Measuring Model Performance

Objectives

Demonstrate various ways to evaluate model performance. Evaluate model performance of the exponential smoothing model. Test the model using a holdout sample.

Data

A catalog company, interested in developing a forecasting model, has collected data on monthly sales of men's clothing. The first value was collected in March 2003.

The data for the current example is collected in *catalog.sav* on Blackboard.

Measures of Model Performance

Prior to forecasting it is necessary to assess the effectiveness and robustness of any time series model. There are a number of ways to test the performance of time series models including:

- Seeing how well the model fits the actual historic series.

- Analysing model error on a sequence chart. For example, if the variance of the error increases towards the end of the series then this might suggest that the model is going to forecast badly beyond the historic series. If the variance of the error changes over time, this is known as *heteroscedasticity*.

- Testing whether the error variable for a particular model has a random pattern. The properties of error variables are usually examined using *autocorrelation* and *partial autocorrelation* functions (to be explained in this lab). If the error term is not random, that is, it exhibits some pattern, then this condition can possibly invalidate the time series model estimated and/or suggest ways that the model specification can be improved.

- Compare the model fit and accuracy. In order to do this; it is standard practice to measure the magnitude of the error. The larger the error the worse the fit of the model. The model with the smallest error statistics provides the best data fit. The model with the smallest margin of error will also probably be the most successful for forecasting *if the future is similar to the past*.

- A final check on model performance is to split the time observations into two sub-samples, one to derive the model fit and the second sample to test the forecasting performance of the model. For example, if you have six years' worth of quarterly data, it may be useful to divide this data into two sub-samples: the first sub-sample would be used to estimate the time series model and the second sub-sample to test the performance of forecasts.

Typically, you use most of the data for model estimation and just one season's worth (if there is seasonality) for validation. This is because when the model is used for true, future forecasts, you will typically forecast only a short interval into the future, so there is no reason to test the model on many time periods. For the catalog example, we will develop a model based upon nearly ten years' worth of data and then see how well the model forecasts the final year of data.

Lab04

Similar to Lab04 develop an exponential smoothing model using the following steps.

- Define the time series dates using the define date procedure in SPSS
- Run a sequence chart for the variable which you wish to forecast, so that general trend and seasonal patterns can be identified
- Specify the appropriate trend and seasonal patterns for the exponential smoothing model or use the Expert Modeler
- Estimate the model parameters for the exponential smoothing model
- Save the predicted and residual variables.

Analysing the Model Error on a Sequence Chart

Continuing from the last lab, we will now first assess the overall performance of the exponential smoothing model that we developed. First, we will assess how the model error changes over time. The model fit and error variables are *Predicted_catalog_Model_1* and *NResidual_catalog_Model_1* respectively. These variables have default names based on the dependent variable (catalog) and the number of the model developed (*Model* 1 because we only developed one model).

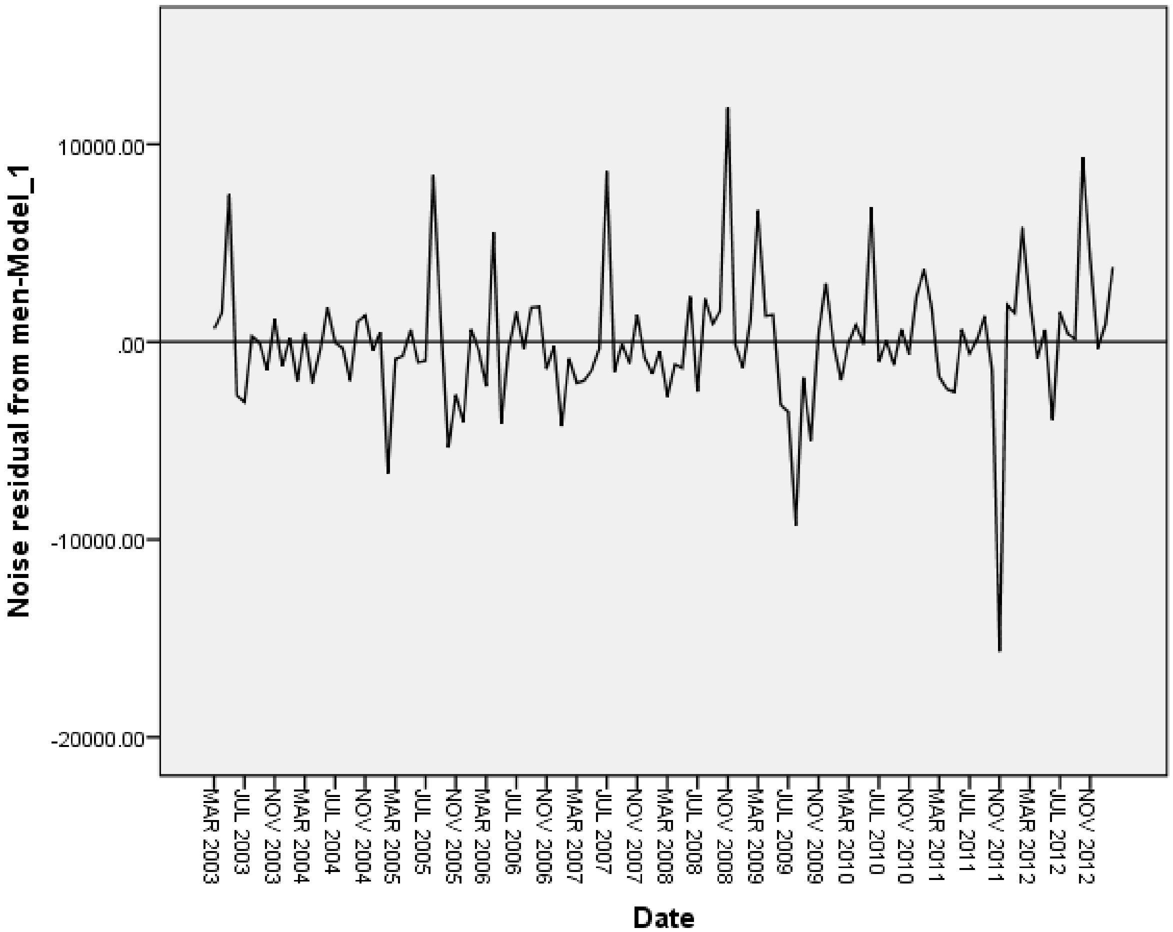
Figure 5.1 Catalog Data with Model Variables

men	YEAR_	MONTH_	DATE_	Predicted_men_Model_1	NResidual_men_Model_1
11357.92	2003	3	MAR 2003	10688.96	668.96
10605.95	2003	4	APR 2003	9148.59	1457.36
16998.57	2003	5	MAY 2003	9538.77	7459.80
6563.75	2003	6	JUN 2003	9266.82	-2703.07
6607.69	2003	7	JUL 2003	9642.23	-3034.54
9839.00	2003	8	AUG 2003	9540.74	298.26
9398.32	2003	9	SEP 2003	9451.63	-53.31
10395.53	2003	10	OCT 2003	11813.65	-1418.12
11663.13	2003	11	NOV 2003	10498.21	1164.92
12805.22	2003	12	DEC 2003	14015.60	-1210.38
13636.25	2004	1	JAN 2004	13445.25	191.00
22849.01	2004	2	FEB 2004	24814.05	-1965.04
12325.80	2004	3	MAR 2004	11893.14	432.66
8273.58	2004	4	APR 2004	10338.13	-2064.55
10061.19	2004	5	MAY 2004	10510.15	-448.96
11497.76	2004	6	JUN 2004	9747.84	1749.92
10363.16	2004	7	JUL 2004	10399.23	-36.07
10194.68	2004	8	AUG 2004	10483.64	-288.96

In order to compare how well the model fits the historic data on catalog sales, select:

Click **Analyze…Forecasting…Sequence Charts**
Click **Reset** and move **NResidual_catalog_Model_1** into the **Variables** list box
Click **Format**
Click **Reference line at mean of series** check box
Click **Continue**, and then click **OK**

Figure 5.2 A Sequence Chart of Model Error



The unexplained variation of the catalog series, picked up by the error term, seems to be fairly random (that is, it does not display any obvious trends or periodicity), and there seems to be no tendency for the model to over predict on more occasions than under predict (which would indicate bias). The model error ranges from under predicting by ten thousand catalog sales a month (shown by the largest positive error on the sequence chart) to over predicting by about fifteen thousand parcels (the most negative error value). During the other periods the error is a lot smaller. Notice that the variance of the error increases slightly further into the series. The errors are potentially heteroscedastic and this might affect the performance of the time series forecast. Some time series models (ARCH and GARCH—not currently available in SPSS) can formally accommodate heteroscedasticity in a time series.

The sequence chart gives some guide as to whether the error is likely to be random or whether there are patterns that can be captured by a different model. A more powerful way investigating this issue is to look at the autocorrelation and partial autocorrelation functions with their associated tests.

Fit Statistics

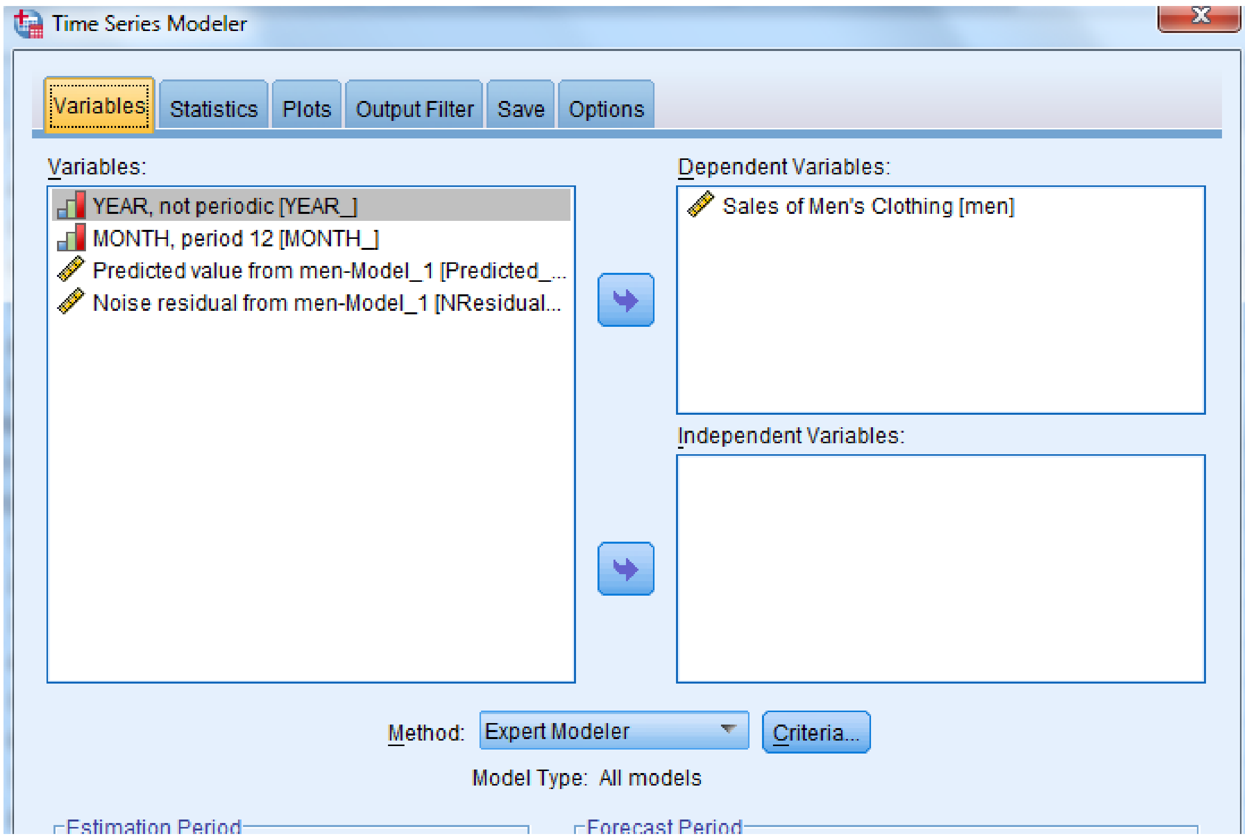
We need to rerun the model to request the fit statistics. We could have requested them when we first created the model previous, but in this instance, we wanted to concentrate on basic output first, then on measures of model performance.

Click **Analyze…Forecasting…Create Traditional Models**

SPSS brings up the main Time Series Modeler dialog box.

Move **men** into the Dependent Variables box

Figure 5.3 Expert Modeler Dialog Box

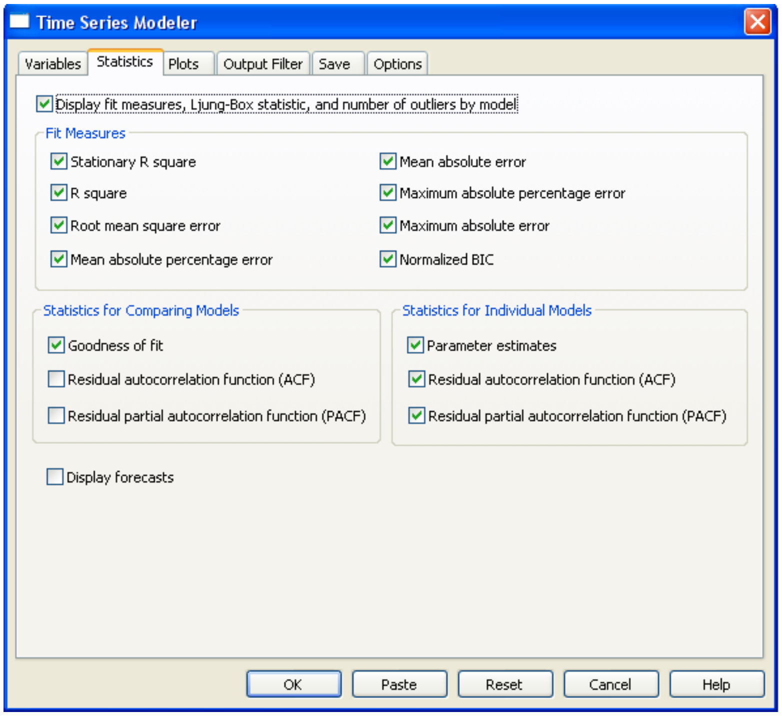


Click the **Statistics** tab

The various fit measures that we discussed in the lecture yesterday are requested in the *Statistics* tab.

Check all the **Fit Measures** boxes
Check all three boxes in the section **Statistics for Individual Models**

Figure 5.4 Statistics Tab



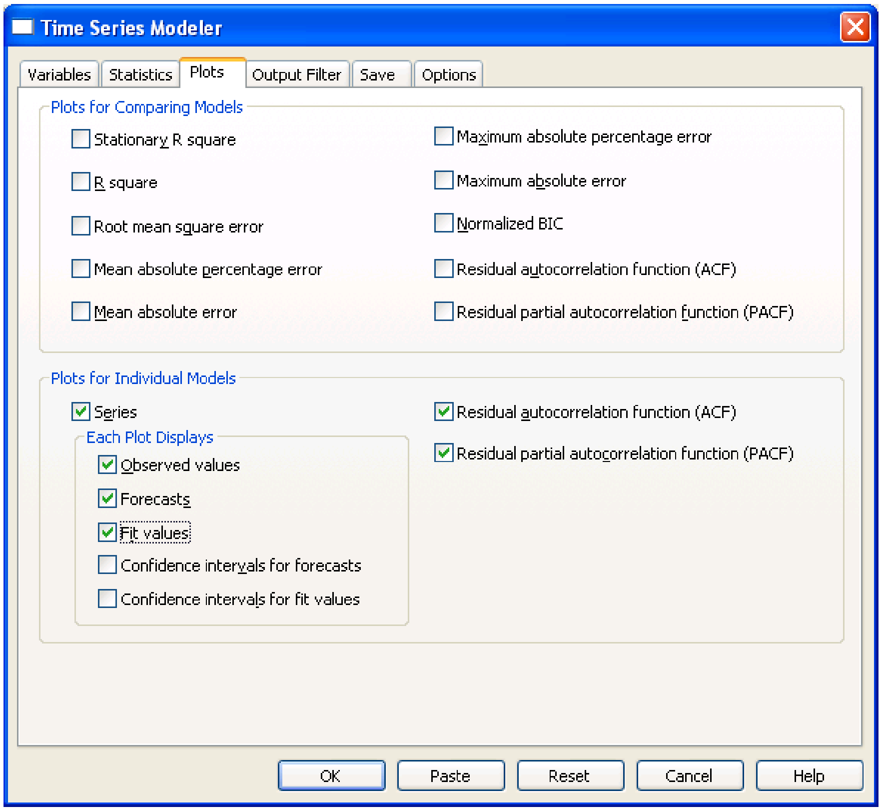
Click the **Plots** tab

In the *Plots* tab we want to request a plot with the fit values, but also plots of the autocorrelation (ACF) and partial autocorrelation (PACF) functions. These functions are excellent tests of whether there is any non-random variation remaining the in the residuals.

Check the boxes for **Fit values**, **Residual Autocorrelation Function**, and **Residual Partial Autocorrelation Function** in the Plots for Individual Models area

We can leave the *Forecasts* check box selected, even though we are not forecasting yet. The Expert Modeler will simply ignore this request.

Figure 5.5 Plots Tab



We don’t need to save any new variables because they are already in the data file from our previous use of Expert Modeler. Since the data are the same, as are the modeling specifications, we will find the same exponential smoothing model again.

Click **OK**

Now we will look at the goodness of fit statistics for our model, which are included in the Model Statistics table. Note that this table was edited in the pivot table editor to transpose the rows and columns in order to make the table easier to view and discuss.

Figure 5.6 Goodness of Fit Statistics

Model Statistics

Number of Predictors		0
Model Fit statistics	Stationary R-squared	.713
	R-squared	.719
	RMSE	3383.634
	MAPE	17.814
	MAE	2166.830
	MaxAPE	482.446
	MaxAE	15656.236
	Normalized BIC	16.373
Ljung-Box Q(18)	Statistics	21.476
	DF	15
	Sig.	.122
Number of Outliers		0

The first line of the model statistics tells us that the number of predictors (independent variables) used in the model was zero. Since we ran this model in the already we know that the model selected by the Expert Modeler was the Winters' Additive model.

The stationary R-squared is a measure that compares the stationary part of the model to a simple mean model. This measure is preferable to ordinary R-squared when there is a trend or seasonal pattern. In this case we have a value of 0.713 which is positive and indicates that we are explaining a good part of the overall variation in the model. However, since there is a seasonal pattern in our data, we can not determine the exact amount explained.

The *Mean Absolute Error* (MAE) is often used as the primary measure of fit. In our data, the average magnitude of error either side of zero is about €2166.83. The mean absolute error must be interpreted with reference to the original units of measurement.

The *Mean Absolute Percentage Error* (MAPE) is sometimes preferred because it is a percentage and thus a relative measure. The value of MAPE is 17.814%. It should be possible, using business knowledge and desired outcomes, to determine whether this level of accuracy is acceptable. Both MAE and MAPE can give an indication of the likely errors when forecasting.

The *Maximum Absolute Error* (MaxAE) is the largest forecast error, positive or negative. It is useful for imagining the worst case scenario for your forecasts. The *Maximum Absolute Percentage Error* (MaxAPE*)* is also useful for worst case scenarios..

Maximum absolute error and maximum absolute percentage error may occur at different series points; for example, when the absolute error for a large series value is slightly larger than the absolute error for a small series value. In that case the maximum absolute error will occur at the larger series value and the maximum absolute percentage error will occur at the smaller series value.

The *Root Mean Squared Error* (RMSE) is the square root of the mean squared error. It can be thought of as the standard deviation of the error terms. A small RMSE is preferred since this signifies that the error terms do not have a large spread, that is, the errors concentrate near zero, which means a good fit.

One of the differences between the measures based on absolute error values (MAE, MAPE) and measures based on the squared value of the error (MSE) is that the latter penalises a forecast much more for extreme deviations (a large error) than it does for small ones. For instance the MAE calculation counts an error of 2 as twice as much as an error of 1, where the MSE calculation squares an error of 2 and thus counts it as 4 times as much as an error of 1. Thus adopting the criterion of minimising mean squared error implies that we would prefer to have several small deviations from the forecast value rather than one large deviation. For this reason, the RMSE is often best used when comparing models on the same series.

The *Normalized BIC* (Bayesian Information Criterion) is a general measure of the overall fit of a model that attempts to account for model complexity. It is a score based upon the mean square error and includes a penalty for the number of parameters in the model and the length of the series. The penalty removes the advantage of models with more parameters, making the statistic easy to compare across different models for the same series. In this data set, we had only one model so this measure cannot be used as a tool for evaluating the Winters' Additive model.

We will discuss the Ljung-Box statistic in the next section.

As review, Figure 5.7 displays the Model Parameters table, reminding us of the form of the exponential smoothing model.

Figure 5.7 Model Parameters

Exponential Smoothing Model Parameters

Model			Estimate	SE	t	Sig.
Sales of Men's Clothing-Model_1	No Transformation	Alpha (Level)	.062	.049	1.264	.209
		Gamma (Trend)	5.379E-009	.001	3.603E-006	1.000
		Delta (Season)	1.933E-005	.042	.000	1.000

Autocorrelation

The autocorrelation (ACF) and partial autocorrelation functions (PACF) are tools to check the performance and specifications of time series models. As with linear regression, the errors in time series models should not be correlated, and the ACF and PACF functions help to test for this. A time series model that is well specified will capture all of the non-random variation, including seasonality, trend, and cyclic and other factors that are important. If this is the case, then the error should not be correlated with itself over time. The ACF and PACF will show whether there are patterns in the errors, and provide clues about how the current model can be improved.

An error can be positively and negatively correlated with itself over time. When there is positive autocorrelation there is a tendency for a positive error to be followed by positive errors, or negative errors to be followed by negative errors. Conversely, if there is negative autocorrelation the errors the error will tend to follow a plus to minus to plus to minus sequence over time.

There are also different orders of autocorrelation. If the error is strongly correlated with the error four periods previous then the type of autocorrelation is known as fourth-order autocorrelation (autocorrelation at lag 4). Similarly if the error is strongly related with the error twelve periods back, this is known as twelfth-order autocorrelation.

When testing for error autocorrelation, or autocorrelation within the original series, it is important to look for lags that are multiples of the periodicity, since they reflect seasonal effects. In this example, the data are yearly and the period is twelve months. Thus the 12th (and 24th , 36th, etc.) order autocorrelation is of special interest. If the error series has a significant autocorrelation at lag 12 (12 months), it suggests that the model was misspecified, having failed to pick up some important monthly factors which relate to demand for parcels.

We will first look at the residual ACF values and plot. The Residual ACF table in Figure 5.8 has been edited by transposing the rows and columns to make it easier to view.

Figure 5.8 Residual ACF Table

Residual ACF

	Model	
	Sales of Men's Clothing-Model_1	
	ACF	SE
1	.095	.091
2	-.005	.092
3	-.077	.092
4	-.091	.093
5	-.019	.093
6	-.039	.093
7	.065	.094
8	-.003	.094
9	-.119	.094
10	-.117	.095
11	-.225	.096
12	-.075	.101
13	.145	.101
14	.030	.103
15	.058	.103
16	.012	.103
17	-.140	.103
18	-.041	.105
19	.016	.105
20	.080	.105
21	.039	.105
22	.045	.106
23	-.006	.106
24	-.057	.106

Time lag appears along the vertical axis. 24 autocorrelations are displayed by default (this value can be changed in the *Options* tab). The second and third columns display the autocorrelation values (which always lie in the range –1 to 1) and their associated standard errors. As a general rule of thumb, to assess the significance of an autocorrelation at a particular time lag, take its autocorrelation value and divide by its standard error. If the ratio is above 2 in absolute value then there is a significant autocorrelation at that lag. Note that the ratio for the autocorrelation for most of the lags is below 2 (in absolute value) – i.e., insignificant. Lag 11 is the only significant one – i.e., |-0.225/0.096| = 2.344. But this lag is not a multiple of the periodicity and so not a concern.

In a second table, the PACF values are listed. A PACF looks at correlations after controlling for the series values at the intervening time points. For example, the fourth order PACF (lag 4) measures the correlation between the error at time t and the error at time t-4, after controlling for the correlations between errors at lag 1, lag 2, and lag3. Thus we are looking at the relationship at a fixed time lag after controlling for relationships at smaller (intervening) lags (note that the standard error is the same at every lag). The ACF and PACF at lag 1 are identical, since there are no intervening time points on which to adjust. You can also use the same general rule of thumb to assess the PACF values.

Figure 5.9 Residual PACF Table

Residual PACF

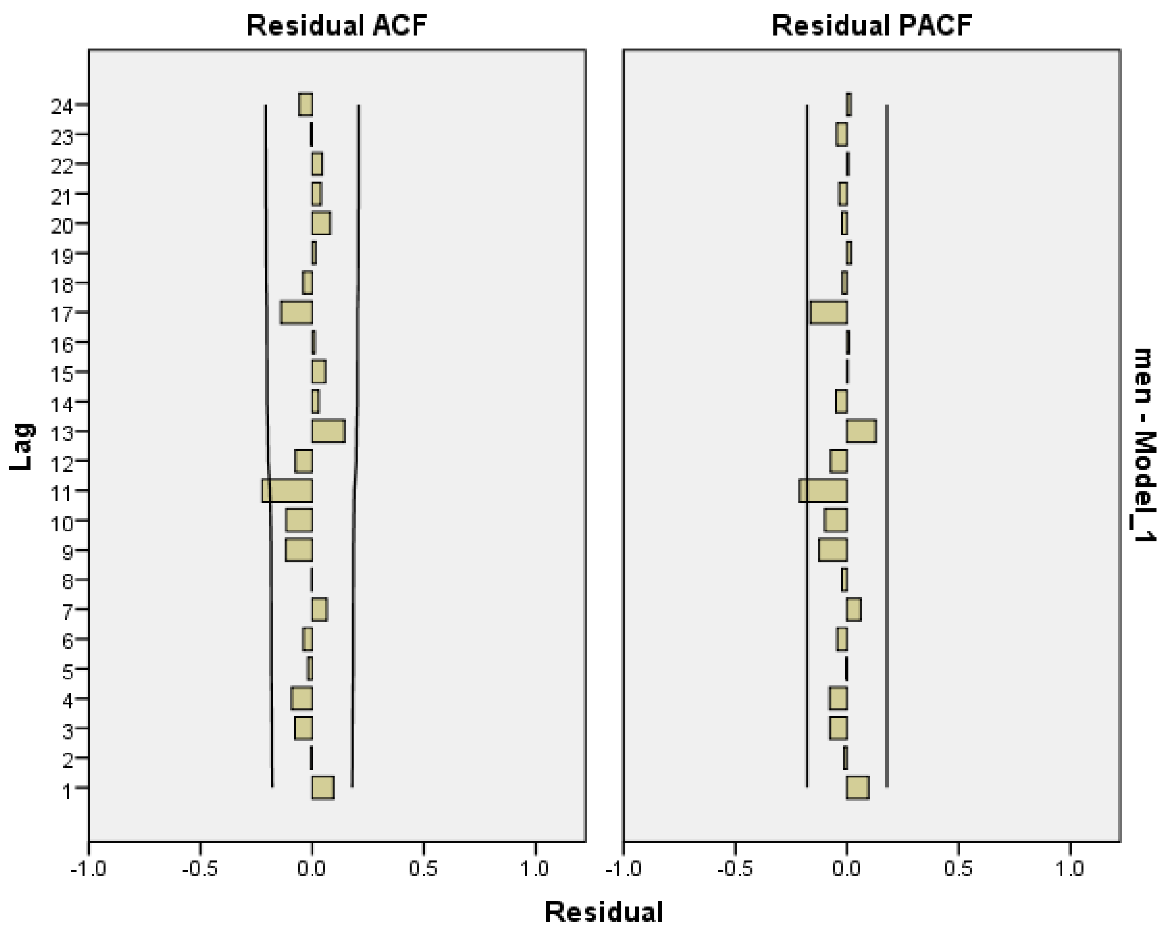
	Model	
	Sales of Men's Clothing-Model_1	
	PACF	SE
1	.095	.091
2	-.014	.091
3	-.076	.091
4	-.078	.091
5	-.005	.091
6	-.044	.091
7	.061	.091
8	-.024	.091
9	-.127	.091
10	-.098	.091
11	-.211	.091
12	-.074	.091
13	.129	.091
14	-.050	.091
15	.004	.091
16	.011	.091
17	-.164	.091
18	-.022	.091
19	.018	.091
20	-.024	.091
21	-.036	.091

22	.008	.091
23	-.047	.091
24	.016	.091

The PACF at lag 11 has ratio of value/standard error above 2 (–.211/.091).

The ACF and PACF values are also displayed graphically by SPSS, as show in Figure 5.10.

Figure 5.10 Residual ACF and PACF Charts



On the vertical axis the lag number is displayed with values ranging (by default) from 1 to 24. For each lag there is a bar. For example, the first bar shows the autocorrelation estimate calculated for the error term at time t and the error term at time t-1 (one time point back) throughout the series. Similarly, the third bar shows the autocorrelation relationship for the error term at time t and the error term at time t-3 (three time points back). In other words, each bar shows the correlation between the error and the error a fixed number of time points back, in accordance to the specified lag number.

The horizontal axis represents the autocorrelation/partial autocorrelation values, which as correlations can range from –1 to 1.

Also displayed on the graph are black vertical lines that represent the 95% confidence interval for the correlations. In order to test for patterns in the error variable, it is necessary to compare each of the bars with the confidence interval lines.

If all of the bars fall within the confidence intervals, then there are no significant autocorrelations in the error series. On the other hand if any of the bars cross the black line then there are significant autocorrelations in the series. The autocorrelation plots suggest that there may be significant eleventh order autocorrelation

If the autocorrelation function shows signs of significant autocorrelation, then it is also important to look at the partial autocorrelation function. Many analysts would routinely view both since each provides information about possible relationships over time. The PACF also shows some evidence of eleventh-order autocorrelation.

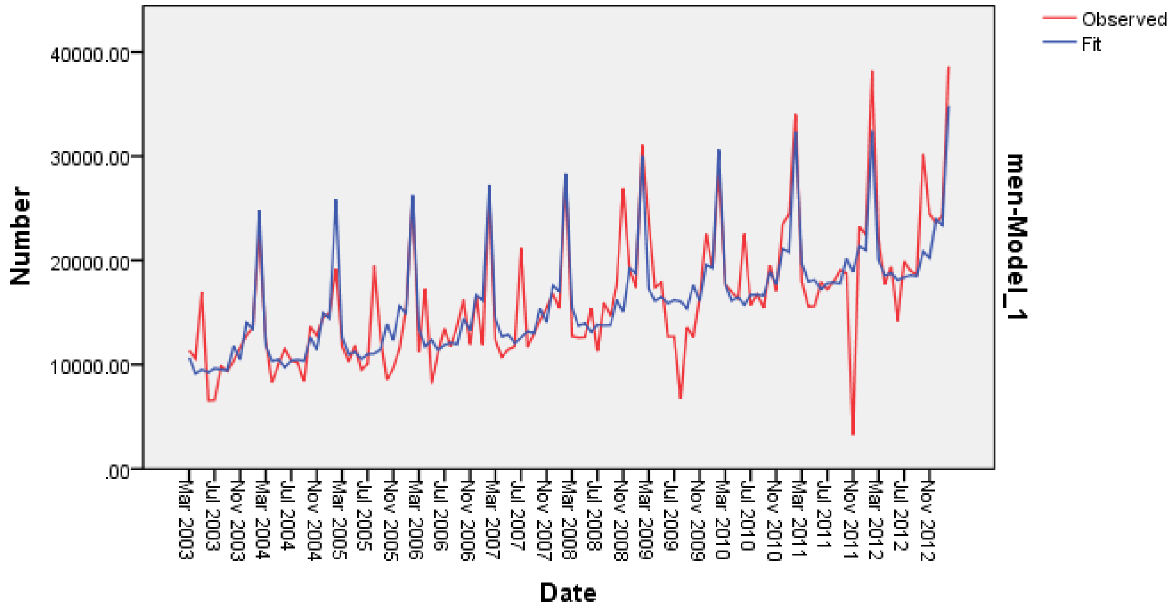
Another measure of autocorrelation is the *Ljung-Box Q* statistic (also sometimes known as the modified Box-Pierce statistic). It can be found in the Model Statistics table (see Figure 5.6). It is a joint test for the overall autocorrelation of the error. At lag t, the Ljung-Box statistic jointly tests whether *any* autocorrelations up to lag t (that is, lag 1 to lag t) are significant. As such it provides an omnibus test for any significant autocorrelations at the given lag or fewer. Some analysts recommend examining the Ljung-Box test at the lag equal to ¼ the length of the series, and if the significance value is below .05, conclude the series is autocorrelated. It provides one overall test (assuming you examine it at a single lag of interest) instead of examining many autocorrelation tests, each at the .05 level.

In this example the sample size is 120, and ¼ the length of this series is about 30. By default, autocorrelations through lag 24 appear in the plots. The Ljung-Box statistic is calculated for the 18th lag for all models; it has a value of 21.476 and the associated probability is 0.122, above the .05 level. This means that about 12.2 percent of all random samples from a population with no autocorrelation at lags 1 through 18 could be expected to have an autocorrelation pattern at least as pronounced as this one. From this result we would conclude that that there are insignificant autocorrelations (through lag 18).

The results of the ACF and PACF plots and the Ljung-Box test are consistent: there is a random pattern in the errors. An insignificant autocorrelation exists. What this implies is that the current model can be used for forecasting.

Now we will look at the sequence chart showing both the actual series and the fit values.

Figure 5.11 Chart Showing both Actual and Fitted Series



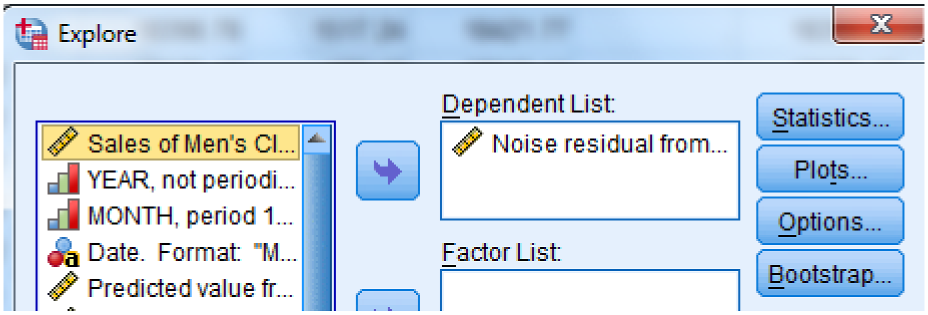
We can see from this chart that while in general the fitted data follows the pattern of the original data, it does not fit the troughs very well.

Testing for Normality of Residuals

Another test of the model is the normality of the errors (residuals). The normality of residuals can be assessed graphically in at least two ways: a histogram or a Q-Q plot. Further, you can do a direct test of whether the errors are normally distributed with the K-S (Lilliefors) and/or Shapiro-Wilk tests, which can be obtained through the Explore procedure. It is important to note that exponential smoothing does not assume normality of the errors, unlike other time-series models (e.g., ARIMA and Regression). However, to demonstrate the general point we will examine the errors for the current model.

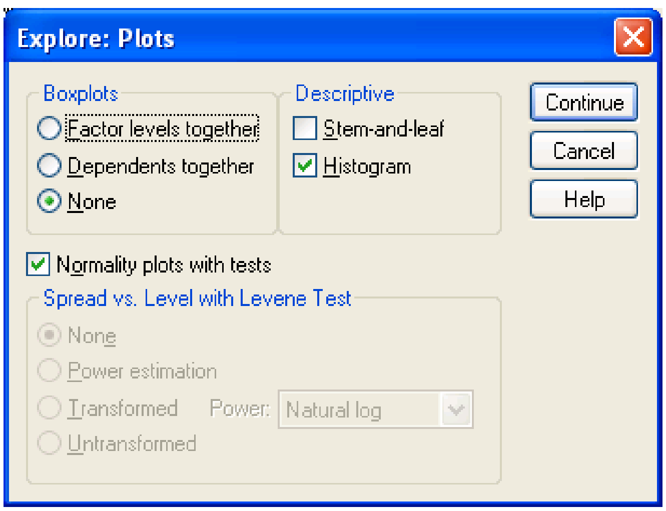
Click **Analyze…Descriptive Statistics…Explore**
Move **NResidual_catalog_Model_1** to the Dependent(s) list box

Figure 5.12 Explore Dialog Box



Click **Plots** pushbutton
Click the Boxplots **None** option button
Click **Stem-and-Leaf** check box to deselect it
Click **Histogram** check box
Click the **Normality plots with tests** option button

Figure 5.13 Explore: Plots Dialog Box



Click **Continue**
Click **OK**

Figure 5.14 Tests of Normality

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Noise residual from men-Model_1	.148	120	.000	.889	120	.000

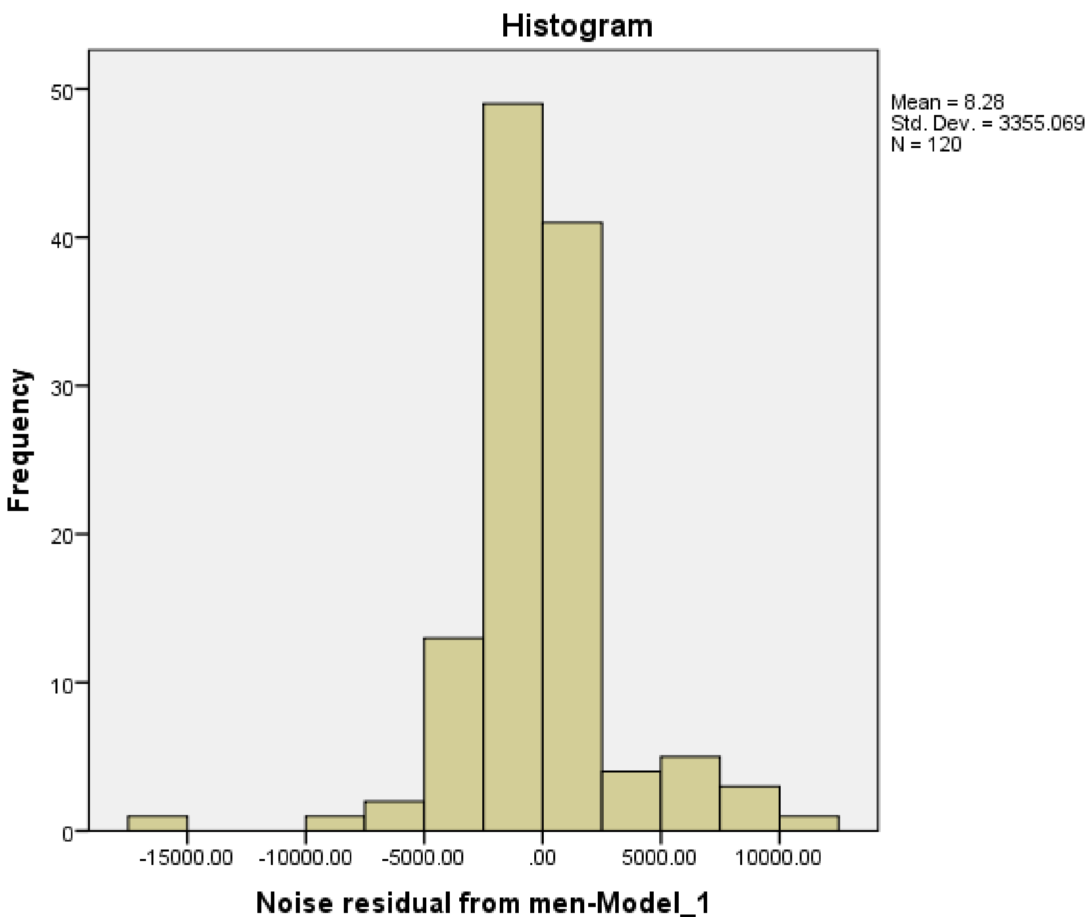
a. Lilliefors Significance Correction

The *KS-Lilliefors* (*Kolmogorov-Smirnov* test with *Lilliefors* correction) statistic tests whether the sample distribution of a variable is consistent with the one expected if the sample were from a normal population. The null hypothesis is that in the population the error is normally distributed. In this case the significance (Sig.) value less than 0.001 suggests that we can reject the null hypothesis of normality.

For an exponential smoothing model, normality of residuals is not an issue. Also, for time series models that do assume normality of error, the violation of normality tends to be a less serious problem than autocorrelation of error.

We can also view the histogram. If the error distribution is not normal, the histogram (or Q-Q plot) can show you where the deviation occurs. The distribution appears to be left skewed, as shown in Figure 5.15. There is an outlier(s) present. We will look at detecting outliers in a future lab.

Figure 5.15 Histogram of Error Variable



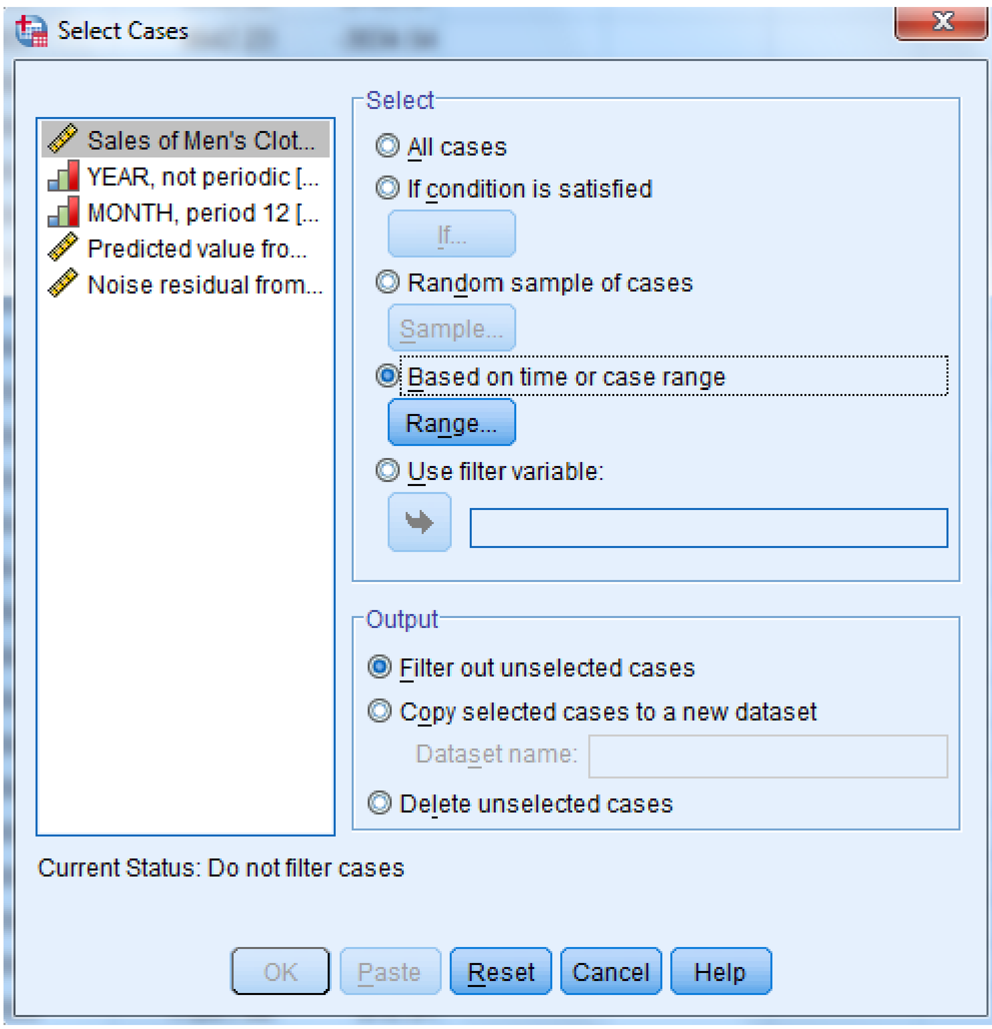
Splitting the Sample into Estimation and Forecasting Sub-Samples

In the catalog data set there are ten years' worth of data, so another useful evaluation of performance is to develop a model based on the first X (here we choose 9) years of data and see how well the model forecasts the remaining (here 1) year of data. The final year will form a "hold-out" sample to instantly assess how well the model might forecast beyond the ninth year.

To split the ten years' worth of data into sub-samples we must use the Select Cases Dialog.

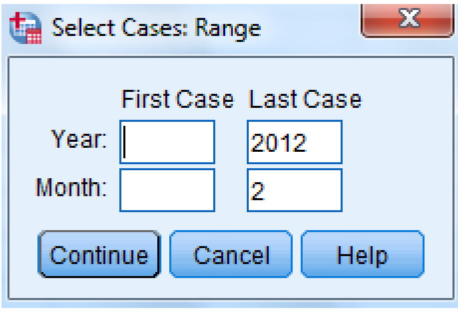
Click **Data…Select Cases**
Click the **Based on time or case range** option button

Figure 5.16 Select Cases Dialog Box



Click the **Range** pushbutton
In the **Last Case** Column enter **2012** in the **Year** text box and **2** in the **Month** text box.

Figure 5.17 Specifying the Use Range



Since the First Case column was left blank, SPSS will use the default, which is the first case in the file.

Click on **Continue**
Click on **OK**

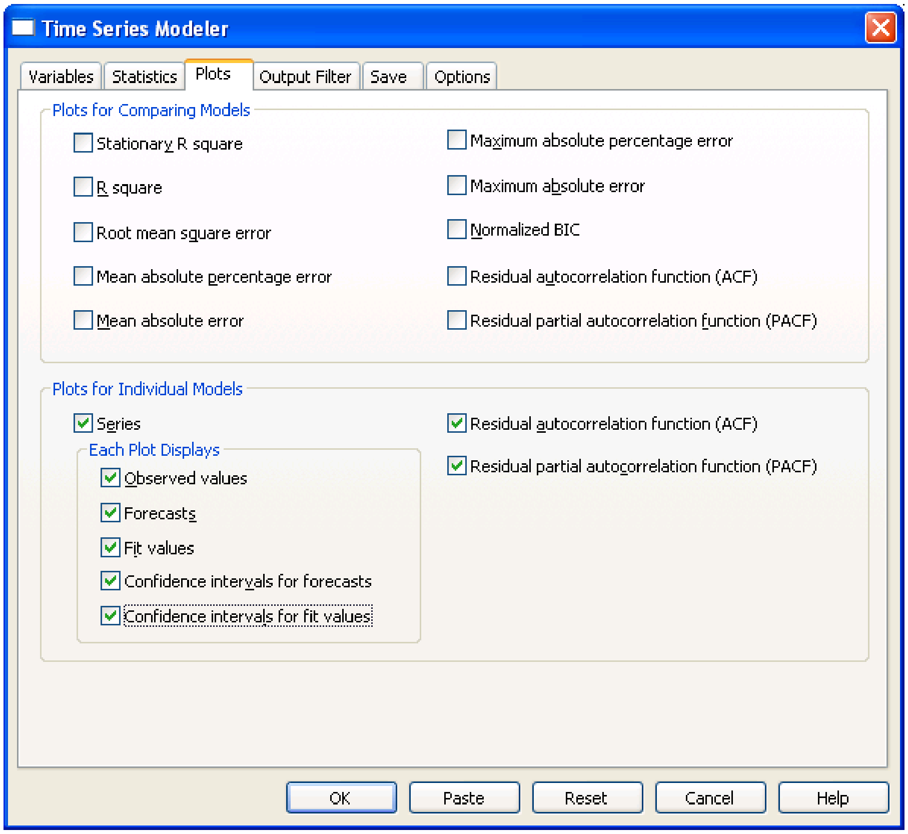
SPSS will use data through the last month of year 9 when estimating time series models.

If you look in the Data Editor, you will notice that all cases after year 9 are filtered out (have diagonal lines across their case sequence numbers). Also, the Status Bar at the bottom of the Data Editor window notes "Use On." The filtered cases comprise the holdout sample.

We must rerun the Expert Modeler on the first nine years to get forecasts for the tenth year. Most of the dialog box specification we need are already selected.

Click **Analyze…Forecasting…Create Traditional Models** (or use the Dialog Recall tool)
Click the **Statistics** tab
Click **Display forecasts** check box (not shown)
Click the **Plots** tab
Click **Confidence intervals for forecasts** and **Confidence intervals for fit values** check boxes

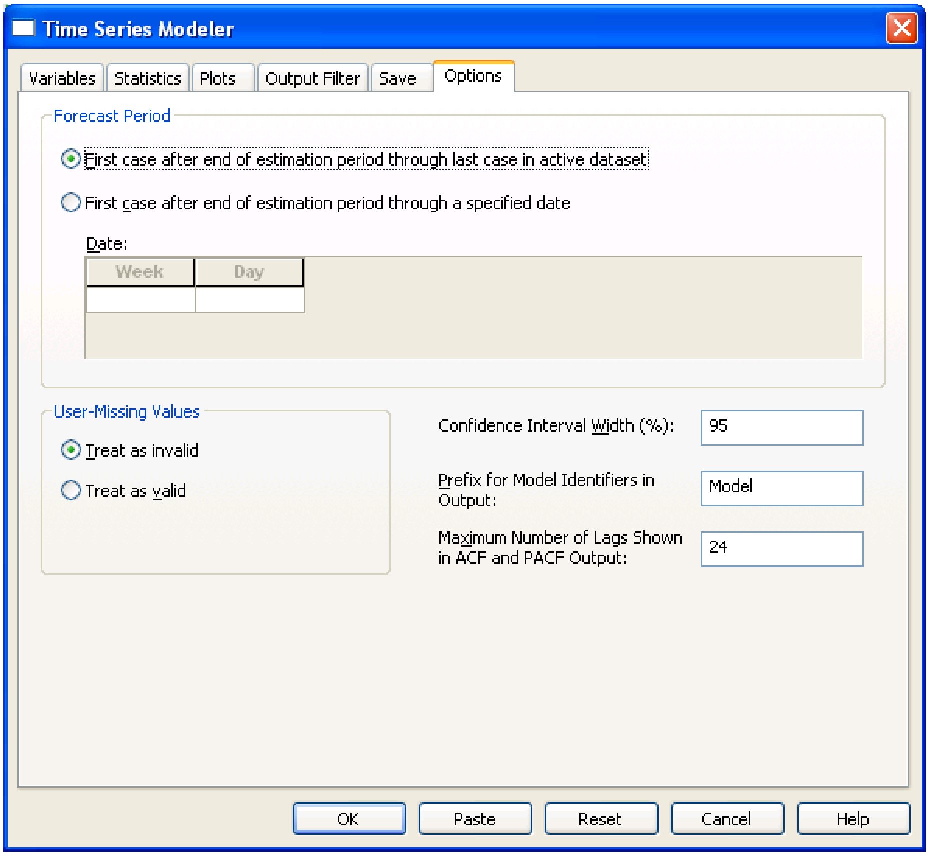
Figure 5.18 Plots Tab



Click the **Save** tab
Click **Predicted Values** and **Noise Residual** check boxes (not shown)
Click **Options** tab

Note that, by default, the *First case after end of estimation period through last case in active dataset* check box is selected. Because of this specification, the Expert Modeler will automatically make forecasts for year 10 since those cases are included in the data file.

Figure 5.19 Output Tab



Click **OK**

Once again, the model type is Winters' Additive. This is not surprising, since we are using 9/10, or 90% of the original series data, to calculate the model.

Are the two models basically the same? Figure 5.20 presents the Model Parameters table, first for the original model, and then second for the 9-year model. We can see that the gamma, alpha and delta parameters are all quite close (particularly when the parameter standard errors are considered). So all in all, we are using basically the same model to make forecasts for the holdout sample.

Figure 5.20 Model Parameters for 9 and 10 Year Models

Exponential Smoothing Model Parameters

Model			Estimate	SE	t	Sig.
Sales of Men's Clothing-Model_1	No Transformation	Alpha (Level)	.062	.049	1.264	.209
		Gamma (Trend)	5.379E-009	.001	3.603E-006	1.000
		Delta (Season)	1.933E-005	.042	.000	1.000

Exponential Smoothing Model Parameters

Model			Estimate	SE	t	Sig.
Sales of Men's Clothing-Model_1	No Transformation	Alpha (Level)	.056	.053	1.058	.293
		Gamma (Trend)	2.866E-007	.001	.000	1.000
		Delta (Season)	.000	.043	.002	.998

In Figures 5.21 and 5.22, we have the results from the Model Statistics tables for the original 10 year model and the new model based on the first 9 years of catalog sales. When using a holdout sample to test a model's fit, it can also be helpful to compare the fit statistics for the two models, before looking at how well the new model forecasts the holdout sample.

Figure 5.21 Model Statistics for 10 year Model

Model Statistics

Number of Predictors		0
Model Fit statistics	Stationary R-squared	.713
	R-squared	.719
	RMSE	3383.634
	MAPE	17.814
	MAE	2166.830
	MaxAPE	482.446
	MaxAE	15656.236
	Normalized BIC	16.373
Ljung-Box Q(18)	Statistics	21.476
	DF	15
	Sig.	.122
Number of Outliers		0

Figure 5.22 Model Statistics for 9 year Model

Model Statistics

		Model
		Sales of Men's Clothing-Model_1
Number of Predictors		0
Model Fit statistics	Stationary R-squared	.695
	R-squared	.686
	RMSE	3346.080
	MAPE	18.471
	MAE	2130.897
	MaxAPE	468.186
	MaxAE	15193.470
	Normalized BIC	16.361
Ljung-Box Q(18)	Statistics	17.520
	DF	15
	Sig.	.289
Number of Outliers		0

Comparing these fit statistics with the fit statistics from the entire data set we find that the 10 year sample has slightly better fit results. However, the 9 year model still has an insignificant Ljung-Box statistic, indicating no autocorrelation. We won't focus on autocorrelation any further in this example.

The results of the analysis also provide us with a table of the forecasted values. They rise in the winter months then fall in spring and summer. Notice the large confidence interval around the forecast values.

Figure 5.23 Forecasted Values

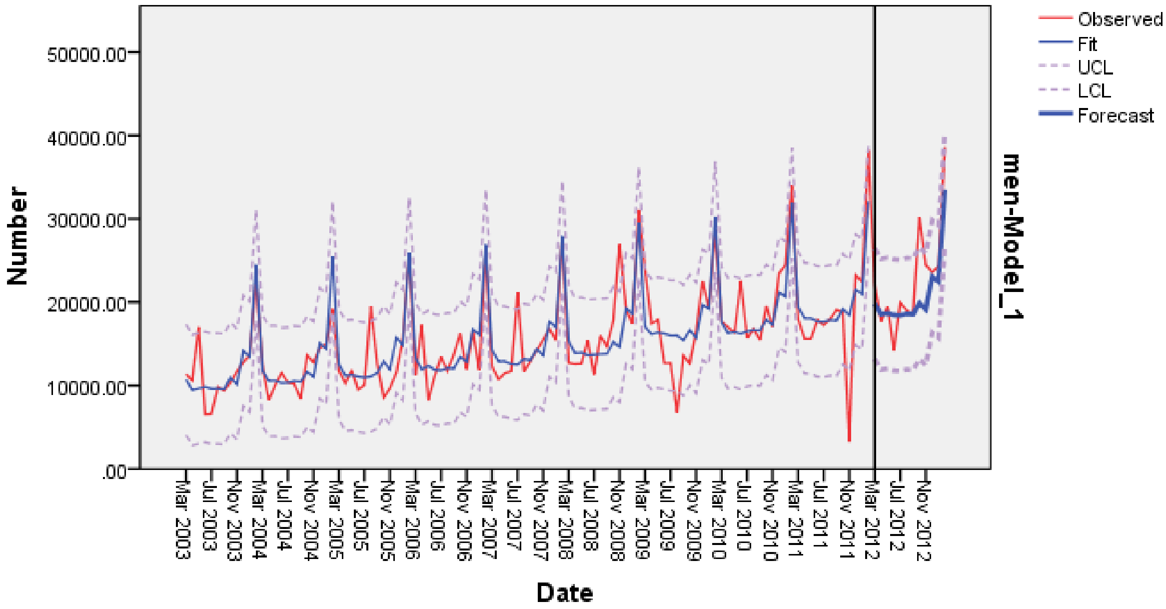
Forecast

	Model		
	Sales of Men's Clothing-Model_1		
	Forecast	UCL	LCL
Mar 2012	19824.41	26459.07	13189.75
Apr 2012	18545.50	25190.62	11900.38
May 2012	18690.92	25346.48	12035.35
Jun 2012	18457.09	25123.09	11791.10
Jul 2012	18421.77	25098.18	11745.36
Aug 2012	18619.44	25306.25	11932.64
Sep 2012	18540.97	25238.15	11843.78
Oct 2012	19882.65	26590.20	13175.10
Nov 2012	19156.38	25874.28	12438.49
Dec 2012	23080.46	29808.69	16352.23
Jan 2013	22453.58	29192.13	15715.03
Feb 2013	33481.12	40229.97	26732.27

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

This may sound quite large, but it isn't necessarily worse than the predictions for the 9 years preceding. To see this, we can look at the sequence chart of observed and fitted values.

Figure 5.24 Sequence Chart of Observed, Fit, and Forecast Values



The point at which forecasting begins is conveniently marked with a vertical line. The blue line is the forecast and the dotted lines are the forecasts confidence intervals, and we can observe that the forecasts for the first 9 years have about as much error as those produced for the 10 year. The chart also illustrates that the exponential smoothing model has captured the basic monthly pattern of catalog sales when forecasting, as the forecasted values follow this pattern

To illustrate that SPSS is truly forecasting for the last year, let’s look at the Data Editor window.

Switch to the Data Editor window, Data View tab
Scroll to the **bottom rows**

Figure 5.25 Data Editor with Predictions for Week 18

107	22452.33	2012	1	JAN 2012	20996.28	1456.05	20978.01	1474.32
108	38236.59	2012	2	FEB 2012	32443.34	5793.25	32087.92	6148.67
~~109~~	22028.80	2012	3	MAR 2012	20003.41	2025.39	19824.41	.
~~110~~	17719.90	2012	4	APR 2012	18547.15	-827.25	18545.50	.
~~111~~	19408.95	2012	5	MAY 2012	18795.71	613.24	18690.92	.
~~112~~	14169.61	2012	6	JUN 2012	18099.51	-3929.90	18457.09	.
~~113~~	19916.02	2012	7	JUL 2012	18398.78	1517.24	18421.77	.
~~114~~	19001.89	2012	8	AUG 2012	18579.42	422.47	18619.44	.
~~115~~	18631.15	2012	9	SEP 2012	18498.01	133.14	18540.97	.
~~116~~	30208.17	2012	10	OCT 2012	20871.42	9336.75	19882.65	.
~~117~~	24467.94	2012	11	NOV 2012	20222.64	4245.30	19156.38	.
~~118~~	23602.00	2012	12	DEC 2012	23931.04	-329.04	23080.46	.
~~119~~	24289.32	2013	1	JAN 2013	23415.29	874.03	22453.58	.
~~120~~	38609.66	2013	2	FEB 2013	34826.36	3783.30	33481.12	.
121								

We can see that there are two new variables in the file, one for the predictions and one for the errors. The predictions have values for every case, including those being filtered, but there are no residuals for these remaining seven cases, as SPSS is treating these cases as nonexistent when constructing the model and forecasting. In other words, it ignores the actual value of catalog for cases in the last week when developing the model.

True Forecasting with the Model

Let's assume for this example that the model is adequate. We want to finish this example by demonstrating how to forecast beyond the last year in the data.

The first step is to use all the data again. Although it was important to use the last week as a holdout sample to test the model, we want to use all available data when making predictions.

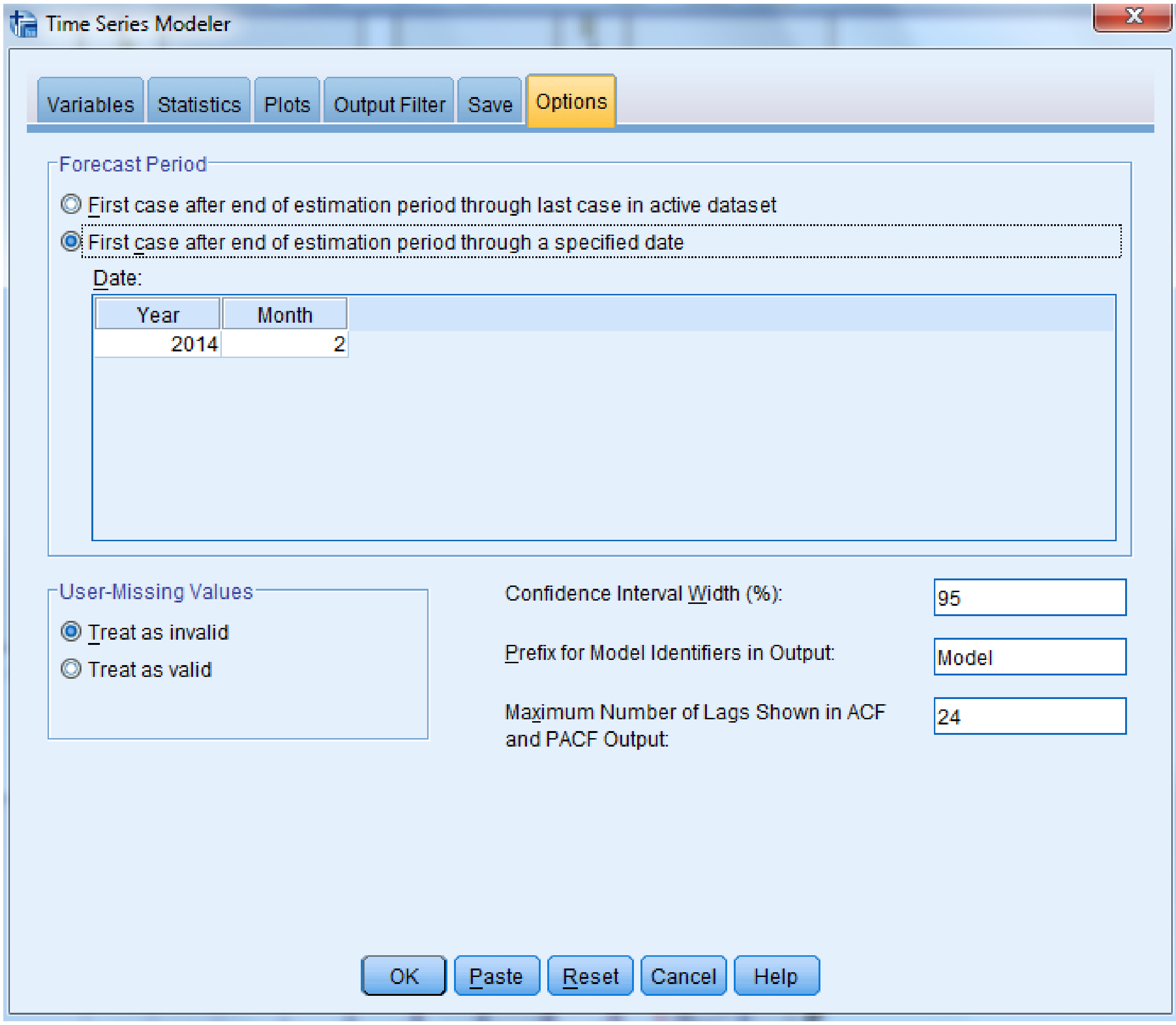
Click **Data…Select Cases**
Click **All cases** option button (not shown)
Click **OK**

Now we rerun the model.

Click **Dialog Recall button**, and then click **Time Series Modeler**
Click **Plots** tab
Click the two check boxes for **Confidence intervals** under Plots for Individual Models area to deselect them
Click **Options** tab
Click **First case after end of estimation period through a specified date** option button
Type **2014** in the Year box; type 2 in the Month box

We can extend the forecast period beyond the end of the data by specifying a date through which forecasts should be made. The date is specified in terms of the current date definition in the data.

Figure 5.26 Specifying a Forecast Period For Future Dates



Click **OK**

The output is not critical to review, except for the Forecast table and the sequence chart. The Forecast table contains the exact forecasts, with upper and lower confidence limits, for year 11. These are the values that the parcel company can use to determine the number of staff required for the upcoming year.

Figure 5.27 Forecasts for Year 11

Forecast

	Model		
	Sales of Men's Clothing-Model_1		
	Forecast	UCL	LCL
Mar 2013	22261.78	28962.89	15560.67
Apr 2013	20679.95	27393.91	13965.98
May 2013	20979.81	27706.61	14253.01
Jun 2013	20245.51	26985.12	13505.90
Jul 2013	20788.45	27540.85	14036.05
Aug 2013	20875.04	27640.19	14109.88
Sep 2013	20767.44	27545.33	13989.55
Oct 2013	23132.76	29923.37	16342.15
Nov 2013	21905.21	28708.51	15101.92
Dec 2013	25350.41	32166.37	18534.45
Jan 2014	24855.08	31683.68	18026.48
Feb 2014	36212.03	43053.25	29370.81

We can see a graphical depiction of this in the sequence chart. As expected, the forecast values for year 11 have the expected growth, with the peak being in the winter months.

Figure 5.28 Sequence Chart of Observed, Fit, and Forecast Values