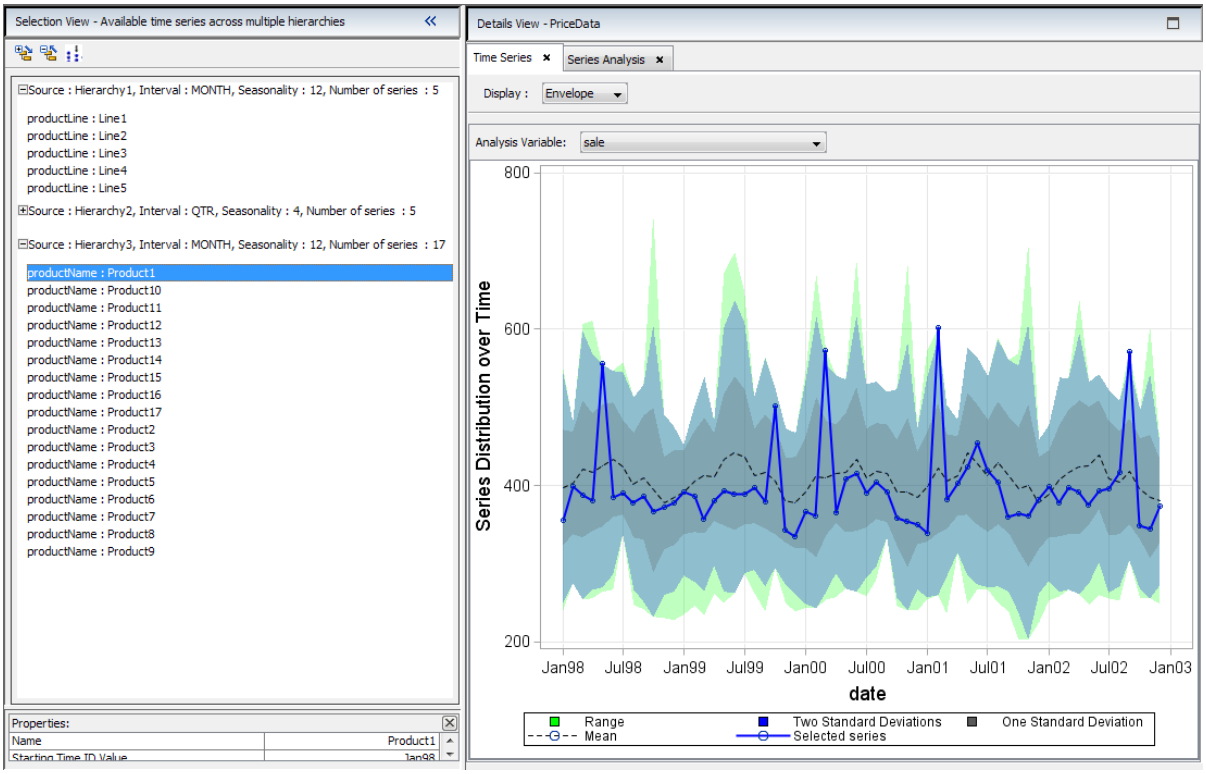


When you select a series in a hierarchy, the graphs in the Details View update to reflect the selected series.

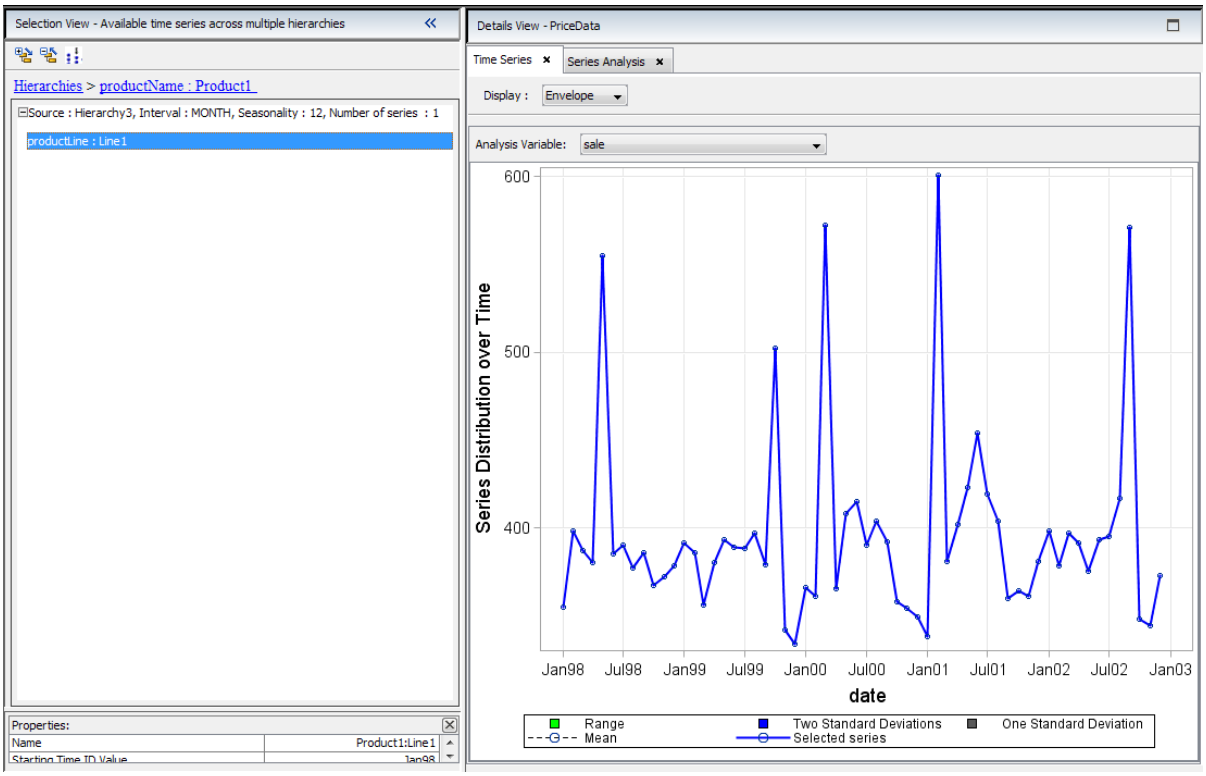
For example, here are the results when you select **productLine: Line1** in Hierarchy1. In the graph for the time series analysis, the selected series appears in blue.



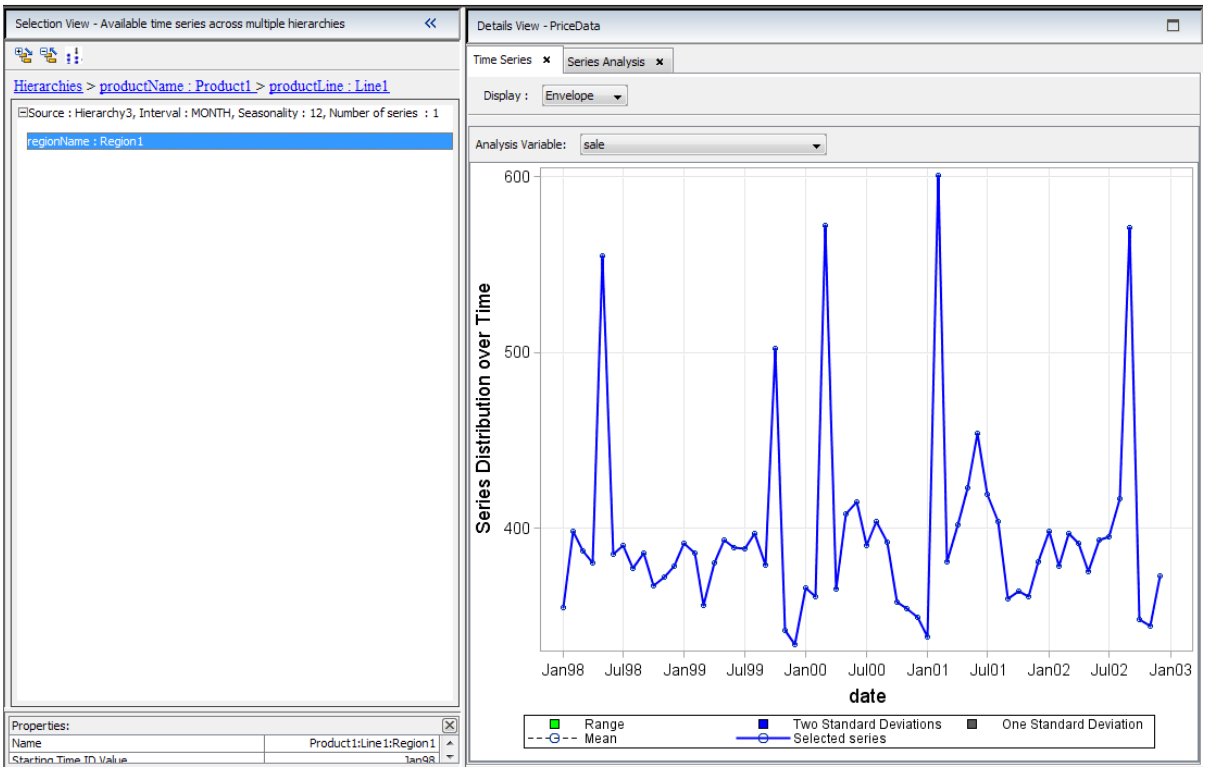
This example shows the productName: Product1 series in Hierarchy3.



When you double-click the name of a hierarchy, you are drilling down through the levels in 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.

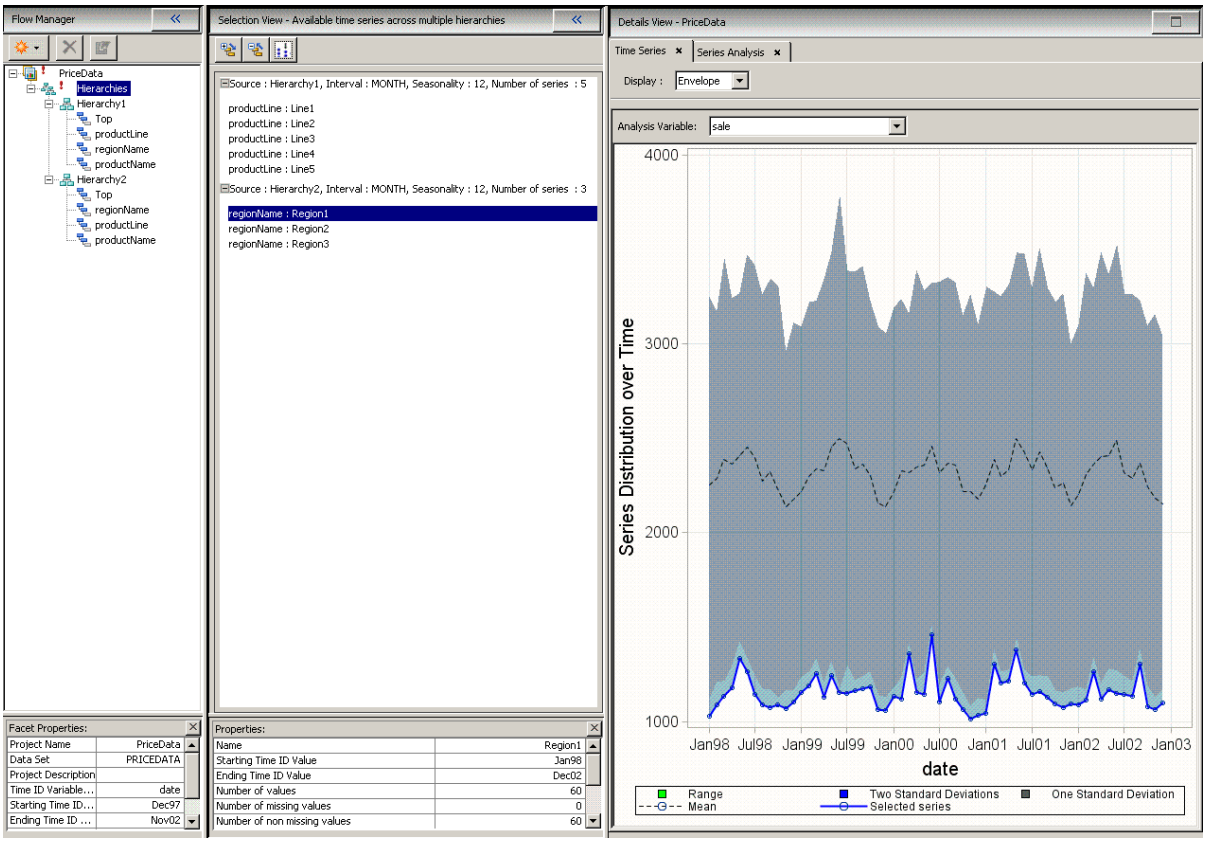
You can dynamically add a new hierarchy by adding new categories.

To add a new category:

1 Click  in the toolbar of the Selection View. The New Category dialog box appears.

2 Select the classification (BY) variable to use to create the hierarchy. Click **OK**.

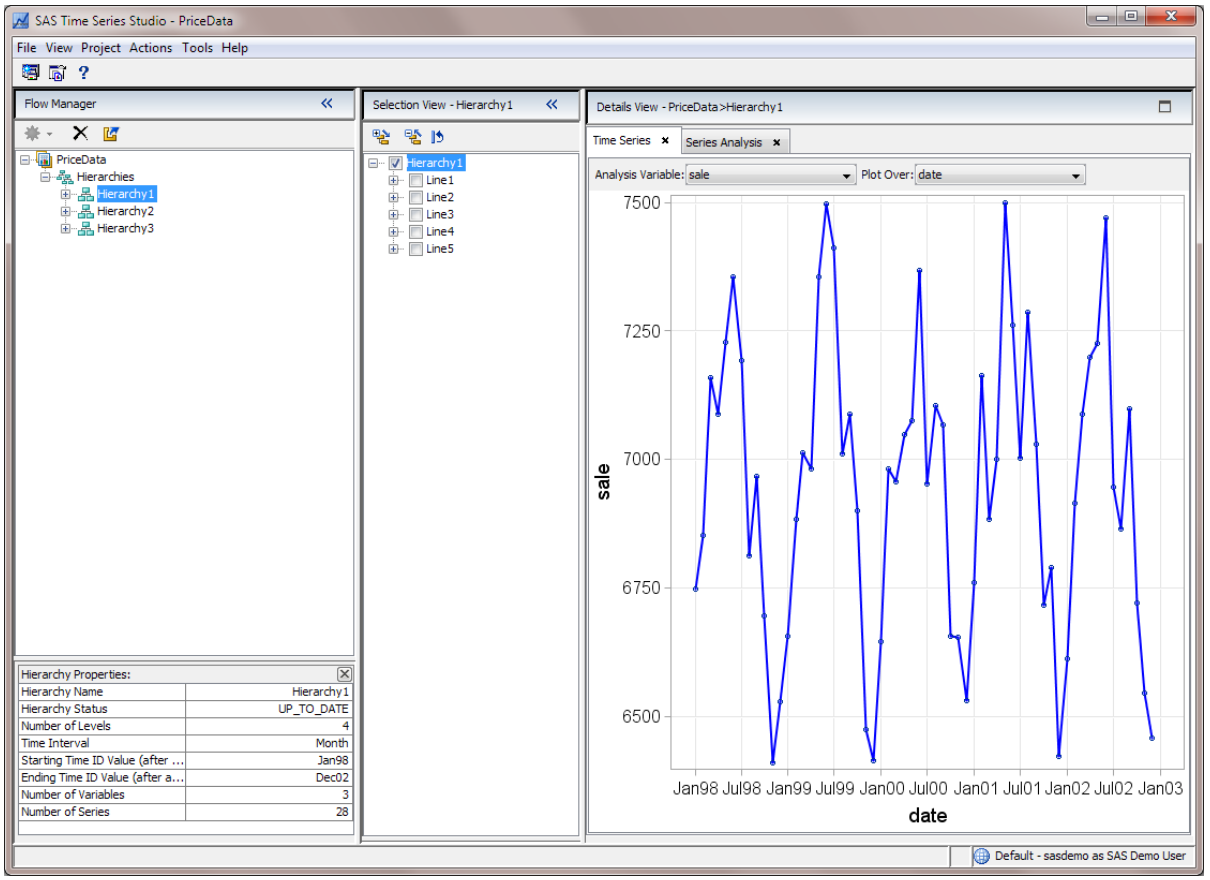
The new hierarchy is now available in the project.



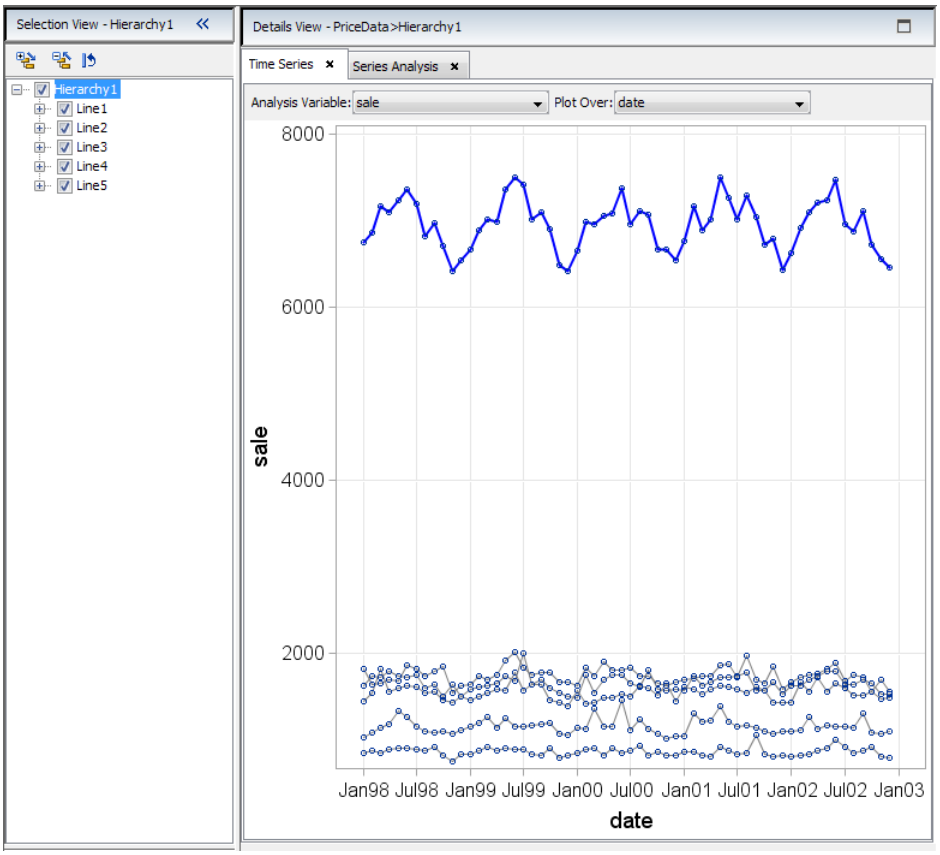
Individual Hierarchies

Using the Flow Manager, you can also view individual hierarchies. For example, if you select 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:



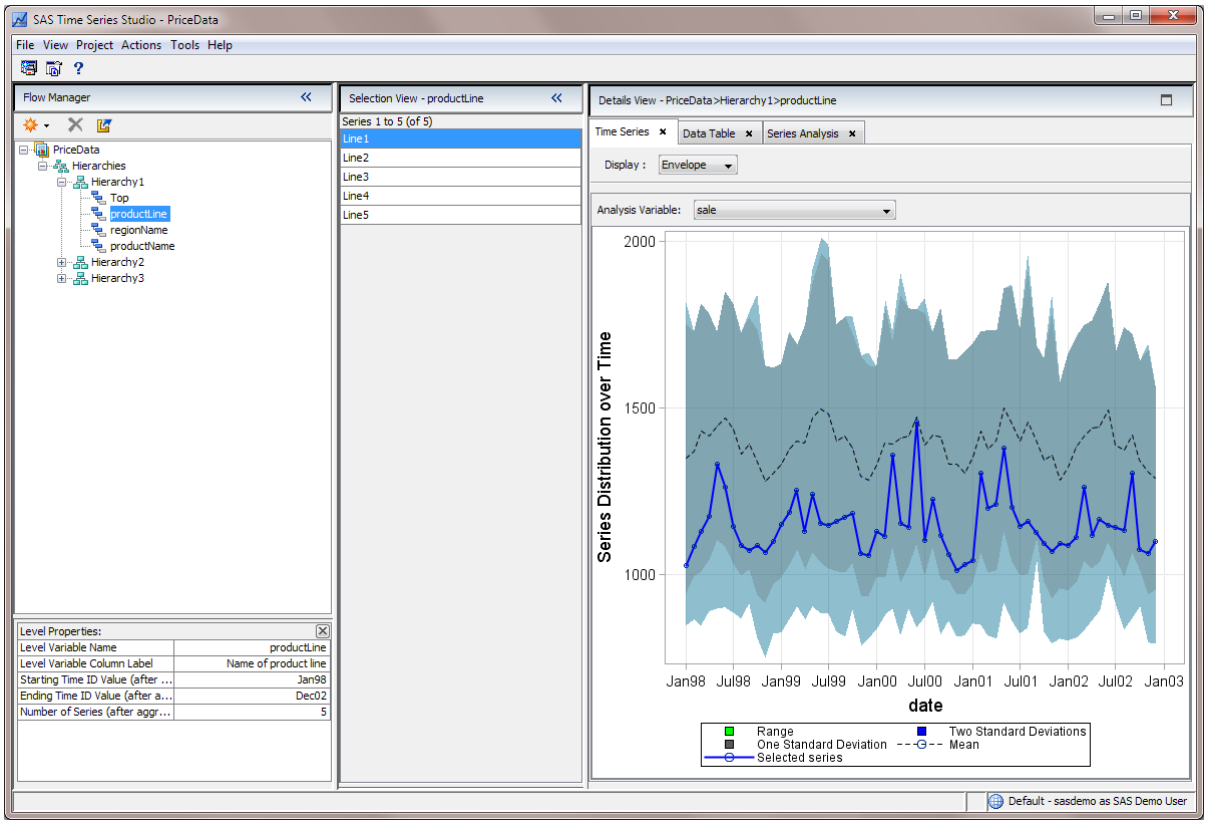
Using the Selection View, you can view the series for all of the children in the hierarchy. For example, right-click **Hierarchy1** and select **Select Children**. In the time series graph, the line for Hierarchy1 appears in blue because Hierarchy1 is selected in the Selection Manager.



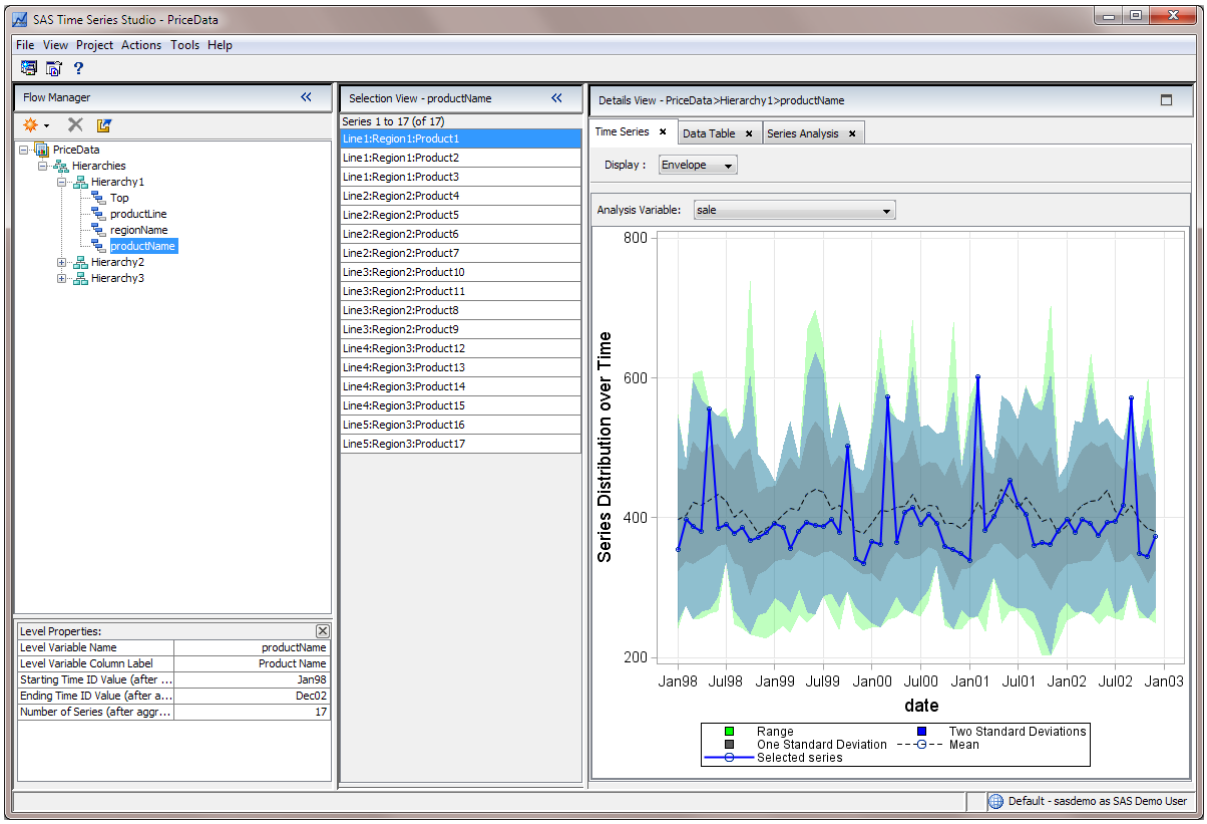
To clear the check boxes, click  .

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 further that you drill down into 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 Selection View.

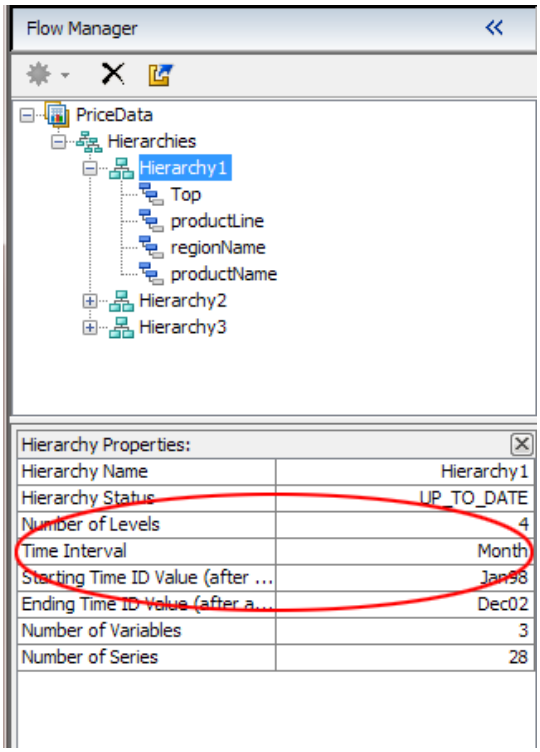


If you select the productName level, then all of the data is displayed in the results.

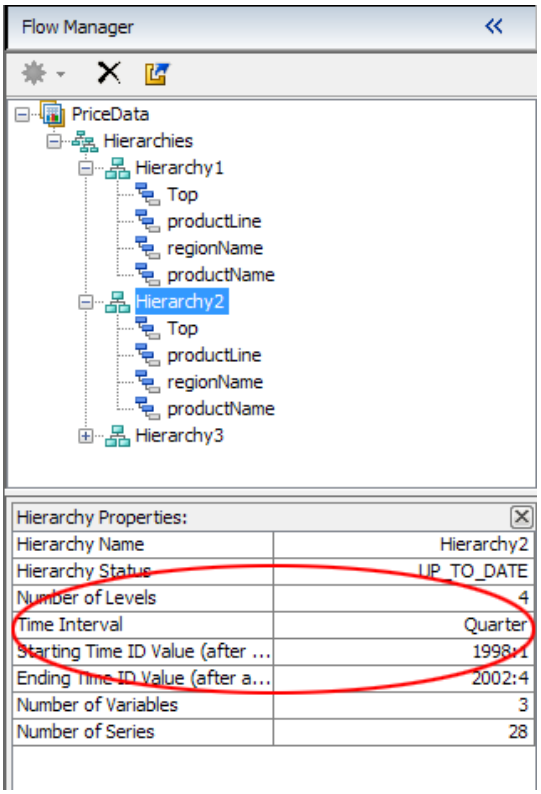


Comparing Hierarchies with Different Time Intervals

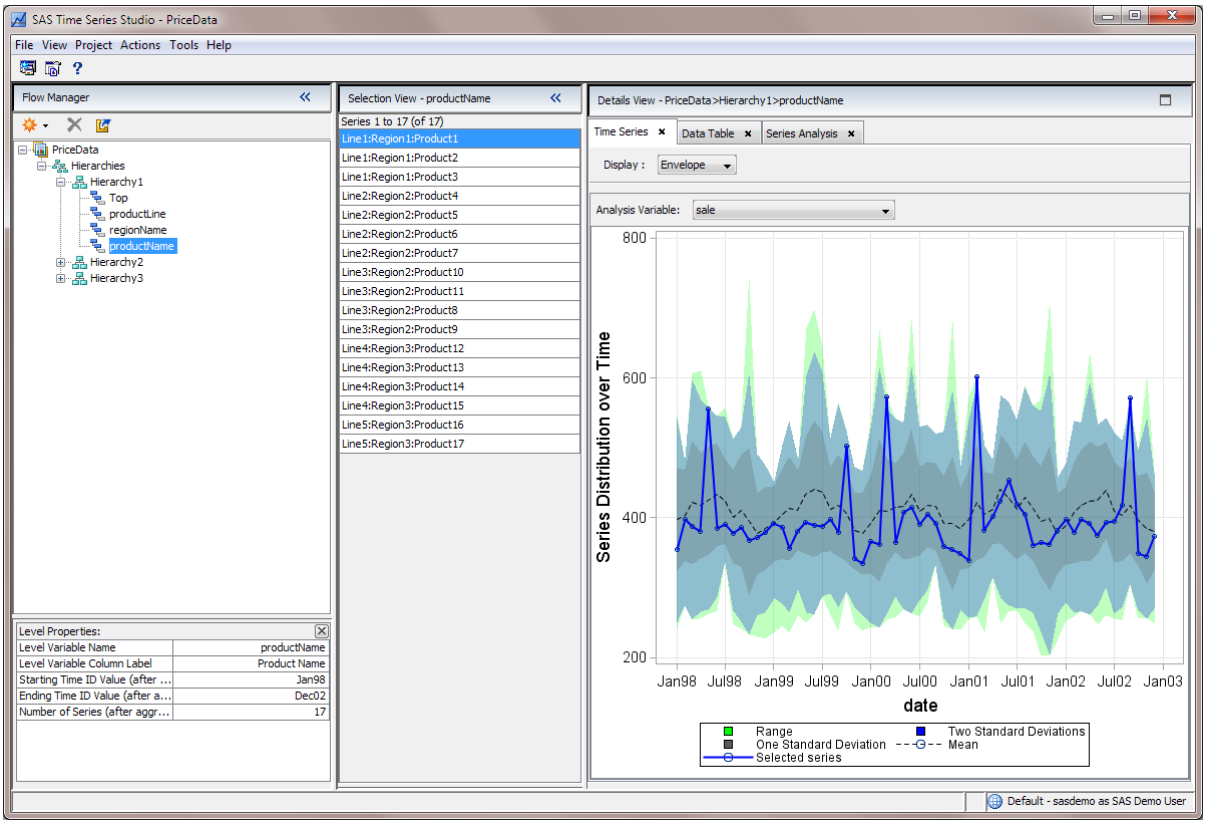
In the example, Hierarchy1 has three classification (BY) variables and is ordered productLine > regionName > productName. The hierarchy uses a monthly time interval with a start date of Jan98 and an end date of Dec02.



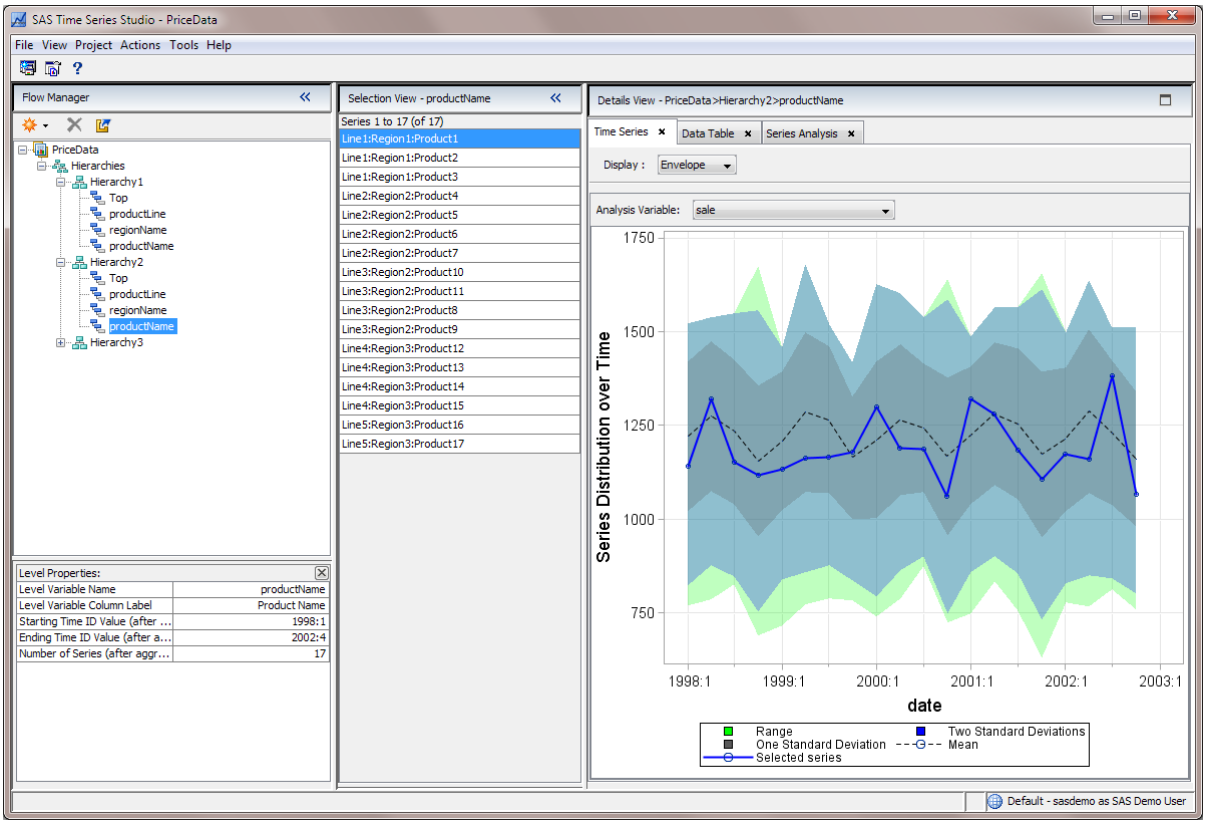
Hierarchy2 has three classification (BY) variables and is ordered by productLine > regionName > productName. This hierarchy uses a quarterly time interval with a start date of 1998:1 (the first quarter in 1998) and an end date of 2002:4 (the fourth quarter in 2002).



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 accumulated at a lower frequency.

9

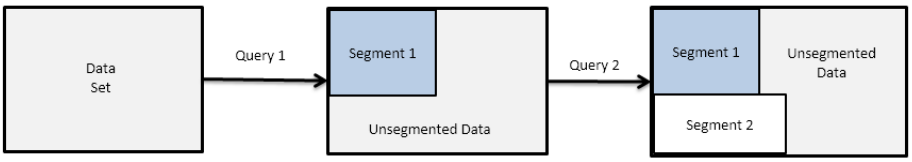
Understanding Segmentation and Subsets

Understanding Segments

What Is a Segment?

Often, you need to partition data into different groups based on the nature of the data (for example, slow moving items, new products, and so on). Segmentation is a process for creating these partitions, also called segments. In SAS Time Series Studio, you start with all of the data. You use queries to create the individual segments. Data that does not satisfy the criteria in the query is grouped into an unsegmented node. All of the segments (including the unsegmented node) make up 100% of the data.

In SAS Time Series Studio, you can create more than one segmentation node to explore different ways of segmenting the data. Each segmentation process starts with all of the data that is available for the selected node.

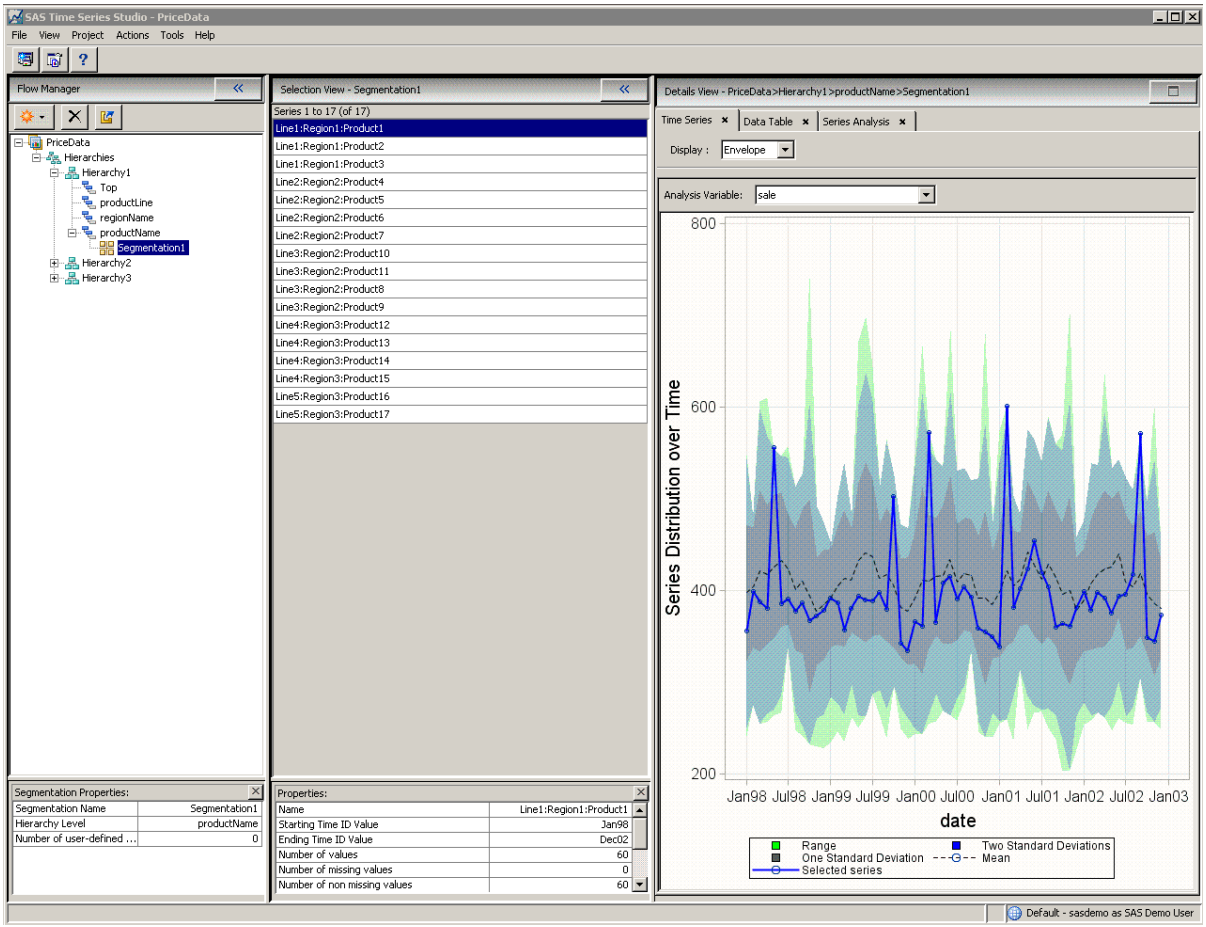


Example 1: Creating the Sales_Less_Than_600 Query

To create this query:

1 Select **productName** in Hierarchy1, and from the main menu, select **Actions** ▶ **Add Segmentation**. The Segmentation1 node is added to the Flow Manager. Because no queries have been defined yet, the number of series in Segmentation1 is the same as the number of series in productName.

Here are the 17 series in Segmentation1:

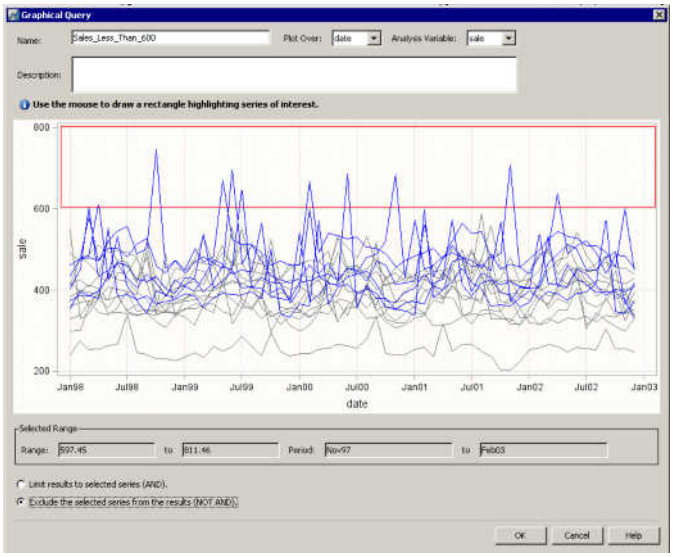


2 In the Flow Manager, select the **Segmentation1** node, and from the main menu, select **Actions** ▶ **New Segment** ▶ **Add Graphical Query**. The Graphical Query dialog box appears.

3 Specify the following information for the graphical query:

- For the name of the query, enter `Sales_Less_Than_600`.
- In the graph, create a rectangular box for sales between 600 and 800.
- Select **Exclude the selected series from the results**.

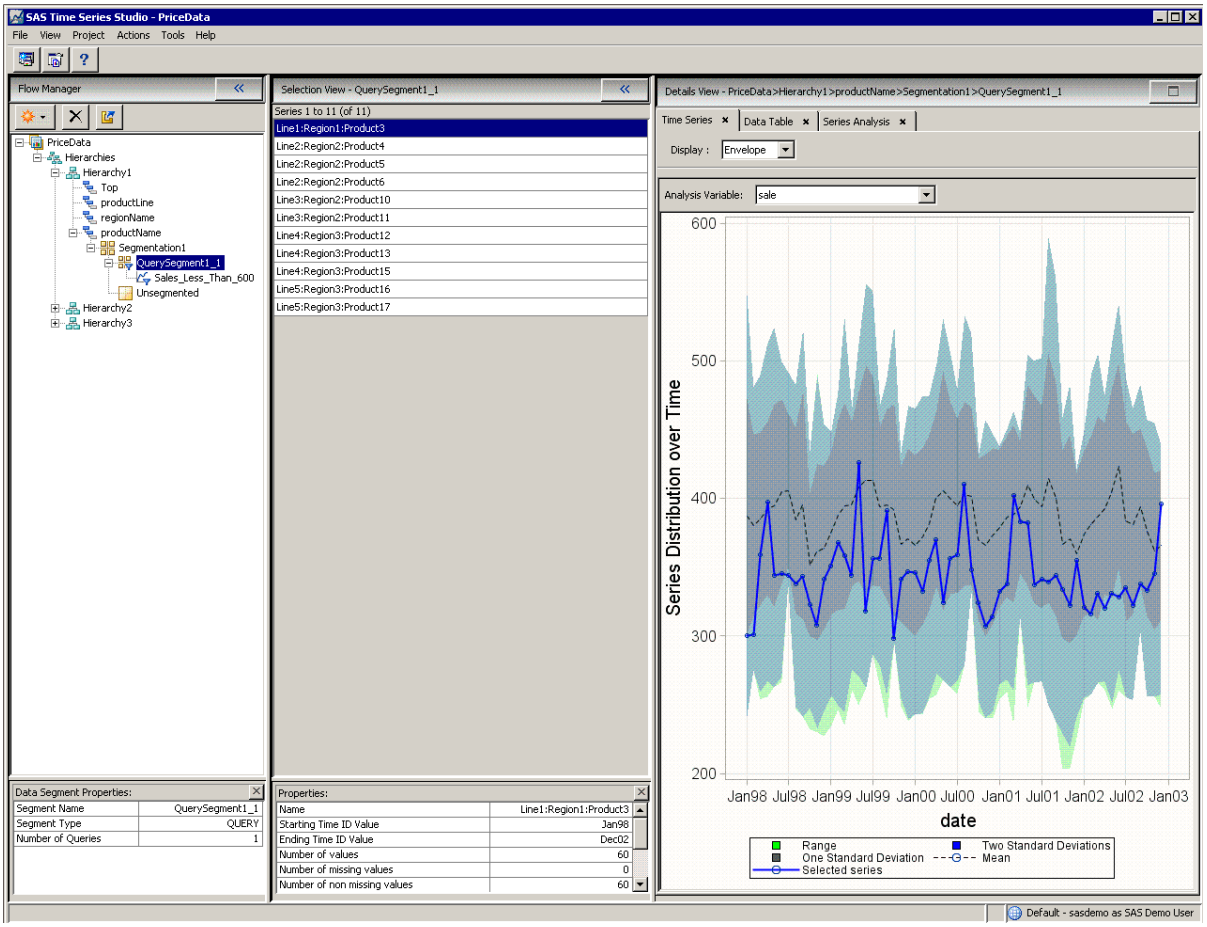
When you are finished, the Graphical Query dialog box should appear similar to the following:



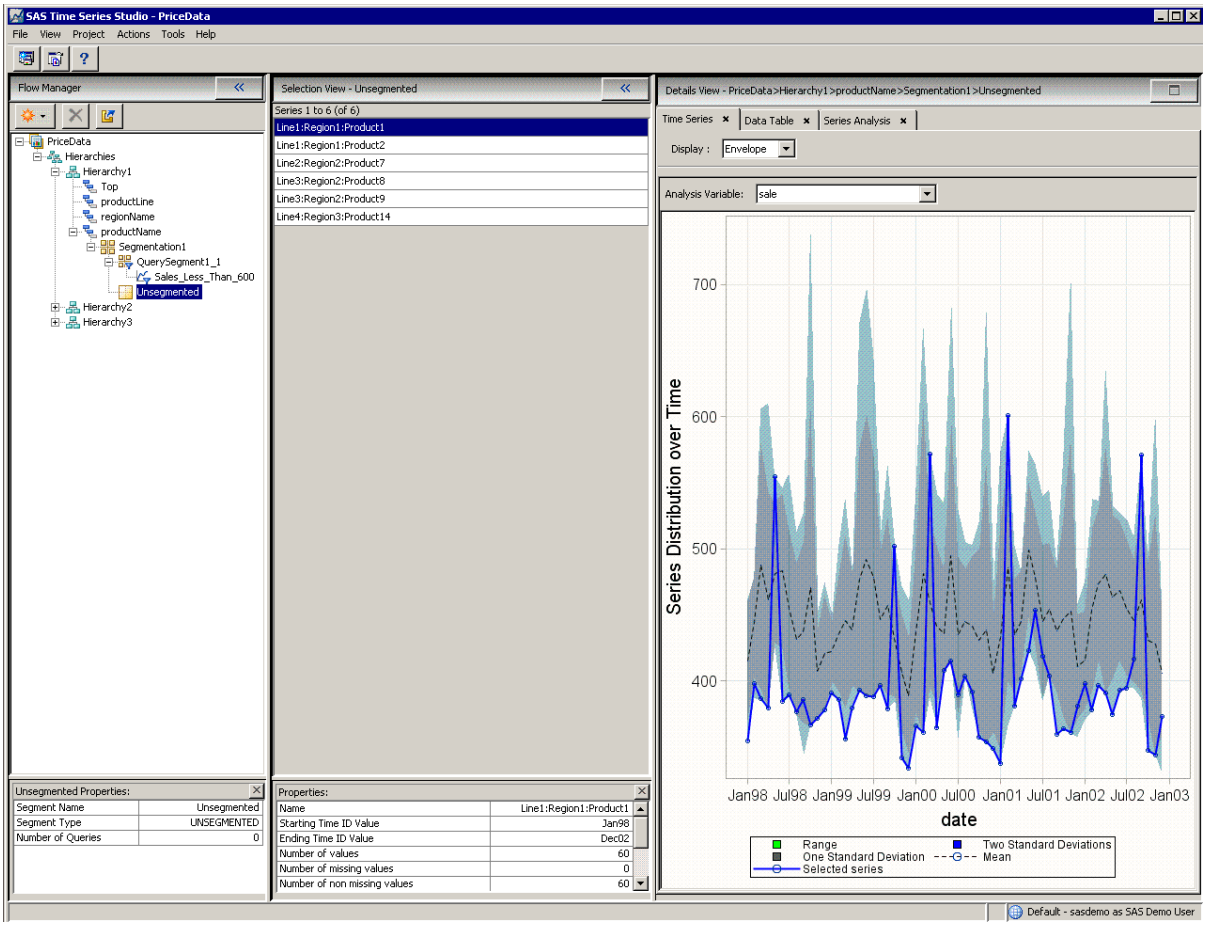
Click **OK**.

In the Flow Manager, the QuerySegment1_1 node appears and contains 11 series. When you created the query, you specified that the selected series (the series that met the criteria) should be excluded from the results.

As a result, the series in QuerySegment1_1 do not include any series that have a sales value between 600 and 800.



The **Unsegmented** node now contains the six series that have a sales value between 600 and 800. These series were excluded from the segment results.



Example 2: Creating the Price_Discount Query

As you continue to explore the data, you decide to create a segment that contains the series with a price discount.

To create the Price_Discount query:

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

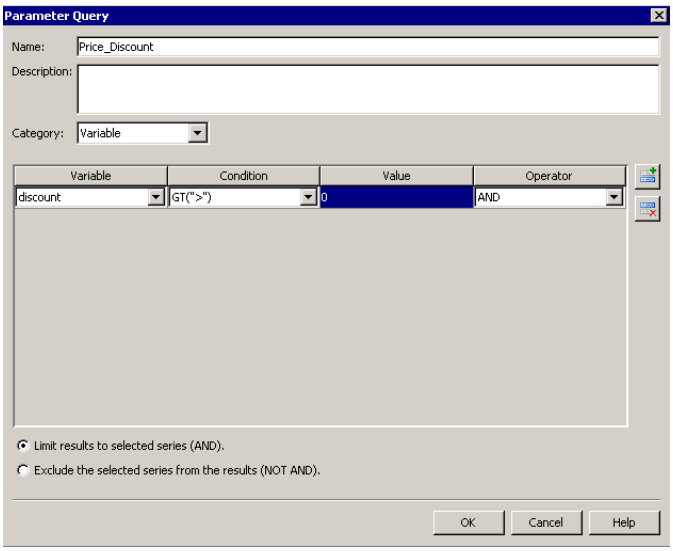
2 Specify the following information for the parameter query:

- For the name of the query, enter **`Price_Discount`**.
- From the **Category** drop-down list, select **Variable**.
- To add a row to the table, click  . In the table, specify these values:

Option Name	Value to Specify for That Option
Variable	**discount**
Condition	**GT(“>”)**
Value	**0**

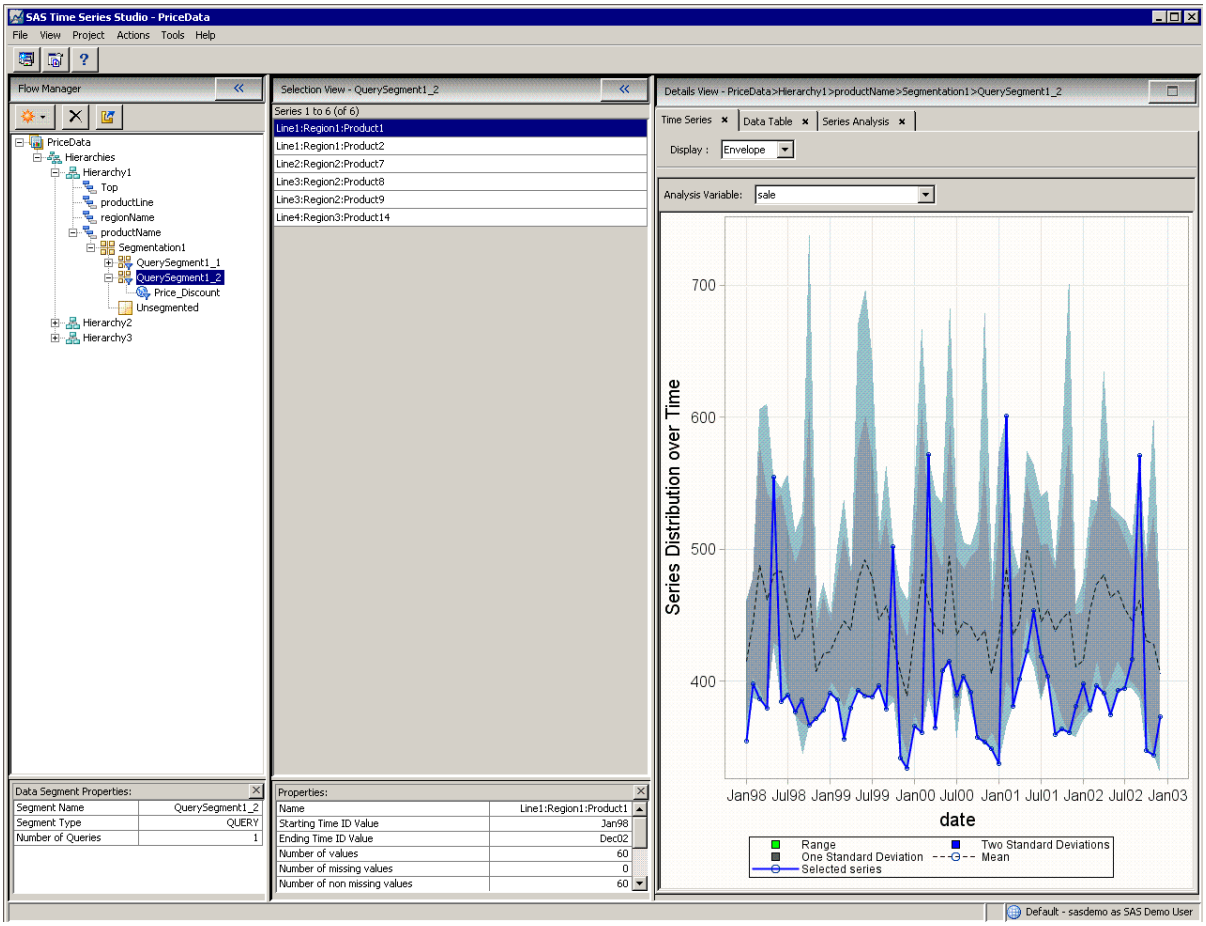
- Select **Limit results to the selected series (AND)** to include only the series that meet the defined criteria in the results.

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

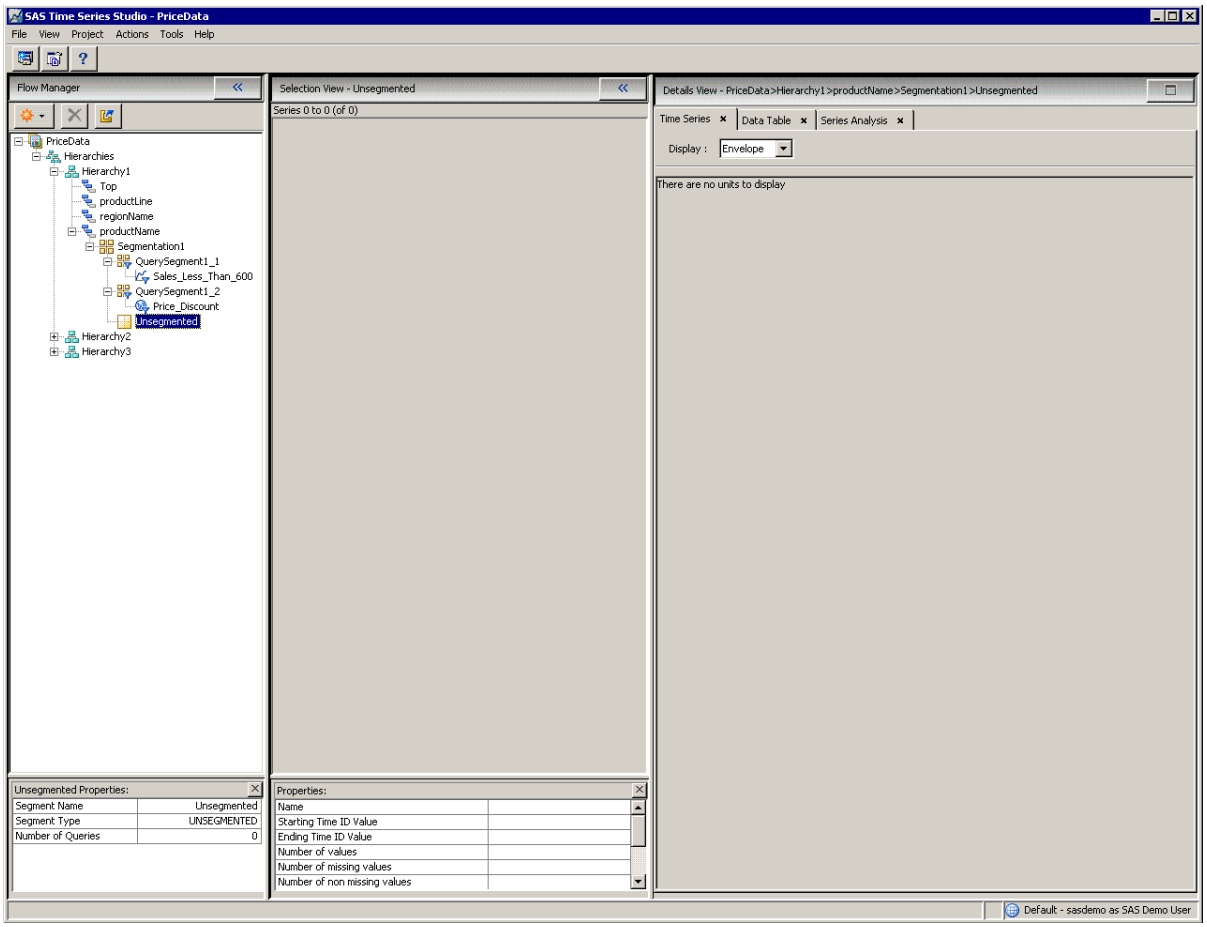


Click **OK**.

In the Flow Manager, QuerySegment1_2 contains the six series that had a discount of greater than zero.


SAS Time Series Studio - PriceData
File View Project Actions Tools Help
Flow Manager
PriceData
Hierarchies
Hierarchy1
Top
productLine
regionName
productName
Segmentation1
QuerySegment1_1
QuerySegment1_2
Price_Discount
Unsegmented
Hierarchy2
Hierarchy3
Selection View - QuerySegment1_2
Series 1 to 6 (of 6)
Line1:Region1:Product1
Line1:Region1:Product2
Line2:Region2:Product7
Line3:Region2:Product8
Line3:Region2:Product9
Line4:Region3:Product14
Details View - PriceData>Hierarchy1>productName>Segmentation1>QuerySegment1_2
Time Series
Data Table
Series Analysis
Display : Envelope
Analysis Variable: sale
Series Distribution over Time
700
600
500
400
Jan98 Jul98 Jan99 Jul99 Jan00 Jul00 Jan01 Jul01 Jan02 Jul02 Jan03
date
Range
One Standard Deviation
Selected series
Two Standard Deviations
Mean
Data Segment Properties:
Segment Name QuerySegment1_2
Segment Type QUERY
Number of Queries 1
Properties:
Name Line1:Region1:Product1
Starting Time ID Value Jan98
Ending Time ID Value Dec02
Number of values 60
Number of missing values 0
Number of non missing values 60
Default - sasdemo as SAS Demo User

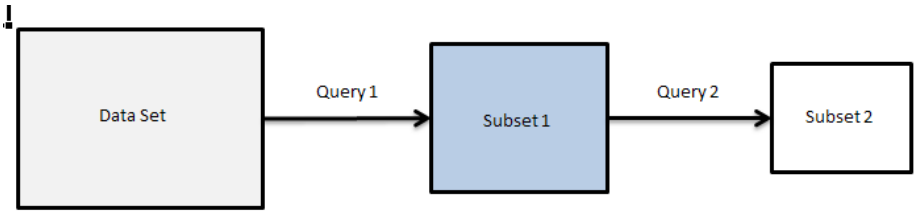
The **Unsegmented** node now contains zero series because all the series are in QuerySegment1_1 or QuerySegment1_2.



Understanding Subsets

What Is a Subset?

A subset is a part of the entire data. To create a subset, you start with all of the data. Then you define a query to select the data of interest. The result of the query will contain only a portion of the original data.



Example 1: Creating the Region_3_Products Query

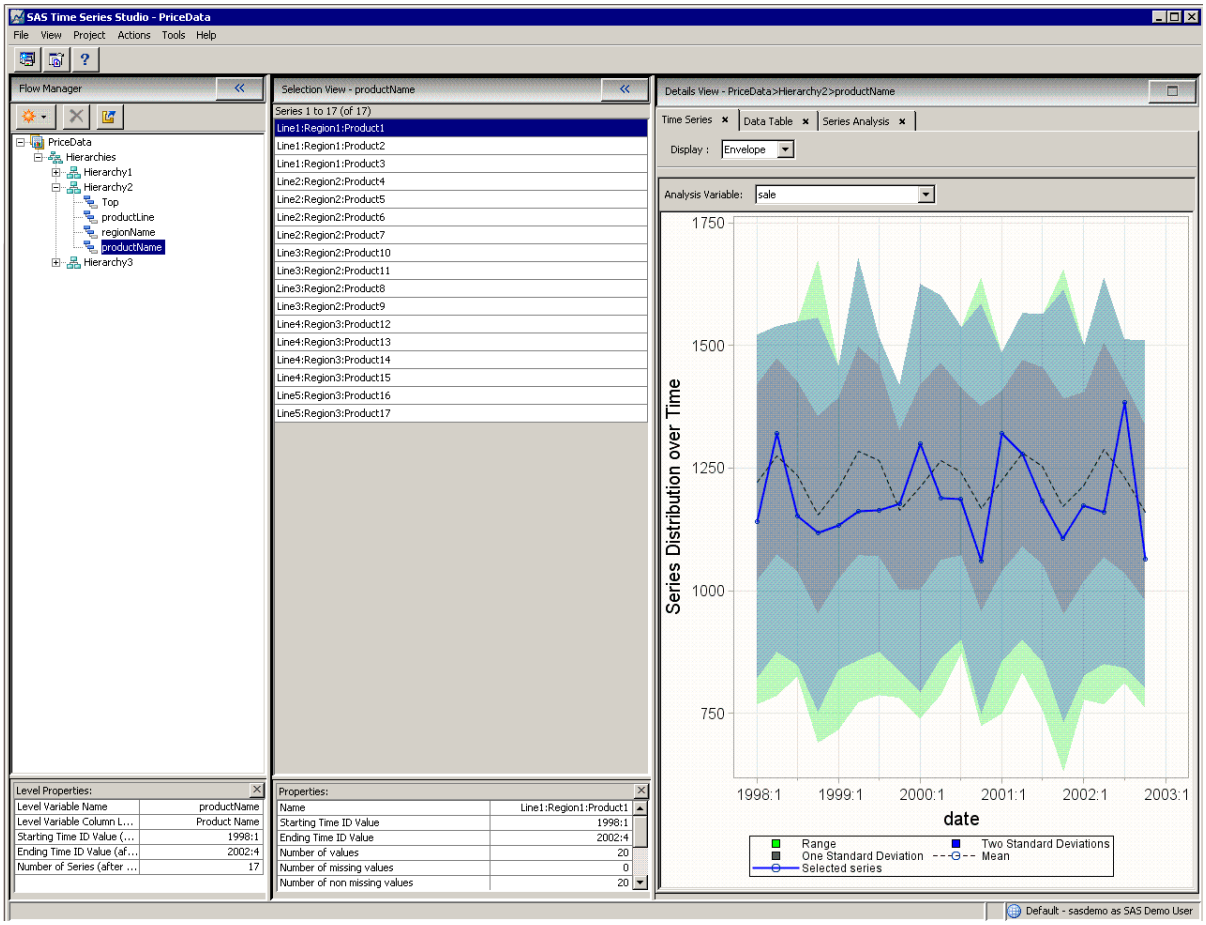
In SAS Time Series Studio, a subset could consist of several cascading queries. How those queries are combined depends on whether you specify to limit the results to the series in the query, exclude the selected series from the results, or include the selected series in the results.

To create the Region_3_Products query:

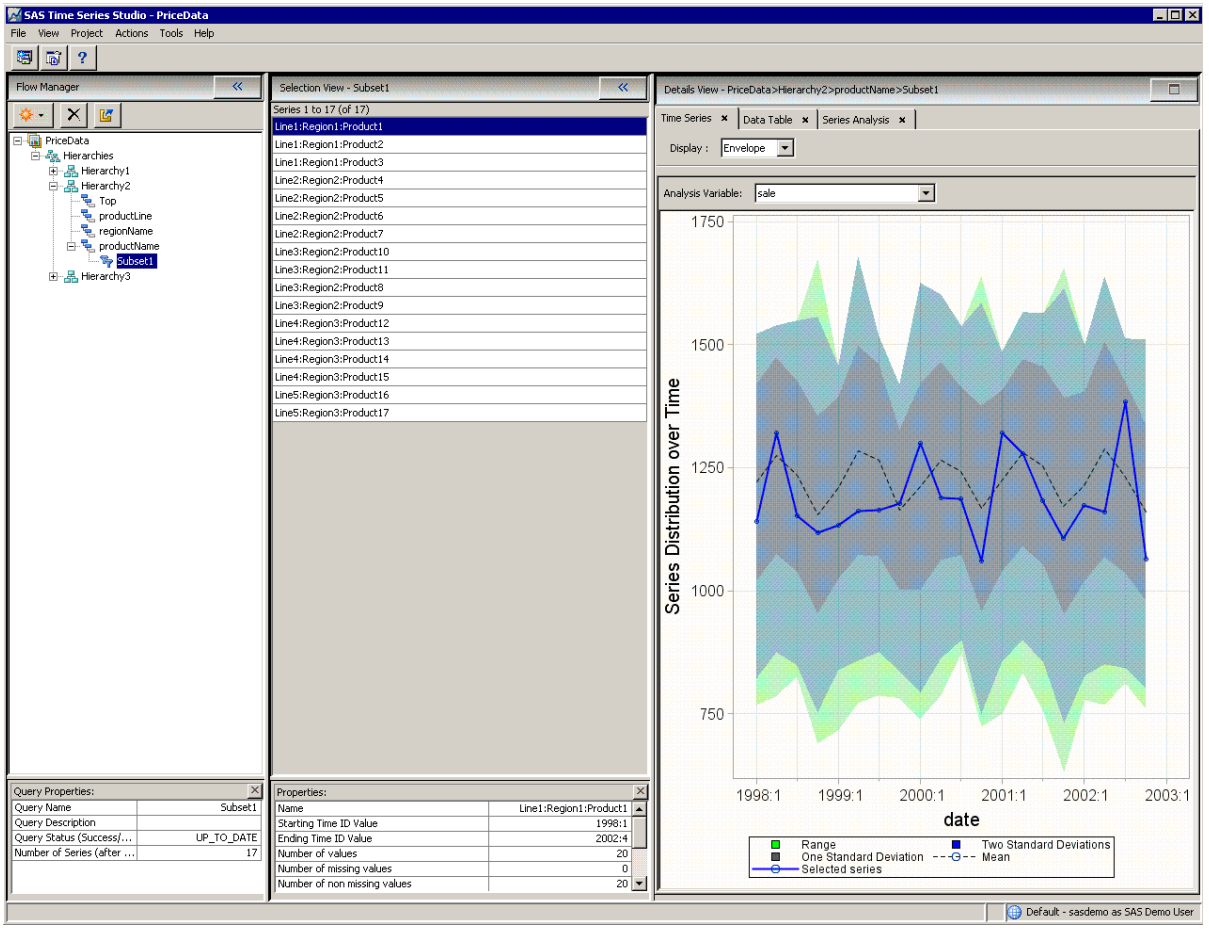
1 Select the **productName** node in Hierarchy2, and from the main menu, select **Actions** ▶ **Add Subset**. Subset1 is added to the Flow Manager.

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 View.

Here are the 17 series for productName:



Here are the 17 series for Subset1:

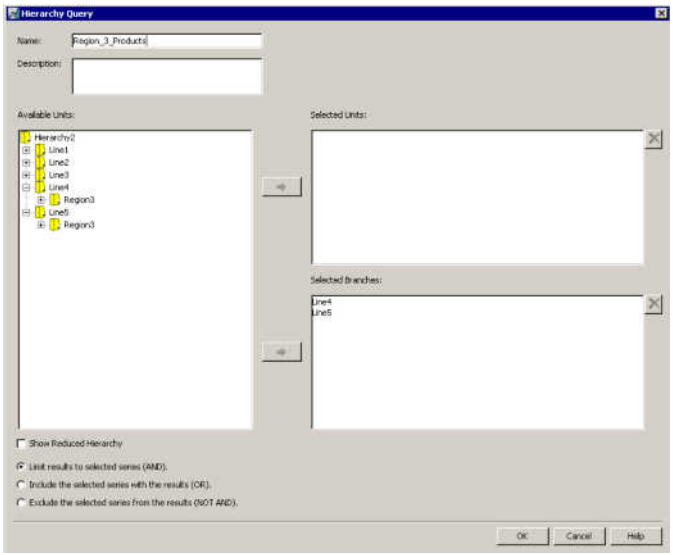


2 In the Flow Manager, select the **Subset1** node, and from the main menu, select **Actions** ▶ **Add Hierarchical Query**. The Hierarchical Query dialog box appears.

3 Specify the following information for the hierarchical query:

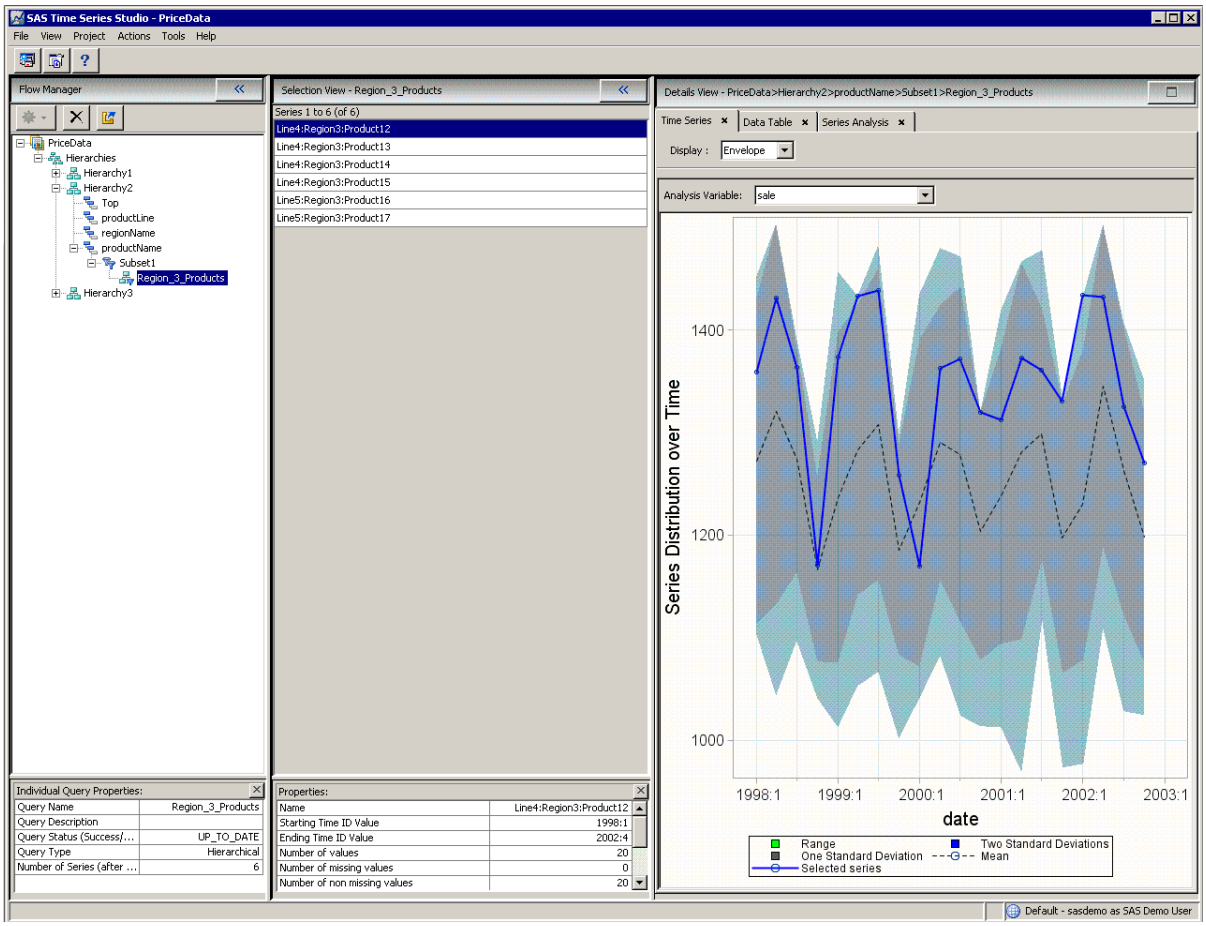
- For the name of the query, enter `Region_3_Products`.
- From the list of available branches, select **Line4** and **Line5**.
- Select **Limit results to the selected series (AND)**.

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 the Region_3_Products query contains only the six series that met the criteria.



Example 2: Creating the No_Line_4_Products Query

You decide to subset the data even further by removing the Line4 products. To create the No_Line4_Products query:

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

2 Specify the following information for the hierarchical query:

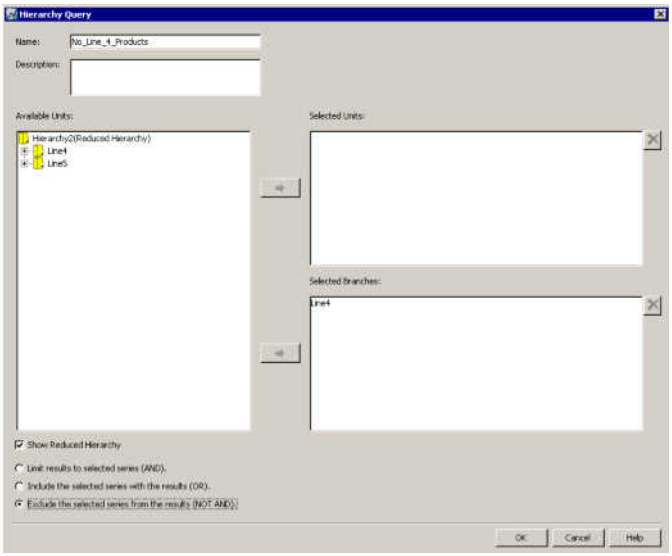
- For the name of the query, enter **`No_Line_4_Products`**.

- From the list of available branches, select **Line4**.

> **TIP** Because you are creating a subset of the results from the Region_3_Products subset, you might want to view only those series that are part of the Region_3_Products subset. To see only these units in the hierarchy, select **Show Reduced Hierarchy**.

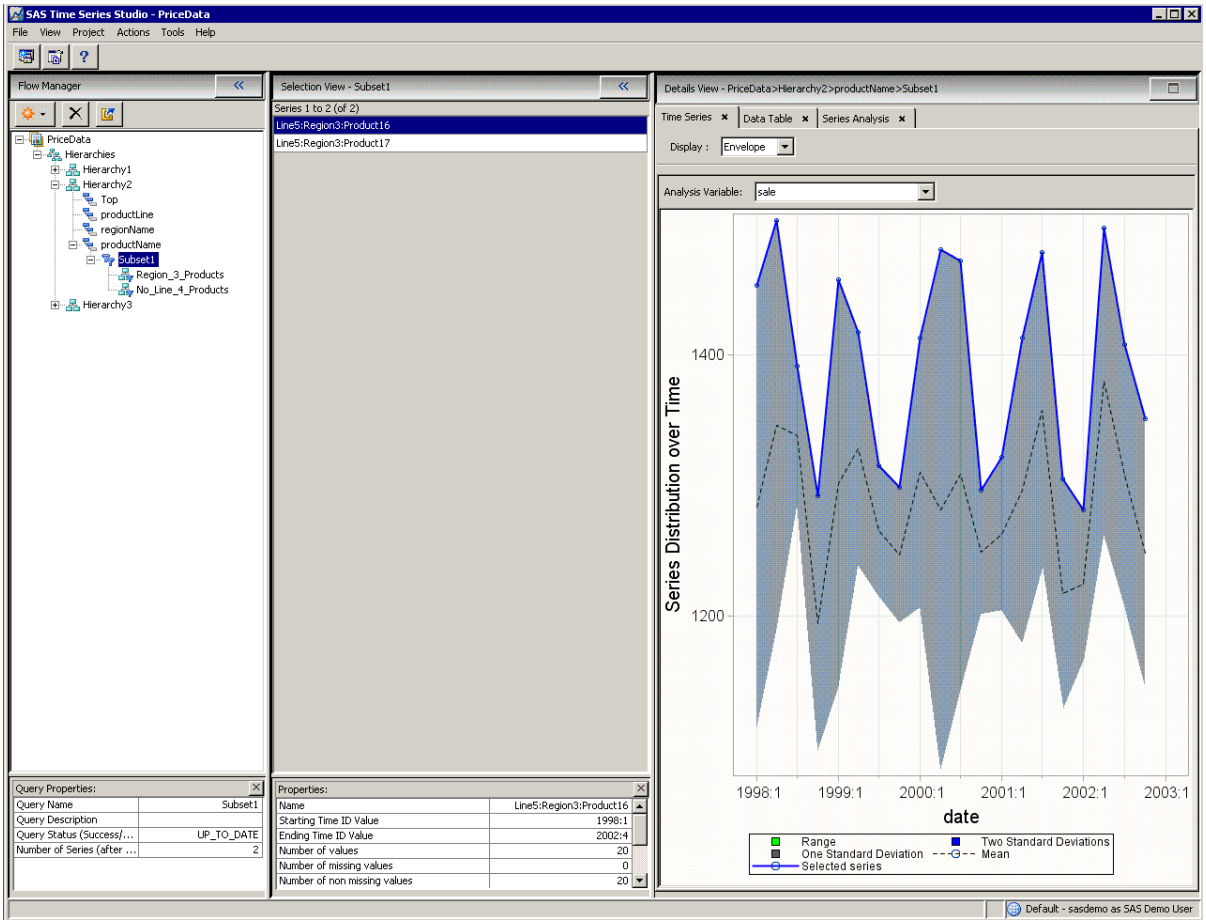
- Select **Exclude the selected series from the results (NOT AND)**.

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 exclude any series for Line4 products from the results, Subset1 now contains two series. These are the only two series that met the criteria in the Region_3_Products and No_Line_4_Products queries.


SAS Time Series Studio - PriceData
File View Project Actions Tools Help
Flow Manager
PriceData
Hierarchies
Hierarchy1
Hierarchy2
Top
productLine
regionName
productName
Subset1
Region_3_Products
No_Line_4_Products
Hierarchy3
Selection View - Subset1
Series 1 to 2 (of 2)
Line5:Region3:Product16
Line5:Region3:Product17
Details View - PriceData>Hierarchy2>productName>Subset1
Time Series
Data Table
Series Analysis
Display : Envelope
Analysis Variable: sale
Series Distribution over Time
1400
1200
1998:1
1999:1
2000:1
2001:1
2002:1
2003:1
date
Range
Two Standard Deviations
One Standard Deviation
Mean
Selected series
Query Properties:
Query Name Subset1
Query Description
Query Status (Success/... UP_TO_DATE
Number of Series (after ... 2
Properties:
Name Line5:Region3:Product16
Starting Time ID Value 1998:1
Ending Time ID Value 2002:4
Number of values 20
Number of missing values 0
Number of non missing values 20
Default - sasdemo as SAS Demo User

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

You can combine these different types of queries to find the data of interest.

Specifying the Results from a Query

When defining a query, you must specify how to include the series that meet the query criteria in your results. The options that are available depend on whether you are creating a segment or a subset.

Limit results to the selected series (AND)
specifies that the results of the segment or subset should contain only those series that meet the query criteria. For example, if you create a graphical query where you specify that sales must be more than 1500 and you select this option, the result contains only those series that sales more than 1500.

Include the selected series with the results (OR)
includes the series that meet the query criteria with the results. For example, if you create a graphical query where you specify that sales must be more than 1500 and you select this option, the result contains the selected data and the series that meet the query criteria.

Exclude the selected series from the results (NOT AND)
specifies that the series that meet the query criteria should not be in the results. For example, if you create a graphical query where you specify that sales must be more than 1500 and you select this option, the result contains the series where sales are less than 1500.

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 limit the results to the selected series, include the selected series in the results, or exclude the selected series from the results.

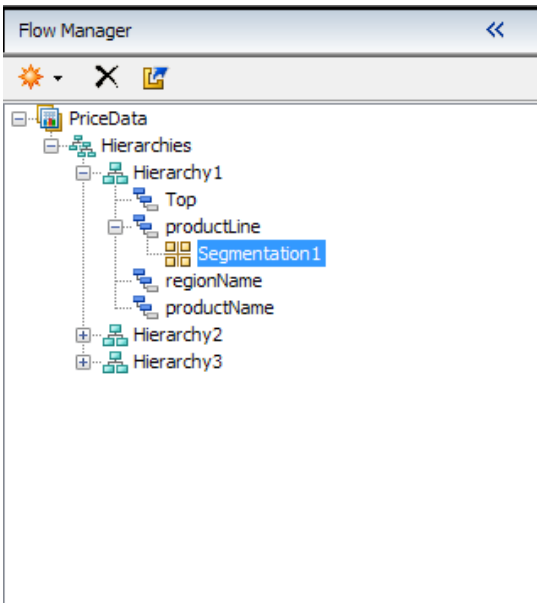
Note: The options that are available depend on whether you are creating a segment or a subset.

6 Click **OK**.

Example: Segmenting productLine by Using a Graphical Query

To create a segment for productLine in Hierarchy1:

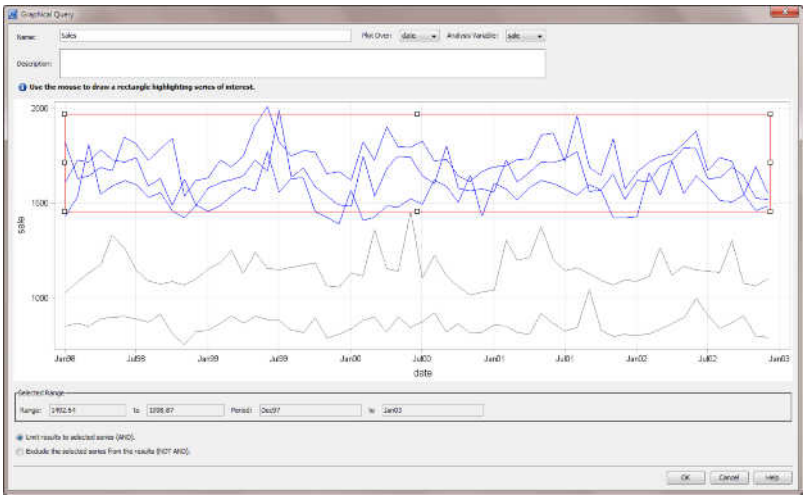
1 Select **productLine** in Hierarchy1 and select **Actions** ▶ **Add Segmentation**. Segmentation1 is added to the Flow Manager. For now, this segment contains the same data as productLine1.



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.

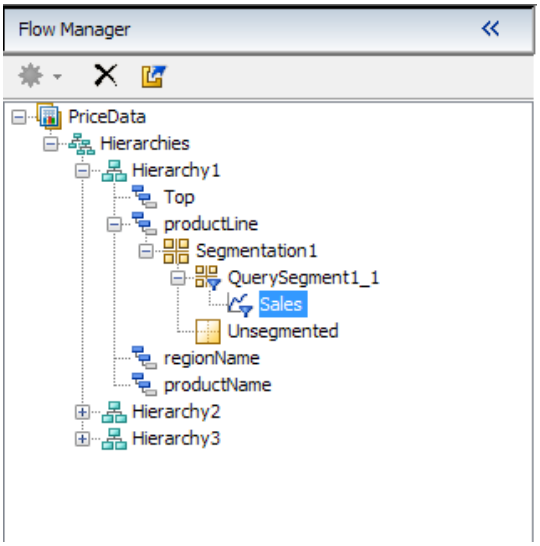
4 In the graph, select the series of interest. In this example, you are interested in the series that have sales more than 1500.



5 Select **Limit results to selected series (AND)**.

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 limit the results to the selected series, include the selected series in the results, or exclude the selected series from the results.

   Note: The options that are available depend on whether you are creating a segment or a subset.
5. Click **OK**.

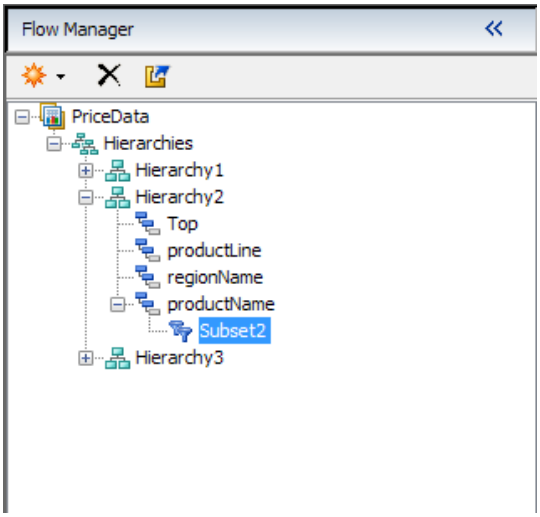
Example: Subsetting Your Data Using a Hierarchical Query

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 Hierarchy2 and select **Actions** ▸ **Add Subset**.

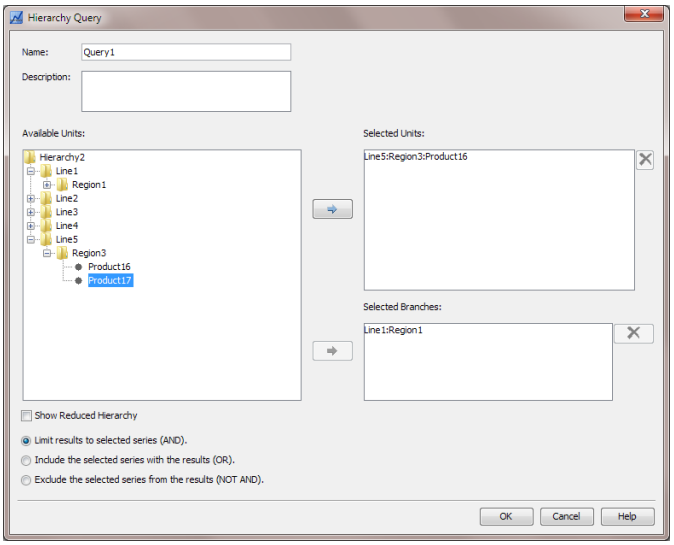
Subset2 now appears in the Flow Manager. In the Selection Manager and Details View, the results of Subset1 are the same as the results for productName.



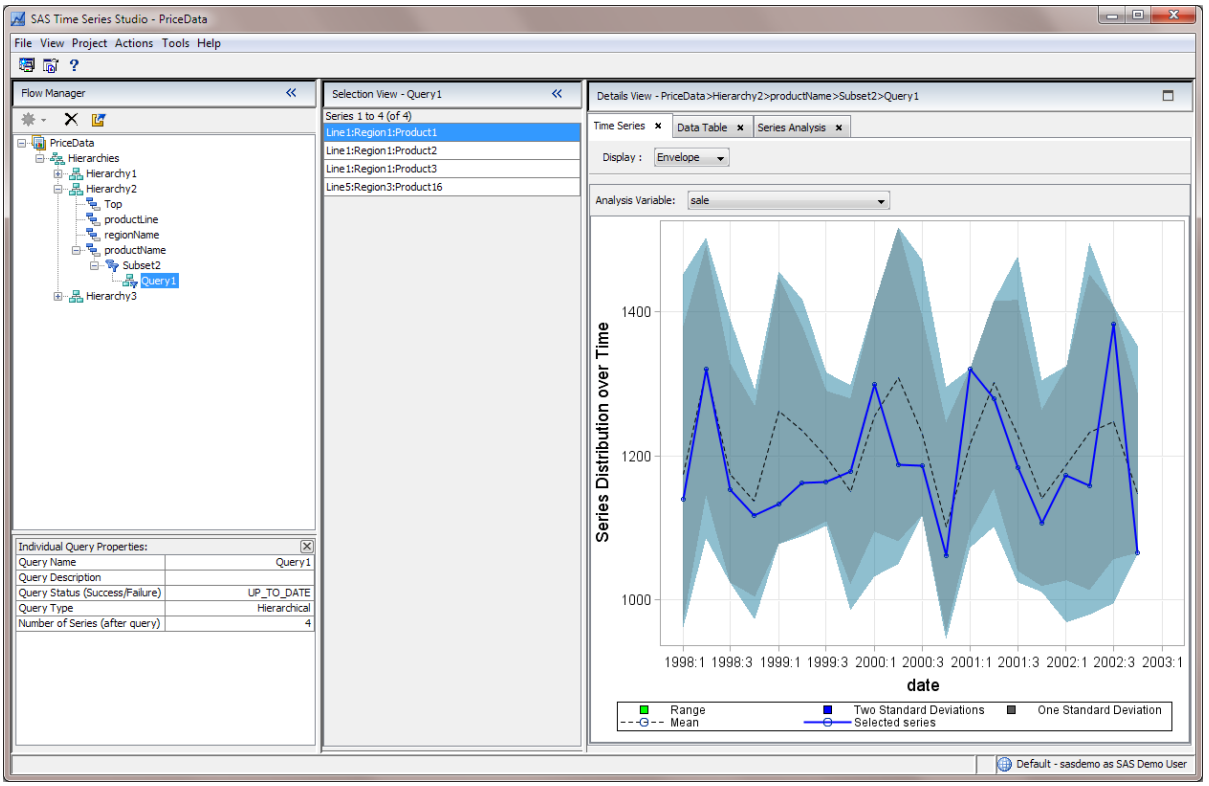
2 Right-click **Subset2** and select **Add Hierarchical Query**. The Hierarchy Query dialog box appears.

3 From the list of available branches, select **Line1:Region1**. From the list of available units, select **Line5:Region3:Product16**. Click  to move them to the list of selected units and branches.

4 Select **Limit results to the selected series (AND)**. 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 segment or subset where you want to add the query and select **Actions** ▶ **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 limit the results to the selected series, include the selected series in the results, or exclude the selected series from the results.

   Note: The options that are available depend on whether you are creating a segment or a subset.
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**.

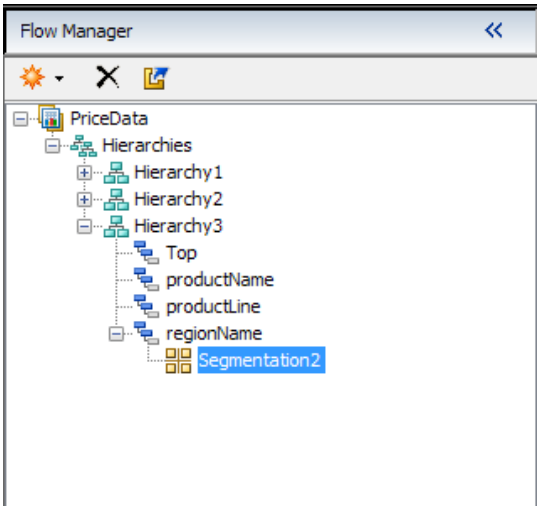
Example: Segmenting the Data by Using a Parameter Query

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.

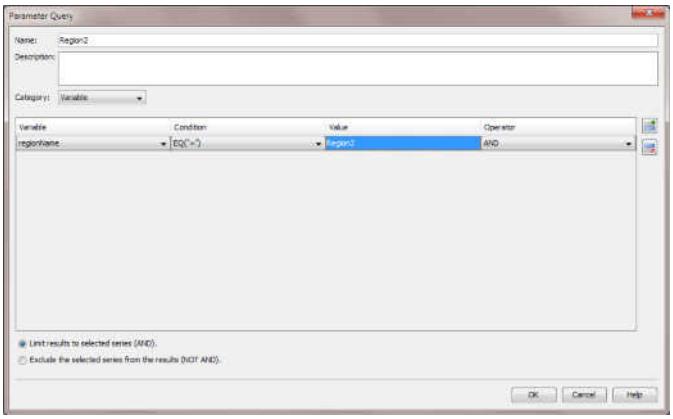
4 Select **Limit results to selected series (AND)** to include only the selected series in the results.

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

6 Click  to add a row to the table.

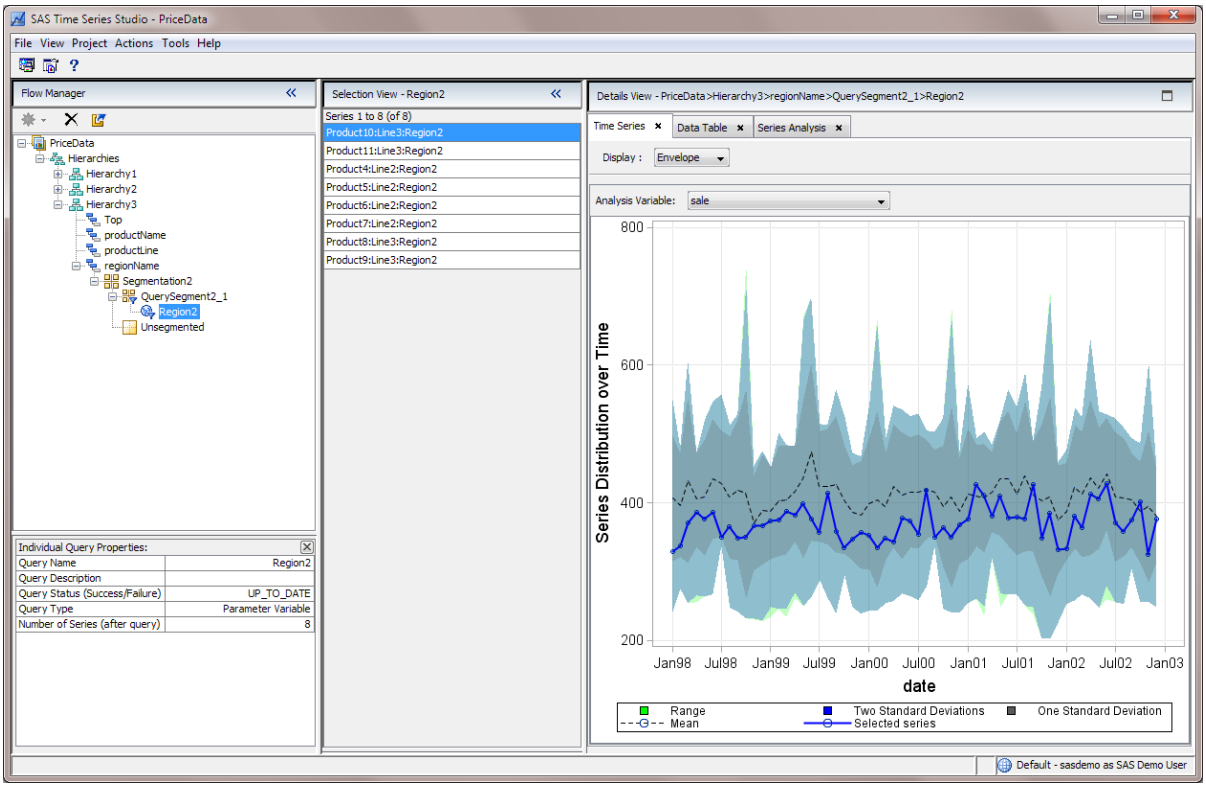
7 In the table, select the following options:

- 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

Exporting Data

11

Exporting Data from SAS Time Series Studio

About the Export Functionality in SAS Time Series Studio

When you use the export functionality, SAS Time Series Studio exports the data and the metadata about the SAS Time Series Studio project. When you import the data into SAS Forecast Studio, the New Project wizard recognizes the variable assignments and hierarchy that you created for the data in SAS Time Series Studio. You use these settings to create a SAS Forecast Studio project, or you can change the settings using the options in the New Project wizard.

Here is the metadata that is included in the exported data set:

Category	Metadata Information
Project Data	■ the name of the project ■ the version of SAS Time Series Studio ■ a description of the project ■ the user ID of the person who created the project ■ the user ID of the person who owns the project ■ the date and time that the project was created ■ the directory path where the project is saved ■ the name of the input data source in the format Libref:Data_Set_Name ■ the name of the environment that contains the project ■ the name of the time ID variable
Hierarchy Information	■ the name of the hierarchy ■ the time interval that is used in the hierarchy ■ the format of the time interval ■ the seasonality ■ the shift for the time interval ■ the multiplier for the time interval
Hierarchy Level Information	BY variables (listed in the order in which they appear in the hierarchy)
Variable Information	■ the role that you assigned to the variable ■ the accumulation method for that variable ■ the aggregation method for the variable ■ the value of the **Missing Interpretation** option, which specifies how to replace missing values in the data ■ the value of the **Missing Trim** option, which specifies how missing values are removed from the accumulated time series ■ the value of the **Zero Miss Interpretation** option, which specifies how beginning and ending zero values are interpreted in the accumulated time series

The output from the export functionality includes the exported data set in the library that you selected. It also includes these two files in the project directory. (For more information about the project directory, see “What Is a Project?” on page 54.)

- a version of the exported data set (including the project metadata) created by the CPORT procedure. This procedure writes SAS data sets to sequential file formats (transport files).
- a code file generated from the settings in your SAS Time Series Studio project. The code includes the %FSCREATE macro, which can be used to create a SAS Forecast Server project. When you run the code, you need to provide the user name, password, and SAS environment. You also need to verify that the library that is associated with the given data set is assigned in the SAS Forecast Studio session. The code is not run automatically, so you can make additional changes to the code, such as specifying multiple dependent variables. For more information about the macros, see *SAS Forecast Server: Administrator's Guide*.

  Here is an example of the %FSCREATE code that is generated by the export functionality in SAS Time Series Studio:

```
/*------------------------------------------------------------
* The following code has been compiled using settings from
* a TSS project. Inputs may need to be updated to ensure the
* creation of a valid forecasting project.
*------------------------------------------------------------*/

/*------------------------------------------------------------*/
/*You MUST provide username, password and environment for login*/
/*------------------------------------------------------------*/
%fslogin(user=<yourusernamehere>,
      password=<yourpasswordhere>,
      sasEnvironment=<yourenvironmenthere>,
      desktop=NO);

%fscreate(projectname=QuerySegment1_1,
/*-------------------------------------------------------------------*/
/*You MUST ensure that library used below is assigned in this session*/
/*-------------------------------------------------------------------*/
     data=TSSLIB.QuerySegment1_1,
      by=regionName productLine,
      hierarchy=YES,
      id=date,
      interval=MONTH,
      idformat=MONYY.,
```

```
            seasonality=12,
            timeshift=1,
            timemultiplier=1,
            accumulate=TOTAL(sale) AVERAGE(line price),
            aggregate=TOTAL(sale) AVERAGE(line price),
            input=price,
            reporting=line,
            var=sale);

      %fslogout();
```

Export Data

To export data from SAS Time Series Studio:

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 a name and description for the exported data.
4. (Optional) Select the level of data to be exported.
5. 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.
6. Click **Export**.

Example: Exporting Data to Create a Project in SAS Forecast Studio

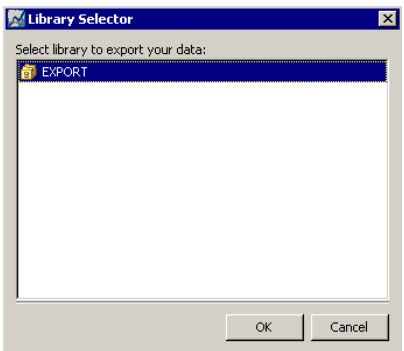
About This Example

In this example, you will export the data for Hierarchy1 from the PriceData project. For more information about creating this hierarchy, see “Example 1: Create the productLine > regionName > productName Hierarchy” on page 74.

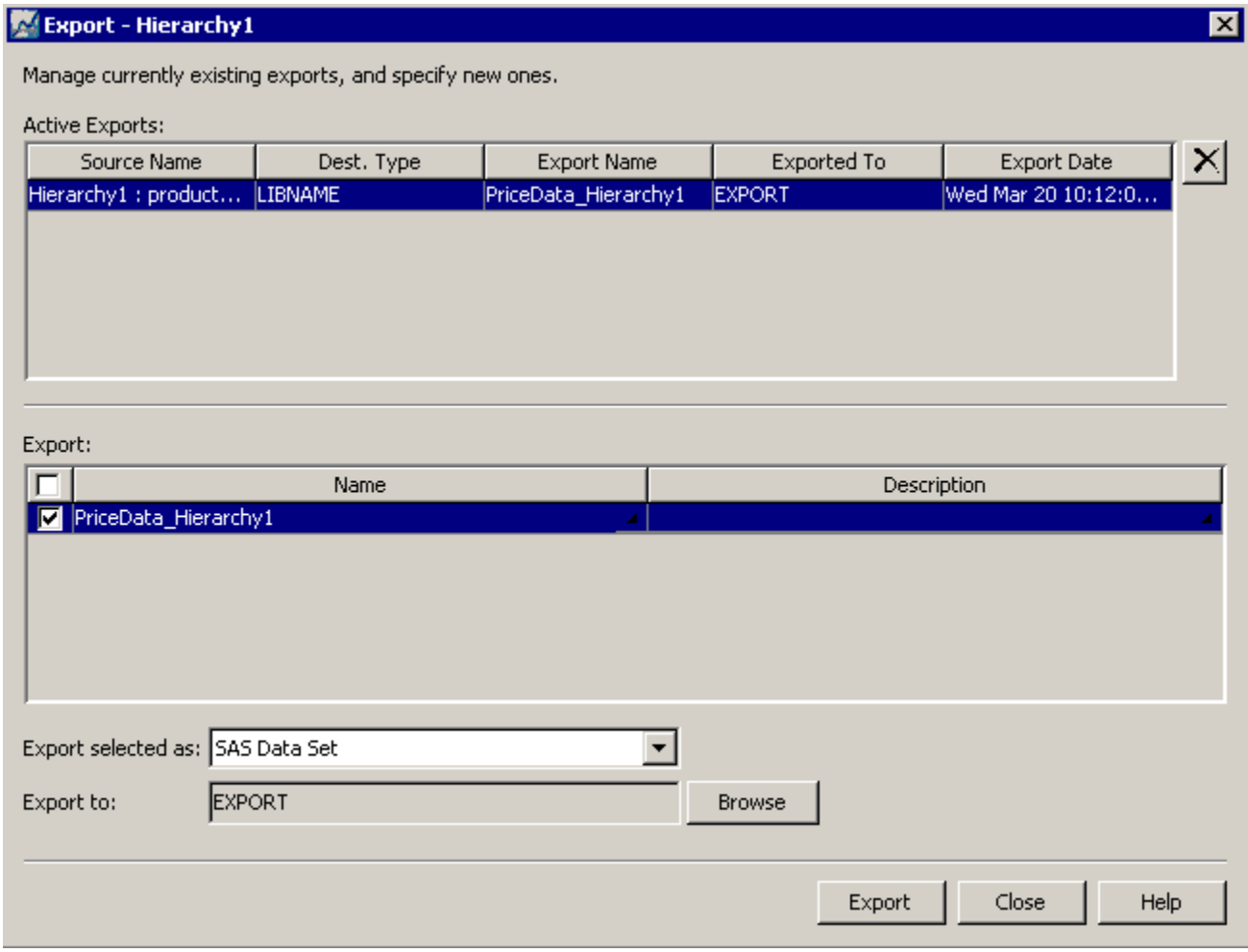
Export the Data from SAS Time Series Studio

To export the data from SAS Time Series Studio:

1. In the Flow Manager, select the **Hierarchy1** node, and select **Actions** ▶ **Export**. The Export - Hierarchy1 dialog box appears.
2. In the **Export** table, select the check box for **Hierarchy1**.
3. Double-click the Hierarchy1 value in the **Name** column. Enter `PriceData_Hierarchy1` as the new name.
4. Specify that the data should be exported as a SAS data set.
5. Click **Browse**. The Library Selector dialog box appears. Select the **EXPORT** library, and click **OK**.



6 Click **Export** in the Export - Hierarchy1 dialog box. An entry appears in the **Active Exports** table when this task is complete.



7 Click **Close** and exit SAS Time Series Studio.

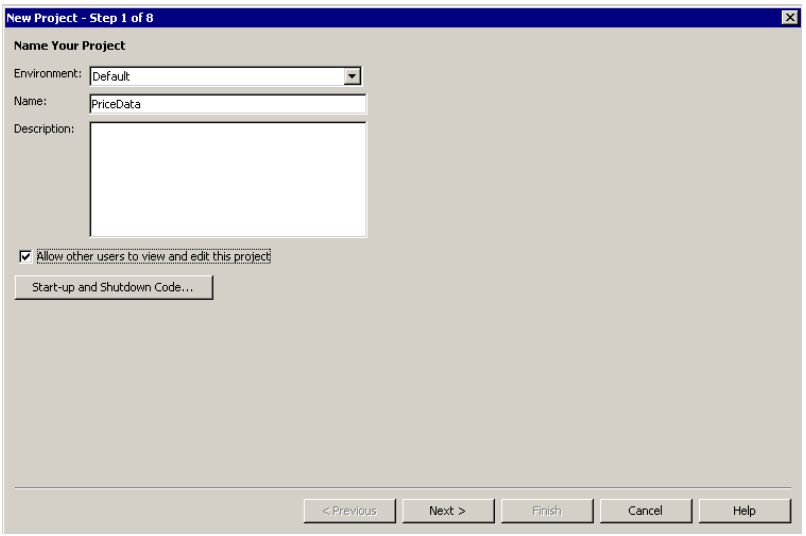
Importing the Data into SAS Forecast Studio

To create a project in SAS Forecast Studio using the exported data:

1 Open SAS Forecast Studio.

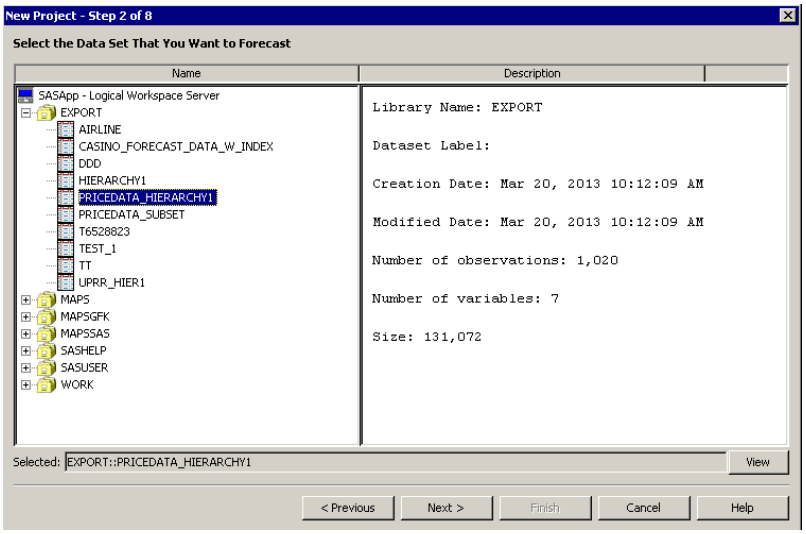
2 In the Projects dialog box, click **New**. The New Project wizard appears.

3 In the first step, enter `PriceData` as the name of your project, and select the **Allow other users to view and edit this project** check box.



Click **Next**.

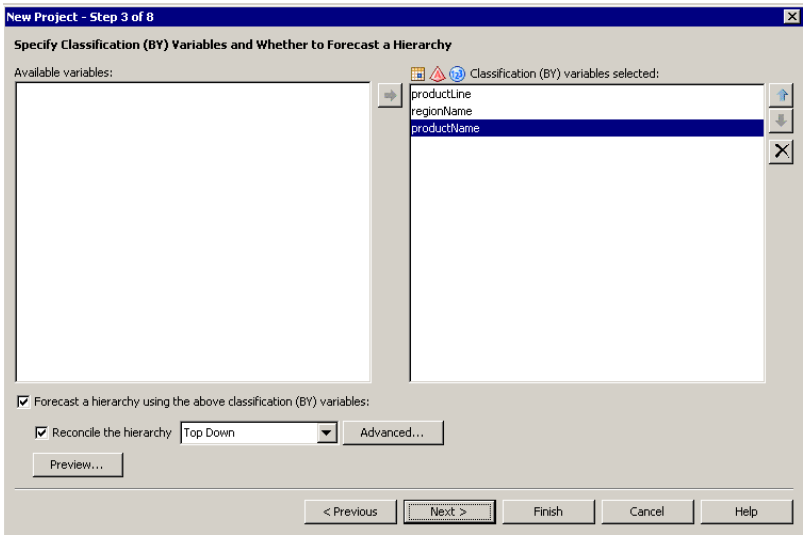
4 In Step 2, select the PriceData_Hierarchy1 data set in the Export library.



Click **Next**.

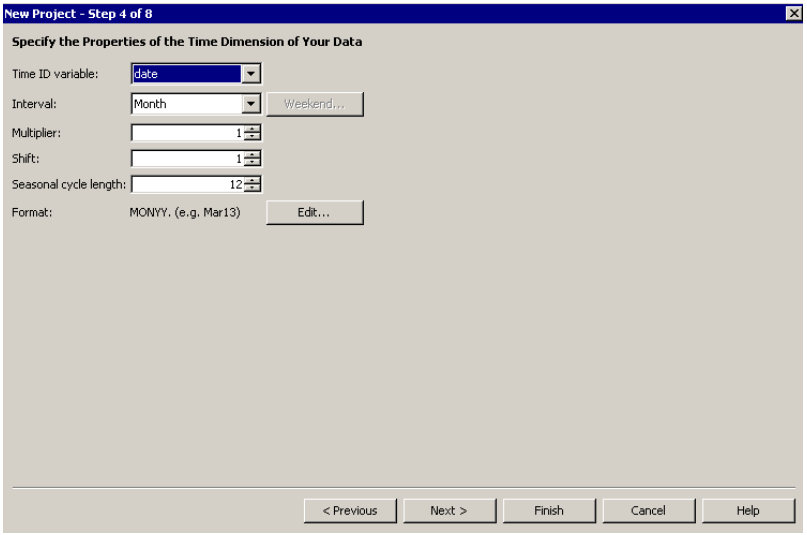
5 In step 3, click **Finish**.

When creating the project, SAS Forecast Studio uses the variable assignments and hierarchy that you specified in SAS Time Series Studio. For example, step 3 in the New Project wizard lists three classification (BY) variables (productLine, regionName, and productName). And the **Forecast a hierarchy using the above classification (BY) variables** check box is selected. The order of the classification variables in the **Classification (BY) variables selected** pane specifies the order of the hierarchy.

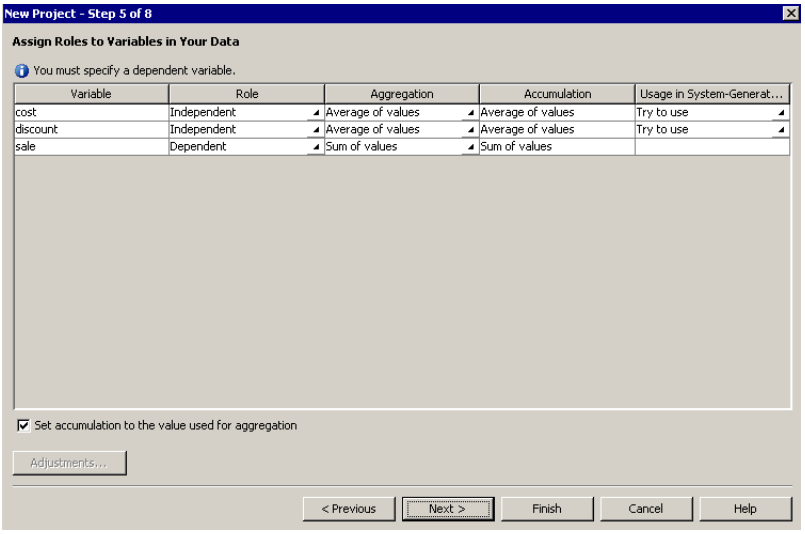


As a result, SAS Forecast Studio creates a project with a hierarchy of productLine > regionName > productName.

If you click **Next** (instead of **Finish)**, you see that SAS Forecast Studio is using **date** as the time ID variable.



In step 5 of the New Project wizard, cost and discount are listed as independent variables, and sale is listed as the dependent variable.



All of this data came from the data set that you exported from SAS Time Series Studio. You can use the options in the New Project wizard to change any of these options before creating your project in SAS Forecast Studio.

Part 5

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), select the node in the hierarchy, and select **Actions** ▶ **View Code**. You can save this code to a file, which can then be shared with SAS Technical Support.

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.

Troubleshooting the Time ID Variable

By default, SAS Time Series Studio reviews the first 10,000 observations in the time series data. If SAS Time Series Studio identifies three unique values for the time ID variable, then SAS Time Series Studio can automatically set the time interval.

Here are some common errors:

- SAS Time Series Studio determines the time interval that best fits the data. If SAS Time Series Studio detects multiple time intervals, SAS Time Series Studio selects the most frequently used base interval. Although SAS Time Series Studio selects the default interval, you can change this interval using the **Interval** option in the New Project wizard.
- SAS Time Series Studio is unable to determine the time interval. When SAS Time Series Studio is unable to determine the time interval for the data, you must select another variable as the time ID variable or to manually set the time interval.

  Here are the reasons why SAS Time Series Studio cannot determine the time interval:

  - The values of the time ID variable are not evenly spaced values at a specific time interval. For example, transactional data is recorded at no particular time interval.

    In this case, you must specify the time interval to use. Then SAS Time Series Studio can aggregate the data using that time interval.

    Note: If the specified time interval is too small, then a large number of observations could have missing values. For example, you select **WEEK** as the interval, but there is no data for many of the weeks. A larger interval, such as MONTH, should be selected.

  - The values of the time ID variable are evenly spaced values at a specific time interval, but SAS Time Series Studio cannot detect the time interval. For example, SAS Time Series STudio cannot detect the time interval if the data represents a multi-weekday interval, such as the WEEKDAY3. interval.

    You must manually specify the time interval to use.

- The wrong variable has been selected as the time ID variable.

  In the New Project wizard, you must select a different variable to use as the time ID.

- The values of the time ID variable are in an unsupported format, such as data that contains SAS time values. SAS Time Series Studio supports only SAS date and datetime values.

  If your data is hourly or uses a smaller time interval, then the values in the time ID variable must use the datetime format. (Datetime format is the number of seconds since January 1, 1960.) If the data uses a SAS time format, then SAS Time Series Studio does not allow you to select that variable as the time ID. If the time ID variable has time values (the number of seconds since midnight) but no format, then SAS Time Series Studio assumes that the time ID values are dates (the number of days since January 1, 1960).

Index

I

M

N

O

P

Q

R

S

T

V