# SAS® Visual Forecasting 8.3: User's Guide

SAS® Documentation
July 10, 2018

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2018. *SAS® Visual Forecasting 8.3: User's Guide*. Cary, NC: SAS Institute Inc.

**SAS® Visual Forecasting 8.3: User's Guide**



SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

July 2018



8.3-P1:vfug

# Contents

# 1

# Setting Up Your Project



## Preparing an Input Data Set

### Overview

In SAS Visual Forecasting, you can create a project using one of two types of data sets:

Time series
: This is the source time-stamped data that you need to form time series vectors and to determine the best modeling strategies to use and to create forecasts. For more information about this data source, see "Preparing Time Series Data" on page 2.

External forecast
: This is an output data set that already contains forecasts along with the historical data. Typically, this data has been exported from another forecasting project as an OUTFOR data set. External forecasts are used primarily to work with any overrides that need to be applied to the forecasts. For more information, see "Working with External Forecast Projects" on page 18.

The data can be imported into a new SAS Visual Forecasting project by loading it into Cloud Analytic Services (CAS). See *SAS Cloud Analytic Services: User's Guide* for more information about using CAS.

Sample data sets are provided, which you can use to experiment with SAS Visual Forecasting. You can access this data at the following location:

- http://support.sas.com/kb/61/960.html

## Preparing Time Series Data

This topic describes the general requirements for getting source data into SAS Visual Forecasting. For further information about time series data, see "Understanding Time Series Data" on page 55.

### Requirements for Time Series Data

SAS Visual Forecasting projects require input data that is in time series format. Each observation or row in the table should be using a SAS date or datetime format.

Some time series data is equally spaced. That is, successive observations are a fixed time interval apart, and the data can be described by a single interval, such as daily, every three weeks, or twice a month. You might already have this time series data, or you might have time-stamped data at irregular intervals. You can use the accumulation options in SAS Visual Forecasting to convert the time-stamped data into a time series.

Your time series data must meet the following requirements:

- The data set contains the dependent variable that needs to be modeled and forecast. If multiple variables need to be forecast, you can set that up in separate projects. Multiple projects can be run using the same source data set.
- The data set contains a variable that is in a SAS date or datetime format. When SAS Visual Forecasting creates a project, the data is sorted by the time variable so that the observations are in order according to time.
- The data is loaded into a table in Cloud Analytic Services (CAS). You can use SAS Data Explorer to import data into CAS, or to load a SAS data set into CAS. See *SAS Data Explorer: User's Guide* for more information.
- The data must have a time series that includes at least ten observations.

**Note:** The names of the variables cannot match any of the reserved variable names that are used in the output data set. For more information, see "Reserved Names for Variables and Models" on page 92.

- For general information about working with time series data, see Working with Time Series Data in the *SAS/ETS User's Guide*.
- For more information about creating time series data from transactional data, see "The TSMODEL Procedure" in the *SAS Visual Forecasting: Programming Guide*.

### Converting Dates from Character to SAS Date Format

Often a data source might have the date or date and time in character format. For example, consider a variable with the name **`Sale_Date`** with this construction:

*`ddd, yyyy mmm dt hh:mm`*

- *ddd* = 3-letter day of the week
- *yyyy* = 4-digit year
- *mmm* = 3-letter month
- *dt* = 1- or 2-digit day of the month
- *hh* = 1- or 2-digit hour in 24-hour notation (0–23)
- *mm* = 2-digit minute

An example of an observation in this format is "`Mon, 2017 May 8 8:08`".

To convert this character string into a SAS date, you would need to use some SAS functions in a SAS DATA step. The following functions extract the year, month, and date portions of the `Sale_Date` variable. The TRIM function concatenates them into a single string that can be read into a new variable, `Project_Date`, using the SAS `date9.` informat.

```
Year = substrn(Sale_Date,6,4);            /* extract the year starting at the sixth position */
Month = substrn(Sale_Date,11,3);          /* extract the month starting at the 11th position */
Date = substrn(Sale_Date,15,2);           /* extract the date  starting at the 15th position */
Datechar = trim(Date)||trim(Month)||trim(Year);
                                          /* trim trailing spaces from each new variable,
                                             concatenate the three variables into ddmmmyyyy format  */
Project_Date = input( Datechar , date9.); /* read in date string as SAS date9. informat */
Drop Year Month Date Datechar;            /* drop temporary variables from the final data set */
Format Project_Date date9.;               /* set date format for time variable               */
```

Using the code in this example, the string `Mon, 2017 May 8 8:08` is converted to the SAS numeric date of 20947. Use the FORMAT statement to format the variable so that SAS Visual Forecasting will properly detect the data when you create a project.

SAS administrators can also use SAS Data Studio to convert string data to a SAS date or datetime format.

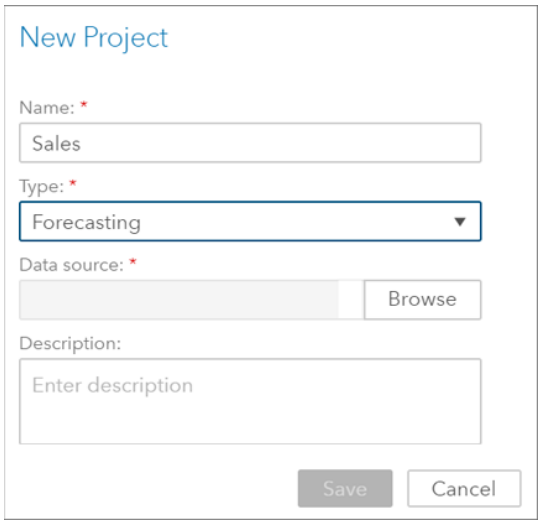
# Creating Your First Project

To open a forecasting project, click on the project in the list on the Model Studio Project page. Follow the steps below to set up a new project.

## Creating a Forecasting Project

1. From the Model Studio Project page, click **New Project**. The New Project dialog box is displayed.

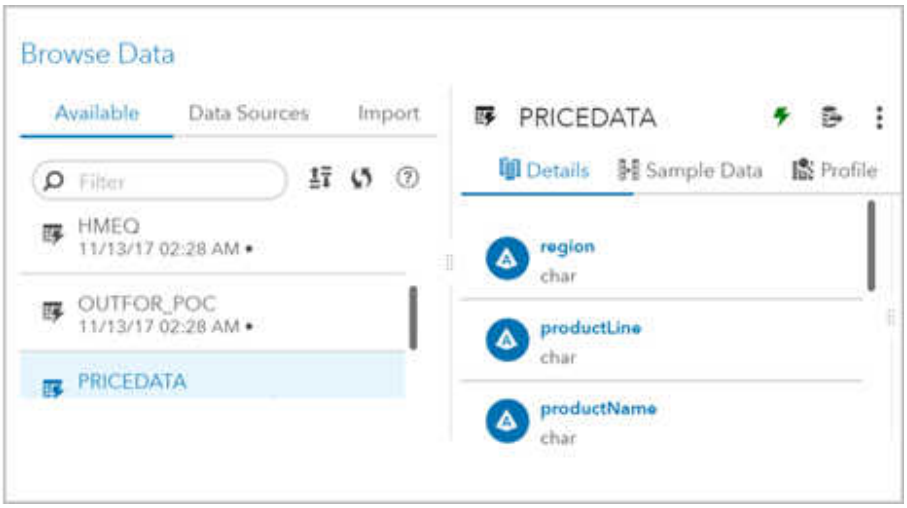
4




2 For **Name**, give the project a unique name.

3 For **Type**, select `Forecasting`.

4 For **Data Source**, click **Browse**. The Browse Data dialog box is displayed.




A list of data sets is displayed in the left-side Available tab. These are data sets that are available in CAS and ready for use in a Model Studio project.

5 If your source data set is not listed, select the Data Sources tab and search through the libraries for your data set. When you select a data set, the variables are displayed to the right.

**Note:** If you have any data sets in SAS7BDAT or SASHDAT format, you can right-click the data set and select **Load**. The data is loaded into CAS memory and is ready to be used in a project.

**Note:** If you have a local file with a valid time variable, select the Import tab and click **Local File** to upload it.

For more information about how to use this dialog box to access data, see "Getting Started with the Choose Data Window" in SAS Data Explorer: User's Guide.

6 Click **OK** to load the selected data set and then click **Save** to create the new project.

The new project that you created is opened with the Data tab selected. Your next steps are to assign the time and dependent variables for your project.

**Note:** If SAS Visual Forecasting cannot successfully create the project, the project is still added to the Model Studio project list. In this case, you should manually delete the project from the Projects view. Typically, project creation is not successful when a valid time variable cannot be detected in the data source.

## Assigning Variable Roles

When you create a project with time series data, the project is opened to the Data tab. In the left panel, the time series data set is selected.

### Overview

Use the middle panel to select variables and the right panel to assign roles to a selected variable.

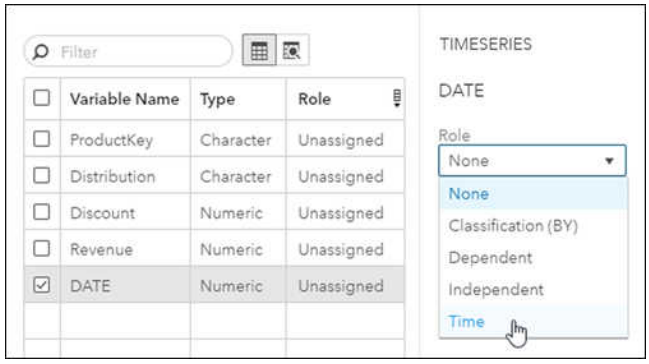
- The time and dependent variables are required and should be assigned before performing any other activities in the project.
- The BY variables are optional but are strongly recommended. These variables are used to set the project up so that you can work with different cross sections of your data based on specific attribute values that you specify.
- You might want to assign some of your variables as independent. Independent variables can be used to model and forecast the dependent variable.
- After making your variable assignments, review the project settings on page 68.

**Note:** After you run one or more pipelines and open the Overrides tab, the variable assignments and settings are committed and cannot be subsequently updated for the project. If you need to change any of these settings after they are committed, you can create a new project using the same data source. For more information, see "Locking the Data Definition" on page 41.

### Assigning the Time Variable

For new projects, if the system detects only a single time variable candidate, then that variable is already assigned. If the system detects multiple time variable candidates, you must choose one of them as the time variable. The variable must be using either a date or datetime format.

1. Select the variable in the middle panel and, in the right panel, select **Time** for **Role**.




   The settings for the time variable are displayed on the right.

2. Review the time variable settings and make sure they are correct for this project. For a complete description of each field, see "Time Variable" on page 62.

If you need to reassign the time role to another variable, you must first remove the assignment for the current time variable.

## Assigning the Dependent Variable

Dependent variables are the variables that you want to model and forecast. You must assign one numeric variable to this role. For example, you want to forecast the sales for each product, you assign the Sales variable to the dependent variable role.

1. Select the variable in the middle panel that you want to use to obtain forecasts and, in the right panel, select **Dependent** for **Role**.
2. Review the dependent variable settings and make sure they are correct for this project. For a complete description of each field, see "Dependent Variable" on page 64.

If you need to reassign the dependent role to another variable, you must first remove the assignment for the current dependent variable.

## Assigning the Default Attributes

BY variables are used to create attributes to uniquely identify each time series. For example, your data source might have variables for Region, Product line, and Product. By assigning these as BY variables, SAS Visual Forecasting collects the dependent variable into individual time series using each unique combination specified by these attributes.

See "BY Variables" on page 66 for more information about assigning default attributes.

You can also order your BY variables to create a forecast hierarchy. For more information about hierarchical forecast projects, see "Defining the Hierarchy" on page 7.

When you are certain about which variables you want to use as default attributes, follow these steps.

1. Select one or more variables in the middle panel that you want to use as an attribute.
2. In the right panel, select **BY Variable** for **Role**.
3. In the right panel, examine order of the BY variables under **Selected hierarchy**. If you use hierarchical forecasting in the project, use ⬇ and ⬆ to arrange selected variables in the best order for the hierarchy. For more information about hierarchical forecasts, see "Defining the Hierarchy" on page 7.
4. Examine the **Reconciliation level** setting. Select the default reconciliation method for hierarchical modeling nodes to use. This setting can be overridden by settings in the modeling node, if that option is provided.

   To use top-down reconciliation, select `Top`. Or you can select a specific BY variable to generate the forecasts and then let the forecasting model reconcile the forecasts for the other BY variables.

   If you make any change to the structure of the hierarchy, either adding or removing BY variables, or changing the order, the **Reconciliation level** is reset to `Top`.

   For more information about reconciliation methods, see "Understanding Hierarchy Reconciliation" on page 59.

When you have completed the default attributes assignments, select **Default attributes** in the left panel. The default attributes are listed in the middle panel. See "Working with Attributes" on page 10 for further instructions.

## Assigning Independent Variables

After you have assigned your time and dependent variables, you might want to assign some independent variables that should be considered for model generation. Independent variables are the explanatory, input, predictor, or causal variables that can be used to model and forecast the dependent variable. You can assign only numeric variables to this role. When creating the system-generated models, SAS Visual Forecasting tries to use the independent variables in the model generation.

Follow these steps to assign independent variables.

1 Select the variable in the middle panel that you want to use to obtain forecasts and, in the right panel, select **Independent** for **Role**.

2 Complete the following fields:

**Hierarchy aggregation**
Select from the following values:

**Standard deviation of values**
Aggregation uses the standard deviation of values for all of the time series in each hierarchy level.

**Minimum of values**
Aggregation uses the minimum of the values for all of the time series in each hierarchy level.

**Maximum of values**
Aggregation uses the maximum of the values for all of the time series in each hierarchy level.

**Average of values**
Aggregation uses the average of the values for all of the time series in each hierarchy level. For example, select this option if your data set contains the price of each product, and you want to know the average price for a product line.

**Sum of values**
Aggregation uses the sum of the values for all of the time series in each hierarchy level.

**Number of non-missing values**
Aggregation uses the number of nonmissing values for all of the time series in each hierarchy level.

**Number of missing values**
Aggregation uses the number of missing values for all of the time series in each hierarchy level.

**Uncorrected sum of squares**
Aggregation uses the uncorrected sum of squares for all of the time series in each hierarchy level.

**Corrected sum of squares**
Aggregation uses the corrected sum of squares for all of the time series in each hierarchy level.

**Time interval accumulation**
Accumulation combines data within the same time interval into a summary value for that time interval. For a complete description of the accumulation options, see "Time Interval Accumulation Settings" on page 66.

**Missing value interpretation**
Once the data has been accumulated to form a time series, missing value interpretation is performed. If any time series contain missing values for variables other than the time variable (such as the dependent or independent variables), you can specify how SAS Visual Forecasting should to interpret these missing values. For a complete description of these options, see "Missing interpretation" on page 65.

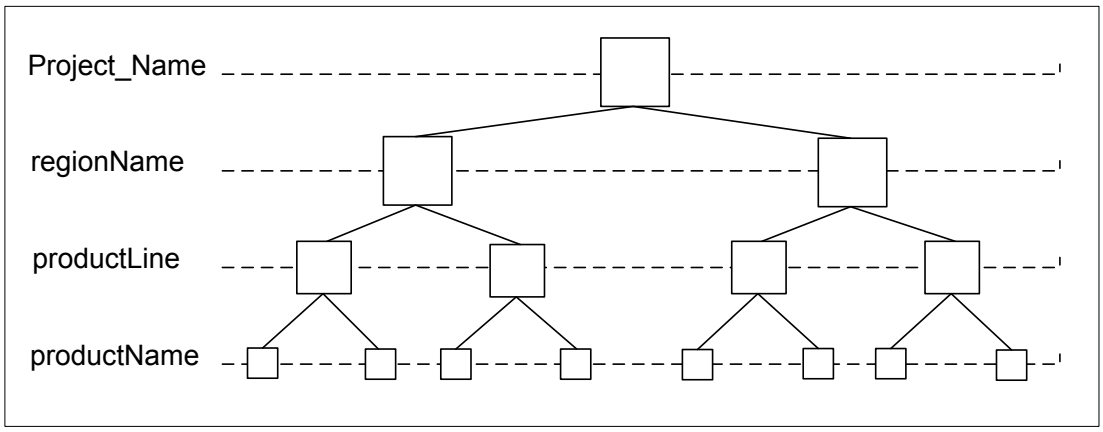
**Usage in system-generated models**
Check this option if SAS Visual Forecasting should use this variable to generate forecasts.

## Defining the Hierarchy

To forecast hierarchically, you must assign BY variables and use one of the hierarchical models in the pipeline for the project. The ordering of the BY variables constructs the forecasting hierarchy.

An example of a hierarchy is Region > Product Line > Product Name.

*Figure 1.1* *Example of Hierarchy*




After you set up a project hierarchy, you can specify how SAS Visual Forecasting reconciles forecasts between levels in the hierarchy. Make this specification in the hierarchical forecast model in the pipeline. See "Understanding Hierarchy Reconciliation" on page 59 for more information.

## Checking the Project Settings

Before working with any pipelines, verify that the global settings for the project are correct.

1 From the Data tab, click ⚙.

   The Project Settings dialog box is displayed.

Changes to these settings can affect your pipelines and overrides. Make sure these settings are correct so you do not have to rerun pipelines or resubmit overrides. For more information about these fields, see "Project Settings" on page 68.

# Using the Data Tab

The Data tab is used to work with the data source for your project. After you create your project, the project is opened to the Data tab so that you can assign roles to the variables that are key to the project. The Data tab displays the data that was loaded into the project CAS library from the original data source.

The Data tab consists of three panels.

## Left Data Panel

The left panel enables you to select different data sources to work with. The toolbar along the top of the panel provides these actions:

Add or change the data source. If you select a time series or external forecast, this action replaces the current data set. For replacement, the data source must match the schema of the original data set that is to be replaced.

For attributes, this action supplements the default attributes table. You must assign BY variables before you can add an attributes table. For more information, see "Working with Attributes" on page 10.

For events, this action supplements the predefined events. For more information, see “Importing Custom Events” on page 13.

After any data is changed, any pipelines that have been run become `Out-of-date` and need to be run again.

Check the data source for updates. You can also perform this action by right-clicking the data source and selecting **Refresh all data**.

Open the selected data source in SAS Visual Analytics for further investigation. This action is not available for predefined events.

You can also perform this action by right-clicking the data source and selecting **Explore and visualize**.

You can work with these data sources in the left panel:

**Time Series**
Time series are the most common input data source for forecasting projects. Each record has a date or date-time stamp that is used to assess the historical record of the data and to create forecasts. For more information, see “Understanding Time Series Data” on page 55.

**External Forecasts**
Sometimes you need to work only with the output data from another forecasting project. For example, you can generate forecasts for a time series data set and then export the data. Using the output data, you can create multiple projects, experiment with different pipelines and modeling strategies, and work with overrides in the forecast horizon.

For more information, see “Working with External Forecast Projects” on page 18.

**Attributes**
BY variables are the initial, default attributes for the project. Attributes provide a way for you to visualize and work with subsets of project data based on specific attribute values that you define. After assigning the default attributes, they are listed when you select the **Default attributes** table under **Attributes**.

If you add an external attributes table, all of the attributes, including the default attributes, are listed using the table name of the external attributes. To add more attributes, see “Working with Attributes” on page 10.

**Events**
An event is an incident that disrupts the normal flow of a process that generates the time series. Examples of events are holidays, retail promotions, and natural disasters. You can create events from a list of predefined events and you can import events from an external data source with custom event definitions. Predefined events are available by selecting **Predefined events** under **Events**. For more information, see “Adding Events to Your Project” on page 12.

## Middle Data Panel

The middle panel shows the content of the selected data set. Click **Filter** to enter characters to focus on a subset of the content.

The icons next to **Filter** enable you to toggle the view. These icons are not available for external forecast projects.

Show the variables in the data source, their role assignments, and other properties. This is the default view for time series projects.

You can modify the columns displayed by the table by clicking .

Show the data source, with the variables as column headings and the observations in each row. This is the only view for external forecast projects.

### Right Data Panel

The right panel shows the properties of the current selection. Use this panel to assign properties to variables in time series, external attributes, or events. After you assign the time and dependent variables, these roles cannot be assigned to another variable unless you first remove the original assignment.

- If a data set is selected in the left panel, the right panel shows the properties of that data set.
- If a variable is selected in the middle panel, the right panel shows the properties of that variable. If you select multiple variables, the right panel shows blank values if the variables do not share the same properties. Variables for external forecasts cannot be selected and their properties cannot be changed.

## Working with Attributes

Attributes provide a way for you to visualize and work with subsets of project data based on specific values of those attributes. BY variables are the initial, default attributes for the project. You can add additional attributes to a project from another data source.

### Adding Attributes to Your Project

You can add variables to a forecasting project from another data source. The additional variables are used as attributes for creating filters. For more information about filters, see "Working with Filters" on page 46.

Before adding attributes to the project, make sure the following tasks are complete.

- Assign BY variables for the project, as described in "Assigning Variable Roles" on page 5. These are the default attributes for the project.
- Make sure the data source with the additional attributes includes the same variable names as the existing BY variables in the original data source for the project. For example, if the original BY variables in the time series project are `Location` and `Category`, then the attribute table must include columns for `Location` and `Category`.
- Make sure the values for each combination of BY variables in the attribute table are unique. For example, the table can have only one row with this combination:
  - Location = Miami
  - Category = Shoes

Follow these steps to add attributes from another data source to your project.

1. From the Data tab, click and select **Attributes**.

   The Browse Data dialog box is displayed.

2. Locate the data source with the additional variables to use as attributes with this project. Select the data source and click **OK** to load it.

When the data source is successfully loaded, the table name of the variables is listed in the left panel, replacing **Default attributes**. The new variables are listed as attributes in the table along with the original default attributes.

## Selecting Attributes for Filters

Use discretion in the number of attributes to use as filters. For performance reasons, it is better not to have more than 20 attributes as filters in a project. When selecting which attributes to use as filters, it is helpful to see how attribute values can be specified as filters. Follow these steps for selecting the best attributes to use as filters by exploring a time series in a pipeline.

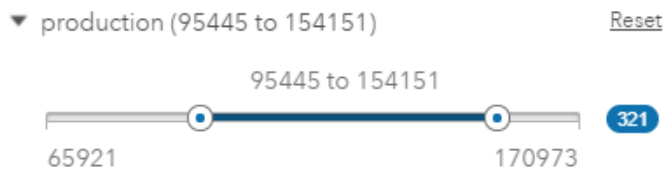
1 In the Data tab, select the table listed under **Attributes** in the left panel.

2 Select an attribute in the middle panel and examine these fields in the right panel.

**Use in faceted filter**
This setting enables you to create filters used to subset the data on the Overrides tab and create overrides for forecasted values based on the filter settings. Filters can also be used to display filtered views of the historical data in the project.

**Display as range**
Turn on this setting for any numeric filters that have a continuous value. Continuous variables can have a limitless number of values within a range. For example, if you are working with a *Weight* attribute that could be any decimal range between 20 and 200 pounds, you would turn this setting on.

With this setting turned on, you use a slider to specify the lowest and highest values of the range. The following image shows a *production* attribute with a range of 65,921 to 170,973. The slider is used to specify a filter with a range of 95,445 to 154,151 for *production*.




3 Select some variables that are good candidates for filtering the project data. In the right panel, select **Yes** for **Use in faceted filter**.

**Note:**

Make sure **Use in faceted filter** is turned on for all default attributes.

4 For any numeric variables with the **Use in faceted filter** setting turned on, examine the **Display as Range** setting and make any necessary changes.

5 Select the Pipelines tab.

For new projects, the default pipeline is loaded.

6 Right-click the Data node at the top of the pipeline and select **Run**.

7 When the Data node has finished running, right-click the Data node again and select **Explore time series**.

This opens the Time Series Aggregation Plot of the project's historical data. To the left of the plot is the Attributes panel. All of the attributes that you selected as filters are displayed. Under each filter are values that you can select. See "Attributes" on page 43 for more information about using Attributes.

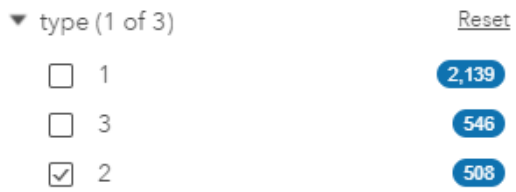
8 Explore different combinations of filter values and observe how the aggregation plot changes to reflect the data that you have selected.

After you run pipelines on the project and have a champion pipeline selected, you can use any combination of these filter values to specify overrides on forecast values. However, after you start working with overrides, your selection of which attributes to use as filters is limited. You can still add attributes to the list of filters, but you cannot remove them.

9 Return to the Data tab and select the attributes table again.

Make any necessary changes to your filter selections. Here are some considerations for setting **Display as Range** for different data types.

- You cannot use **Display as Range** for attributes with a character data type.
- Attributes with integer values with a very limited range are not good candidates for **Display as Range**. The following example shows a numeric attribute, *type*, with only three values. Using a slider to specify a range of values for this attribute is not useful.




- Attributes with decimal values or a very large integer range are good candidates for **Display as Range**.
- In general, you should use **Display as Range** for attributes with a SAS date or datetime type.

10 Review your settings carefully. After your assignments are locked down, the **Use in faceted filter** setting cannot be turned off for an attribute. For more information, see "Locking the Data Definition" on page 41.

# Adding Events to Your Project

## Overview

The most likely problem with your initial forecast model is that the model ignores a known event. You can improve the accuracy of your model by adding events. When you add events in a project, SAS Visual Forecasting must reforecast the project in order for the models to reflect these changes.

An event is an incident that disrupts the normal flow of a process that generates the time series. Examples of events are holidays, retail promotions, and natural disasters. Defining an event enables you to model the effect that special events have on the dependent time series. When you apply an event to a time series, SAS Visual Forecasting creates an indicator variable. This variable indicates the occurrence of the event at the time period for which the event is defined. This indicator variable is used as a regressor variable for time series modeling and forecasting.

For example, retail sales data follows a fairly steady pattern depending only on the day of the week. However, when you include events such as a one-time New Year's Day promotion, the forecasting model can predict a temporary increase in sales and then return to normal sales levels after the event. Some yearly events that occur on a fixed date, such as New Year's Day, can be modeled as part of the regular seasonal model. Seasonal events that are most effectively modeled as regressors are those that occur on a different date each year, such as Thanksgiving.

You can use SAS Visual Forecasting to predict the effect of a future event. Before the forecasting model can predict the effects of a future event, your data must include the effects of past occurrences of the event. For example, after reviewing your December sales data in the United States, you decide to create an event for next Christmas. SAS Visual Forecasting must have data that explains how retail sales were impacted by previous Christmas events. Without this data, SAS Visual Forecasting cannot predict the effect of the next Christmas event on retail sales.

SAS Visual Forecasting includes a list of defined events. However, you can provide a customized list of events. This is described in "Creating a Customized List of Events" on page 14.

**Note:**

- You cannot add events to an external forecast project.
- The Naive Model Forecasting modeling strategy does not support events.

## Adding Predefined Events

SAS Visual Forecasting is shipped with a list of predefined events that you can add to any project. You assign events at the project level.

Follow these steps to select from this list which events to add to your project.

1. On the Data tab for a project, select **Predefined events** under **Events** in the left panel.

   An empty table is shown in the middle panel.

2. Click **Add Predefined Events** at the top of the middle panel. If this button is not visible, click ⋮ and select **Add Predefined Events**.

   The Add Predefined Events dialog box is displayed.

3. Select events from the list and click **Add**.

   The events are added to the table in the middle panel. The **Event Name** is a system name, for example, `MOTHERS` is the name for `Mothers Day`.

After a predefined event is added to the project, it cannot be removed. If you do not want to use it in the project, you can change the event usage to **Do not use**. For any predefined events that you add to the project, the default is **Use if significant**.

To change settings for predefined events, follow the steps in "Changing Event Settings" on page 14. You can change these settings for specific models in a pipeline. For more information, see "Changing Event Usage in Modeling Strategies" on page 28.

**Note:** If you import custom events that include an event with the same name as a predefined event, you cannot add the predefined event.

## Importing Custom Events

If you have your own set of custom events to add to the project, you can add an event definition table to your project. Follow these steps to import your custom events.

If a custom event definition table is already added, these steps are used to replace it. The data set that is used to replace the custom events must include the same event definitions. More events can be added in the replacement, but existing custom events cannot be removed.

**Note:** If the data set with custom events includes an event with the same name as one that you have already added from predefined events, the data set cannot be added.

You can also add custom events from a Forecast Server project. You will need to import the data set that contains the event definitions. For more information, see Save Events to a SAS Data Set in the *SAS Forecast Studio: User's Guide*.

1. In the left panel of the **Data** tab, click ✱ and select **Events**.

   The Browse Data dialog box is displayed.

2. Locate the data source with the custom events to use with this project. Select the data source and click **OK** to load it.

   When the data source is successfully loaded, the new events are listed in the middle panel.

By default, each custom defined event that you add is set to **Do not use**. To change settings for custom events, follow the steps in "Changing Event Settings" on page 14. You can change these settings for specific models in a pipeline. For more information, see "Changing Event Usage in Modeling Strategies" on page 28.

You can click to replace the custom events data set. The new data set must contain at least the same event definitions as the original custom events data set. More event definitions can be added. If you update the original event data set, use instead to update the events.

## Changing Event Settings

In the left panel of the Data tab, selecting a data source under **Events** shows the list of events. The data source could be the predefined events provided by SAS Visual Forecasting or a data set with custom event definitions that was imported for the project.

In the middle panel, each event for the data source is listed along with the event settings, under the column **Usage in System Models**. These settings are the default settings used by the models in each pipeline, unless you change the event setting for any individual models within the pipeline, as described in "Changing Event Usage in Modeling Strategies" on page 28.

Follow these steps to change the default event settings.

1. With the event data source selected in the left panel, select an event in the middle panel. You can select multiple events if you want to assign the same setting to more than one event.
2. In the right panel, click **Usage in system-generated models** to view the settings that you can apply to the event. If the right panel is not visible, click « to display it.
3. Select one of the following settings:

   **Do not use**
   Do not include the event in the model.

   **Try to use**
   Include the event in the model as long as the parameters of the events are significant and the increment of the value of the criterion exceeds a specified threshold.

   **Use if significant**
   Include the event in the model as long as the parameters of the events are significant.

   **Force use**
   Include the event in the model as long as the model does not fail to be diagnosed.

The new settings are displayed in the middle panel under **Usage in System Models**. If you have run any pipelines, they become out-of-date and must be run again.

## Creating a Customized List of Events

SAS Visual Forecasting includes a list of predefined events. However, you might need to add to this list of events or create a customized list of events for your company. To define an event, you must create the event using the HPFEVENTS procedure.

To create a customized list of events:

1. From the SAS Drive menu bar, click ≡ and select **Develop SAS Code**.
2. Create an events data set using HPFEVENTS procedure. Here is an example.

***Example Code 1.1*** *HPFEVENTS Procedure*

```
proc hpfevents data=sashelp.air lead=12;        /* a Start the HPFEVENTS procedure */
   id date interval=month;
```

```
   eventdef fallpromo=    '01oct2018'd ;        /* b EVENTDEF statement for the event definitions */
   eventdef back2school=  '07aug2018'd to '30aug2018'd by day;
   eventdef superbowl =   '06FEB2011'D  '05FEB2012'D  '03FEB2013'D
                          '02FEB2014'D  '01FEB2015'D  '07FEB2016'D
                          '05FEB2017'D  '04FEB2018'D  '03FEB2019'D
                          '02FEB2020'D;
   eventdata out=eventds (label='Event List');  /* c Specify the name of the output event data set */
run;
```

a The DATA option requires an input data set. For this example, `sashelp.air` is used because it is already sorted. The goal of this step is to get the output data set with event definitions. Any sorted data set is sufficient.

b Use the EVENTDEF statement for each custom event definition. Each definition is in the form of *SAS-variable-name=timing-value-list*.

In this case, the `fallpromo` defines a single date. Most events are likely to be repeated, such as the `superbowl` definition. Since the Superbowl is always on a Sunday, there is a specific date for each year. The `back2school` event definition spans each day from August 7 to August 30. Like `fallpromo`, this event is defined only for a single year. Adding another `fallpromo` event for the following year would require another definition.

The event definition can also include optional *qualifier-options*. See the *SAS Forecast Server Procedures 14.3: User's Guide* for more information about event definitions.

c The event data set is created by specifying the data set name using the OUT option in the EVENTDATA statement.

For more information about defining events using the EVENTDEF statement or using the HPFEVENTS procedure, see The HPFEVENTS Procedure in the *SAS Forecast Server Procedures 14.3: User's Guide*.

# Segmenting Project Data

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 segments that you can run through different modeling strategies in a pipeline.

## Defining Project Segments

To segment your data, you first have to define how you want the data segmented. For example, you might want to run different modeling strategies on seasonal products as opposed to products that are sold throughout the year.

After you have your segmentation strategy defined, you need to setup an attributes data set with a _SEG_ variable. This is a simple numerical variable that assigns a segmentation number to your data, based on the attribute values that are part of your segmentation strategy.

The following code provides a simple example for adding segments to the SKINPRODUCT_ATTRIBUTES data set. This data set is available with other sample data sets that you can download, as described in “Preparing an Input Data Set” on page 1.

```
cas casauto;                            /* 1 - Start a CAS session          */
libname mycas cas;                      /* 2 - Create a CAS engine libref   */
data mycas.segment_attr (promote=yes);  /* 3 - Specify the new data set     */
  set public.skinproduct_attributes;    /* 4 Read in the attributes table   */
  if Venue = 'Catalog' then _seg_ = 1;  /* Assign Segments                  */
```

```
  if Venue = 'Grocery Store' then _seg_ = 2;
  if Venue = 'Internet' then _seg_ = 3;
  if Venue = 'Outlet Store' then _seg_ = 4;
  if Venue = 'Third Party Vendor' then _seg_ = 5;
run;
```

This code simply shows how to assign _SEG_ numbers based on the value of a single variable, `Venue`. How you segment your project data depends on the characteristics and complexity of your data source, and how you might need to process different segments.

When assigning _SEG_ values, make sure the values fall within the valid range for segmentation. Segment numbers can be any integer between 1 and 25 inclusive. Any other value, as well as missing _SEG_ values, places the data into a segment called Remaining Series.

Follow the steps in "Working with Attributes" on page 10 to add the data set with the segments and attributes to the project.

After the data set with the segment definitions have been added, you can create segmented pipelines to run the project data. For more information, see "Working with Segmented Pipelines" on page 34.

# Working with Projects

When you log in to Model Studio, the page displays a list of projects that have already been created. To create a new project, see "Creating Your First Project" on page 3.

To open a project, click the link on the project name.

## Exporting Projects

You can download projects when you need to export them to another SAS Visual Forecasting server. The remote server must have access to the same data sources as the original project. If the project has any pending overrides, they will not be included in the downloaded project.

**Note:** Only users in the **`SAS Administrators`** group can export projects.

1. From the Model Studio Projects page, select one or more projects to export.
2. Click the menu icon ( ⋮ ) over the right side of the project list and select **Export**.

   A message indicates that the project is being prepared for export.
3. When the project is ready for export, select the location on your local drive for the project. If multiple projects are selected, repeat this step until all selected projects are downloaded.

## Importing Projects

You can import projects that have been created from another forecasting application.

- Projects that have been archived from SAS Forecast Studio have a FAR file extension. The imported project is placed in your private or personal CASUSER directory. Scenario analyses are not supported in this release of SAS Visual Forecasting and are not migrated. Overrides and custom models are supported, but are not migrated. Events must be imported separately. See "Importing Custom Events" on page 13 for more information.

  The data set for the project must contain a variable that is in a SAS date or datetime format. For more information, see "Requirements for Time Series Data" on page 2.

- Projects from another SAS Visual Forecasting installation must be in ZIP format. Projects from SAS Visual Forecasting 8.2 must be imported using the transfer command. See "Promoting Projects and Templates to SAS Visual Forecasting 8.3" on page 97 for complete instructions.

  **Note:** Only users in the `SAS Administrators` group can import SAS Visual Forecasting projects.

- The source data for the project must already be loaded to the CAS server before importing the project. If the project has additional attributes or event definitions from an external data source, those items must also be loaded to the CAS server.

1. From the Projects page, click the menu icon ( ⋮ ) over the right side of the project list and select **Import** ⇨ **Forecasting**.

   The Import Forecasting Project dialog box is displayed.

2. Click **Browse** to provide the filename and location of the project to be uploaded.

3. Click **Browse** to provide the location on the target server where the project data is located. The variables in the selected data set must match the variables in the uploaded project.

4. Click **Import** to complete the process. The uploaded project is opened to the Data tab. The variable assignments on the Data tab are the same as the original project.

5. Depending on the type of project that you are importing, follow these steps.

   SAS Visual Forecasting
   After importing a project, the pipelines must be run again. If a champion pipeline was manually selected in the original project, that selection is not preserved in the uploaded project. After uploading a project from another SAS Visual Forecasting server and running the pipelines, any existing overrides must be resubmitted.

   SAS Forecast Studio
   When importing projects from SAS Forecast Studio, the following hierarchy aggregation and time accumulation values are not supported. The imported project sets these values to `Sum of values`.

   - **Median**
   - **First Occurrence**
   - **Last Occurrence**
   - **Number of Observations**
   - **None**

   In the right panel, review the settings for the dependent and independent variables and make any adjustments as needed.

After a project ZIP file is imported, only the original owner of the project can access it. The owner can share the project with other user groups. If the owner is not available, the SAS administrator can share it.

## Sharing Projects

When you create a project, you become the project owner and the project is kept in private mode so that only you can view or edit it. When you need to allow other users to access the project, you can use Model Studio to share the project. You can designate groups of users to share projects, not individuals.

The Model Studio implementation of sharing is separate and distinct from project sharing as performed in SAS Drive. Any projects that you share using SAS Drive do not retain the same settings for user groups in Model Studio. Also, any projects that you share using Model Studio do not retain the same settings for users in SAS Drive.

When you share a project using Model Studio, you must specify how it can be accessed, either read-only or read-write mode.

- When a project is in read-only mode, it can be viewed but nobody can make any changes, including the project owner. If you need to make changes to the project, you can change the access to read-write and make any necessary adjustments to the user groups who can access the project. You can also make the project private again while you work on it.
- When a project is in read-write mode, anyone with access to it can make changes to the project. However, the project can be accessed only by a single user at a time. If another user tries to access the project while it is being edited, the user receives a message indicated that the project cannot be accessed until it is released by the current editor.

To share a project that you own, or to change the project back to private access, follow these steps.

1. Open the Projects page in Model Studio and select the project in the table.
2. Click the menu icon ( ⋮ ) over the right side of the project table and select **Share project**. The Share Project dialog box is displayed.
3. Select the project sharing options.
   - To share the project with other users, click **Shared Project** and select the user groups that should have access.
   - To remove sharing for a project, click **Private project**. The project is now available only for you to view and edit.
4. For shared projects, set the access level for all users. Select **Read-only** if you want to prevent any changes to the project. If you want all users to be able to make changes to the project, make sure **Read-only** is not selected.

## Working with External Forecast Projects

External forecasts are projects created from data that already has forecasts along with the historical data for each time series. The data set must follow this schema.

- _NAME_ - the name of the dependent variable in each row
- ACTUAL - values of the dependent variable in the historical data
- PREDICT - predicted values of the dependent variable in the forecast period (horizon)
- ERROR - prediction errors
- STD - prediction standard errors
- UPPER - upper confidence limits
- LOWER - lower confidence limits
- The time variable from the original time series project
- All BY variables in the hierarchy of the original time series project

**Note:**

When you create an external forecast project, the input data set must include both historical and forecast time periods.

After an external forecast project is created, the forecast period is already defined. Refreshing the input data source with actual values placed in the forecast period changes the start of the horizon, which can invalidate the external forecast project.

You can create an external forecast data set from a time series project in SAS Visual Forecasting and SAS Forecast Studio. For SAS Visual Forecasting, see “Exporting Output Data from a Project” on page 52 for more information.

To create an external forecast project, follow the instructions in "Creating a Forecasting Project" on page 3, and specify the external forecast data set as the data source.

When the project is created, it is opened to the Data tab with this table shown. The project pipeline is initialized with the External Forecasts modeling node. This is the only model that can be used in the pipeline for an external forecast project. You can add additional attributes as described in "Working with Attributes" on page 10.

## Refreshing Project Data

SAS Visual Forecasting automatically checks the source data from which a project is created for updates. If the source data has been changed, you are notified and can update the project data. If you choose not to update the project data when you are prompted, you can submit a request to refresh the data at a later time as follows:

1 Navigate to the project's Data tab.

2 In the left panel, click .

All data sources listed are refreshed if their content has changed.

**Note:** In the original data source, if any of the variable formats are changed or if variables are added or removed, data refresh will fail.

If you choose not to refresh data when prompted, the project will still display updated metadata in the right data panel when the data source is selected.

## Running Batch Code

You can download code to run a SAS Visual Forecasting project in batch mode.

1 From the Model Studio Project page, select a project to run in batch.

2 Click the menu icon ( ⋮ ) and select **Download batch API**.

The Batch API dialog box is displayed with the code for the project inline.

**Note:** You can also download batch API code from within a project by clicking the settings icon in the project title bar.

3 Select the **Download Type** at the bottom of the dialog box. The options are as follows:

- **SAS**: Downloads code to run in a SAS programming environment.
- **Python**: Downloads code to run in a Python programming environment.
- **REST**: Downloads a text file with example REST calls that you can use in an application.

4 Click **Download** and select the location to save the file.

To run the batch code, you need to supply the host name of the SAS Visual Forecasting server and the user name and password.

**SAS**

Update the `host`, `username`, and `password` macro variables at the end of the code.

```
%let protocol = http;
%let host = test.example.com;
%let port = 80;
%let username = my_username;
%let password = my_password;
%let projectId = a0548b2f-a669-4a10-a4dd-052c671c0c00;
```

**Python**

Set these parameters on the command line. For example, if **download.py** is the file name of the Python batch code, invoke the command:

```
$ python download.py --host test.example.com --username my_username --password my_password
```

If you have overrides in the project, they might generate conflicts when they are submitted by the batch code. You can update the batch code to automatically resolve these conflicts by adding `autoresolve="true"` to the code.

**SAS**

Find the PROC HTTP procedure that includes the following IN option with `firstTransaction` and `lastTransaction`.

```
in="{""firstTransaction"":""@first"",""lastTransaction"":""@last""}"
```

Add the autoResolve setting, as follows:

```
in="{""firstTransaction"":""@first"",""lastTransaction"":""@last"",""autoResolve"":true}"
```

**Python**

Find the resubmit_overrides function:

```
def resubmit_overrides(env):
```

Update the body statement with autoResolve assignment.

```
body = '{\"firstTransaction\":\"@first\",\"lastTransaction\":\"@last\",\"autoResolve\":true}'
```

For more information about running SAS code in batch mode, see *SAS Companion for UNIX Environments*.

# 2

# Working with Pipelines

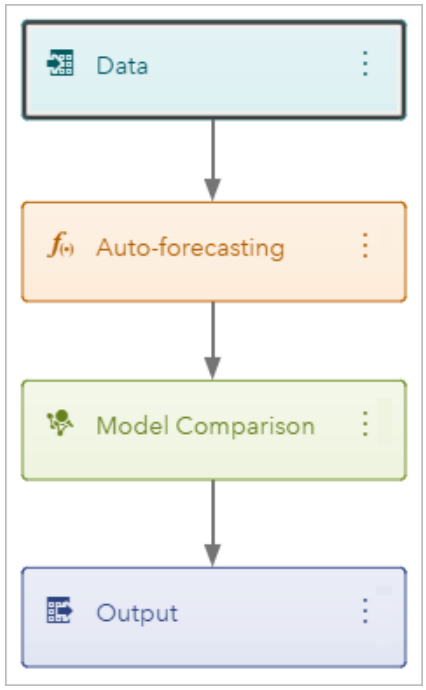
# Using the Pipelines Tab

## Initial Pipeline for a New Project

Pipelines consist of a series of nodes that are run in a sequence to produce forecasts based on the models and settings that you choose for the project. After you first create a project, the pipeline is set up in the Pipelines tab. The pipeline is labeled **`Pipeline 1`**. For time series projects, the Auto-forecasting node is selected for the modeling strategy.

The pipeline cannot be run until you assign the required variable roles as described in "Assigning Variable Roles" on page 5.

For a general description of the nodes in a pipeline, see "Pipelines" in SAS Visual Forecasting: Overview.




## Creating a New Pipeline

1 Click **+** to create a new pipeline. The New Pipeline dialog box is displayed.

   If this is the first pipeline that you are creating for the project, it is named **`Pipeline 2`**.

2 You can change the name to any meaningful string. After the pipeline is created, you can change the pipeline name by double-clicking it the name in the tab.

3 Select a template and click **Save**.

## Node Status

Hold your pointer over each node to see its status.

- When you first create the pipeline, all nodes have a status of **`Initialized`**.
- After a node in the pipeline has successfully finished execution, the node shows a status of **`Successful`**.
- A node with a status of **`Pending`** is waiting for other nodes in the pipeline to complete before starting.
- If a pipeline run fails, check to see which node has a **`Failed`** status.

- If you make any changes to the project settings or variable assignments, the nodes will change to `Out-of-date` status and the pipeline needs to be run again.
- When running a pipeline, if a node fails for any reason, the status of the subsequent nodes in the pipeline show `Canceled`.

## Pipeline Details

Right-click any node in a pipeline to open a menu of actions.

- Select **Run** to run the node. Any preceding nodes in the pipeline that do not have a status of `Successful` are run first. After the preceding nodes complete successfully, the selected node runs.
- Select **Rename** to rename a node. The new name applies only to the node on this pipeline.

Right-click the data node to open these actions in the menu.

- Select **Add below** to add a forecasting modeling node to the pipeline. If the pipeline is segmented, this action is available only from each segment.
- After the data node has completed a run, select **Explore time series** to open a plot of the historical data.

If the pipeline is segmented, right-click the External Segmentation node and select **Manage Segments** to change the number of segments in the pipeline. See "Changing the Number of Segments" on page 35 for more information.

Right-click any modeling node to get these actions in the menu.

- Select **Delete** to remove the node from the pipeline. If it is the only modeling node in the pipeline, then the Model Comparison and Output nodes are also removed.
- Select **Open** to open the code editor for the node. This option is not available for all modeling strategy nodes.
- Select **Modify Event Usage** to change the settings for events for the node in this pipeline or segment. See "Changing Event Usage in Modeling Strategies" on page 28 for more information.
- Select **Save as** to save the node, with any settings you have made in the Options panel on the right, to The Exchange. This makes the saved node available for other projects.
- Select **Results** to view the results of the modeling strategy, including an execution summary, MAPE distribution, and other relevant data. For more information, see "Viewing the Results for a Modeling Strategy Node" on page 29.
- Select **Log** to view a log of the node's processing.
- Select **View code** to get a read-only view of the code for the modeling strategy.

Right-click the **Model Comparison** node to get these actions in the menu.

- Select **Results** to view the MAPE distribution and summary statistics for each modeling node along with the selected champion model in the pipeline.
- Select **Select champion model** to choose a champion model other than the one selected by the Model Comparison. If you choose a different champion model, you need to rerun the Model Comparison and Output nodes.

Use the following icons on the right side of the pipeline content area to manage your pipelines.

Saves the pipeline as a template to The Exchange for use with other forecasting projects. External Forecasts should not be saved as a template because the new template cannot be used in any project.

Displays a smaller scale view of the pipeline in the upper left corner. Hold your mouse down on this view and move it around to change the area of the pipeline that is displayed.

▶

Runs the pipeline.

■

Stops a running pipeline.

🗑

Deletes a pipeline. This is disabled if you have only one pipeline.

## Nodes Panel

On the left side of the pipeline is the Nodes panel. This panel lists the modeling strategy nodes that you can add to any pipeline.

If the Nodes panel is not displayed, click » on the left side of the pipeline to open it. You can collapse the panel again by clicking «.

You can drag a modeling node over the Data node in the pipeline to add it. To remove a modeling node, right-click the node and select **Delete**. Each pipeline requires at least one modeling node.

## Options Panel

The Options panel on the right side of the pipeline displays options that you can set for a selected node in the pipeline.

If the Options panel is not displayed, click « to open it. You can collapse the panel again by clicking ».

- All nodes in the pipeline have a **Description** option that you can update. Any description you provide is saved for the node only in the pipeline. The description for the node template in The Exchange is not affected by any updates that you make in the pipeline.
- Many modeling nodes in the pipeline have an **Open** button that you can use to view and edit the code inline. Use *SAS Visual Forecasting: Programming Guide* as a reference when you are working with this code. You cannot edit code for the following nodes:
  - Hierarchical Forecasting
  - External Forecasts
  - Panel Series Neural Network
  - Stacked Model (NN + TS) Forecasting

After running a pipeline, any subsequent changes to the options for a modeling node move the pipeline back to `Out-of-date` state. This is also true if you make a change and then change the setting back to the original value.

When a pipeline is running, the options for each modeling node in the pipeline cannot be changed until the pipeline completes. If you need to make further updates to a modeling node after starting the pipeline, click ■ and wait for the pipeline to stop operation before you make any changes.

# Modeling Strategy Nodes

Every pipeline must have at least one modeling strategy node. You can select from the models provided by SAS Visual Forecasting or any custom models that might have been created for your site.

To add a modeling node to the pipeline, choose a node from the Nodes panel on the left and drag it over the data node in the pipeline. If the Nodes panel is collapsed, click » to reveal the nodes that are available.

## Modeling Strategy Nodes Provided by SAS Visual Forecasting

You can use one or more of the following nodes for your pipelines. These modeling strategies are listed in the Nodes panel on the left of the Pipelines tab. They are also available as templates in The Exchange.

Some of the nodes contain a code editor that you can use to customize its settings. After running a pipeline, any change to the settings for a modeling node requires the pipeline to be run again.

### Auto-forecasting

For each time series, the Auto-forecasting node automatically diagnoses the statistical characteristics of the time series and generates a list of appropriate time series models. Next, it automatically selects the model, and generates forecasts. This node evaluates and selects the following models for each time series:

- Autoregressive integrated moving average with exogenous inputs (ARIMAX)
- Exponential smoothing (ESM)

You can change the following settings for auto-forecasting.

**Open code editor**
Click **Open** to view and change the code for this node. For more information, see "Customizing a Modeling Strategy" on page 32.

**Model selection criteria**
The model selection criteria are statistical values that are used to evaluate how well a forecasting model performs by comparing the actual data to the predictions. These are also known as statistics of fit. For a complete description of all of the options, see "Descriptions of Statistics of Fit" on page 89.

This setting overrides the **Champion selection criteria** setting in Project Settings.

### Hierarchical Forecasting

For each time series, the Hierarchical Forecasting node performs the following tasks.

- Diagnoses the statistical characteristics of the time series for each level of the hierarchy
- Generates a list of appropriate time series models based on the diagnostic settings that you specified
- Optimizes model parameters
- Selects the champion model from candidate list of models
- Generates forecasts
- Reconciles the forecasts from the reconciliation setting

To use this modeling node, you need at least one BY variable assigned.

**Note:** The code used for this node cannot be edited. If you download this node, it does not include the code to create the node.

For a complete description of the settings for Hierarchical Forecasting, see "Hierarchical Forecasting Settings " on page 69.

### Hierarchical Forecasting (Pluggable)

The Hierarchical Forecasting (Pluggable) node provides similar functionality as the Hierarchical Forecasting node. The main difference is you can use the code editor to make changes to the model inline. Also, if you download this node, it includes the code for further customizing. For a complete description of the settings for this modeling node, see "Hierarchical Forecasting (Pluggable) Settings" on page 74.

## Naive Model Forecasting

The Naive Model Forecasting node uses one of the naive models that you select to generate forecasts for each time series.

You can change the following settings for auto-forecasting.

**Open code editor**
Click **Open** to view and change the code for this node. For more information, see "Customizing a Modeling Strategy" on page 32.

**Naive model type**
Select from the following.

- **Moving average**

  For this option, indicate the size of the moving average window.

- **Random walk**

  For this option, specify if you want the drift option.

- **Seasonal random walk**

  For this option, specify if you want the drift option.

**Drift option for random walk models**
If you choose a *Random walk* or *Seasonal random walk* model, you can add the drift option.

**Moving average window size**
Specify an integer, greater than 1, for the window size for the moving average model.

**Note:** The Naive Model Forecasting node does not support events.

## External Forecasts

The External Forecasts node works only with external forecast projects. It cannot be added to the pipeline for projects created from any other type of data source. External forecast projects can use only External Forecasts in the pipeline.

**Note:** The code used for this node cannot be edited. If you download this node, it does not include the code to create the node.

External forecasts do not support events

Use External Forecasts in a pipeline to generate the summary statistics and to work with overrides in the forecast horizon. For more information, see "Working with External Forecast Projects" on page 18.

## Modeling Strategy Nodes Using Neural Networks

If your site has a license for SAS Visual Data Mining and Machine Learning, these modeling strategies are available for use in the pipelines.

**Panel Series Neural Networks**
The Panel Series Neural Network modeling strategy provides forecasts by training a neural network based on user settings and develops a model to extract salient features across multiple time series. Neural networks consist of predictors (input variables), hidden layers, an output layer, and the connections between each of those items.

The Panel Series Neural Network modeling strategy is best suited for panels of time series with several interval independent variables. Independent variables for the neural network can be assigned in the data tab or generated within the Feature Generation section of the node options. BY variables are automatically

included as categorical inputs. This neural network cannot generate a model for the project without an independent variable.

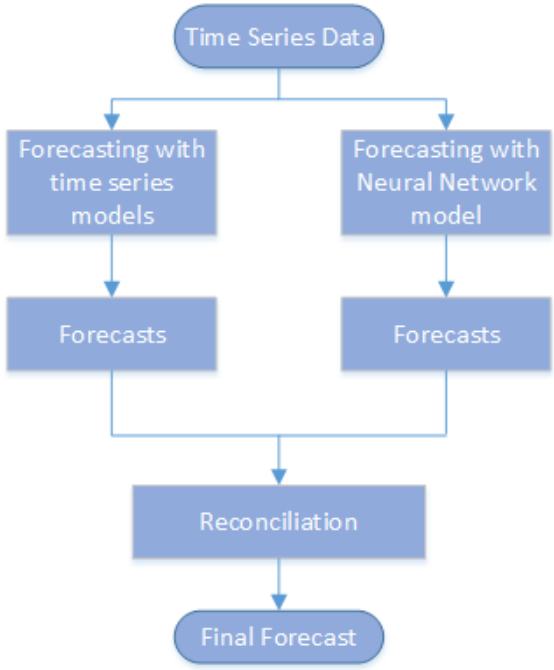
**Note:** The code used for this node cannot be edited. If you download this node, it does not include the code to create the node.

For a description of the settings for this node, see "Panel Series Neural Network Settings" on page 80.

The other two modeling strategies combine neural network modeling with time series models.

### Multistage-forecasting

The Multistage-forecasting modeling strategy provides a general framework that combines time series models and feature extraction techniques to build a hierarchy-based forecasting system in three stages. This diagram shows the process flow of the Multistage-forecasting modeling strategy.




- In the first stage, the modeling strategy applies time series models to generate forecasts at the aggregated level of the hierarchy.
  - If **Stage 1 high level** is specified, forecasts for both stage 1 high level and stage 1 low level are generated using the time series model. The forecast at **Stage 1 low level** is the top-down reconciled forecast from these two results.
  - If **Stage 1 high level** is not specified, forecasts at stage 1 low level is generated using the time series model.
- In the second stage, the feature extraction technique is applied, using regression or neural network, to generate forecasts for each individual time series at lower levels of the hierarchy.

  **Note:** The first and second stages run and generate forecasts in parallel.
- In the third stage, the forecasts for low level is generated by reconciling the stage 1 and stage 2 forecasts using top-down reconciliation.

This modeling strategy is intended for a project with a defined hierarchy. If no BY variables are defined for the project, the forecast is generated using stage 1 time series model. For more information, see "Defining the Hierarchy" on page 7.

For more information about the individual settings for Multistage-forecasting, see "Multistage Forecasting Settings" on page 76.

### Stacked Model (NN + TS) Forecasting

The stacked modeling strategy generates forecasts using stacked models that include a neural network model (NN) and a time series model (TS). This modeling strategy captures the nonlinear relationship between the dependent and independent variables as well as time series characteristics in the data, such as seasonality and trend. As this diagram shows, it models the time series in two steps:

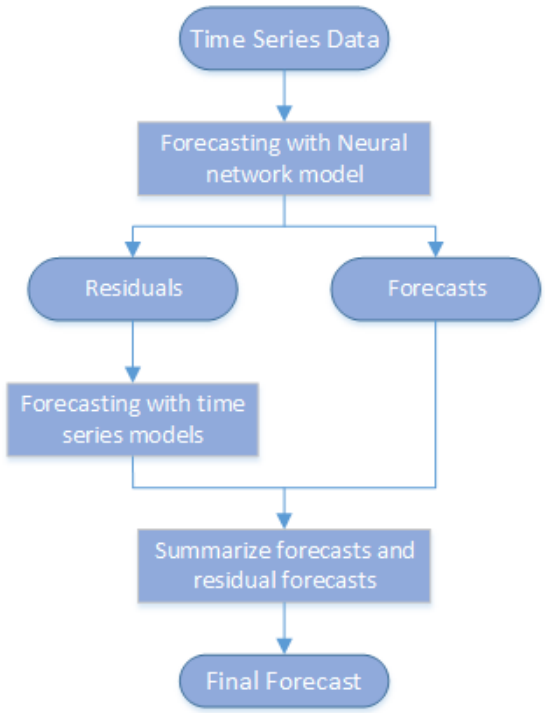
- In the first step, the neural network model is used to generate the forecasts.
- In the second step, the residuals from the first step are passed to the time series model to generate residual forecasts.

The first and second steps run sequentially.

The final forecasts are the sum of the forecasts from the neural network and the residual forecasts from the time series model.

**Note:** The code used for this node cannot be edited. If you download this node, it does not include the code to create the node.

For more information about the settings for Stacked Model (NN + TS) Forecasting, see "Stacked Model (NN + TS) Forecasting Settings" on page 85.




## Changing Event Usage in Modeling Strategies

Events that have been added to a project can be further customized for each model in each pipeline. For predefined events, you can change only the settings for predefined events that have been added to the project. For custom events that have been added to the project, you can change the settings for all of the events.

1. Right-click the modeling node and select **Modify event usage**.

   The Modify event usage page is displayed.

2. Select either **Predefined events** or the custom event table from the list in the left panel.

3. Select an event from the table in the middle panel.

4. In the right side panel, change the **Usage in system-generated models** setting.

   **Do not use**
   Do not include the event in the model.

   **Try to use**
   Include the event in the model as long as the parameters of the events are significant and the increment of the value of the criterion exceeds a specified threshold.

   **Use if significant**
   Include the event in the model as long as the parameters of the events are significant.

   **Force use**
   Include the event in the model as long as the model does not fail to be diagnosed.

The setting that you apply is effective for this modeling strategy in the pipeline. If the pipeline is segmented, the other modeling strategies are not effected by this change. If you save the modeling strategy to The Exchange, the event definitions are not included.

Changing the event usage makes the modeling strategy out-of-date. The pipeline must be run again.

## Viewing the Results for a Modeling Strategy Node

When a modeling node has finished running successfully, it has a green check mark indicating a status of Successful. Right-click the node and select **Results** to view the output of the node. Results displays a Summary and Output Data tab for a modeling strategy.

### Summary Tab

The following information is shown on the Summary tab. For the Hierarchical Forecasting modeling strategy, this information is provided for each level within the defined hierarchy.

**MAPE Distribution**
This graph shows the MAPE distribution of results.

**Model Family**
This graphs shows the percentage of time series model by each model family. This graph shows how each model type (ARIMA, Combined, Exponential Smoothing, Intermittent Demand, and Unobserved Components) fits the series. Neural network models are included in the `Other` family.

**Model Type**
This graphs shows the percentage of time series used for by each model type. This graph shows whether the model included a dependent transformation, a seasonal component, a trend component, inputs (such as independent variables), events, and outliers. This table is not displayed for non-time series based modeling strategies, such as the Panel Series Neural Network, Stacked Model (NN + TS) Forecasting, and Multistage-forecasting modeling.

**Execution Summary**
This table shows the results of the following measures.

**Number of series**
the total number of time series accumulated from the data. For a Hierarchical Forecasting modeling strategy, this corresponds to the number of time series in each level of the hierarchy.

**Number of failures**
the number of forecasts that failed. For a Hierarchical Forecasting modeling strategy, this corresponds to the number of failed forecasts in each level of the hierarchy.

**Number of forecasts equal to zero**
the number of forecasts that are 0

**Number of intermittent series with flat forecasts**
the number intermittent time series that have a constant forecast value over the length of the forecast horizon

**Number of seasonal series with flat forecasts**
the number in seasonal time series that have a constant forecast value over the length of the forecast horizon

**Number of short series with flat forecasts**
the number of short time series that have a constant forecast value over the length of the forecast horizon

**Number of all other series with flat forecasts**
the number of all other time series that have a constant forecast value over the length of the forecast horizon. This number does not include short, intermittent, or seasonal time series with flat forecast values.

**Number of forecasts below 70%**
the number of time series with forecast values that are below the minimum range of the mean value for that time series. The forecasts for these series can be problematic. The minimum range value is set to 70% by default. You can change this setting using the **Forecast exception boundaries** in the Project Settings on page 68.

**Number of forecasts above 300%**
the number of time series with forecast values that are above the maximum range of the mean value for that time series. The forecasts for these series can be problematic. The maximum range value is set to 300% by default. You can change this setting using the **Forecast exception boundaries** in the Project Settings on page 68.

## Output Data

The following output tables are generated by modeling strategies in a pipeline or segment. If the table does not initially load, click **View Output Data** in the content area.

### Forecasts

The OUTFOR table provides the forecasted values generated by the modeling strategy. This table includes the BY variables for the project and the following information.

- **_NAME_** - the name of the dependent variable
- **Time ID Values** - the Time ID interval values for each time series in English format
- **Actual Values** - the values of the dependent variable for each time series in the historical record
- **Predicted Values** - the predicted values for each time series in the forecast horizon
- **Prediction Standard Errors** - the standard error estimate from the models
- **Lower Confidence Limits** - the lower confidence limit as determined by the model
- **Upper Confidence Limits** - the upper confidence limit as determined by the model
- **Prediction Errors** - residuals between the actual values and the model predictions

### Model Information

The OUTMODELINFO table contains detailed information about the selected forecast model. For each BY variable combination, this table contains the following information.

**Note:** This table is not available for non-time series based modeling strategies such as the Panel Series Neural Network, Stacked Model (NN + TS) Forecasting, and Multistage-forecasting.

- **Variable Name** - the name of the dependent variable
- **Model** - the name of the selected model specification
- **Model Family** - the type of model used ( ESM, ARIMA, UCM, or IDM)
- **Dependent Variable Transform** - the type of transformation used, or NONE if not used
- **Seasonal Model** - a value of 1 indicates a seasonal model was used
- **Trend Model** - a value of 1 indicates a Trend model was used
- **Inputs Present** - a value of 1 indicates inputs present in the model
- **Events Present** - a value of 1 indicates events present in the model
- **Outliers Present** - a value of 1 indicates outliers present in the model
- **Model Status** - the execution status of the model
- **Model Source**

### Forecast Statistics

The section contains the following tables.

OUTSTAT

contains the statistics for the current model for the most recent data update. In addition to the dependent and BY variables, this table includes the following information.

- **Variable Name** - the name of the dependent variable.
- **Region** - shows the region in which the statistics are calculated.
  - FIT indicates that fit statistics were calculated over the fit region.
  - FORECAST indicates that fit statistics were calculated over the fit region.
- **Selection List** - the name of the model selection list to which the specified **Model** belongs.
- **Model** - the name of the model specification.
- **Degrees of Freedom Error**
- **Number of Observations**
- **Number of Observations Used**
- **Number of Missing Actuals**
- **Number of Missing Predicted Values**
- **Number of Model Parameters**
- There are many columns not shown by default in this table. These columns include the statistics of fit, which are used as model selection criteria. To add these columns to the display, click . For more information, see "Model Selection Criteria" on page 88.

OUTSUM

The OUTSUM data set contains the summary statistics for the dependent variable as determined by the value of the forecast horizon. For each time series, the following statistics are provided.

**Time Series Length**
the number of observations in each time series

**Number of Nonmissing Values**
the number of observations in each time series that do not have missing values

**Number of Missing Values**
the number of observations in each time series that have missing values

**Minimum Value of Series**
the smallest value of observations in the time series

**Maximum Value of Series**
the largest value of observations in the time series

**Mean Value of Series**
the mean value of observations in the time series

**Standard Deviation of Series**
the standard deviation of observations in the time series

**Condition Code for Series**
a value that indicates whether analysis was successful

- 0 - Analysis was successful
- 3000 - Accumulation failed
- 4000 - Missing value interpretation failed
- 6000 - Series is all missing
- 9000 - Descriptive statistics could not be computed

### Logs

This section contains the OUTLOG table, VALIDATIONLOG table, or both. The tables contain text messages that arise from the processing of the BY group's time series data. If no exceptions occur during processing, the content for this table is empty. In addition to the BY variables, this table includes the following information.

- **_ERRNO_ Status** - a numeric variable that stores the _ERRNO_ variable for the BY group. The value of _ERRNO_ might be set by the user-defined program directly or might be set implicitly by calling a function or method that sets the _ERRNO_ value.
- **Length of log content** - a numeric variable that stores the length of the _LOG_ variable text (byte count).
- **Log content** - messages that are logged from the execution of the user-defined program on the BY group's time series data. All messages from the BY group are concatenated into the variable. End-of-line characters separate the individual messages.

## Working with Modeling Strategy Templates

You can customize many modeling strategies and save them as templates in The Exchange. If you are familiar with SAS code for SAS Visual Forecasting, you can create your own modeling strategy templates and upload them to The Exchange.

### Customizing a Modeling Strategy

Use the Options panel on the Pipelines tab to customize the settings for a selected modeling strategy in your pipeline. You can also open the code editor to make changes to some of the modeling strategies. See *SAS Visual Forecasting: Programming Guide* for more information about the code used for modeling strategies.

The changes that you make to the modeling strategy affect only the current pipeline.

Click to save the modeling strategy to The Exchange. This makes the modeling strategy available to be selected in other pipelines and projects, or to be downloaded for use in other deployments of SAS Visual Forecasting.

For more information about The Exchange, see "Working with Model Studio" in SAS Visual Forecasting: Overview.

**Note:** The code for Hierarchical Forecasting cannot be edited inline or downloaded. However, the settings that you choose in a pipeline can be saved and downloaded. External Forecasts has no settings that can be customized and saved and the code is also not available.

Event definitions for modeling strategies are not saved to The Exchange.

### Downloading Modeling Strategies

To download a modeling strategy, follow these steps.

1. Go to The Exchange, as described in "Working with Model Studio" in SAS Visual Forecasting: Overview.

   The types of modeling strategies are listed on the left under Nodes.

2. In the list of nodes, select **Forecasting Modeling**.
3. Select the modeling strategy that you want to download. You can select multiple modeling strategies to download.
4. Click the **Download** option from the menu icon ( ⋮ ).

Each modeling strategy is downloaded as a ZIP file with the following contents:

**`code.sas`**
This file contains the SAS code to run the model for the input data.

**`template.json`**
The JSON file defines the name, description, identifier, and other properties of the modeling strategy. This file defines properties that users can select in the Options panel on the Pipeline tab for the modeling node.

**`validation.xml`**
Contains the validation model for the modeling strategy.

**Note:** This file is not downloaded for the following nodes:

- Hierarchical Forecasting
- External Forecasts
- Panel Series Neural Network
- Stacked Model (NN + TS) Forecasting

### Uploading Modeling Strategies

The modeling strategy must be packaged in a ZIP file. The contents must include the same files that the download package includes (**`code.sas`**, **`validation.xml`**, and **`template.json`**).

To upload a modeling strategy, go to The Exchange and click the **Upload** option. When the upload operation is complete, the time stamp is displayed by the modeling strategy.

For more information about using The Exchange, see "The Exchange" in SAS Visual Forecasting: Overview.

# Generating Forecasts

## Running the Forecast Pipeline

Before running a pipeline, be certain that your settings for variable assignments and project settings are correct. After you run one or more pipelines successfully and open the Overrides tab, many of these settings are committed and cannot be changed. Before running the complete pipeline, you can run just the data node and review the results.

1. Right-click the Data node and select **Run**.
2. When the Data node has finished processing successfully, right-click the Data node and select **Explore Time Series**.

The Time Series Aggregation Plot is shown for the historical data. The data is accumulated using the time interval settings that you specified in the variable assignments. The BY variables that you specified for the project are listed on the left. If you need to make any final changes to these settings, make those changes before running the full pipeline.

When you are certain that the variable assignment and project settings are good, click ▶ to run the pipeline.

While the pipeline is running, you can create another pipeline or exit the project. If you remain on the Pipelines tab long enough, a message is displayed when the pipeline completes.

If there are any problems running the pipeline, check the logs of the modeling nodes. To open the log, right-click a modeling node and select **Log**. The log shows the code for the node along with any notes, errors, and warnings.

### Changes That Invalidate Pipelines

If you make certain changes to a project, all pipelines must be run again to correctly reflect these changes:

- changing the forecast settings in the Project Settings dialog box.
- changing the data source for a time series, attributes, or external forecast project. Any change to the data source must follow the same schema as the original.
- adding attributes to the project using the Attributes view of the Data tab.
- adding or removing any independent variables or changing their settings.
- changing the modeling nodes in a pipeline. This includes adding or removing nodes, or changing certain options for a node. Only the affected pipeline needs to be run again.

After you run a pipeline and open the Overrides tab, variable assignments and project settings are committed and cannot be changed in the project. For more information, see "Locking the Data Definition" on page 41.

If you need to make any changes to the committed settings, for example, setting a different time interval, you can start a new project using the same data source.

## Working with Segmented Pipelines

After creating a pipeline with segments, you can make many adjustments to customize the output of your forecasts.

### Creating Segmented Pipelines

Follow these steps to create a pipeline for segmented data. The project must already have an external attributes table with defined segments. See "Defining Project Segments" on page 15 for more information.

1. From the Pipelines tab, click **+** to create a new pipeline.

   This opens the New Pipeline dialog box.

2. For **Name**, enter a meaningful name or accept the default.
3. Click the drop-down menu for **Template** and select **Browse templates**.
4. In the template list, select **External Segmentation** and click **OK**.
5. Click **Save** on the New Pipeline dialog box.

The new pipeline is created with the number of segments defined by the attributes table. There is also a Remaining Series segment to contain any data with missing values or segment values that do not fall within the valid range. The Merged Segments node collects and passes the merged results from each segment to the Output node.

### Customizing Each Segment

When you create a segmented pipeline, each segment is created with a top-level segment node, modeling node, and Model Comparison node.

- Each segment node is numbered, for example, `Segment 1` and `Segment 2`. You can rename the segment nodes to a more meaningful name to reflect the nature of the data in each segment, for example, *Low Volume Sales* and *High Volume Sales*.

- You can run each segment node individually. After the segment has completed, right-click the segment and select **Explore time series** to open a plot of the historical data for the segment.
- You can add or remove any number of modeling strategies for any segment. From the Nodes panel, you can drag the modeling strategy over the segment node to add it. To remove a modeling strategy, right-click the node and select **Delete** from the menu. You can change event usage for each node within a segment.
- After the Model Comparison for a segment has been run, right-click the node and select **Results** to view the results for each modeling strategy.

### Changing the Number of Segments

The number of segments are defined by the attributes table that has been added to the project. In many cases, you do not need to change the number of segments. Following are some reasons why you might need to change the number.

- **Updated segmentation assignments in the attributes table**

  For example, if the attributes table originally contained eight segments but after updating the table it contains fewer or more segments, you might want to change the number of segments in the pipeline to match.

- **Upper level segments can be grouped**

  For example, the pipeline has eight segments, but you want to group segments 6, 7, and 8 and run them using the same modeling strategy. To do this, reduce the number of segments to five. The data from the top level segments is moved to the Remaining Series segment, along with any other data that does not fall within the numbered segments.

Follow these steps to change the number of segments.

1. Select the External Segmentation node in the pipeline.
2. Click **Manage Segments** in the Options panel to the right of the pipeline. The Manage Segments dialog box is displayed.

   If the Options panel is not displayed, click ‹‹ to open it. You can collapse the panel again by clicking ››.

3. For **Number of segments**, enter the number of segments that you want to use for this pipeline and click **Save**.

   The number of segments must be between 1 and 25, inclusive.

The pipeline is refreshed with the new number of segments.

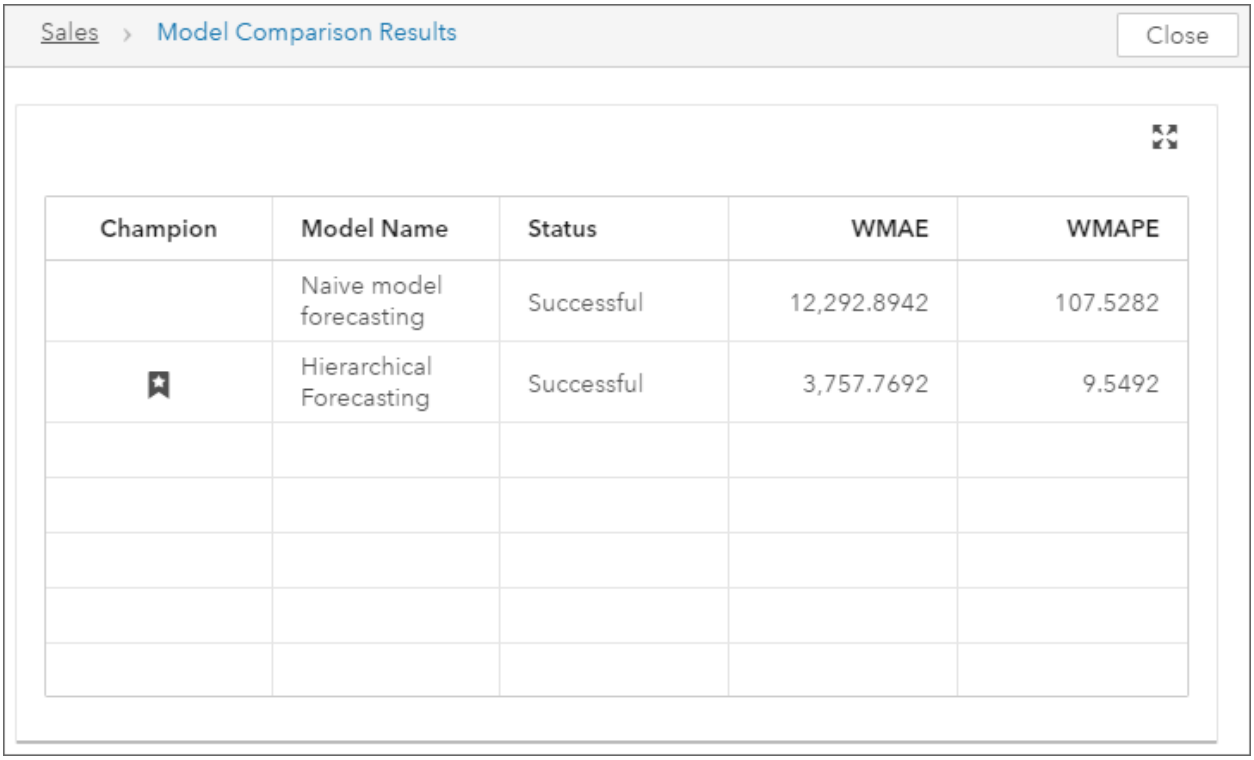
## Selecting a Champion Pipeline

Before you work on any overrides for your forecasting project, you need to make sure that you are working with the best pipeline for your data. SAS Visual Forecasting selects the best fit model in each pipeline and in each segment of a pipeline. After each pipeline is run, the champion pipeline is selected based on the statistics of fit that you choose for the selection criteria. If necessary, you can change the selected champion pipeline.

Changes to the project or other settings force your pipelines and the selected champion pipeline to become out of date. Make sure all pipelines are up-to-date and the best fit model and pipeline is selected as champion before working with overrides.

## Model Comparison

Each pipeline has a Model Comparison node. When this node has completed successfully, you can view the results by right-clicking the node and selecting **Results** from the menu. The results show the MAPE distribution and summary statistics for each modeling node along with the selected champion model in the pipeline.




| Champion | Model Name | Status | WMAE | WMAPE |
|---|---|---|---|---|
| | Naive model forecasting | Successful | 12,292.8942 | 107.5282 |
| | Hierarchical Forecasting | Successful | 3,757.7692 | 9.5492 |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

The table that shows the selected champion modeling strategy also displays the selection criteria that are available for comparing different pipelines. Not all columns are shown. To add or remove columns from this table, follow these steps.

1 Click in the upper left over the table.

2 In the right most table header, click and select **Manage columns**.

3 To add columns to the table display, select the column name in the **Hidden columns** list and click .

   To remove columns from the table display, select the column name from the **Displayed columns** list and click .

The table also shows the status of each modeling node in the pipeline. If any models are out of date, rerun the pipeline before selecting the final champion pipeline and working with overrides.

If you determine that another model should be selected as champion, you can manually select the champion model for the pipeline. To select a champion model, right-click the Model Comparison node in the pipeline and select the model from the **Select champion model** list. If you choose a different champion model, you need to rerun the Model Comparison and Output nodes.

## Model Comparison Results

When a pipeline or pipeline segment has finished running successfully, the Model Comparison node has a green check mark indicating a status of Successful. Right-click the node and select **Results** to view the output of the node. Results displays a Summary and Output Data tab for the pipeline or segment.

## Summary Tab

The following information is available on the Summary tab for Model Comparison Results.

Model Names

The table at the top displays all of the model strategies used in the pipeline or segment, the selected **Champion** modeling strategy, the

- **Champion** - The champion modeling strategy is marked with ★. See "Pipeline Details" on page 23 for information about manually selecting another modeling strategy as the champion.
- **Model Name** - Each model in the pipeline or segment is listed.
- **Status** - Indicates the status of each modeling node in the pipeline or segment. Make sure each node has a status of Successful before working in the Pipeline Comparison or Overrides.
- Weighted metrics - For a complete description, see "Weighted Model Comparison Selection Criteria" on page 91.

**MAPE Distribution**

There is a MAPE Distribution for each model in the pipeline or segment.

This graph shows the MAPE distribution of results.

**Execution Summary**

There is an Execution Summary for each model in the pipeline or segment.

**Number of series**
the total number of time series accumulated from the data. For a Hierarchical Forecasting modeling strategy, this corresponds to the number of time series in each level of the hierarchy.

**Number of failures**
the number of forecasts that failed. For a Hierarchical Forecasting modeling strategy, this corresponds to the number of failed forecasts in each level of the hierarchy.

**Number of forecasts equal to zero**
the number of forecasts that are 0

**Number of intermittent series with flat forecasts**
the number intermittent time series that have a constant forecast value over the length of the forecast horizon

**Number of seasonal series with flat forecasts**
the number in seasonal time series that have a constant forecast value over the length of the forecast horizon

**Number of short series with flat forecasts**
the number of short time series that have a constant forecast value over the length of the forecast horizon

**Number of all other series with flat forecasts**
the number of all other time series that have a constant forecast value over the length of the forecast horizon. This number does not include short, intermittent, or seasonal time series with flat forecast values.

**Number of forecasts below 70%**
the number of time series with forecast values that are below the minimum range of the mean value for that time series. The forecasts for these series can be problematic. The minimum range value is set to 70% by default. You can change this setting using the **Forecast exception boundaries** in the Project Settings on page 68.

**Number of forecasts above 300%**
the number of time series with forecast values that are above the maximum range of the mean value for that time series. The forecasts for these series can be problematic. The maximum range value is set to 300% by default. You can change this setting using the **Forecast exception boundaries** in the Project Settings on page 68.

### Output Data

The tables on this tab reflect the output generated by the current champion modeling strategy in the pipeline or segment. If the table does not initially load, click **View Output Data** in the content area. For a full description of these tables, see "Output Data" on page 30.

## Pipeline Comparison

The Pipeline Comparison tab shows the different pipelines for a project and selects a champion pipeline. The selected champion pipeline, along with the selected model in that pipeline, are used for the final forecasts for the project. When you create overrides for the forecasts, they are based on forecasts generated from the champion pipeline. You do not want to start creating overrides until you are confident that the best pipeline and model has been selected as champion.

If you determine for other reasons that another pipeline should be used as champion, follow these steps to select a different champion.

1 Select the Pipeline Comparison tab.

2 Select a pipeline from the list.

3 Click ⋮ over the right side of the table and select **Set as champion**.

The table is updated to show the new champion pipeline, using . When you go to the Overrides tab, the page is updated with the forecasts from the new champion.

**Note:** After selecting a new champion, all applied overrides must be resubmitted.

After manually selecting a champion pipeline, if you change the settings for any pipelines, the pipeline must be run again and the champion pipeline can be reassigned.

# Working with Pipeline Templates

When you create a new pipeline, Model Studio opens a dialog box that you can use to select a template for your pipeline. Pipeline templates are typically distinguished by the models that are included. You can also view the list of all pipeline templates in The Exchange.

Any changes that you make to a pipeline can be saved as a template to The Exchange. This includes any models that you add to the pipeline and any changes to the settings, names, code, or descriptions that you apply to any of the pipeline nodes. To save a pipeline, click  in the toolbox next to the pipeline.

## Pipeline Templates Provided by SAS Visual Forecasting

The following pipeline templates are provided by SAS Visual Forecasting.

You can select any of these templates as the default template to use when creating projects.

Auto-forecasting
: Creates a full pipeline using the Auto-forecasting node as the modeling strategy.

Base Forecasting
: Creates a pipeline with just the Data node. Add a modeling node to complete the pipeline.

Hierarchical Forecasting
: Creates a full pipeline using the Hierarchical Forecasting node as the modeling strategy.

You should never set the following templates as the default because they are used only for specific projects.

External Forecasts
: This template is valid only for external forecast projects. If this template is set to default, any attempt to create a new time series project causes the default pipeline to use the Base Forecasting template instead. The schema for external forecasts is explained in "Working with External Forecast Projects" on page 18.

  This template is automatically loaded as the pipeline for any external forecast projects that you create.

  **Note:**

  Do not save this pipeline as a new pipeline template. The new template cannot be used in any project.

External Segmentation
: This template is valid only for projects with defined segments. If this template is set to default, any attempt to create a new project without segment definitions causes the default pipeline to use the Base Forecasting template instead. For more information, see "Defining Project Segments" on page 15.

## Setting the Default Pipeline Template

SAS Visual Forecasting is deployed with the Auto-forecasting pipeline template set as the default. Follow these steps to choose another pipeline template as the default.

1. Open The Exchange. A list of node and pipeline templates is displayed.
2. Expand the **Pipelines** templates on the left side and select **Forecasting**.
3. Select the template that you want to use as the default whenever you create a new pipeline.
4. Click ⋮, and select **Set as default template**.

# 3

# Working with Overrides



## Locking the Data Definition

There are two important events that occur when you open Overrides.

- When you first open Overrides after running a pipeline, the data definition for the project is locked. The data definition includes the role assignments for the project and many of the project settings. After the role assignments are committed, you can no longer change these assignments for this project.

  Make sure you are satisfied with the project settings and variable assignments before opening Overrides for the first time. After the data definition is locked, if you determine that you need to change these role assignments, you can open a new project using the same data source and change the settings.

- Every time Overrides is opened, it checks for any changes to the champion pipeline output. If updates are found, you are prompted to refresh the overrides data. If you refresh the data, any applied overrides must be resubmitted and any pending overrides are deleted.

  If you have made any changes to BY variables or attributes, you should select **Yes** to update the data in Overrides. Selecting **No** in this situation generates errors in the Overrides tab.

After Overrides has been opened and the data definition is locked, you can still work with pipelines, adding or removing new models, or changing settings for the models. After completing your changes, make sure all of your

pipelines show a status of `Successful` and the best champion pipeline is selected before returning to the Overrides tab and refreshing the data.

## Overrides Overview

Use Overrides to view a plot of the historical and forecast data from your champion pipeline and to work with overrides for forecasts. The Overrides tab consists of two views.

### Overrides

The Overrides view is divided into three sections.

## Attributes

The Attributes panel is available in two locations:

- On the left side of the Overrides view of the Overrides tab.
- On the left side of the Explore time series view. This view is accessed from the Pipelines tab by right-clicking the Data or any Segment node and selecting **Explore time series**.

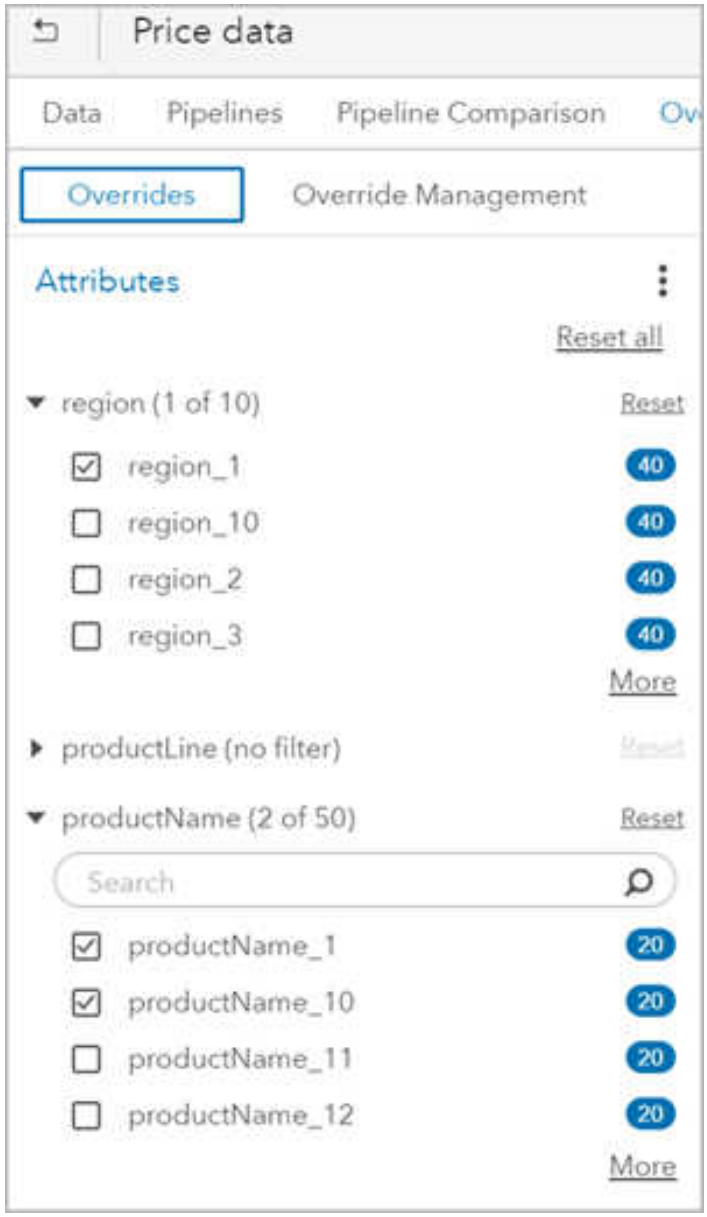
Use the Attributes panel to browse attributes of your data set and create filters based on specific values of those attributes. Filters enable you to work with a subset of the data set. The plot and data on the page are updated as you make your attribute selections.

For attributes that are character variables, there is a search field that enables you to search for specific values for that attribute. When using the search field to find attribute values, use either whole words or add a wildcard (*) to the string. For example, using the string `Shamp` is not going to find any results for products with *Shampoo* in the name. Instead, you can use `Shamp*`. The search field is not available for attributes that are of a numeric or date type.

**Note:**

There can be a discrepancy between the number of distinct values displayed for an attribute and the corresponding number shown in the Attributes view of the Data tab. In this case, the value on the Data tab is based on the formatted value of the variable. The values shown in the Overrides view are based on the raw, unformatted values of the variable.

If user-defined formats are used for any of the attribute variables in a project, they are not shown in the Attributes panel for Overrides. Instead, they are listed using the unformatted, raw data format.




For more information, see "Working with Filters" on page 46.

## Time Series Aggregation and Overrides Plot

The Time Series Aggregation and Overrides Plot provides a graphical view of your historical and forecast data. You can focus the plot on a subset of the time series by selecting **Show overview axis**. This displays a slider that you can use to set a beginning and ending date for the plot.

You can also use the scroll wheel to zoom in or out of the plot time line. When you are zoomed into a portion of the time line, you can scroll the view horizontally or vertically as follows.

- Hold your cursor over the bottom of the plot to display a horizontal scroll bar so that you can view time periods before or after those that are in view.

- Hold your cursor over the right side of the plot to display a vertical scroll bar so that you can view values above or below those that are in view.

These icons are located above the Time Series Aggregation and Overrides Plot:

- Click to save any selected attribute values as a filter. You can also use this to recall any filters you have saved. See "Working with Filters" on page 46 for more information.
- Click to export the project along with applied overrides. See "Exporting Output Data from a Project" on page 52 for more information.

### Forecasts and Overrides Table

This table shows the historical data for the time series and forecasts for the designated forecast horizon. Use this table to create, edit, and submit, override values for a time series, based on external factors that are not included in the forecast models. For time series within the forecast horizon, the table shows the statistical forecast values from the champion pipeline along with the lower and upper confidence limits (Forecast LCL and Forecast UCL).

In the Override row, right-click an empty cell for a time period to open the menu. Depending on the status of any overrides, the following menu options are available. These options are also available from the menu icon over the table for a selected cell.

**Override Calculator**
Use this dialog box to set override specifications on a filter for one or more time periods. This action is not available for a time period that already has an applied lock on the current filter. For complete information, see "Override Calculator" on page 51.

**Impact Analysis**
Open this view on a cell to see information about overrides from other filters in the same time period and their impact on any override you might create. This action is available when overrides from other filters are applied on the same time period. **Impact Analysis** is not available if another filter has a locked statistical forecast for the same time period.

See "Impact Analysis" on page 46 for more information.

**Submit All**
After creating pending overrides, this action is available on any cell in the Overrides row. All pending overrides that you have created for this filter are submitted.

**Delete Pending Override**
This action is available on a cell with a pending override. Only the selected override is deleted.

## Override Management

Use this view to edit overrides that have already been submitted or to resolve conflicts between overrides. You can also remove overrides that have already been applied or resubmit overrides that have been impacted by other activity.

**Filter**
Provides a way to show a subset of the overrides based on character data in **Filter** and **Status** columns.

View menu
Click this menu to select the overrides to display, based on the override status.

**View pending**
View overrides that have been created but not submitted, or those that need to be resubmitted (Resubmit Pending).

**View applied**
View overrides that have been successfully submitted.

**View conflicts**
View overrides that have conflicts with other overrides for one or more time periods. Click the conflict link to view details about the conflict and resolution options. See "Working with Override Conflicts" on page 50 for more information.

**View archived**
View overrides that are no longer in the current forecast horizon. Archived overrides cannot be edited or removed. There are two types of archives.

- **`Archive - Expired`** — The data has been refreshed and overrides that were created in the forecast horizon are now included in the historical data.
- **`Archive - Blank`** — The forecast horizon has been reset so that the last time interval ends on a date that precedes the time period for the override. The override is no longer valid for this filter and time period.

**View active**
View all overrides except for those that are archived.

**View all**
View all overrides, regardless of status.

Use the following features to manage overrides.

**Submit Pending**
Submits all pending overrides listed in Override Management. This performs the same action as submitting pending overrides in the Forecasts and Overrides table. After reconciliation, the overrides are moved to applied status if no conflicts are found. If conflicts are found in any overrides, they are moved to conflict status and you must resolve the conflict before you can proceed any further with overrides.

**Undo Last Submitted**
Click this button to move the last submitted overrides, which are currently in **`Applied`** status, to **`Pending`** status. This essentially removes the submit action and returns the override to **`Pending`** status. After performing this action, you can either delete or resubmit specific pending overrides.

- If the last submitted override has been archived, this action is disabled.
- If you have any overrides with **`Pending`** status, they must be removed before you can undo the last submission.

If you have selected a single override in the table, this opens the Override Calculator with the settings for the selected override. You can make changes to these settings using the Override Calculator. For more information, see "Override Calculator" on page 51.

Deletes one or more selected overrides.

Opens a menu with the following actions:

**Impact analysis**
Open this view on a selected pending or applied override to see information about overrides in the same time period. For more information, see "Impact Analysis" on page 46.

**Automatically resolve all conflicts**
If one or more override conflicts are listed, select this option to let SAS Visual Forecasting resolve the conflicts for you.

**Resubmit overrides**
Under some circumstances, applied overrides must be resubmitted. This can occur when certain applied overrides are removed, conflicts are resolved, or override data has been refreshed.

**View overrides logs**
Opens the output log for the overrides listed in Override Management. Click **Download Logs** to download the log to your local drive. You can select any transaction group on the left, or you can view and download all override logs.

## Impact Analysis

Use Impact Analysis to view information about overrides from other filters and their impact on your current filter and time slice selection either before or after submitting an override. It can be used to review the impacts of previous overrides in history or inform a decision on an upcoming set of overrides. It shows the number of time series that intersect with the selected override. It also shows the difference in statistical forecast and final forecast as denoted by the **Delta** value. Click on any filter link to view the override specification in the Overrides tab or click on a transaction group to view the override in context on the Override Management tab.

# Working with Filters

## What Is a Filter?

Use filters to segment forecast data based on specific value combinations of attributes in your project. At the minimum, the default attributes for the project can be used to create filters. You can also include additional attributes to your project. For more information, see "Working with Attributes" on page 10.

For example, consider a project with the following attributes.

- `Region`
- `Product line`
- `Product`

You could create the following filters using these attributes:

- `Region` = `District_12`
- `Productline` = `Shoes`
- `Region` = `District_12` and `District_8` and `Productline` = `Swimwear`
- `Productline` = `Shoes` and `Product` = `Women's Addidas Tennis Shoes`

You can use these filters to view the historical and forecast values for data that meets any of these attribute combinations. You can also override forecast values based on time series for these filters.

## Creating Filters

One or more default attributes must be assigned to the project before you can create filters. See "Assigning Variable Roles" on page 5 for more information.

Use the Attributes panel on the left side of the Overrides view to create and manage filters. See "Attributes" on page 43 for more information about how the Attributes panel works.

1. Click ⋮ to open the Filter Settings dialog box and customize which attributes are displayed in the Attributes panel. Use the Filter Settings dialog box to show, hide, or rearrange the order of the attributes. These settings are saved for the duration of your session.
2. Select values from the attributes in the list to create a subset of the data to work with. You can select multiple values from any set of attributes. The time series aggregation plot and forecasts in this view are updated to reflect the selections that you make.

For example, consider a project with the attribute, `Product Name`. You could use the search field for the *Product Name* attribute and search for all products with *Shampoo* in the name. Select all of the products in the search results. The time series aggregation plot and forecasts in this view are updated to reflect the selected products that contain the string *Shampoo*.

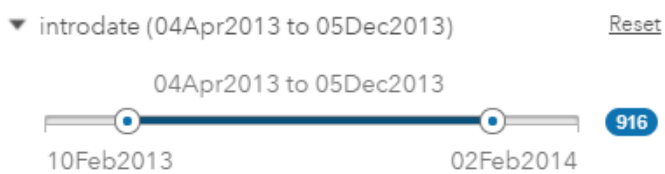
The search field is available only for character variables.

You can make selections from a single attribute or from multiple attributes, using any combination of values. For example, you could further subset the Shampoo filter by selecting a brand or geography.

**Note:** The Time Series Aggregation and Overrides Plot is updated as you make selections. If no time series exists in the set of attributes that you have selected for a filter, then no data is shown.

3 Click the **More** link to get a complete list of all attributes that match your search. To select all of the results in the search request, be sure to manually select each item shown. Do not use the **Select all** field because this selects all values for the attribute.

4 When working with the slider to indicate a range of values, it can be difficult to set the slide control to a specific number. The following image shows a date range for the attribute, *introdate*. Dates are numerical values that indicate the number of days between January 1, 1960, and a specified date.




a Move the slider control as close to the date that you want to set as the upper or lower range.

b Click the slider control. The numerical value for the date is shown. In this figure, 19,452 is the SAS date numerical equivalent to April 4, 2013.


c Change the number in the field, adding or subtracting the number of days you need to set the exact date. In this figure, five days is subtracted from the number to change the date to March 30, 2013.


5 On the Overrides tab, after creating a filter that you can use, click over the Time Series Aggregation and Overrides Plot and click **Add Filter to Favorites**.

6 Give the filter a unique name that you can use to retrieve this filter again and click **OK**. The length of the name cannot be more than 50 characters.

# Working with Overrides

## Overview

An override is used to provide adjustments based on the statistical forecasts and assess the impact of overrides on the final forecast results. Use the Overrides view to create, edit, or delete pending overrides. When you have finished working with an override, you can submit. After submission, the override is moved to applied state and the Final Forecast for that time period reflects the reconciled value.

The statistical forecasts in the time series in Overrides are based on the champion pipeline selected from all of your pipelines. All pipelines should be finished running and the champion pipeline should be selected before you start working with overrides. If a new champion pipeline is selected after you start working with overrides, you are prompted to refresh the overrides data and resubmit any overrides that have been applied.

## Creating Overrides

Before you can add any overrides, you must first run one or more pipelines and a champion pipeline must be selected. Any changes you make to the project settings can invalidate the selected champion pipeline. In this case, the pipelines must be run again and a new champion must be selected. When you are certain that the final champion pipeline is selected, you are ready to start working on applying overrides to the statistical forecasts from that pipeline. See "Running the Forecast Pipeline" on page 33 for more information.

When you are specifying overrides, it is important to remember the following:

- You can enter overrides only for the time periods in the forecast horizon. The forecast horizon is the period of time for which forecasts are computed. It starts after the end of the historical data and continues for the number of periods specified by the **Number of forecast periods (horizon)** option in the Project Settings dialog box. The end date for the series is the largest time ID value with a nonmissing value for the dependent variable.
- You cannot add a negative override if negative forecasts are not allowed in the project. To allow negative forecasts, you must select the **Allow negative values for forecasts and overrides** check box in the Overrides panel of the Project Settings dialog box. Changes to this setting require you to rerun the pipelines.
- If you create one or more overrides for one filter, you must submit those overrides before you can create overrides on another filter.
- Forecast values of `0` cannot be overridden.
- Override values cannot be pasted from the clipboard into the override cell. You can paste the value into the **Set to a value** field in the Override Calculator.

Follow these steps to add overrides:

1. In SAS Visual Forecasting, select the Overrides tab for the project.
2. Select one or more filters from the Attributes section. For example, you could select a single product name if you want to review the forecasts and submit any overrides for that product. Or you might select several store locations within a focused geographical location that often have similar results. Typically, you want to have some filter specifications already bookmarked. See "Creating Filters" on page 46 for more information.
3. To add an override for any forecasted time series, right-click the Override row for that time series and select **Override Calculator**. For more information about using this tool to create overrides, see "Override Calculator" on page 51. You also have the option of entering a specific value directly in that table cell.

For each override, you can specify whether the override is locked or unlocked. This locking determines how SAS Visual Forecasting treats the override during the reconciliation process. For more information, see "Lock and Unlock Overrides" on page 49.

4 When you have finished specifying overrides for a filter, click ⋮ and select **Submit All**.

When you submit a pending override, SAS Visual Forecasting reconciles the override value with other overrides and constraints imposed on that time period. If no conflicts are found, the override's status is applied and the final forecast shows the reconciled value. If conflicts are found, you must resolve them as described in "Working with Override Conflicts" on page 50.

## Override Status

After you have created and submitted new values for overrides, the cells change depending on the status of the override. The following are status indicators shown on table cells to describe the status of an override:

Pending override: 8,000,000.00

After you create an override, the cell in the Override row displays the override value and the pending status icon. When you are ready for one or more pending overrides to be calculated in the final forecast, right-click the cell and select **Submit All**.

If you set a lock on this time period, the cell background is gray. You cannot enter an override value directly in the cell, but you can still open the Override Calculator and make changes while the status is in pending.

Applied override: 2,000,000.00

After you submit an override, the override is reconciled and, if no conflicts are found, the status becomes applied. The applied value is shown on the Final Forecast row for that time period.

If the **Aggregate final forecast lock** is turned on for an applied override, the cell in the Override row is disabled.

Locked statistical forecast: 6,520,800.13

If the **Statistical forecast lock** is turned on for a time series, the value is shown as locked in the Forecast row. The icon remains in the Overrides cell until the statistical forecast lock is applied.

Applied locked statistical forecast: 31,914,676.76

After a locked statistical forecast is applied, the status is indicated in the Final forecast row. The value shown matches the value in the Forecast row and the Override cell is disabled.

## Lock and Unlock Overrides

You can choose to lock an override value or the statistical forecast for a time period.

A locked override is a user-supplied value for a forecast that is honored when the hierarchy is reconciled. When you lock an override, SAS Visual Forecasting changes the final forecast to the override value and ensures that the reconciled forecasts add up to the locked override value. If the locked override value cannot be honored, this creates an override conflict that you must resolve, as described in "Working with Override Conflicts" on page 50.

Two types of locks exist in SAS Visual Forecasting:

Statistical forecast lock

The statistical forecast lock preserves the statistical forecast for a time period and filter, ignoring any overrides that might impact the values for that time period. For example, if you lock the statistical forecast for

one store in the month of June, any overrides that might apply to the products sold in that store are ignored for the June time period. Those overrides might be applied to other stores for the month of June instead.

Aggregate final forecast lock
: The aggregate final forecast is the value calculated for an override that is applied to a time period for a specific filter. The lock is applied to the aggregated value.

To lock an individual override, right-click the Overrides cell and select **Override Calculator**. Select the lock that you want to use on the right side of the calculator.

If SAS Visual Forecasting cannot resolve any override conflicts, then you can also unlock overrides from the Override Conflicts dialog box. For more information, see "Working with Override Conflicts" on page 50.

## Remove Overrides

After you remove an override, you must manually re-enter that override. There is no other way to restore any overrides that you previously specified.

- To remove overrides in pending state, right-click the override cell in the Forecasts and Overrides table and select **Delete pending override**. You can also delete pending overrides in the Override Management tab by selecting the overrides and clicking 🗑.
- To remove overrides in applied state, select the overrides that you want to remove from the Overrides Management tab and click 🗑.

When the operation is complete, any remaining applied overrides are moved to `Resubmit pending` state. You must resubmit all of these overrides before you can perform any further work with overrides.

## Working with Override Conflicts

### What Is an Override Conflict?

Override conflicts are detected when you submit pending overrides or resubmit applied overrides. When you submit an override, SAS Visual Forecasting reconciles the override with other overrides and forecasts in the time series. If a conflict is found, you are notified and the conflict must be resolved immediately.

Conflicts can be caused by the following conditions:

- Two or more overrides cannot be reconciled for the same time period.
- An override falls outside of the **Maximum percentage change of final forecast** setting in the Project Settings.
- Any combination of constraints caused by locked statistical or aggregate values, maximum allowed changes, or overrides from filters with overlapping filter specifications.

When the conflict is found, you are prompted to make a choice between **Resolve Automatically** (letting SAS Visual Forecasting resolve the conflict) or **Resolve Manually** (resolving it yourself). You cannot create any new overrides until conflicts are resolved.

### Resolving Override Conflicts Automatically

If you do not choose to resolve conflicts automatically when you are prompted, you can still choose to take this option using the Overrides Management tab. Click ⋮ over the right corner of the table and select **Automatically resolve all conflicts**. SAS Visual Forecasting resolves the conflicts and updates the table.

### Resolving Override Conflicts Manually

If you choose to resolve override conflicts yourself, the Override Management tab is opened. The override with the conflict has a conflict link in the Status column. Each conflict in the table is labeled as **Conflict** plus a number.

| Filter | Time Period | Status |
|---|---|---|
| Filter 2 | 22Mar2015 | Conflict 1 |
| Group B | 22Mar2015 | Applied |
| Filter 1 | 22Mar2015 | Applied |

Click the conflict link in the Status to open the dialog box to resolve the conflict. You are provided with several resolution options for resolving the conflict. The following figure shows a conflict with two filters that are already applied with an aggregate final forecast lock.

| Filter | Time Period | Override Status | Statistical Forecast | Override | Solution 1 Optimized Final Forecast | Solution 2 Optimized Final Forecast | Aggregated Final Forecast Lock | Delete |
|---|---|---|---|---|---|---|---|---|
| Group B | Sun, Mar 2... | Applied | 4,297,5.... | 2,477,0.... | 1 2,477,0.... | 3,517,5.... | | |
| Filter 2 | Sun, Mar 2... | Pending | 5,884,0.... | 10,414,... | 7,861,5.... | 2 10,414,... | | |
| Filter 1 | Sun, Mar 2... | Applied | 4,343,7.... | 6,515,6.... | 1 6,515,6.... | 8,178,0.... | | |

**Solution 1**
This option gives priority to locked overrides that have already been applied. Choosing this solution replaces your pending override value with the best alternative that can be applied.

**Solution 2**
This option is available when you try to apply a locked override that cannot be reconciled with a previously locked value. This solution gives priority to preserving the pending locked override and replaces one or more previously applied locked overrides.

**Update lock(s) and/or delete overrides(s)**
Choose this option to either update a lock or delete an override altogether.

In some cases where the **Optimized Final Forecast** value is the same for both solutions, only **Solution 1** is offered.

After you make your selection and resolve the conflict, go to the Override Management tab and resubmit the overrides.

## Override Calculator

Use the Override Calculator to create and edit specifications for overrides. Each override is used to update the final forecast for a filter on a given time period.

### Properties

Use the Properties tab to set the specifications of the override. You can choose from the following specifications.

**Adjust based on an existing forecast value**
Select the forecast value for which you want the override to be evaluated during reconciliation.

- **Statistical Forecast** — Reconcile the override with the statistical forecast as determined by the model pipeline.
- **Final Forecast** — Reconcile the override with the final forecast along with any other overrides or constraints that affect this time period and filter selection.

**Adjustment** — Select an up or down (+ or –) change to the forecast based on a percentage or number of units from the statistical or final forecast value.

**Set to a value**
Provide a specific integer value to be considered for the final forecast value.

**Set to a range**
Specify a lower and upper confidence limit to be considered during reconciliation for the final forecast value.

You can add multiple time periods to your override specifications. Click the + icon in the **Selected time periods(s)** field to move time periods from the list of available items to the list of selected items. When you click **OK**, the time periods are added to the selected field.

### Filter

Use the filter tab to work with the current filter for the bookmark.

**Filter**
Browse through the selected values and attributes for the current filter. If you have not selected any attributes, this field is not available.

**Name**
Provide a name for the filter that is reflected in the override transaction as displayed in Override Management and Impact Analysis. If you do not provide a name, SAS Visual Forecasting creates a name by default. The length of the name cannot be more than 50 characters.

**Add this filter to favorites**
Select this to add the filter to the favorites list, which is accessed from over the Time Series Aggregation and Overrides Plot.

# Exporting Output Data from a Project

You can export project data that can be used in another project. The exported data is saved as an OUTFOR data set. When this is used to create a new project, that project can be used to view the historical time series and to work with overrides.

There are two places that you can export project data.

## Exporting from a Selected Project Pipeline

You can export the project along any pipeline that you choose.

1. Select the Pipeline Comparison tab from your project.
2. Select a pipeline from the list.

3 Click ⋮ and select **Export output data**.

   The Export Data dialog box is displayed.

4 Under **Data Sources**, select a library to export the data.

5 Enter a table name for the output data set and click **Export**.

You can now create a new project using the output data as a data source.

## Exporting Applied Overrides with Project Data

You can export your project data along with applied overrides. The OUTFOR data set will include an FF column for final forecasts.

**Note:**

If the project is shared as read-only, this action is disabled.

1 Select the Overrides tab from your project.

2 Click ⋮ and select **Export all data**.

   The Export All Data dialog box is displayed.

3 Under **Data Sources**, select a library to export the data.

4 Enter a table name for the output data set and click **Export**. Make sure the name has no spaces.

You can now create a new project using the output data as a data source.

# 4

# Forecasting Concepts



## Understanding Time Series Data

### How SAS Visual Forecasting Creates Time Series Data

SAS Visual Forecasting creates the time series data through the following process:

1. The data is sorted by the BY variables and the time variable.
2. The data is accumulated to the appropriate time interval if the input is one of the following types:
   - time-stamped data that is recorded at no particular frequency (also called transactional data)
   - data recorded at a higher time interval frequency than needed for forecasting (for example, data recorded on a daily frequency but weekly interval forecast is desired)
3. Your input data set might contain BY variables. If you do not use all of the BY variables in your project, the observations are aggregated.
4. Any gaps in the data are filled in. Gaps appear when there is not an observation for each time period or when the data is not equally spaced. The added observations have the required values of the time variable and the value that you specified for missing values. For more information, see “Missing interpretation” on page 65.

When you create a project, you select the input data set to use, and you must assign variables to the time and dependent variable roles. You can also specify the default attributes (BY variables) and independent variable roles. SAS Visual Forecasting uses this information to create the time series data.

### Examples of Input Data Sets

Here are two examples of input data sets:

- This input data set contains monthly sales revenue and price information for the past 12 months. In this example, the time variable is Date, the dependent variable is Revenue, and the independent variables are Avg. Price and Holiday. The variable Holiday indicates the number of holidays during the month.

*Table 4.1 Monthly Sales for 12 Months*

| Date | Revenue | Avg. Price | Holiday |
|---|---|---|---|
| JAN2011 | 18817 | 26.3 | 0 |
| FEB2011 | 52573 | 25.3 | 0 |
| … | … | … | … |
| DEC2011 | 44205 | 20.3 | 1 |

- This input data set contains monthly retail sales information for different regions and product categories over the past 12 months. You can use the Region and Product variables to create a hierarchy for the sales forecasts.

*Table 4.2 Monthly Sales for Different Regions and Categories*

| Date | Sales | Avg. Price | Holiday | Region | Product |
|---|---|---|---|---|---|
| JAN2011 | 355 | 25.3 | 0 | Region 1 | Product 1 |
| FEB2011 | 398 | 25.3 | 0 | Region 1 | Product 1 |
| … | … | … | … | … | … |
| JAN2011 | 555 | 19.8 | 0 | Region 1 | Product 2 |
| FEB2011 | 390 | 25.3 | 0 | Region 1 | Product 2 |
| … | … | … | … | … | … |
| JAN2011 | 301 | 27.1 | 0 | Region 2 | Product 1 |
| FEB2011 | 350 | 25.3 | 0 | Region 2 | Product 1 |
| … | … | … | … | … | … |
| JAN2011 | 314 | 27.2 | 0 | Region 2 | Product 2 |
| FEB2011 | 388 | 25.3 | 0 | Region 2 | Product 2 |
| … | … | … | … | … | … |
| DEC2011 | 518 | 20.3 | 1 | Region 2 | Product 2 |

## Including Future Input Values in the Input Data Source

To include future input (independent) values for the time series in your project, include these future values in the input data set. For example, suppose that you are running a book store. For the next three months, the publisher will sell you copies of a book for $20. Because you know this future input value for the series, you want to use

the values from your input data set rather than have SAS Visual Forecasting generate a forecast. Future values for dependent variables must be set to missing.

If future independent variable values are not provided, SAS Visual Forecasting extends the future independent variable series using a smoothing model.

# Understanding Time Intervals

## Overview

All time intervals 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 $t_i$.

SAS Visual Forecasting analyzes the variable assigned to the time variable role to detect the time interval of the data. SAS assumes that all of the values in the time variable are either date or datetime values and distinguishes between the values by their magnitude.

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.

***Table 4.3*** *Equally Spaced Time Intervals*

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

SAS Visual Forecasting 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 Visual Forecasting supplies the missing values. A validation routine checks the values of the time variable to determine whether they are spaced according to the interval that you specified.

## Specifying Time Intervals

Statistical forecasting requires that the historical input data be organized as time series, where successive measurements are taken at points equally spaced in time. SAS Visual Forecasting satisfies this requirement in three ways:

- The input data is already in time series form, and is used without any alteration in the time dimension.

- The input data is already in time series form, but is aggregated over time to a different interval (lower frequency), which you specify.
- The input data is transactional (recorded at no particular interval) and is aggregated over time to form time series at an interval which you specify.

Define the time interval specifications when you first create your project using the Assign Variables dialog box. For more information, see "Assigning Variable Roles" on page 5.

SAS Visual Forecasting represents time intervals using a very flexible notation based on a set of basic intervals and optional settings. These settings include the multiplier (affects the length of the interval) and the shift (affects the starting point of the interval). For most time series, the start, length, and end of every observation is identifiable once the time interval is specified in this way. See "Time Variable" on page 62 for a description of each setting.

For the time variable, SAS Visual Forecasting detects the best default settings for the time interval, multiplier, and shift. They are determined by analysis of the time variable values. For input data that are already in time series form, you can often proceed with these default values. You can change them if you know the interval and disagree with the defaults. You can choose any time interval specification that is of lower frequency than the default setting. For transactional data, choose an interval specification in which most or all time periods encompass times at which transactions were recorded.

The examples in the following table show how the values that you specify for the interval, multiplier, and shift work together. For the interval name using SAS code, the `m.s` suffix indicates the multiplier and shift settings for that time interval. In the examples shown, YEAR2.7 indicates a time interval set to YEAR, multiplier set to 2, and shift set to 7.

***Table 4.4*** *Time Interval, Multiplier, and Shift Examples*

| Interval Name (in SAS code format) | Default Starting Point | Shift Period | Example |
|---|---|---|---|
| YEARm.s | January 1 | Months | YEAR2.7 specifies an interval of every two years with the first month of the year starting in July. |
| SEMIYEARm.s | January 1 July 1 | Months | SEMIYEAR.3 specifies six-month intervals spanning March-August and September-February. |
| QTRm.s | January 1 April 1 July 1 October 1 | Months | QTR.3 specifies three-month intervals starting on March 1, June 1, September 1, and December 1. |
| SEMIMONTHm.s | First and 16th of each month | Semimonthly periods | SEMIMONTH2.2 specifies intervals from the 16th of one month through the 15th of the next month. |
| MONTHm.s | First of each month | Months | MONTH2.2 specifies February-March, April-May, June-July, August-September, October-November, and December-January of the following year. |

| Interval Name (in SAS code format) | Default Starting Point | Shift Period | Example |
|---|---|---|---|
| TENDAYm.s | First, 11th, and 21st of each month | Ten-day periods | TENDAY4.2 specifies four 10-day periods starting at the second 10-day period. |
| WEEKm.s | Each Sunday | Days (1=Sunday . . . 7=Saturday) | WEEK6.3 specifies six-week intervals starting on Tuesdays. |
| DAYm.s | Each day | Days | DAY3 specifies three-day intervals starting on Sunday. |
| HOURm.s | Start of the day (midnight) | Hours | HOUR8.7 specifies eight-hour intervals starting at 6:00 a.m., 2:00 p.m., and 10:00 p.m. |

# Understanding Hierarchy Reconciliation

When data is organized in a hierarchical fashion, there are often accounting constraints that link the data at different levels of the hierarchy. Typically, for any historical time period, the data in a parent node is either the sum or the average of the data of its child nodes. For example, the total sales of a product by a retail company are the sum of the sales of the same product in all stores that belong to the company. With forecast data, however, time series are often forecast independently at different levels.

As a result, the forecast values do not abide by the constraints that bind the original series. However, you can enforce these constraints by using an after-the-fact process known as reconciliation of hierarchical forecasts. In the hierarchical model, you can select from these reconciliation methods:

Top-down
: aggregates the data from the lowest levels in the forecast and then uses these values to generate the forecasts at the highest level. SAS Visual Forecasting then uses this forecast and the disaggregation method that you specified to reconcile the forecasts for lower levels in the hierarchy. The top-down method enables you to remove the excessive noise from the data at the lower levels of the hierarchy. However, you also might lose some components of the pattern (such as the seasonality) in the forecast.

  You can also specify the type of loss function for top-down reconciliation:

  - **Difference** - bases the loss function on the root mean square error (RMSE), which results in adjustments that are the (possibly weighted) mean difference of the aggregated child nodes and the parent node.
  - **Proportions** - uses a loss function that results in reconciled forecasts that are the (possibly weighted) proportional disaggregation of the parent node.

Bottom-Up
: uses the data at the lowest level of the hierarchy to generate the forecasts. These forecasts are then used to reconcile the forecasts for the higher levels in the hierarchy.

  The bottom-up method enables you to see any patterns (such as seasonality) in the data. However, because you are using the lowest level of the hierarchy as a reference level for the forecasts, you can also have too much noise or randomness in the data. Also, these forecasts might fail because the data at the lowest level of the hierarchy can be sporadic or too sparse.

Middle-out
: aggregates the data from the lower levels and then uses these values to generate the forecasts for the middle level. SAS Visual Forecasting uses the forecasts at the middle level to reconcile the forecasts for both the higher and lower levels. Some hierarchies have more than one middle level, so you need to specify the level that you want to use.

The following example shows how the forecasts are generated for the hierarchy Region > Product Category > Product Line > Product, based on the reconciliation method that you choose. Aggregation forecasts for the higher levels in the hierarchy are created based on the aggregation statistic that you select. Disaggregation forecasts for the lower levels in the hierarchy are created based on the disaggregation method that you select.

***Table 4.5*** *Reconciliation Methods*

| | **Reconciliation Method** | | | |
|---|---|---|---|---|
| **Hierarchy Level** | **Top Down** | **Middle Out – Product Category** | **Middle Out – Product Line** | **Bottom Up** |
| Region | Forecast | Aggregation Forecast | Aggregation Forecast | Aggregation Forecast |
| Product Category | Disaggregation Forecast | Forecast | Aggregation Forecast | Aggregation Forecast |
| Product Line | Disaggregation Forecast | Disaggregation Forecast | Forecast | Aggregation Forecast |
| Product | Disaggregation Forecast | Disaggregation Forecast | Disaggregation Forecast | Forecast |

# 5

# Reference

# Assign Variable Roles

Use the Data tab to define the roles for the variables in your data source that are required for any forecasting project. The role assignments are committed after you run one or more pipelines and then open the Overrides tab. After the role assignments are committed, you can no longer change these assignments for this project.

The following fields are provided to provide specific settings for each role assignment.

## Time Variable

A time variable is required for all forecasting projects. See "Understanding Time Intervals" on page 57 for more information about the other settings for the time variable.

**Time variable**
If a single candidate time variable is detected in your data source, it is automatically selected. If multiple candidates are detected, you must choose one.

**Time interval**
Select the time interval that you want to use to collect the data. 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. Smaller intervals generate a large number of observations and should be avoided. Each time interval has its own **Shift** interval setting and some have specific starting subintervals. For example, by default, a weekly time interval starts on Sunday (**Shift** = 1), which you can change to another starting day of the week using the **Shift** setting.

Choose from the following options:

**Second**
specifies second intervals. The shift interval is in seconds.

**Minute**
specifies minute intervals. The shift interval is in minutes.

**Hour**
specifies hourly intervals. The shift interval is in hours.

**Day**
specifies daily intervals. The shift interval is in days.

**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. There are five weekday intervals in the default calendar week: Monday, Tuesday, Wednesday, Thursday, and the three-day period Friday, Saturday, and Sunday. The shift interval is in weekdays.

- **Weekend** This button is enabled if you select **Weekday**. Click **Weekend** if you need to define the days of the week as the weekend. The defaults are Saturday and Sunday.

**Week**
specifies weekly intervals of seven days. The shift interval is days with the starting subinterval on Sunday.

The days of the week are numbered by the **Shift** setting as follows:

*Table 5.1 Number Values for Days of the Week*

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

**ISO 8601 week**
specifies ISO 8601 weekly intervals of seven days. The shift interval is days with the starting subinterval on Monday.

**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. The shift interval is 10-day periods.

**Semimonth**
specifies semimonthly intervals. SEMIMONTH breaks each month into two periods, starting on the first and 16th days. The shift period is in semimonthly intervals.

**Month**
specifies monthly intervals. The shift interval is in 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. The shift interval is in retail 4-4-5 months.

**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. The shift interval is in retail 4-5-4 months.

**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. The shift interval is in retail 5-4-4 months.

**Quarter**
specifies quarterly intervals (every three months). The shift interval is in months.

**Retail 4-4-5 quarter**
specifies retail 4-4-5 quarterly intervals (every 13 ISO 8601 weeks). Some fourth quarters will contain a leap week. The shift interval 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 will contain a leap week. The shift interval 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 will contain a leap week. The shift interval is in retail 5-4-4 months.

**Semiyear**
specifies intervals every six months. The shift interval is in months.

**Year**
specifies yearly intervals. The shift interval is in months.

**ISO 8601 year**
specifies ISO 8601 yearly intervals. The ISO 8601 year starts on the Monday on or immediately preceding January 4 (or on January 4 if it is a Monday). 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. For more information about ISO weeks, see Technical Committee ISO/TC 154 (Processes, Data Elements, and Documents in Commerce, Industry, and Administration) (2004). The shift interval is in ISO 8601 weeks (WEEKV).

**Retail 4-4-5 year**
is the same as YEARV except that the shift interval is in retail 4-4-5 months.

**Retail 4-5-4 year**
is the same as YEARV except that the shift interval is in retail 4-5-4 months

**Retail 5-4-4 year**
is the same as YEARV except that the shift interval is in retail 5-4-4 months

**Multiplier**
Specify an integer that is used to multiply the time interval for customizing the interval. For example, if time interval is set to MONTH and multiplier is 2, then the interval used by the project is every two months. This value can be any positive number.

**Shift**
Specify an integer to shift the starting point for the day that the interval begins. Specify the starting point for the interval. By default, this value is 1. 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, if shift is set to 4 and the time interval is YEAR, this shifts the start of the time interval to the fourth month. In this case, a year is from April 1 through March 31 of the following year. The starting point is dependent on the interval.

For a list of default starting points for each interval, see Table 4.4 on page 58.

**Seasonal cycle length**
Specify the length of a season in the units set by the time interval. This value is populated automatically if SAS Visual Forecasting can determine the seasonal cycle length from the time 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 Visual Forecasting.

## Dependent Variable

If only a single candidate dependent variable is detected in your data source, it is automatically selected. If multiple candidates are detected, you must choose one.

**Dependent variable**
Specify the variable in your data set that you want to forecast. This variable must be numeric.

**Hierarchy aggregation**
Aggregation is the process of combining data from more than one time series to form a single series. Select the aggregation method that you want to use for all of the time series in each level of the hierarchy. The

dependent variable has fewer aggregation options than independent variables. The following examples explain when you might want to use an aggregation method:

- Your data set contains the sales for a group of products and you want to know the total sales for a category.
- Your data contains the price of each product and you want to know the average price for a product line.

Select from one of the following options.

**Sum of Values**
aggregates the vector values based on the number of values.

$$a = Q$$

**Average of Values**
aggregates the vector values based on the average of their values.

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

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

**Note:** Average of values applies only to hierarchical modeling. The plot and table in Overrides shows a dynamic sum of values of the time series captured by the selected filters.

**Time interval accumulation**
Accumulation combines data within the same time interval into a summary value for that time interval. For a complete description of the accumulation options, see "Time Interval Accumulation Settings" on page 66.

**Missing interpretation**
Once the data has been accumulated to form a time series, missing value interpretation is performed. If any time series contain missing values for the dependent or independent variables, you can specify how SAS Visual Forecasting should interpret these missing values. Choose from the following options:

**First**
Specifies that missing values are set to the first nonmissing value of all accumulated nonmissing values in the span of the series.

**Last**
Specifies that missing values are set to the last nonmissing value of all accumulated nonmissing values in the span of the series.

**Maximum**
Specifies that missing values are set to the maximum value of all accumulated nonmissing values in the span of the series

**Median**
Specifies that missing values are set to the median value of all accumulated nonmissing values in the span of the series

**Minimum**
Specifies that missing values are set to the minimum value of all accumulated nonmissing values in the span of the series

**Average**
Specifies that missing values are set to the average value of all accumulated nonmissing values in the span of the series.

**Missing**
Specifies missing values to remain missing. Use this option if a missing value indicates an unknown value.

**Next**
Specifies that missing values are set to the next period's accumulated nonmissing value. Missing values at the end of the accumulated series remain missing.

**Previous**
Specifies that missing values are set to the previous period's accumulated nonmissing value. Missing values at the beginning of the accumulated series remain missing.

**0**
Specifies missing values to be set to zero. This setting is often used for transactional data, because no recorded data usually implies no activity.

## BY Variables

Assign this role to one or more variables that should be used as the default attributes of the project. Use attributes to create filters to work with subsets of the project data. You can also order your BY variables for hierarchical forecasting.

After you run a pipeline with the hierarchy defined, you can no longer remove variables from the selected list.

For more information, see "Assigning the Default Attributes" on page 6.

The order of the default attributes is also used for hierarchical forecasting. For more information, see "Defining the Hierarchy" on page 7.

# Time Interval Accumulation Settings

Specify how you want the data to be accumulated within each time interval. Accumulation combines data within the same time interval into a summary value for that time period. Accumulation can be used in the following situations:

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

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:

**Average of Values**
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.

**Corrected Sum of Squares**

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

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

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

**Maximum of Values**

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

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

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

**Minimum of Values**

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

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

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

**Number of non-missing values**

accumulates the vector values based on the number of nonmissing values.

$$a = Q_N$$

**Number of Missing values**

accumulates the vector values based on the number of missing values.

$$a = Q_{NMISS}$$

**Standard Deviation of Values**

accumulates the vector values based on their standard deviation.

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

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

**Sum of Values**

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.

**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.

# Project Settings

Use the Project Settings dialog box to set properties that apply to the project as a whole. The Project Settings are available along the top of each tab by clicking .

Review the values selected in Project Settings before you run any pipelines for your project.

## Forecast Settings

These settings for your project should be carefully considered before running pipelines and specifying overrides. Any changes to these require all pipelines to be run again.

**Number of forecast periods (horizon)**

Specify the number of periods to forecast (also called the horizon). For example, if the time interval for this project is set to WEEK, specify 12 to generate 12 weeks of forecasts for the dependent variable.

**Confidence limit**

Specify the size of the confidence level for the forecasts. By default, this confidence level is 0.05, which is a 95% confidence limit.

**Champion selection criteria**

Choose the statistics of fit analysis that is used to determine the champion pipeline for this project. The statistics of fit are statistical values that are used to evaluate how well a forecasting model performs by comparing the actual data to the predictions. For a given forecast model that has been fitted to the time series data, the model should be checked or evaluated to see how well it fits or forecasts the data.

Each modeling node in a pipeline has its own model selection setting to determine the best model to use from within that node. The **Champion selection criteria** is used to select the best modeling node when multiple nodes are used within a pipeline, or to select the best pipeline when more than one pipeline is run.

For a full description of the statistics of fit options that are available, see “Model Selection Criteria” on page 88.

**Allow negative values for forecasts and overrides**

Select this option to allow negative values for forecasts and overrides. If you clear this check box, then any negative values in the forecast model are set to 0 and any overrides with negative values are removed.

If **Missing interpretation** for the dependent variable is set to 0, you could still see some negative values for the lower confidence limits (Forecast LCL) in the project time plots.

Changes to this setting invalidate any existing overrides that have been created. Changes also invalidate the pipelines.

**Note:** To allow negative overrides, you must also set the **Maximum decrease** percentage on the Overrides tab to greater than 100%.

## Reporting Settings

**Forecast exception boundaries**

This range is used to identify forecast values that are problematic. Set the boundary below and above the historical average. Values that fall outside of the boundaries that you set are identified as problematic.

## Overrides Settings

**Disaggregation method**

When you submit an override, SAS Visual Forecasting must reconcile the override with other forecasts in the same time period. Specify the default type of disaggregation method for reconciling overrides. Select one of the following methods:

**Proportional**
reconciles the forecast based on the proportion of forecast contribution.

**Equal**
reconciles the forecast based on equal split of the differences between aggregated forecast of the time series and their individual forecasts.

**Override filter warning threshold**

A large number of time series in an override can have adverse effects on performance. Use this setting to issue a warning message when an override contains an extremely high number of time series.

**Maximum percentage change of final forecast**

Specify the maximum percentage change that is allowed during reconciliation for final forecasts.

**Note:** Both values, **Maximum increase** and **Maximum decrease**, must be zero or greater.

## Output Library Settings

Browse through the libraries to specify the default location for storing project output data.

# Settings for Modeling Strategy Nodes

## Hierarchical Forecasting Settings

The following settings are available in the Options panel for the Hierarchical Forecasting model node.

### Task Settings

**Forecast task**

Specify how to forecast the project. These tasks must be run sequentially. For example, the Diagnose task can be run independently, but the Fit task requires that you run the Diagnose task first.

Choose from one of the following tasks:

- **Diagnose**: performs model selection, estimates parameters of the selected model, and produces forecasts. This is the default.
- **Fit**: estimates parameters for the models that you select and then forecasts. No model selection is performed. You must run the Diagnose task successfully before running the Fit task.
- **Forecast**: forecasts using model parameters estimates. You must run the Diagnose task successfully before running the Forecast task.
- **Select**: performs model selection from the models that you select, estimates parameters of the selected model, and produces forecasts. You must run the Diagnose task successfully before running the Select task.

- **Update**: estimates parameters from the models that you select and then forecasts. No model selection is performed. Update differs from Fit in that the estimated parameters are used as starting values in the estimation. You must run the Diagnose task successfully before running the Update task.

## Diagnostic Settings

Changes to these settings require that you rerun the pipeline.

**Intermittency test**
Turn this setting on to perform an intermittency test and use the IDM model for intermittent series

**Sensitivity for intermittency test**
Specify an integer greater than one. This setting is used to determine whether a time series is intermittent. If the demand interval is equal to or greater than this number, then the series is assumed to be intermittent.

**Seasonality test**
Turn this setting on to perform seasonality testing for the time series

**Sensitivity for seasonality test**
Specify the significance probability value to use in testing whether seasonality is present in the time series. The value must be between 0 and 1. A smaller value means that a stronger evidence of a seasonal pattern in the data is required before seasonal models are used to forecast the time series.

**Diagnose independent variable**
Select from the following options:

- **Transform**: specifies that a transformation is applied to the time series
- **Trend**: specifies that trend analysis is performed on the time series based on the setting for minimum nonmissing values. Incorporation of a trend is checked only for smoothing, UCM, and ARIMA models. For the smoothing case, only simple smoothing is a non-trend model. For UCM, the absence of a slope component qualifies it as a non-trend model. For ARIMA, there must be no differencing of the dependent variable for the time series forecasting engine to consider it as a non-trend model.
- **Both**: specifies that transformation and trend analysis is performed for the time series
- **None**: specifies that no transformation or trend analysis is performed for the time series

**Transformation**

**Functional transformation (dependent)**
specify the type of functional transformation:

- **Auto**: Selects between logarithmic or no transformation as determined by the model selection criteria
- **Box-Cox**: Box-Cox transformation
- **Log**: Logarithmic transformation
- **Logistic**: Logistic transformation
- **None**: No transformations are processed on the time series
- **Square root**: Square-root transformation

**Box-Cox parameter**
Specify a number for the exponent, lambda (λ), which varies from -5 to 5, exclusive. This setting is enabled only if **Functional transformation (dependent)** is set to `Box-Cox`.

**Forecast**
Specify the forecast method when functional transformation is enabled. Forecasts can be based on the mean or median. By default the mean value is provided. This setting is disabled if **Functional transformation (dependent)** is set to `None`.

**Specify the minimum number of observations for a non-mean model**
Specify a minimum value that a time series must meet to be fit using the models in the selection list. Time series that do not meet this minimum value are forecast as the mean of the observations in the series. This value must be greater than or equal to one. The default value is 2.

**Specify the minimum number of observations for a trend model**
Specify that a trend model is not fitted to any series with fewer nonmissing observations than the value specified. The value must be an integer that is greater than or equal to 1. The default value is 2. Trend models are not included for any series with fewer nonmissing observations than this value.

Incorporation of a trend is checked only for smoothing, UCM, and ARIMA models. For the smoothing case, only simple smoothing is a non-trend model. For UCM, the absence of a slope component qualifies it as a non-trend model. For ARIMA, there must be no differencing of the dependent variable for it to be considered as a non-trend model.

**Specify the minimum number of seasonal cycles for a seasonal model**
Seasonal models are not included for any series with fewer nonmissing observations than this value, multiplied by the seasonal length. Specify an integer greater than or equal to one.

## Model Generation Settings

Changes to these settings require that you rerun the pipeline. Make sure that at least one of the models are selected. Otherwise, the node will fail.

**ARIMAX**
Turn this setting on to include an ARIMAX model for diagnosis. With this setting on, you can set the control options for ARIMAX model parameter refinement. You can set the order in which events, inputs, or ARIMA components are included in the models. You can also set a significance level between 0 and 1.

The following settings are enabled when ARIMAX is turned on.

**Identification order for inputs and events**
Select one of the options for the order in which input or ARIMA components are included in the model.

**Refinement**
Specify ARIMA parameter refinement options. These options enable the refinement of insignificant parameters of the final model, identification of the factors to refine, and identification of the order of factors.

**Factor option**: Select from the available options to determine the order for diagnosing model components. For example, if you select **ARMA:INPUT**, ARMA coefficients are tested before input variable coefficients.

**Significance level**: Enter a number between 0 and 1 that specifies the cutoff value for refining all insignificant parameters

**ESM**
Turn this setting on to include an ESM model for diagnosis. With this setting on, specify how the best smoothing model candidate is chosen, either seasonal (**BESTS**), nonseasonal (**BESTN**), or both (**BEST**).

**UCM**
Turn this setting on to include an UCM model for diagnosis.

**External Models**
Turn this setting on to include an external model to use for diagnosis.

**`Models from an external data source`** - Enter the CAS library and table name for the external model, for example: `mycas.externalmodel`.

**Note:** If the external model uses a dummy data set, any independent variables declared in the dummy data set must use the same names as the independent variables in the project time series. Otherwise, the external model is not used.

**Number of levels to fit only the system-generated ESM models**

Specify an integer ranging from 1 to the number of levels in the project hierarchy. A bigger number means the higher level in the hierarchy. The default is 0.

For example, if the hierarchy is **regionName** ⇨ **productLine** ⇨ **productName**, then the valid value could be one of the following:

- 0 — not applicable
- 1 — apply for productName
- 2 — apply for the productLine and the productName
- 3 — apply for the regionName, the productLine, and the productName

**Model Combination**

Specify whether to combine the selected models other than the external ones

**Method for combination**

Specify the method for determining the combination weights used in the weighted average of the candidate forecasts in the combination list. Choose from one of these options:

- **Akaike weights using AICC values** — computes the combination weights based on corrected AIC weights. By default, all AICC scored candidate forecasts are combined. Frequently, there is considerable disparity between the weights because of the exponential weighting scheme, so additional arguments are provided to affect the scaling and to cull low-scoring candidates from consideration for computational efficiency.
- **Average** — computes the simple average of the forecasts selected for combination
- **Equality constrained least squares** — computes the combination weights based on a constrained least squares problem to minimize the $\ell_2$ norm of the combined forecast residuals subject to the constraint that the weights sum to 1.
- **Equality constrained, non-negative least squares** — computes the combination weights based on a constrained least squares problem to minimize the $\ell_2$ norm of the combined forecast residuals subject to the constraints that the weights sum to 1 and be nonnegative.
- **Least absolute deviations** — computes the combination weights that result from the ordinary least squares problem to minimize the $\ell_2$ norm of the combined forecast residuals.
- **Non-negative least squares** — is equivalent to `Equality constrained, non-negative least squares` except that the resulting combination weights are not constrained to summing up to 1.
- **Ranked weighting** — assigns weights by using the rank of the candidate forecasts when the combination is performed.
- **Scaled RMSE weights** — computes the combination weights based on the RMSE statistic of fit for the forecast contributors. The weights are normalized to sum to 1.
- **User-defined weights** — assigns weights by using the list of user-specified values.

**Type for the encompassing test**

specify the encompassing test type. The encompassing test attempts to eliminate from consideration forecasts that fail to add significant information to the final forecast. Select one of the following values.

- **HLN** — uses the Harvey-Leybourne-Newbold (HLN) test to estimate pairwise encompassing between candidate forecasts
- **NONE** — performs no encompassing tests
- **OLS** — uses an OLS-based regression test to estimate pairwise encompassing between candidate forecasts

**Encompass test parameters**

- **Significance level for the encompass test** — specifies the encompassing test significance level. The default value is 0.05.
- **Rank criterion** — specifies the forecast combination criterion (statistic of fit) to be used when ranking forecast candidates in the context of the model combination. For descriptions for each option, see "Descriptions of Statistics of Fit" on page 89.

**Missing value interpretation**
specifies a method for treating missing values in the forecast combination. In a given time slice across the combination ensemble, one or more combination contributors can have a missing value. This setting determines the treatment of those in the final combination for such time indices.

- **Rescale** — rescales the combination weights for the nonmissing contributors at each time index to sum to 1.
- **Missing** — generates a missing combined forecast at each time index with one or more missing contributors.

**How to compute the prediction error variance series**
specifies the method for computing the prediction error variance series. This series is used to compute the prediction standard error, which in turn is used to compute confidence bands on the combined forecast. Select from the following options.

- **DIAG** — computes the prediction error variance by assuming the forecast errors at time *t* are uncorrelated so that the simple diagonal form of $\Sigma_t$ is used. This is the default method for computing prediction error variance.
- **ESTCORR** — computes the prediction error variance by using estimates of $\rho_{i,j,t}$, the sample cross-correlation between $e_{i,t}$ and $e_{j,t}$ over the time span $t = 1,...,T$, where *t* denotes the last time index of the actual series $y_t$. Of course, this option implies that the error series $e_{i,t}$ and $e_{j,t}$ are assumed to be jointly stationary.

**In-sample missing percentage**
specifies a threshold for the percentage of missing forecast values in the combination estimation region that is used to exclude a candidate forecast from consideration in the final combination. By default, no missing percentage test is performed on candidate forecasts. If specified, the admissible range is 1 to 100.

**Horizon missing percentage**
specifies a threshold for the percentage of missing forecast values in the combination horizon used to exclude a candidate forecast from consideration in the final combination. By default, no horizon missing percentage test is performed on candidate forecasts. If specified, the admissible range is 1 to 100.

## Model Selection Settings

Changes to these settings require that you rerun the pipeline.

**Size of data to be used for holdout**
Enter an integer to be used as the size of the holdout sample. The actual holdout sample will be the minimum between this value and the **Percentage of data to be used for holdout**.

**Percentage of data to be used for holdout**
Enter a value between 0 and 100 to specify the percentage of the sample that will be used for the holdout sample. The actual holdout sample will be the minimum between this value and the **Size of data to be used for holdout**.

**Model selection criterion**
Choose the statistics of fit to use for selecting the best model. This setting overrides the **Model selection criteria** value selected in the Project Settings. For descriptions for each option, see "Descriptions of Statistics of Fit" on page 89.

### Forecast Settings

Changes to these settings require that you rerun the pipeline.

**Calculate statistics of fit over an out of sample range**
Specify an integer for the number of observations to remove before the end of the data to test the champion model.

**Extend the lead beyond the out of sample range**
Select `Yes` to extend the forecast horizon beyond the out of sample range.

### Reconciliation Settings

Setting the reconciliation level overrides the default setting selected when assigning the BY variable roles for the project. For more information, see "Assign Variable Roles" on page 62.

**Specify the reconciliation level**
Specify a positive integer that corresponds to the levels in your hierarchy, as follows:

- Set the value to 0 to perform top-down reconciliation.
- Specify an integer that matches the bottom level in the hierarchy to use bottom-up reconciliation. For example, if the hierarchy consists of two BY variables, specify 2 to perform bottom-up reconciliation.
- Specify an integer that corresponds to any middle level in the hierarchy to use middle-out reconciliation. For example, if the hierarchy consists of four BY variables, specify 2 to generate forecasts for that level and then reconcile forecasts for the upper and lower levels in the hierarchy.

If you leave this value blank, the reconciliation level set on the Data tab for the BY variables is used.

If the value provided does not match the hierarchy (for example, specifying 5 when the hierarchy is only 4 levels deep), then bottom-up reconciliation is performed.

For a description of these reconciliation methods, see "Understanding Hierarchy Reconciliation" on page 59.

**Disaggregation method during top-down disaggregation**
specifies the type of disaggregation method and type of loss function for top-down reconciliation. Select one of the following methods for top-down disaggregation:

`Proportions`
uses a loss function that results in reconciled forecasts that are the (possibly weighted) proportional disaggregation of the parent node.

`Difference`
bases the loss function on the root mean squared error (RMSE). This results in adjustments that are the (possibly weighted) mean difference of the aggregated child nodes and the parent node.

## Hierarchical Forecasting (Pluggable) Settings

The following settings are available in the Options panel for Hierarchical Forecasting (Pluggable).

**Set the forecast task**
Select one of the following options.

- **Diagnose**: performs model selection, estimates parameters of the selected model, and produces forecasts. This is the default.
- **Fit**: estimates parameters for the models that you select and then forecasts. No model selection is performed. You must run the Diagnose task successfully before running the Fit task.
- **Forecast**: forecasts using model parameters estimates. You must run the Diagnose task successfully before running the Forecast task.

- **Select**: performs model selection from the models that you select, estimates parameters of the selected model, and produces forecasts. You must run the Diagnose task successfully before running the Select task.
- **Update**: estimates parameters from the models that you select and then forecasts. No model selection is performed. Update differs from Fit in that the estimated parameters are used as starting values in the estimation. You must run the Diagnose task successfully before running the Update task.

**Specify whether to include ARIMAX model for diagnosis**
Turn this setting on to include an ARIMAX model for diagnosis.

**Specify whether to include ESM model for diagnosis**
Turn this setting on to include an exponential smoothing model (ESM) for diagnosis.

**Specify whether to include IDM model for diagnosis**
Turn this setting on to include an intermittent demand model (IDM) for diagnosis.

**Intermittency Test**
Specify an integer greater than one. This setting is used to determine whether a time series is intermittent. If the demand interval is equal to or greater than this number, then the series is assumed to be intermittent. This setting is disabled if the IDM model setting is turned off.

**Specify whether to include UCM model for diagnosis**
Turn this setting on to include a UCM model for diagnosis.

**Specify whether to combine the models for diagnosis**
Select `Yes` if you want to combine the selected models for diagnosis.

**Minimum number of observations for a non-mean model**
Specify a minimum value that a time series must meet to be fit using the models in the selection list. Time series that do not meet this minimum value are forecast as the mean of the observations in the series. This value must be greater than or equal to one. The default value is 2.

**Minimum number of observations for a trend model**
Specify that a trend model is not fitted to any series with fewer nonmissing observations than the value specified. The value must be an integer that is greater than or equal to 1. The default value is 2. Trend models are not included for any series with fewer nonmissing observations than this value.

Incorporation of a trend is checked only for smoothing, UCM, and ARIMA models. For the smoothing case, only simple smoothing is a non-trend model. For UCM, the absence of a slope component qualifies it as a non-trend model. For ARIMA, there must be no differencing of the dependent variable for it to be considered as a non-trend model.

**Minimum number of observations for a seasonal model**
Seasonal models are not included for any series with fewer nonmissing observations than this value, multiplied by the seasonal length. Specify an integer greater than or equal to one.

**Size of data to be used for holdout**
Enter an integer to be used as the size of the holdout sample. The actual holdout sample will be the minimum between this value and the **Percentage of data to be used for holdout**.

**Percentage of data to be used for holdout**
Enter a value between 0 and 100 to specify the percentage of the sample that will be used for the holdout sample. The actual holdout sample will be the minimum between this value and the **Size of data to be used for holdout**.

**Model selection criterion**
Choose the statistics of fit to use for selecting the best model. This setting overrides the **Model selection criteria** value selected in the Project Settings. For descriptions for each option, see "Descriptions of Statistics of Fit" on page 89.

**Calculate statistics of fit over an out of sample range**
Specify an integer for the number of observations to remove before the end of the data to test the champion model.

**Extend the lead beyond the out of sample range**
Select `Yes` to extend the forecast horizon beyond the out of sample range.

**Reconciliation level**
Specify a positive integer that corresponds to the levels in your hierarchy, as follows:

- Set the value to 0 to perform top-down reconciliation.
- Specify an integer that matches the bottom level in the hierarchy to use bottom-up reconciliation. For example, if the hierarchy consists of two BY variables, specify 2 to perform bottom-up reconciliation.
- Specify an integer that corresponds to any middle level in the hierarchy to use middle-out reconciliation. For example, if the hierarchy consists of four BY variables, specify 2 to generate forecasts for that level and then reconcile forecasts for the upper and lower levels in the hierarchy.

If you leave this value blank, the reconciliation level set on the Data tab for the BY variables is used.

For a description of these reconciliation methods, see "Understanding Hierarchy Reconciliation" on page 59.

## Multistage Forecasting Settings

The following settings are available in the Options panel for the Multistage Forecasting modeling node.

**Note:**

This modeling strategy requires a license for SAS Visual Data Mining and Machine Learning.

### General Settings

**Stage 1 high level**
Specifies an integer indicating the BY variable in the modeling hierarchy (for example, CATEGORY) to use as the high level for forecasts in stage 1. Use 0 as the top level node in the modeling hierarchy. If no value is specified for this parameter, no high-level forecasts in stage 1 are generated.

**Stage 1 low level**
Specifies an integer indicating the BY variable in the modeling hierarchy (for example, PRODUCT) to use as the lowest level for forecasts in stage 1. Use 0 as the top level node in the modeling hierarchy. By default, this parameter value is the BY variable that is the second from the last in the hierarchy.

**Calculate statistics of fit over an out of sample range**
Specify an integer for the number of observations to remove before the end of the data to test the champion model.

**Extend the lead beyond the out of sample range**
Select `Yes` to extend the forecast horizon beyond the out of sample range.

**Size of data to be used for holdout**
Enter an integer to be used as the size of the holdout sample. The actual holdout sample will be the minimum between this value and the **Percentage of data to be used for holdout**.

**Percentage of data to be used for holdout**
Enter a value between 0 and 100 to specify the percentage of the sample that will be used for the holdout sample. The actual holdout sample will be the minimum between this value and the **Size of data to be used for holdout**.

### Time Series Forecast Model Settings

**Transformation**
**Functional transformation (dependent)**
specify the type of functional transformation:

- **Auto**: Selects between logarithmic or no transformation as determined by the model selection criteria
- **Log**: Logarithmic transformation
- **Logistic**: Logistic transformation
- **None**: No transformations are processed on the time series
- **Square root**: Square-root transformation

**Forecast**
Specify the forecast method when functional transformation is enabled. Forecasts can be based on the mean or median. By default the mean value is provided. This setting is disabled if **Functional transformation (dependent)** is set to `None`.

**Specify whether to include ARIMAX model for diagnosis**
Turn this setting on to include an ARIMAX model for diagnosis.

**Specify whether to include ESM model for diagnosis**
Turn this setting on to include an exponential smoothing model (ESM) for diagnosis.

**Specify whether to include IDM model for diagnosis**
Turn this setting on to include an intermittent demand model (IDM) for diagnosis.

**Intermittency Test**
Specify an integer greater than one. This setting is used to determine whether a time series is intermittent. If the demand interval is equal to or greater than this number, then the series is assumed to be intermittent. This setting is disabled if the IDM model setting is turned off.

**Specify whether to include UCM model for diagnosis**
Turn this setting on to include a UCM model for diagnosis.

**Specify whether to combine the models for diagnosis**
Select `Yes` if you want to combine the selected models for diagnosis.

**Minimum number of observations for a non-mean model**
Specify a minimum value that a time series must meet to be fit using the models in the selection list. Time series that do not meet this minimum value are forecast as the mean of the observations in the series. This value must be greater than or equal to one. The default value is 2.

**Minimum number of observations for a trend model**
Specify that a trend model is not fitted to any series with fewer nonmissing observations than the value specified. The value must be an integer that is greater than or equal to 1. The default value is 2. Trend models are not included for any series with fewer nonmissing observations than this value.

Incorporation of a trend is checked only for smoothing, UCM, and ARIMA models. For the smoothing case, only simple smoothing is a non-trend model. For UCM, the absence of a slope component qualifies it as a non-trend model. For ARIMA, there must be no differencing of the dependent variable for it to be considered as a non-trend model.

**Minimum number of observations for a seasonal model**
Seasonal models are not included for any series with fewer nonmissing observations than this value, multiplied by the seasonal length. Specify an integer greater than or equal to one.

**Model selection criterion**
Choose the statistics of fit to use for selecting the best model. This setting overrides the **Model selection criteria** value selected in the Project Settings. For descriptions for each option, see "Descriptions of Statistics of Fit" on page 89.

## Feature Extraction Settings

**Feature extraction model**
Select either a **Regression** or **Neural network** model to extract the features.

**Dependent variable transformation**
Specify **Log** for a logarithmic transformation for the dependent variable or **None** for no transformation.

**Seasonal dummy variables**
Select this setting to generate seasonal dummy variables during feature extraction. The number of seasonal dummy variables corresponds to the **Seasonal cycle length** specified for the time variable.

**Independent variable lags**
Specify a positive integer for the number of independent variable lags to generate and include as independent variables. For example, setting the value to 3 computes three variables with lagged values for each independent variable defined in the project.

> **TIP** Setting this field equal to the dependent variable lags enables the neural network to better detect the interactions between the variables.

## Regression Model Settings

This setting is enabled only if **Feature extraction model** is set to **Regression**.

**By level**
Specify the level in the hierarchy for BY variables to use in the regression model during feature extraction. Specify an integer between 0 and the number of BY variables assigned for this project. If not specified, or if this is set to 0, there will be no BY statement in the regression model.

## Neural Network Model Settings

These settings are enabled only if **Feature extraction model** is set to **Neural network**.

**Dependent variable trend**
Specify the method to create a dependent variable trend as an independent variable. You can choose a **Linear** trend or **Damped trend**. If you select **None**, no trend variable is created.

**ESM Forecast of dependent variable**
Specify **Yes** to use an ESM forecast of the dependent variable as an independent variable.

**Dependent variable lags**
Specify a positive integer for the number of dependent variable lags to generate and include as independent variables. Dependent variable lags are required for the neural network to learn the order of the time series.

For lags of missing values in the horizon, the previous forecasted values will be used to generate new lag values and extend the forecast recursively. For example, setting the value to 3 computes three independent variables with lagged values of the dependent variable.

**Output layer direct connections**
Specify whether direct connections from nodes in the input layer to nodes in the output layer should be included in the neural network. By default, this is not selected. This setting is disabled if **Number of hidden layers** is 0.

**Number of hidden layers**
Specify 0, 1, or 2 hidden layers to include in the neural network model. If the number of hidden layers is 0, a GLIM model is trained.

**Layer one neurons**
Specify an integer between 0 and 100 for the number of neurons in the first hidden layer. This is required if **Number of hidden layers** is greater than 0.

**Hidden layer one activation function**
Specify the activation function for the first hidden layer. This is required if **Number of hidden layers** is greater than 0.

Select from these options:

- **Exponential**
- **Identity**
- **Rectifier**
- **Sine**
- **Tanh** (hyperbolic tangent)

**Layer two neurons**
Specify an integer between 0 and 100 for the number of neurons in first second layer. This is required if **Number of hidden layers** is 2.

**Hidden layer two activation function**
Specify the activation function for the second hidden layer. This is required if **Number of hidden layers** is 2.

Select from these options:

- **Exponential**
- **Identity**
- **Rectifier**
- **Sine**
- **Tanh** (hyperbolic tangent)

**Random seed**
Specify a positive integer to use for generating random numbers to initialize the network.

**Dependent variable activation function**
Specify the activation function to use on the output layer of the network. If **Error function** is not set to **Normal**, this setting is disabled and the **Exponential** function is used.

Specify the target layer activation function for interval targets. Select from these options:

- **Identity**
- **Sine**

**Input standardization**
Specify the method that is used to standardize the interval input variables. Select from these options:

- **Midrange**
- **Std**
- **None**

**Dependent variable standardization**
Specify the method that is used to standardize the dependent variable. Select from these options:

- **Midrange**
- **Std**
- **None**

**Error function**
Specify the error function for the dependent variable output layer. Select one of the following options:

- **Gamma** — Selecting this value disables the **Dependent variable activation function**. The Exponential activation function is used with the Gamma error function.
- **Normal** — When there are no hidden layers, the normal error function is used.
- **Poisson** — Selecting this value disables the **Dependent variable activation function**. The Exponential activation function is used with the Poisson error function.

**Maximum training iterations**
Specify the maximum number of training iterations within each try.

## Panel Series Neural Network Settings

The following settings are available in the Options panel for the Panel Series Neural Network modeling strategy node.

**Note:**

This modeling strategy requires a license for SAS Visual Data Mining and Machine Learning.

### Task Settings

**Forecast Task**
Choose from the following tasks:

- **Diagnose**: trains a neural network, computes the results, forecasts the time series, and sends the forecasts to output. This is the default.
- **Fit**: retrains a model with updated data. Parameters required for the fit task are copied from the most recently trained network when the Fit task is run. This includes variable names (inputs, nominals and target), activation functions, and the number of hidden layers and neurons. You must run the Diagnose task successfully before running the Fit task.
- **Forecast**: generates forecasts using input model data. You must run the Diagnose task successfully before running the Forecast task.
- **Update**: retrains a model with updated data. Hyperparameters and initial connection weights are copied from the most recently trained neural network. You must run the Diagnose task successfully before running the Update task.

### Feature Generation Settings

To get a good model, enable some of the **Feature Generation** settings, which includes setting the number of dependent and independent variable lags. The generated features are included as independent variables when training the model.

**Dependent variable lags**
Specify a positive integer for the number of dependent variable lags to generate and include as independent variables. Dependent variable lags are required for the neural network to learn the order of the time series.

For lags of missing values in the horizon, the previous forecasted values will be used to generate new lag values and extend the forecast recursively. For example, setting the value to 3 computes three independent variables with lagged values of the dependent variable.

**Independent variable lags**
Specify a positive integer for the number of independent variable lags to generate and include as independent variables. For example, setting the value to 3 computes three variables with lagged values for each independent variable defined in the project.

> **TIP** Setting this field equal to the dependent variable lags enables the neural network to better detect the interactions between the variables.

**Seasonal dummy variables**
Select this setting to generate seasonal dummy variables during feature extraction. The number of seasonal dummy variables corresponds to the **Seasonal cycle length** specified for the time variable.

**ESM Forecast of dependent variable**
Specify **Yes** to use an ESM forecast of the dependent variable as an independent variable.

**Dependent variable trend**
Specify the method to create a dependent variable trend as an independent variable. You can choose a **Linear** trend or **Damped trend**. If you select **None**, no trend variable is created.

## Model Generation Settings

### Model Initialization Settings

**Input standardization**
Specify the method that is used to standardize the interval input variables. Select from these options:

- **Midrange**
- **Std**
- **None**

**Number of hidden layers**
Specify 0, 1, or 2 hidden layers to include in the neural network model. If the number of hidden layers is 0, a GLIM model is trained.

**Layer one neurons**
Specify an integer between 0 and 100 for the number of neurons in the first hidden layer. This is required if **Number of hidden layers** is greater than 0.

**Hidden layer one activation function**
Specify the activation function for the first hidden layer. This is required if **Number of hidden layers** is greater than 0.

Select from these options:

- **Exponential**
- **Identity**
- **Rectifier**
- **Sine**
- **Tanh** (hyperbolic tangent)

**Layer two neurons**
Specify an integer between 0 and 100 for the number of neurons in first second layer. This is required if **Number of hidden layers** is 2.

**Hidden layer two activation function**
Specify the activation function for the second hidden layer. This is required if **Number of hidden layers** is 2.

Select from these options:

- **Exponential**
- **Identity**
- **Rectifier**
- **Sine**
- **Tanh** (hyperbolic tangent)

**Output layer direct connections**
Specify whether direct connections from nodes in the input layer to nodes in the output layer should be included in the neural network. By default, this is not selected. This setting is disabled if **Number of hidden layers** is 0.

**Dependent variable transformation**
Specify **Log** for a logarithmic transformation for the dependent variable or **None** for no transformation.

**Dependent variable standardization**
Specify the method that is used to standardize the dependent variable. Select from these options:

- **Midrange**
- **Std**
- **None**

**Error function**
Specify the error function for the dependent variable output layer. Select one of the following options:

- **Gamma** — Selecting this value disables the **Dependent variable activation function**. The Exponential activation function is used with the Gamma error function.
- **Normal** — When there are no hidden layers, the normal error function is used.
- **Poisson** — Selecting this value disables the **Dependent variable activation function**. The Exponential activation function is used with the Poisson error function.

**Dependent variable activation function**
Specify the activation function to use on the output layer of the network. If **Error function** is not set to **Normal**, this setting is disabled and the **Exponential** function is used.

Specify the target layer activation function for interval targets. Select from these options:

- **Identity**
- **Sine**

**Neuron connection distribution**
Specify the distribution of randomly generated initial neuron connection weights. Select from these options:

- **Cauchy**
- **MSRA**
- **Normal**
- **Uniform**
- **Xavier**

**Random seed**
Specify a positive integer to use for generating random numbers to initialize the network.

### Model Training Settings

**Algorithm**
Specify the optimization method used to train the neural network. Select one of the following options:

- **LBFGS** - limited memory version of Broyden-Fletcher-Goldfarb-Shanno (BFGS)
- **SGD** - Stochastic gradient descent

**SGD Options**
If you select the SGD algorithm, the following settings apply.

**Learning rate**
Specify the learning rate parameter for SGD optimization. The default value is 0.001.

**Annealing rate**
Specify the annealing rate parameter for SGD optimization. The default value is 0.000001.

**SGD seed**
Specify the random seed to use for the SGD algorithm.

**Input layer dropout ratio**
Specify the dropout ratio for the input layer when SGD optimization is used. The default value is 0. Enter a nonnegative number that is less than 1.

**Hidden layer dropout ratio**
Specify the dropout ratio for the hidden layers when SGD optimization is used. The default value is 0. Enter a nonnegative number that is less than 1.

**Number of tries**
Specify the number of times to train the network with different initial estimates for connection weights. The network with the smallest error is chosen as the optimal network.

**Maximum training iterations**
Specify the maximum number of training iterations within each try.

**Maximum time (minutes)**
Specify in minutes the maximum time allowed for each try. Training continues until the all tries have completed. Zero indicates not to use time to limit the training.

**L1 regularization**
Specify a positive number for the weight decay for L1 regularization. The default is 0.

**L2 regularization**
Specify a positive number for the weight decay for L2 regularization. The default is 0.1.

**Enable early stopping**
Select this option to stop training when the model begins to overfit. The training stops after a number of consecutive iterations without improvement in the holdout region. Set the number of consecutive iterations in **Stagnation limit for early stopping**.

**Stagnation limit for early stopping**
Specify the number of consecutive iterations without improvement in validation error before stopping the optimization. Specifying 0 has the same effect as deselecting **Enable early stopping**.

This field is available only when **Enable early stopping** is selected.

### Autotune Settings

**Enable Autotune** — Turn this setting on to enable autotuning of neural network parameters. Autotuning is applied only when running the **Diagnose** task.

**Note:** Autotuning can significantly increase the amount of time required to train the model.

**Hidden layer tuning**
Specify whether to autotune the number of hidden layers. If you enable this setting for autotuning, specify the initial value and the lower and upper bounds for the number of hidden layers. This setting overrides **Number of hidden layers** in the **Model Initialization** settings.

**Hidden neuron tuning**
Specify whether to autotune the number of neurons in each hidden layer. Specify the initial value and the lower and upper bounds for the number of neurons. This setting overrides **Number of hidden layers** in the **Model Initialization** settings.

This setting is enabled if **Hidden layer tuning** is enabled.

**L1 regularization**
Specify whether to autotune the weight decay for the L1 regularization parameter. If you enable this setting for autotuning, specify the initial value and the lower and upper bounds for the weight decay. This setting overrides **L1 regularization** in **Model Training** settings.

**L2 regularization**
Specify whether to autotune the weight decay for the L2 regularization parameter. If you enable this setting for autotuning, specify the initial value and the lower and upper bounds for the weight decay. This setting overrides **L2 regularization** in **Model Training** settings.

**Learning rate**
Specify whether to autotune the learning rate for the hidden layers. If you enable this setting for autotuning, specify the initial value and the lower and upper bounds for the learning rate. This setting overrides **Learning Rate** in the **Model Training** settings.

This setting is disabled if the **Algorithm** under **Model Training** is set to LBFGS.

**Annealing rate**
Specify whether to autotune the annealing rate for the hidden layers. If you enable this setting for autotuning, specify the initial value and the lower and upper bounds for the learning rate. This setting overrides **Annealing Rate** in the **Model Training** settings.

This setting is disabled if the **Algorithm** under **Model Training** is set to LBFGS.

**Maximum autotuning iterations**
Specify the maximum number of iterations for autotuning. Autotuning continues until all iterations are completed or the **Maximum autotuning time (minutes)** is reached, whichever is first.

**Maximum autotuning time (minutes)**
Specify the maximum time allowed for autotuning in minutes. Autotuning continues until this time is reached or the **Maximum autotuning iterations** are completed, whichever is first.

**Autotune seed**
Specify the random seed for the autotuning.

**Autotune objective function**
Specify the objective function to optimize when tuning parameters. You can select one of these options:

- **ASE** — Average squared error
- **MAE** — Mean absolute error
- **MSE** — Mean squared error
- **MSLE** — Mean squared logarithmic error
- **RASE** — Root average squared error
- **RMAE** — Root mean absolute error
- **RMSLE** — Root mean squared logarithmic error

## Model Validation Settings

The holdout sample is the minimum calculated value between these two settings.

**Validation observation number**
Specify an integer to be used as the size of the holdout sample.

**Validation observation percent**
Specify a value between 0 and 100 indicating the percentage to be used as the holdout sample.

### Forecast Properties

**Calculate statistics of fit over an out of sample range**
Specify an integer for the number of observations to remove before the end of the data to test the champion model.

## Stacked Model (NN + TS) Forecasting Settings

The following settings are available in the Options panel for the Stacked Model (NN + TS) Forecasting modeling node.

**Note:**

This modeling strategy requires a license for SAS Visual Data Mining and Machine Learning.

### Feature Generation Settings

**Dependent variable lags**
Specify a positive integer for the number of dependent variable lags to generate and include as independent variables. Dependent variable lags are required for the neural network to learn the order of the time series.

For lags of missing values in the horizon, the previous forecasted values will be used to generate new lag values and extend the forecast recursively. For example, setting the value to 3 computes three independent variables with lagged values of the dependent variable.

**Independent variable lags**
Specify a positive integer for the number of independent variable lags to generate and include as independent variables. For example, setting the value to 3 computes three variables with lagged values for each independent variable defined in the project.

> **TIP** Setting this field equal to the dependent variable lags enables the neural network to better detect the interactions between the variables.

**Seasonal dummy variables**
Select this setting to generate seasonal dummy variables during feature extraction. The number of seasonal dummy variables corresponds to the **Seasonal cycle length** specified for the time variable.

**ESM Forecast of dependent variable**
Specify **Yes** to use an ESM forecast of the dependent variable as an independent variable.

**Dependent variable trend**
Specify the method to create a dependent variable trend as an independent variable. You can choose a **Linear** trend or **Damped trend**. If you select **None**, no trend variable is created.

### Model Generation Settings

#### Model Initialization

**Input standardization**
Specify the method that is used to standardize the interval input variables. Select from these options:

- **Midrange**
- **Std**
- **None**

**Number of hidden layers**
Specify 0, 1, or 2 hidden layers to include in the neural network model. If the number of hidden layers is 0, a GLIM model is trained.

**Layer one neurons**
Specify an integer between 0 and 100 for the number of neurons in the first hidden layer. This is required if **Number of hidden layers** is greater than 0.

**Hidden layer one activation function**
Specify the activation function for the first hidden layer. This is required if **Number of hidden layers** is greater than 0.

Select from these options:

- **Exponential**
- **Identity**
- **Rectifier**
- **Sine**
- **Tanh** (hyperbolic tangent)

**Layer two neurons**
Specify an integer between 0 and 100 for the number of neurons in first second layer. This is required if **Number of hidden layers** is 2.

**Hidden layer two activation function**
Specify the activation function for the second hidden layer. This is required if **Number of hidden layers** is 2.

Select from these options:

- **Exponential**
- **Identity**
- **Rectifier**
- **Sine**
- **Tanh** (hyperbolic tangent)

**Output layer direct connections**
Specify whether direct connections from nodes in the input layer to nodes in the output layer should be included in the neural network. By default, this is not selected. This setting is disabled if **Number of hidden layers** is 0.

**Dependent variable transformation**
Specify **Log** for a logarithmic transformation for the dependent variable or **None** for no transformation.

**Dependent variable standardization**
Specify the method that is used to standardize the dependent variable. Select from these options:

- **Midrange**
- **Std**
- **None**

**Error function**
Specify the error function for the dependent variable output layer. Select one of the following options:

- **Gamma** — Selecting this value disables the **Dependent variable activation function**. The Exponential activation function is used with the Gamma error function.
- **Normal** — When there are no hidden layers, the normal error function is used.

- **Poisson** — Selecting this value disables the **Dependent variable activation function**. The Exponential activation function is used with the Poisson error function.

**Dependent variable activation function**
Specify the activation function to use on the output layer of the network. If **Error function** is not set to **Normal**, this setting is disabled and the **Exponential** function is used.

Specify the target layer activation function for interval targets. Select from these options:

- **Identity**
- **Sine**

**Neuron connection distribution**
Specify the distribution of randomly generated initial neuron connection weights. Select from these options:

- **Cauchy**
- **MSRA**
- **Normal**
- **Uniform**
- **Xavier**

**Random seed**
Specify a positive integer to use for generating random numbers to initialize the network.

### Model Training Settings

**Algorithm**
Specify the optimization method used to train the neural network. Select one of the following options:

- **LBFGS** - limited memory version of Broyden-Fletcher-Goldfarb-Shanno (BFGS)
- **SGD** - Stochastic gradient descent

**SGD Options**
If you select the SGD algorithm, the following settings apply.

**Learning rate**
Specify the learning rate parameter for SGD optimization. The default value is 0.001.

**Annealing rate**
Specify the annealing rate parameter for SGD optimization. The default value is 0.000001.

**SGD seed**
Specify the random seed to use for the SGD algorithm.

**Input layer dropout ratio**
Specify the dropout ratio for the input layer when SGD optimization is used. The default value is 0. Enter a nonnegative number that is less than 1.

**Hidden layer dropout ratio**
Specify the dropout ratio for the hidden layers when SGD optimization is used. The default value is 0. Enter a nonnegative number that is less than 1.

**Number of tries**
Specify the number of times to train the network with different initial estimates for connection weights. The network with the smallest error is chosen as the optimal network.

**Maximum training iterations**
Specify the maximum number of training iterations within each try.

**Maximum time (minutes)**
Specify in minutes the maximum time allowed for each try. Training continues until the all tries have completed. Zero indicates not to use time to limit the training.

**L1 regularization**
Specify a positive number for the weight decay for L1 regularization. The default is 0.

**L2 regularization**
Specify a positive number for the weight decay for L2 regularization. The default is 0.1.

**Enable early stopping**
Select this option to stop training when the model begins to overfit. The training stops after a number of consecutive iterations without improvement in the holdout region. Set the number of consecutive iterations in **Stagnation limit for early stopping**.

**Stagnation limit for early stopping**
Specify the number of consecutive iterations without improvement in validation error before stopping the optimization. Specifying 0 has the same effect as deselecting **Enable early stopping**.

This field is available only when **Enable early stopping** is selected.

#### Model Selection Options

**Model selection criterion for the second stage model**
Specify the model selection criteria for the second stage. For a description of each option, see "Descriptions of Statistics of Fit" on page 89.

# Model Selection Criteria

## What Are Statistics of Fit?

The statistics of fit are statistical values that are used to evaluate how well a forecasting model performs by comparing the actual data to the predictions. For a given forecast model that has been fitted to the time series data, the model should be checked or evaluated to see how well it fits or forecasts the data. Commonly used statistics of fit are Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE), Akaike Information Criteria (AIC), and many others. The statistics of fit can be computed from the model residuals or the prediction errors.

When the full range of data is used to both fit and evaluate the model, this is referred to as in-sample evaluation. When the most recent data is excluded for parameter estimation (holdout) and this holdout sample is used for evaluation, this is referred to as holdout sample evaluation. Holdout sample analysis is similar to the training and testing of neural networks. A portion of the data is withheld from training (fit) and the withheld data (holdout) is used to test performance.

When a particular statistic of fit is used for forecast model selection, it is referred to as the model selection criterion. For example, if the MAPE (an often recommended choice) is used as a model selection criterion, the forecast model with the smallest MAPE in the evaluation region (in-sample or holdout-sample) is chosen as the best model.

When a particular statistic of fit is used to judge how well the forecasting process is predicting the future, it is referred to as the performance statistic.

## Understanding the Formulas and Definitions

To understand the formulas and definitions for the statistics of fit, review the following information:

- $n$ is the number of nonmissing observations.

- $k$ is the number of fitted parameters in the model.
- $APE = |100 * (y_t - \hat{y}_t)/y_t|$ is the absolute percent error.
- $ASPE = |100 * (y_t - \hat{y}_t)/0.5(y_t + \hat{y}_t)|$ is the absolute symmetric percent error.
- $APPE = |100 * (y_t - \hat{y}_t)/\hat{y}_t|$ is the absolute predictive percent error.
- $RAE = |(y_t - \hat{y}_t)/(y_t - y_{t-1})|$ is relative absolute error.

## Descriptions of Statistics of Fit

The definitions and formulas for the statistics of fit that are available in SAS Visual Forecasting are described below. You can use statistics of fit to measure how well different models fit the data. The statistics of fit for the various forecasting models can be printed or stored in a data set. Choose from one of the following options.

**Adjusted R-square** (ADJRSQ)

The adjusted $R^2$ statistic, $1 - (\frac{n-1}{n-k})(1 - R^2)$

**Akaike information Corrected criterion** (AICC)

Akaike's information criterion with an empirical correction for small sample sizes, $AIC + (\frac{2k(k+1)}{n-k-1})$

**Akaike information criterion** (AIC)

Akaike's information criterion, $n \quad \ln(\frac{SSE}{n}) + 2k$

**Amemiya's adjusted R-square** (AADJRSQ)

Amemiya's adjusted $R^2$, $1 - (\frac{n+k}{n-k})(1 - R^2)$

**Amemiya's prediction criterion** (APC)

Amemiya's prediction criterion, $\frac{1}{n}SST(\frac{n+k}{n-k})(1 - R^2) = (\frac{n+k}{n-k})\frac{1}{n}SSE$

**Geometric mean absolute error percent of standard deviation** (GMAPES)

The geometric mean of the absolute error as a percentage of the standard deviation

**Geometric mean percent error** (GMAPE)

The geometric mean percent error

**Geometric mean predictive percent error** (GMAPPE)

The geometric mean absolute predictive percent prediction error

**Geometric mean relative absolute error** (GMRAE)

The geometric mean of the relative absolute errors

**Geometric mean symmetric percent error** (GMASPE)

The geometric mean of the absolute symmetric percent errors

**Maximum Absolute Error Percent of Standard Deviation** (MAXAPES)

The maximum of the absolute error as a percentage of the standard deviation

**Maximum error** (MAXERR)

The largest prediction error

**Maximum percent error** (MAXPE)

The largest percent prediction error, $100\max(\frac{(y_t - \hat{y}_t)}{y_t})$. The summation ignores observations where $y_t = 0$.

**Maximum predictive percent error** (MAXPPE)

The maximum of the predictive percent errors

**Maximum relative error** (MAXRE)
The maximum of the relative errors

**Maximum symmetric percent error** (MAXSPE)
The maximum of the symmetric percent errors

**Mean absolute error** (MAE)
The mean absolute prediction error, $\frac{1}{n}\Sigma_{t=1}^{n}|y_t - \hat{y}_t|$.

**Mean absolute error percent of standard deviation** (MAPES)
The mean of the absolute error as a percentage of the standard deviation.

**Mean absolute percent error** (MAPE)
The mean of the absolute percent errors.

**Mean absolute predictive symmetric percent error** (MAPPE)
The mean of the absolute symmetric predictive percent error.

**Mean absolute scaled error** (MASE)
The mean of the absolute scaled errors, where $MAE_{in-sample,\ naive}$ is the mean absolute error produced by a naive forecast calculated on the in-sample data:

$$\frac{MAE}{MAE_{in-sample,\ naive}}$$

For a non-seasonal series, the naïve forecast is generated by using the actual value from previous record. For a seasonal series, the naïve forecast is generated by using the actual value from previous season.

**Mean error** (ME)
The mean prediction error, $\frac{1}{n}\Sigma_{t=1}^{n}(y_t - \hat{y}_t)$

**Mean percent error** (MPE)
The mean percent prediction error, $\frac{1}{n}\Sigma_{t=1}^{n}\frac{(y_t - \hat{y}_t)}{y_t}$. The summation ignores observations where $y_t = 0$.

**Mean predictive percent error** (MPPE)
The mean of the predictive percent error.

**Mean relative absolute error** (MRAE)
The mean of the relative absolute errors.

**Mean relative error** (MRE)
The mean of the relative errors.

**Mean squared error** (MSE)
The mean squared prediction error calculated from the one-step-ahead forecasts, $MSE = \frac{1}{n}SSE$. This formula enables you to evaluate small holdout samples.

**Mean symmetric percent error** (MSPE)
The mean of the symmetric percent errors.

**Mean absolute symmetric percent error** (SMAPE)
The symmetric mean of the absolute percent error.

**Median absolute error percent of standard deviation** (MDAPES)
The median of the absolute error as a percentage of the standard deviation.

**Median percent error** (MDAPE)
The median of the percent errors.

**Median predictive percent error** (MDAPPE)
The median of the predictive percent errors.

**Median relative absolute error** (MDRAE)
The median of the relative absolute errors.

**Median symmetric percent error** (MDASPE)
The median of the symmetric percent errors.

**Minimum Absolute Error Percent of Standard Deviation** (MINAPES)
The minimum of the absolute error as a percentage of the standard deviation.

**Minimum error** (MINERR)
The smallest prediction error.

**Minimum percent error** (MINPE)
The smallest percent prediction error, $100\min(\frac{(y_t - \hat{y}_t)}{y_t})$. The summation ignores observations where $y_t = 0$.

**Minimum predictive percent error** (MINPPE)
The smallest predictive percent error.

**Minimum relative error** (MINRE)
The smallest relative error.

**Minimum symmetric percent error** (MINSPE)
The smallest symmetric percent error.

**R-square** (RSQUARE)
The $R^2$ statistic, $R^2 = 1 - \frac{SSE}{SST}$. If the model fits the series badly, the model sum of square error, *SSE*, might be larger than *SST* and the $R^2$ statistic will be negative.

**Random walk R-square** (RWRSQ)
The random walk $R^2$ statistic (Harvey's $R^2$ statistic using the random walk model for comparison), $1 - (\frac{n-1}{n})\frac{SSE}{RWSSE}$, where $RWSSE = \Sigma_{t=2}^{n}(y_t - y_{t-1} - \mu)^2$, and $\mu = \frac{1}{n-1}\Sigma_{t=2}^{n}(y_t - y_{t-1})$.

**Root mean squared error** (RMSE)
The root mean square error, $\sqrt{MSE}$.

**Schwarz Bayesian information criterion** (SBC)
Schwarz Bayesian information criterion, $n * \ln(\frac{SSE}{n}) + k * \ln(n)$.

**Sum of square error** (SSE)
The sum of the squared prediction errors. $SSE = \Sigma_{t=1}^{n}(y_t - \hat{y}_t)^2$, where $\hat{y}$ is the one-step predicted value.

**Unbiased mean squared error** (UMSE)
The unbiased mean squared error.

**Unbiased root mean squared error** (URMSE)
The unbiased root mean squared error.

# Weighted Model Comparison Selection Criteria

Set the selection criteria for choosing the champion pipeline on the Pipeline Comparison tab and for choosing the champion model within a pipeline.

- Use the Project Settings to set the selection criteria used for pipeline comparison.
- The Model Comparison node has a property to set the selection criteria for model comparison within the pipeline.

To understand the absolute percent error or APE, where *n* is the number of nonmissing observations and *k* is the number of fitted parameters in the model, the following formula applies:

- $APE = |100 * (y_t - \hat{y}_t)/y_t|$

**Note:** For the following statistics of fit, the final computation is based on weighted measurements from each time series, where more weight is given to time series with a higher average of the dependent variable.

**Mean absolute percent error** (MAPE)
The mean of the absolute percent errors.

**Mean absolute error** (MAE)
The mean absolute prediction error, $\frac{1}{n}\Sigma_{t=1}^{n}|y_t - \hat{y}_t|$.

**Mean absolute scaled error** (MASE)
The mean of the absolute scaled errors, where $MAE_{in-sample,\ naive}$ is the mean absolute error produced by a naive forecast calculated on the in-sample data:

$$\frac{MAE}{MAE_{in-sample,\ naive}}$$

For a non-seasonal series, the naïve forecast is generated by using the actual value from previous record. For a seasonal series, the naïve forecast is generated by using the actual value from previous season.

**Root mean squared error** (RMSE)
The root mean square error, $\sqrt{MSE}$.

**Note:**

For the following statistics of fit, *a* is the actual value.

**Weighted absolute percent error** (WAPE)
The weighted APE is computed as $\frac{\Sigma(a \times APE)}{\Sigma a}$.

**Weighted absolute scaled error** (WASE)
WASE is the weighted ASE for all time series. The WASE comparison metric is computed as $\frac{\Sigma(a \times ASE)}{\Sigma a}$.

# Reserved Names for Variables and Models

## Reserved Variable Names

For each project, SAS Visual Forecasting creates several output data sets. The variable names in your input data set cannot match any of the variable names in these output data sets. The variable names in your input data set also cannot start with an underscore. If you try to assign a variable to a role and the variable name matches either of these conditions, then an error message appears.

The following table lists alphabetically the variables that are used by SAS Visual Forecasting.

***Table 5.2*** *Reserved Variable Names*

| Variable Name | Description |
|---|---|
| *_VariableName* | Any variable name that begins with an underscore |
| AADJRSQ | Amemiya's adjusted R-square |

| Variable Name | Description |
|---|---|
| ACTUAL | Dependent series value |
| ADJRSQ | Adjusted R-square |
| AGGCHILDPREDICT | Aggregated prediction of child nodes |
| AIC | Akaike Information Criterion |
| AICC | Finite sample corrected Akaike Information Criterion |
| APC | Amemiya's Prediction Criterion |
| DFE | Degrees of freedom error |
| END | Ending value of the time variable |
| ENDOBS | Number of the last observation in the data |
| ERROR | Prediction errors |
| FINALPREDICT | Predicted value for the parent node |
| GMAPE | Geometric mean percent error |
| GMAPPE | Geometric mean predictive percent error |
| GMAPES | Geometric mean absolute error percent of standard deviation |
| GMASPE | Geometric mean symmetric percent error |
| GMRAE | Geometric mean relative absolute error |
| ISRECONCILED | If the node is reconciled, then this variable is 1;<br>if the node is not reconciled, then this variable is 0. |
| LEAF | Keyword used in model generation |
| LLOCK | Lock level for lower bound on the forecast |
| LOWBROVR | Lower confidence limits before override reconciliation |
| LOWER | Lower confidence limits |
| LOWERBD | Lower bound on the forecast |
| MAE | Mean absolute error |
| MAPE | Mean absolute percent error |
| MAPPE | Symmetric mean absolute predictive percent error |

| Variable Name | Description |
|---|---|
| MASE | Mean absolute scaled error |
| MAX | Maximum value |
| MAXAPES | Maximum absolute error percent of standard deviation |
| MAXERR | Maximum error |
| MAXPE | Maximum percent error |
| MAXPPE | Maximum predictive percent error |
| MAXRE | Maximum relative error |
| MAXSPE | Maximum symmetric percent error |
| MDAPE | Median percent error |
| MDAPES | Median absolute error percent of standard deviation |
| MDAPPE | Median predictive percent error |
| MDASPE | Median symmetric percent error |
| MDRAE | Median relative absolute error |
| ME | Mean error |
| MEAN | Mean value |
| MIN | Minimum value |
| MINAPES | Minimum absolute error percent of standard deviation |
| MINERR | Minimum error |
| MINPE | Minimum percent error |
| MINPPE | Minimum predictive percent error |
| MINRE | Minimum relative error |
| MINSPE | Minimum symmetric percent error |
| MPE | Mean percent error |
| MPPE | Mean predictive percent error |
| MRAE | Mean relative absolute error |
| MRE | Mean relative error |

| Variable Name | Description |
|---|---|
| MSE | Mean square error |
| MSPE | Mean symmetric percent error |
| N | Number of nonmissing observations or number of variance products |
| NAME | Variable name |
| NMISS | Number of missing observations |
| NMISSA | Number of missing actuals |
| NMISSP | Number of missing predicted |
| NOBS | Number of observations |
| NONMISSCHLD | Number of nonmissing children in the current AGGBY group |
| NOTE | Text that user specifies for a series |
| NPARMS | Number of model parameters |
| OLOCK | Lock level for equality constraint |
| OVERRIDE | Equality constraint on the forecast |
| PREBFOVR | Predicted values before override reconciliation |
| PREDICT | Predicted values |
| RECDIFF | Reconciliation difference |
| RMSE | Root mean square error |
| RSQUARE | R-square |
| RWRSQ | Random walk R-square |
| SBC | Schwarz Bayesian information criterion |
| SMAPE | Symmetric mean absolute percent error |
| SSE | Sum of squares error |
| SST | Corrected total sum of squares |
| START | Beginning value of the time variable |
| STARTOBS | Number of the first observation |
| STD | Prediction standard errors |

| Variable Name | Description |
|---|---|
| STDBFOVR | Standard deviation before override reconciliation |
| STDDEV | Standard deviation |
| SUM | Summation value |
| TOP | Keyword used in model generation |
| TSS | Total sum of squares |
| ULOCK | Lock level for the upper bound on the forecast |
| UMSE | Unbiased mean square error |
| UNLOCK | For locked overrides, the value of this variable is 0.<br>For unlocked overrides, the value is 1. |
| UPPEFOVR | Upper confidence limits before override reconciliation |
| UPPER | Upper confidence limits |
| UPPERBD | Upper bound on the forecast |
| URMSE | Unbiased root mean square error |
| XML | User preferences for a series in XML format |

## Reserved Model Names

SAS Visual Forecasting uses the following keywords in the names of the automatically generated models:

- LEAF
- TOP
- HPF

These words are reserved for SAS Visual Forecasting models. If the name of the model that you are creating contains these keywords, then a warning message appears, and you are prompted to select a different name.

**Note:** If you are copying a system-generated model that contains one of these keywords, then the copied name can contain one of these keywords. For example, if you copy a model named LEAF, then SAS Visual Forecasting names the copied model LEAFCOPY**n**, where **n** is the lowest available integer value. Because this name is unique, SAS Visual Forecasting allows this name.

# Promoting Projects and Templates to SAS Visual Forecasting 8.3

## Overview

An administrator can use the transfer command line interface (CLI) to promote projects and templates from one SAS Visual Forecasting installation to another.

- The source data for the project must already be loaded to the CAS server before promoting the project. This includes any data sources for additional attributes or, for SAS Visual Forecasting 8.3 projects, data sources with custom event definitions.
- For SAS Visual Forecasting 8.3 projects, both the transfer CLI and the Model Studio user interface can be used to promote the project. Any custom pipeline or node templates used by the project are included in the promotion. To use Model Studio to promote 8.3 projects, see "Importing Projects" on page 16 for complete instructions.
- For SAS Visual Forecasting 8.2 projects, the transfer CLI is the only way to promote the project to 8.3.

  You must first apply the latest software update on the SAS Visual Forecasting 8.2 server. See SAS Note 62339 to obtain information about this hot fix.

  Any custom pipeline or node templates used by the project must already be promoted as described below.
- For SAS Visual Forecasting 8.2 pipeline and node templates, the transfer CLI is the only way to promote the template.

For general information about the transfer CLI, see "CLI Examples: Transfer" in SAS Viya Administration: Using the Command-Line Interfaces.

Before you start the promotion process, you must first get the unique ID for the resource that you need to promote.

## Getting the ID for a Project

To promote a project to another server, you must first get the unique ID for the project. Follow these steps to get the ID for the project that you want to export.

1. From the Model Studio project list, select the project that you want to export to the target server.
2. Click ⋮ and select **Download batch code** or **Download batch API**.
3. Copy the project ID from the Batch Code dialog box.
   - If SAS is selected for **Download Type**, the project ID is in a macro at the end of the code. For example:

     ```
     %let projectId = fef185c3-c0c2-46fe-bf73-417784c74bba;
     ```
   - If Python is selected for **Download Type**, the project ID is listed with other arguments near the top of the code. For example:

     ```
     parser.add_argument('--projectId', default='fef185c3-c0c2-46fe-bf73-417784c74bba')
     ```

## Getting the ID for a SAS Visual Forecasting 8.2 Template

Follow these steps to get the ID for a pipeline or node template that you want to export from SAS Visual Forecasting 8.2.

1 Click ≡ at the top left corner of SAS Home and select **Manage Environment**. SAS Environment Manager is opened.

2 Click ☰ on the left side.

3 Open **Users** ⇨ ***username*** ⇨ **My Folder** ⇨ **Analytics Toolbox**

For *username*, provide the user ID for the person who created the template.

Custom templates created by the user are listed on the left.

4 Select the template that you want to export.

The properties of the selected template are displayed to the right. Each will have a URI value, which includes a path followed by the ID for the template, for example:

Pipeline template

```
/analyticsComponents/templates/84271a16-a22b-4b0a-b01d-100f806758c4
```

The ID for this pipeline template is `84271a16-a22b-4b0a-b01d-100f806758c4`.

Node template

```
/analyticsComponents/templates/e26bb3e0-cc68-4153-b42b-a3df6ab1ccc8
```

The ID for this pipeline template is `e26bb3e0-cc68-4153-b42b-a3df6ab1ccc8`.

## Promoting Projects and Templates

In the steps below, the following conventions are used.

*source server*
refers to the SAS Visual Forecasting installation from which the project is exported. These steps use the example `test01.example.org`.

*target server*
refers to the SAS Visual Forecasting installation to which the project is imported. These steps use the example `test02.example.org`.

You need a directory on each server to store the JSON transfer file for the project or template. The examples below use the path **`/tmp/projects`**.

After you have the project or template ID, follow these steps to promote the resource to the target server using the transfer CLI for SAS Viya. You must have administrator access for both installations.

1 From a command prompt, issue an SSH connection to the source server for the project.

```
ssh root@test01.example.org
```

2 Change to the **`/opt/sas/viya/home/bin`** directory.

3 Follow the instructions in "Create at Least One Profile" in SAS Viya Administration: Using the Command-Line Interfaces to create a profile for the source server. Use these responses when prompted by the profile initialization:

| | |
|---|---|
| `Service Endpoint>` | Specify the URL for the source server that contains the project that you want to export. For example: `http://test01.example.org` |
| `Output type (text\|json\|fulljson)>` | Specify *json*. |
| `Enable ANSI colored output (y/n)?>` | Specify *y*. |

4 Sign in to the server as described in "Use a Profile to Sign In" in SAS Viya Administration: Using the Command-Line Interfaces.

5 Export the project specifying the path and ID of the project or template.

```
./sas-admin transfer export --resource-uri path/ID
```

where *path* is one of the following:

project
**/analyticsGateway/projects/**

pipeline template
**/analyticsGateway/pipelineTemplates/**

node template
**/analyticsGateway/componentTemplates/**

The following example shows the output when a project is successfully exported.

```
[root@test01 bin]# ./sas-admin transfer export --resource-uri /analyticsGateway/projects/3fdde180-
be4c-4bc9-9385-f1db9d0a3069
running ........ completed
{
    "items": [
        {
            "message": "Transfer object processing complete.",
            "name": "project",
            "state": "completed"
        }
    ]
}

Package created: 4458628a-47b1-4eaa-846a-b7ff50d7c811
```

The output of this command includes the ID for the transfer package. You need this ID for the next step.

6 Download the transfer package as a JSON file. Specify a valid path and give the package file a name. The ID for the package from the previous step must be specified for the `--id` argument. In the following example, the download path and filename for the transfer file is **/opt/projects/sales01.json**.

```
./sas-admin transfer download -file /tmp/projects/sales01.json --id 4458628a-47b1-4eaa-846a-b7ff50d7c811
```

7 Copy the transfer file to the target server, specifying a valid target path.

```
scp /tmp/projects/sales01.json root@test02.example.org:/tmp/projects/sales01.json
```

8 Start a new SSH session with the target server.

```
ssh root@test02.example.org
```

9 Change to the **/opt/sas/viya/home/bin** directory.

10 Follow the instructions in "Command-Line Interface: Preliminary Instructions" in SAS Viya Administration: Using the Command-Line Interfaces to create a profile and sign in to the target server. Use these responses when prompted by the profile initialization:

| | |
|---|---|
| `Service Endpoint>` | Specify the URL for the SAS Visual Forecasting environment where you want to import the project. For example:<br>`http://test02.example.org` |
| `Output type (text\|json\|fulljson)>` | Specify *json*. |

| | |
|---|---|
| `Enable ANSI colored output (y/n)?>` | Specify *y*. |

11 Sign in to the server as described in "Use a Profile to Sign In" in SAS Viya Administration: Using the Command-Line Interfaces.

12 Upload the transfer file, for example:

```
./sas-admin transfer upload --file /opt/projects/sales01.json
```

```
{
    "id": "09206455-906b-4ae6-805b-6eb7c69f4f20",
    "name": "Export"
}
```

The output of this command includes the ID for the transfer package. You need this ID for the next step.

13 Import the transfer file using the package ID output from the previous step.

```
./sas-admin transfer import --request
     "{\"packageUri\":\"/transfer/packages/09206455-906b-4ae6-805b-6eb7c69f4f20\"}"
```

```
running ......................completed

Summary results of import:

        Total tasks: 5  Succeeded: 5  Failed: 0  Completed with errors: 0  Completed with warnings: 0
```

14 Log on to the target server and open Model Studio.

Verify that you can access the project or template.

# Troubleshooting

Here are some issues that you might find using SAS Visual Forecasting, along with steps to take to work around the problem.

## Project Import Fails with Missing Request File Message

### Description

An "`Import Failed`" message is displayed when attempting to import a project. The message contents indicates "`Zip file missing request file`".

This issue can occur after upgrading a SAS Visual Forecasting 8.2 system to 8.3. Existing projects from 8.2 require an upgrade before they can be used. If you export the project before upgrading them, an import will fail with this message.

### Solution

Upgrade all 8.2 projects to 8.3 before exporting them. After the system has been upgraded to 8.3, you are prompted to upgrade the project.

## Pipeline Fails to Run for Project with User-Defined Formats

### Description

If you create a project using a data set that uses formats created with the FORMAT procedure, running the pipeline fails.

### Solution

There are some steps that you need to take to make the format library available to SAS Visual Forecasting and CAS.

1. Make sure the formats are in a path accessible by CAS, for example, ***cas_path*/default/formats/**.
2. Add the format library to CAS. You can use the CAS `addFmtLib` action, for example:

   ```
   cas.addfmtlib="myformats1.sashdat myuserformats2.sashdat"
   ```

   For more information about this action, see addFmtLib Action in *SAS 9.4 and SAS Viya 3.4 Programming Documentation*.
3. Set the FMTSEARCH= system option to include the format library, for example:

   ```
   cas casauto fmtsearch=(myformats1 myuserformats2);
   ```

   For more information about this system option, see FMTSEARCH= System Option in *SAS 9.4 and SAS Viya 3.4 Programming Documentation*.
4. Restart the CAS server.

For more information about user-defined formats in CAS, see these resources.

- Manage Your User-Defined Formats in *SAS 9.4 and SAS Viya 3.4 Programming Documentation*
- *SAS Cloud Analytic Services: User-Defined Formats*
- Add a Global Format Library to a Session's Format Search List in *SAS 9.4 and SAS Viya 3.4 Programming Documentation*

## Cannot Access the Bottom of the Overrides Tab

### Description

You open the Overrides tab but the content at the bottom is cut off and you cannot scroll to access it.

### Solution

Resize the browser window. This forces the vertical scroll bar to appear so that you can scroll to the bottom.

## Cannot Open Project When Source Data Tables Are Not Loaded

### Description

You attempt to open a project and receive this error message:

```
The project cannot be opened because the required data tables for the
project are not loaded. Contact your system administrator.
```

This issue often occurs for projects that are shared as read-only. If the CAS server has been restarted, the source data tables cannot be reloaded for projects that are read-only.

### Solution

The project owner can remove the read-only setting from the Model Studio project page and then open the project. The source data tables are reloaded and the read-only setting can then be applied back to the project.

## Pipeline Fails with Hierarchical Forecasting Model

### Description

You run a pipeline after setting the forecast task to a value other than Diagnose. The pipeline run fails. The forecast tasks must be run sequentially. Each task requires the preceding task to be run first. These settings are described in "Task Settings" on page 69.

### Solution

Create a new pipeline and start running the Diagnose task first, followed by the other tasks in the order provided.

## Error Replacing the Time Series Data Source with a Compatible Data Set

### Description

On the Data tab, you click , select **Time series**, and then select a valid data set that uses the same schema as the current time series for the project. As the new data set is loading, an If-Match Header error is generated.

### Solution

This often occurs with a new project. The pipeline is run and the data definition is updated, but not from the Data tab.

1. Exit the project and open it again. This refreshes the data definition.
2. Repeat the steps to replace the time series.

## Error Exploring Time Series for a Data Node

### Description

For an external forecast, right-click a Data node that has been run in the pipeline and select **Explore time series**. Typically, this should show the time series aggregation plot for the data. Instead, the action generates an error and the plot cannot open.

This problem can occur after the data source for the external forecast project has been refreshed with the changes to the data source.

### Solution

You can view the time series plot by clicking the **Overrides** tab.