* [Forecasting Catalog Sales (Time Series)](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/clementine/example_catalog_forecast.htm)

# Forecasting Catalog Sales (Time Series)

A catalog company is interested in forecasting monthly sales of its men's clothing line, based on their sales data for the last 10 years.

This example uses the stream named catalog\_forecast.str, which references the data file named catalog\_seasfac.sav. These files are available from the Demos directory of any IBM® SPSS® Modeler installation. This can be accessed from the IBM SPSS Modeler program group on the Windows Start menu. The catalog\_forecast.str file is in the streams directory.

We've seen in an earlier example how you can let the Expert Modeler decide which is the most appropriate model for your time series. Now it's time to take a closer look at the two methods that are available when choosing a model yourself--exponential smoothing and ARIMA.

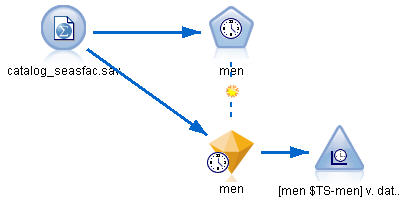
To help you decide on an appropriate model, it's a good idea to plot the time series first. Visual inspection of a time series can often be a powerful guide in helping you choose. In particular, you need to ask yourself:

* Does the series have an overall trend? If so, does the trend appear constant or does it appear to be dying out with time?
* Does the series show seasonality? If so, do the seasonal fluctuations seem to grow with time or do they appear constant over successive periods?

# Creating the Stream

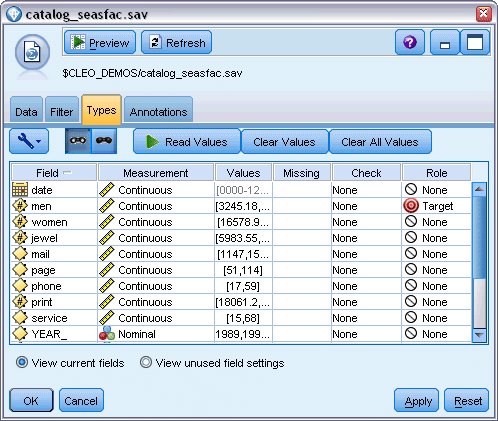
1. Create a new stream and add a Statistics File source node pointing to catalog\_seasfac.sav.

*Figure 1. Forecasting catalog sales*



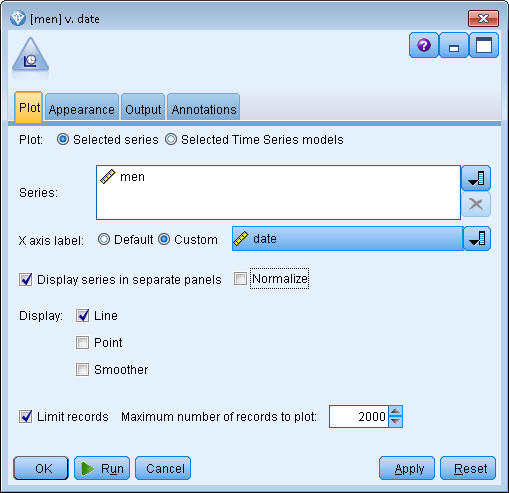
1. Open the IBM® SPSS® Statistics File source node and select the Types tab.
2. Click **Read Values**, then **OK**.
3. Click the **Role** column for the men field and set the role to **Target**.

*Figure 2. Specifying the target field*



1. Set the role for all the other fields to **None**, and click **OK**.
2. Attach a Time Plot graph node to the IBM SPSS Statistics File source node.
3. Open the Time Plot node and, on the Plot tab, add men to the **Series** list.
4. Set the **X axis label** to **Custom**, and select date.
5. Clear the **Normalize** check box.

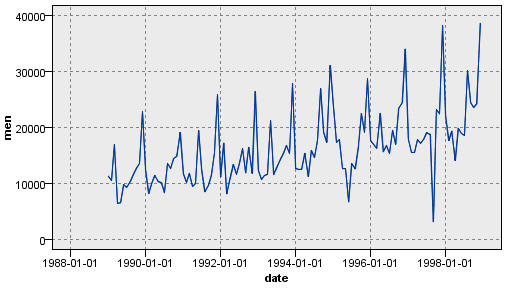
*Figure 3. Plotting the time series*



1. Click **Run**.

**Examining the Data**

*Figure 1. Actual sales of men's clothing*



The series shows a general upward trend; that is, the series values tend to increase over time. The upward trend is seemingly constant, which indicates a linear trend.

The series also has a distinct seasonal pattern with annual highs in December, as indicated by the vertical lines on the graph. The seasonal variations appear to grow with the upward series trend, which suggests multiplicative rather than additive seasonality.

1. Click **OK** to close the plot.

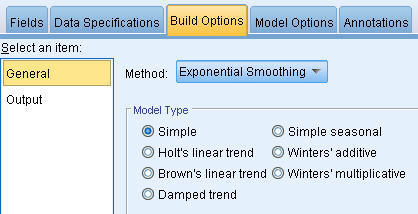
Now that you've identified the characteristics of the series, you're ready to try modeling it. The exponential smoothing method is useful for forecasting series that exhibit trend, seasonality, or both. As we've seen, your data exhibit both characteristics.

**Exponential Smoothing**

Building a best-fit exponential smoothing model involves determining the model type (whether the model needs to include trend, seasonality, or both) and then obtaining the best-fit parameters for the chosen model.

The plot of men's clothing sales over time suggested a model with both a linear trend component and a multiplicative seasonality component. This implies a Winters' model. First, however, we will explore a simple model (no trend and no seasonality) and then a Holt's model (incorporates linear trend but no seasonality). This will give you practice in identifying when a model is not a good fit to the data, an essential skill in successful model building.

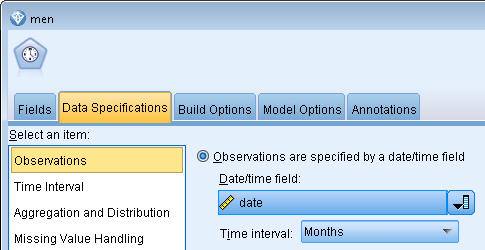
*Figure 1. Specifying exponential smoothing*



We'll start with a simple exponential smoothing model.

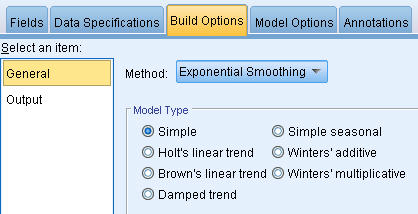
1. Add a Time Series node to the stream and attach it to the source node.
2. On the Data Specifications tab, in the Observations pane, select date as the **Date/time field.**
3. Select Months as the **Time interval**.

*Figure 2. Setting the time interval*



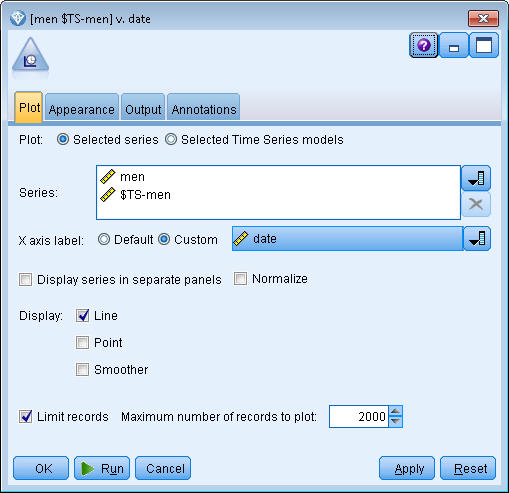
1. On the Build Options tab, in the General pane, set **Method** to **Exponential Smoothing**.
2. Set **Model Type** to **Simple**.

*Figure 3. Setting the model building method*



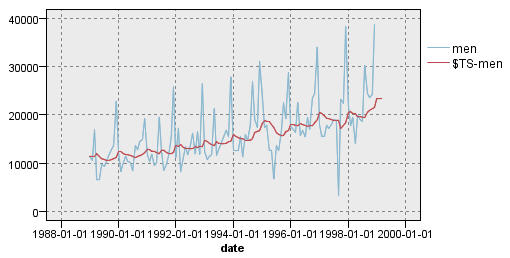
1. Click **Run** to create the model nugget.

*Figure 4. Plotting the Time Series model*



1. Attach a Time Plot node to the model nugget.
2. On the **Plot** tab, add men and $TS-men to the **Series** list.
3. Set the **X axis label** to **Custom**, and select date.
4. Clear the **Display series in separate panels** and **Normalize** check boxes.
5. Click **Run**.

*Figure 5. Simple exponential smoothing model*

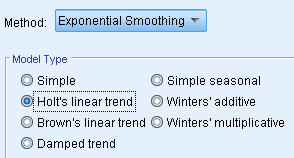


The **men** plot represents the actual data, while **$TS-men** denotes the time series model.

Although the simple model does, in fact, exhibit a gradual (and rather ponderous) upward trend, it takes no account of seasonality. You can safely reject this model.

1. Click **OK** to close the time plot window.

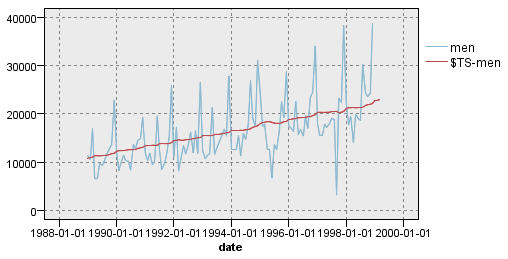
*Figure 6. Selecting Holt's model*



Let's try Holt's linear model. This should at least model the trend better than the simple model, although it too is unlikely to capture the seasonality.

1. Reopen the Time Series node.
2. On the Build Options tab, in the General pane, with **Exponential Smoothing** still selected as the **Method**, select **Holts linear trend** as the **Model Type**.
3. Click **Run** to re-create the model nugget.
4. Re-open the Time Plot node and click **Run**.

*Figure 7. Holt's linear trend model*

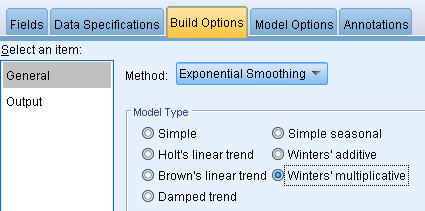


Holt's model displays a smoother upward trend than the simple model but it still takes no account of the seasonality, so you can discard this one too.

1. Close the time plot window.

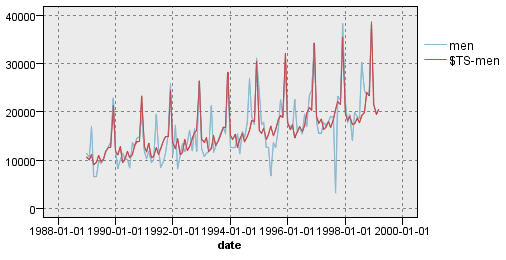
You may recall that the initial plot of men's clothing sales over time suggested a model incorporating a linear trend and multiplicative seasonality. A more suitable candidate, therefore, might be Winters' model.

*Figure 8. Selecting Winters' model*



1. Reopen the Time Series node.
2. On the Build Options tab, in the General pane, with **Exponential Smoothing** still selected as the **Method**, select **Winters' multiplicative** as the **Model Type**.
3. Click **Run** to re-create the model nugget.
4. Open the Time Plot node and click **Run**.

*Figure 9. Winters' multiplicative model*



This looks better; the model reflects both the trend and the seasonality of the data.

The dataset covers a period of 10 years and includes 10 seasonal peaks occurring in December of each year. The 10 peaks present in the predicted results match up well with the 10 annual peaks in the real data.

However, the results also underscore the limitations of the Exponential Smoothing procedure. Looking at both the upward and downward spikes, there is significant structure that is not accounted for.

If you are primarily interested in modeling a long-term trend with seasonal variation, then exponential smoothing may be a good choice. To model a more complex structure such as this one, we need to consider using the ARIMA procedure.

**ARIMA**

With the ARIMA procedure you can create an autoregressive integrated moving-average (ARIMA) model that is suitable for finely tuned modeling of time series. ARIMA models provide more sophisticated methods for modeling trend and seasonal components than do exponential smoothing models, and they have the added benefit of being able to include predictor variables in the model.

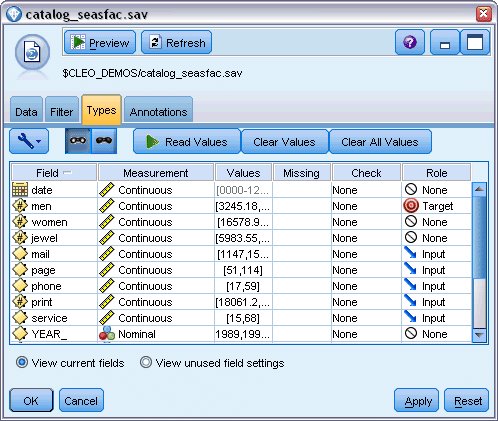
Continuing the example of the catalog company that wants to develop a forecasting model, we have seen how the company has collected data on monthly sales of men's clothing along with several series that might be used to explain some of the variation in sales. Possible predictors include the number of catalogs mailed and the number of pages in the catalog, the number of phone lines open for ordering, the amount spent on print advertising, and the number of customer service representatives.

Are any of these predictors useful for forecasting? Is a model with predictors really better than one without? Using the ARIMA procedure, we can create a forecasting model with predictors, and see if there is a significant difference in predictive ability over the exponential smoothing model with no predictors.

With the ARIMA method, you can fine-tune the model by specifying orders of autoregression, differencing, and moving average, as well as seasonal counterparts to these components. Determining the best values for these components manually can be a time-consuming process involving a good deal of trial and error so, for this example, we'll let the Expert Modeler choose an ARIMA model for us.

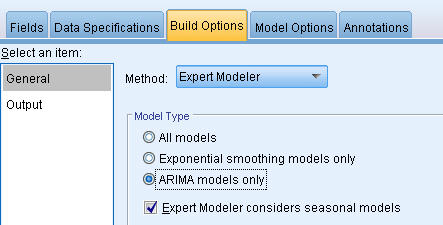
We'll try to build a better model by treating some of the other variables in the dataset as predictor variables. The ones that seem most useful to include as predictors are the number of catalogs mailed (mail), the number of pages in the catalog (page), the number of phone lines open for ordering (phone), the amount spent on print advertising (print), and the number of customer service representatives (service).

*Figure 1. Setting the predictor fields*



1. Open the IBM® SPSS® Statistics File source node.
2. On the Types tab, set the **Role** for mail, page, phone, print, and service to **Input**.
3. Ensure that the role for men is set to **Target** and that all the remaining fields are set to **None**.
4. Click **OK**.
5. Open the Time Series node.
6. On the Build Options tab, in the General pane, set **Method** to **Expert Modeler**.
7. Select the **ARIMA models only** option and ensure that **Expert Modeler considers seasonal models** is checked.

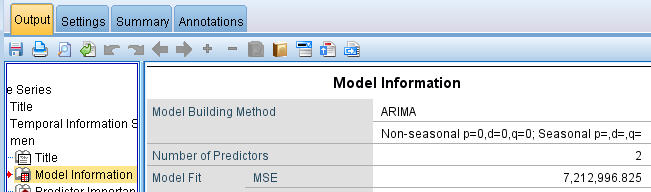
*Figure 2. Choosing only ARIMA models*



1. Click **Run** to re-create the model nugget.
2. Open the model nugget.

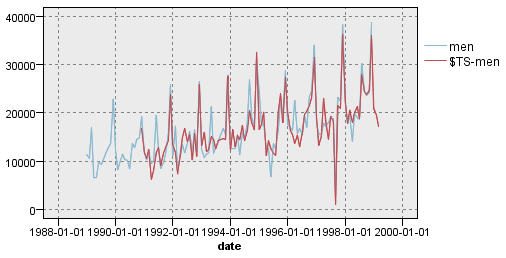
On the Output tab, in the left column, select the **Model information**. Notice how the Expert Modeler has chosen only two of the five specified predictors as being significant to the model.

*Figure 3. Expert Modeler chooses two predictors*



1. Click **OK** to close the model nugget.
2. Open the Time Plot node and click **Run**.

*Figure 4. ARIMA model with predictors specified*



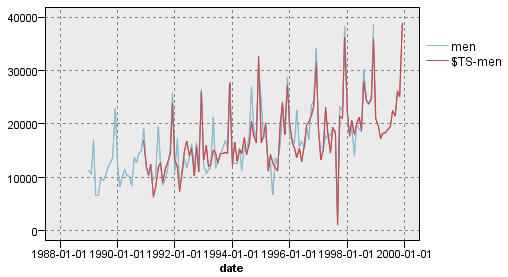
This model improves on the previous one by capturing the large downward spike as well, making it the best fit so far.

We could try refining the model even further, but any improvements from this point on are likely to be minimal. We've established that the ARIMA model with predictors is preferable, so let's use the model we have just built. For the purposes of this example, we'll forecast sales for the coming year.

1. Click **OK** to close the time plot window.
2. Open the Time Series node and select the Model Options tab.
3. Select the **Extend records into the future** checkbox and set its value to 12.
4. Select the **Compute future values of inputs** checkbox.
5. Click **Run** to re-create the model nugget.
6. Open the Time Plot node and click **Run**.

The forecast for 1999 looks good; as expected, there's a return to normal sales levels following the December peak, and a steady upward trend in the second half of the year, with sales in general above those for the previous year.

*Figure 5. Sales forecast extended by 12 months*



# Summary

You have successfully modeled a complex time series, incorporating not only an upward trend but also seasonal and other variations. You have also seen how, through trial and error, you can get closer and closer to an accurate model, which you have then used to forecast future sales.

In practice, you would need to reapply the model as your actual sales data are updated--for example, every month or every quarter--and produce updated forecasts. See the topic [Reapplying a Time Series Model](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/clementine/example_broadband_applymodels.htm#example_broadband_applymodels) for more information.