ARIMA Models

Objectives

- Create a custom ARIMA model using Time Series Modeler.
- Then use the Expert Modeler automatically to create an ARIMA model for the same series.

Data

The data used in this lab are seasonally adjusted unemployment rate (%) of females over 25 years old, since January 1983. Open the "Unemployment rate.sav" data file.

Introduction

In this lab we will consider time series models in which future values of the series are predicted from previous series values at specified lags. Such models are called autoregressive (regression of lags of a time series variable on the variable itself).

Since autoregressive models represent one component of the more general ARIMA model, this chapter serves as an introduction to ARIMA modeling. We will apply autoregressive models to a series containing unemployment data. First we will try to create an ARIMA model from information gained in exploratory analysis, including autocorrelation plots. Then we will let the Expert Modeler create an ARIMA model and compare the two.

Identifying an ARIMA Model for Unemployment Rate

You understand the structure of an ARIMA model and the process of model identification (from the lecture on ARIMA models). We will use this knowledge to predict the monthly female unemployment rate. Note: The data have already been seasonally adjusted.

After the dates are defined, we will request a sequence chart for the unemployment data.

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

Figure 9.1 Sequence Chart Dialog Box

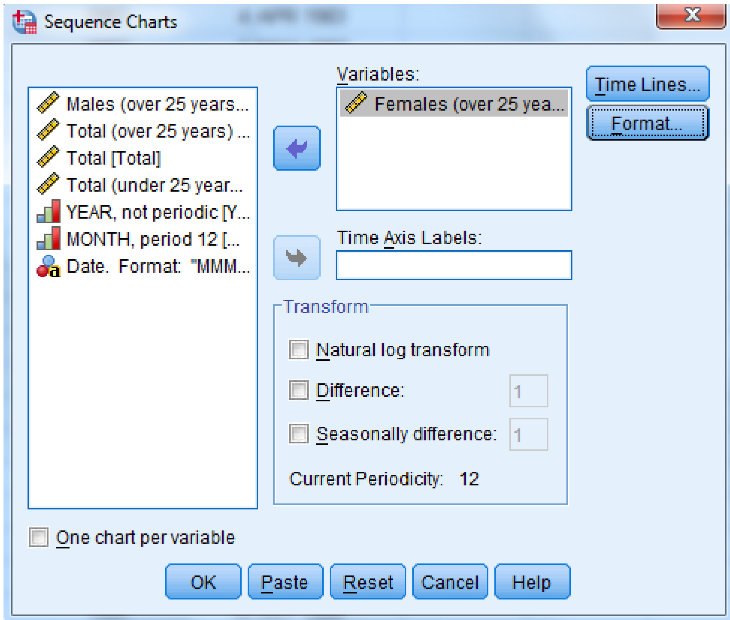
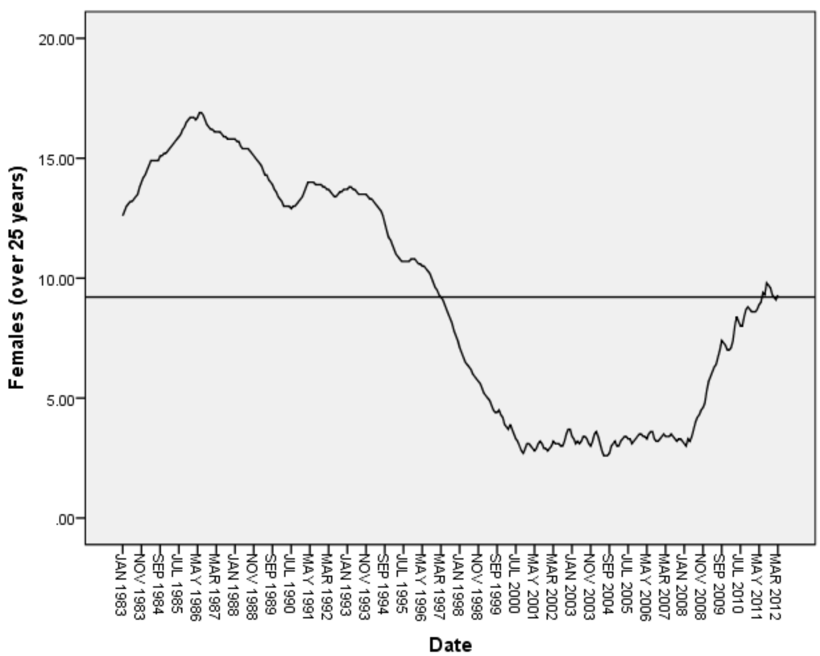


Figure 9.2 Sequence Chart for Unemployment Rate Series

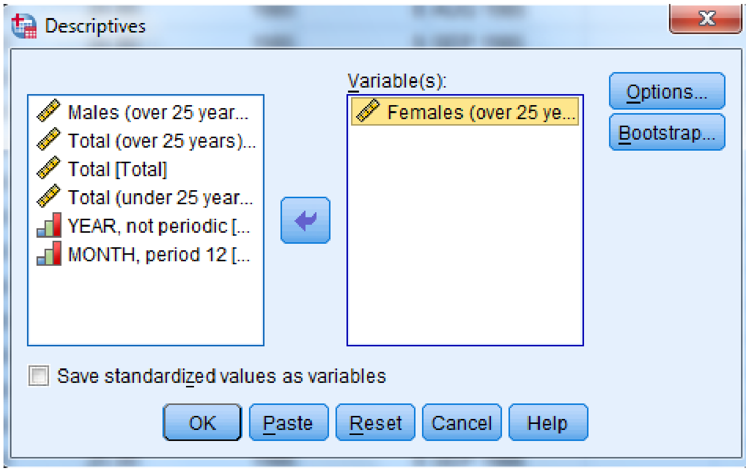


The series has an obvious trend. The series increases, decreases and then increases again. The mean about 9 percent or so. It has several peaks and valleys within the overall trend, but it is not clear what the periodicity is. Seasonally adjusted unemployment reaches its historic high in middle of 1986, and the values in that year are much higher than in other years. At the end of the series, unemployment is back meandering around the 9 percent level.

Next we will request basic descriptive information about the unemployment data.

Select **Analyze…Descriptive Statistics…Descriptives**
Move **Female** into the Variables list box

Figure 9.3 Descriptives Dialog Box



Click **OK**

Figure 9.4 Descriptive Statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Females (over 25 years)	351	2.60	16.90	9.2120	4.92036
Valid N (listwise)	351				

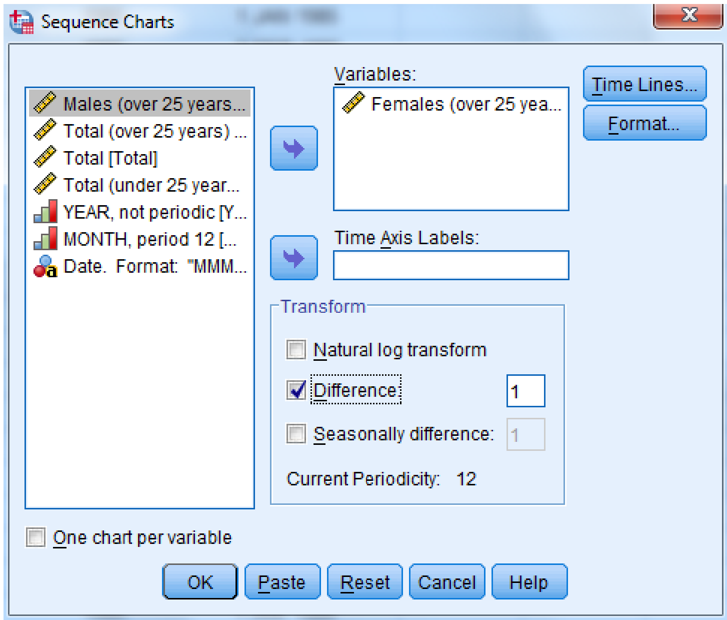
As we noted, the overall mean of the series is just over 9% percent. We see that the maximum value in a quarter (in 1986) was 16.9 percent.

Due to there been trend in the series, the next step is to create the differenced series and plot. This can be done within the Sequence Charts dialog.

Click Dialog Recall tool, and then click **Sequence Charts**
Click **Difference** check box

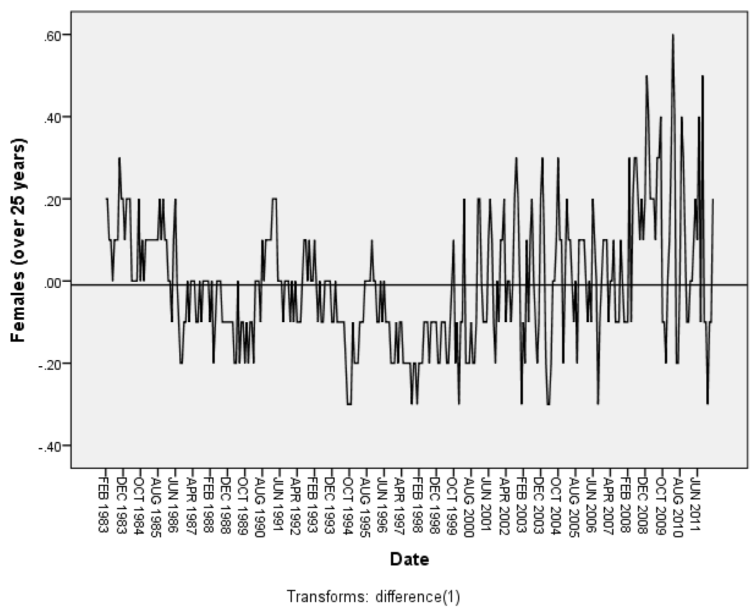
By default, first order differencing will be done.

Figure 9.5 Sequence Chart Dialog Requesting First Order Differencing



Click **OK**

Figure 9.6 Sequence Chart of Differenced Unemployment Rate



There are at least two things to note about the chart. First, the series mean appears to be constant and hover around 0. It is clearly more stable than the original series. Second, there are several spikes in the data, especially in the late 2000's. The variance of the series appears to be greater in that period, or there are outliers at particular time points. We will keep this in mind as we move forward.

The results from the sequence chart indicate that the order of differencing required is 1, and that we will need to estimate models of the form ARIMA(p,1,q).

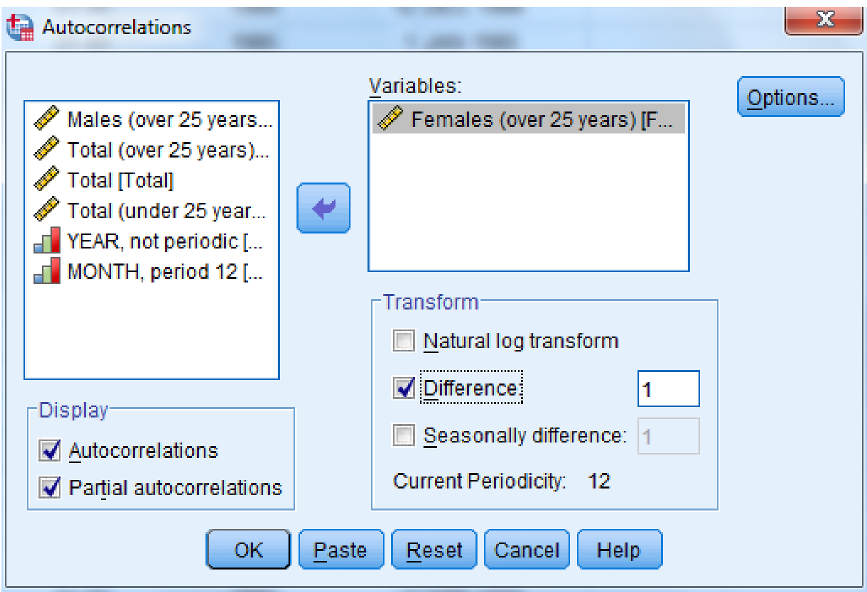
Autocorrelations and Model Identification

The next step in the study of the unemployment rate series will be to look at the autocorrelation function plots.

Previously we used these plots to examine the errors from time series models. Now we want to use them to help determine the correct form of our ARIMA model.

Select **Analyze…Forecasting…Autocorrelations**
Move **Female** into the Variables list box
Click **Difference** check box

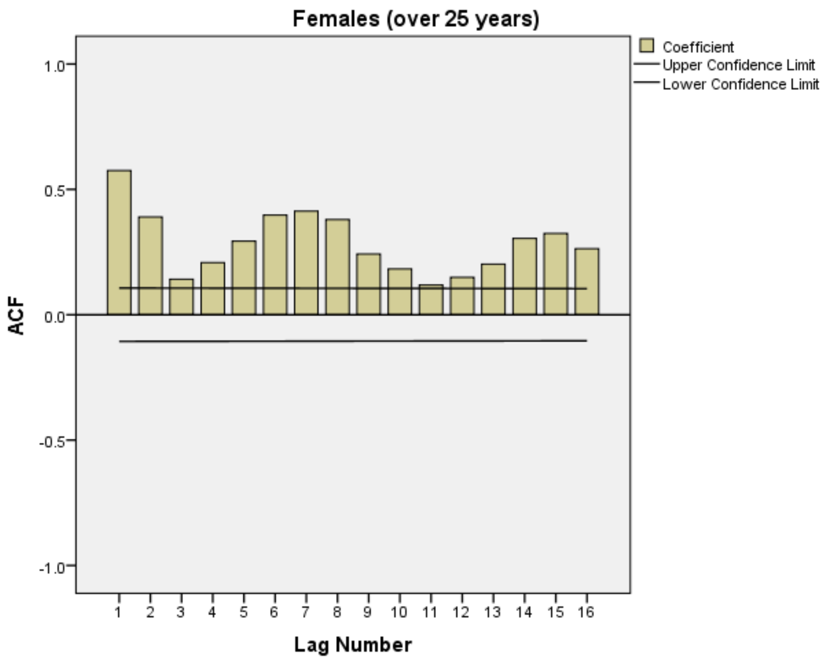
Figure 9.7 Autocorrelations Dialog



We difference *Female* when creating the autocorrelations.

Click **OK**

Figure 9.8 Autocorrelation Plot of Unemployment Rate

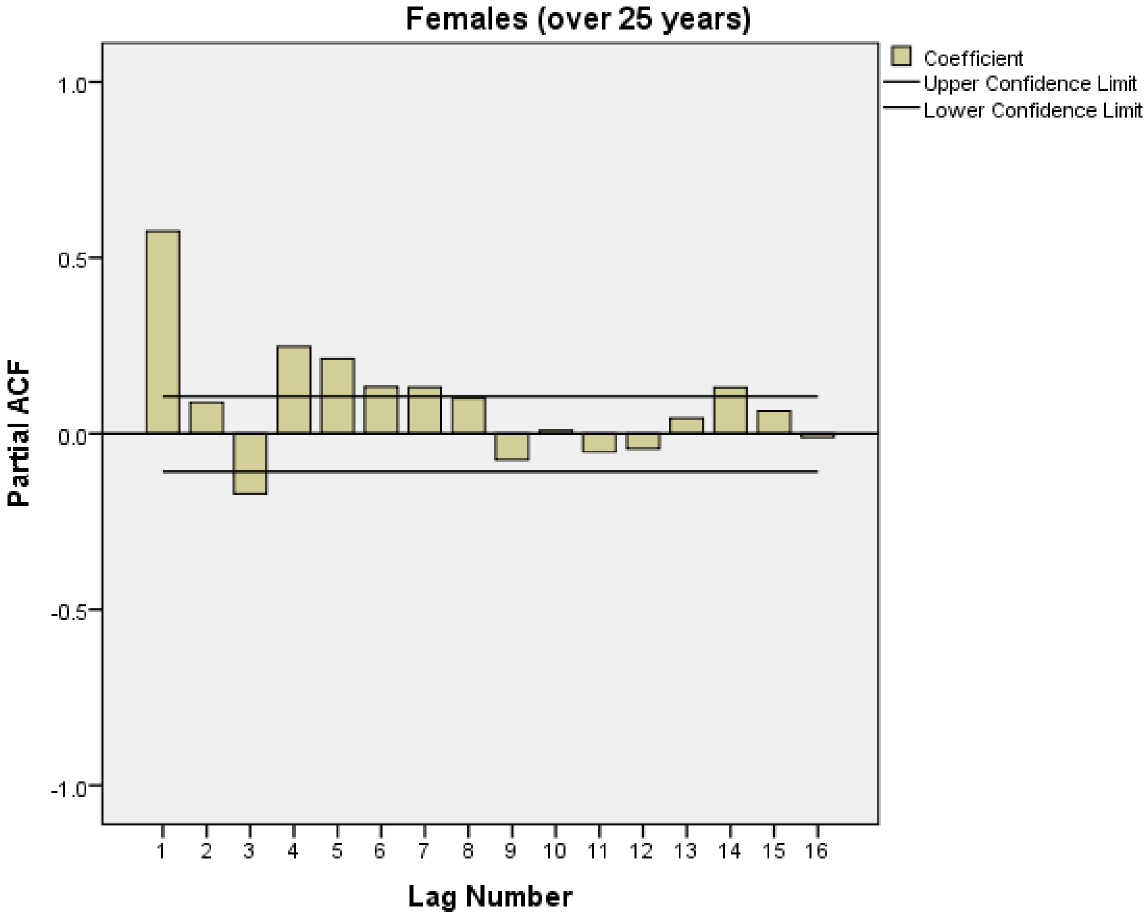


The first autocorrelation is 0.575, which is very large. The pattern is not one of exponential decline to zero, or a sinusoidal pattern. Nor is it of one or two significant lags, with all others being nonsignificant.

Since the data are monthly, you should check the autocorrelations at the seasonal lags of 12, 24, and so on. Although the data have been seasonally adjusted, there is no harm in checking for seasonality. In this plot, there does not seem to be evidence of seasonality. However, lag 12 is significant.

Let’s next look at the PACF plot in Figure 9.9. Here we find another pattern that isn’t simple. The first lag is significant and so it the third. But the first lag is positive, the third is negative. There is no sinusoidal pattern, but higher order lags, including 4 and 7, are significant. Again, there is a bit of evidence for seasonality.

Figure 9.9 Partial Autocorrelation Plot of Unemployment Rate



What does one do in a case like this? Since both the ACF and PACF plots did not fit the simple pattern we outlined previously, the best guess is that both autoregressive and moving average terms are necessary in the model. How many terms is very uncertain, although one could imagine including terms of order 2 at least to create a ARIMA(2,1,2) model. And then there is the matter of seasonality.

Since there are several combinations of the values of p and q that could be tried, it is best to have a plan in mind and begin with simpler models. And, in this case, we will ignore the potential seasonality because we know the government economists have carefully adjusted these data for that factor.

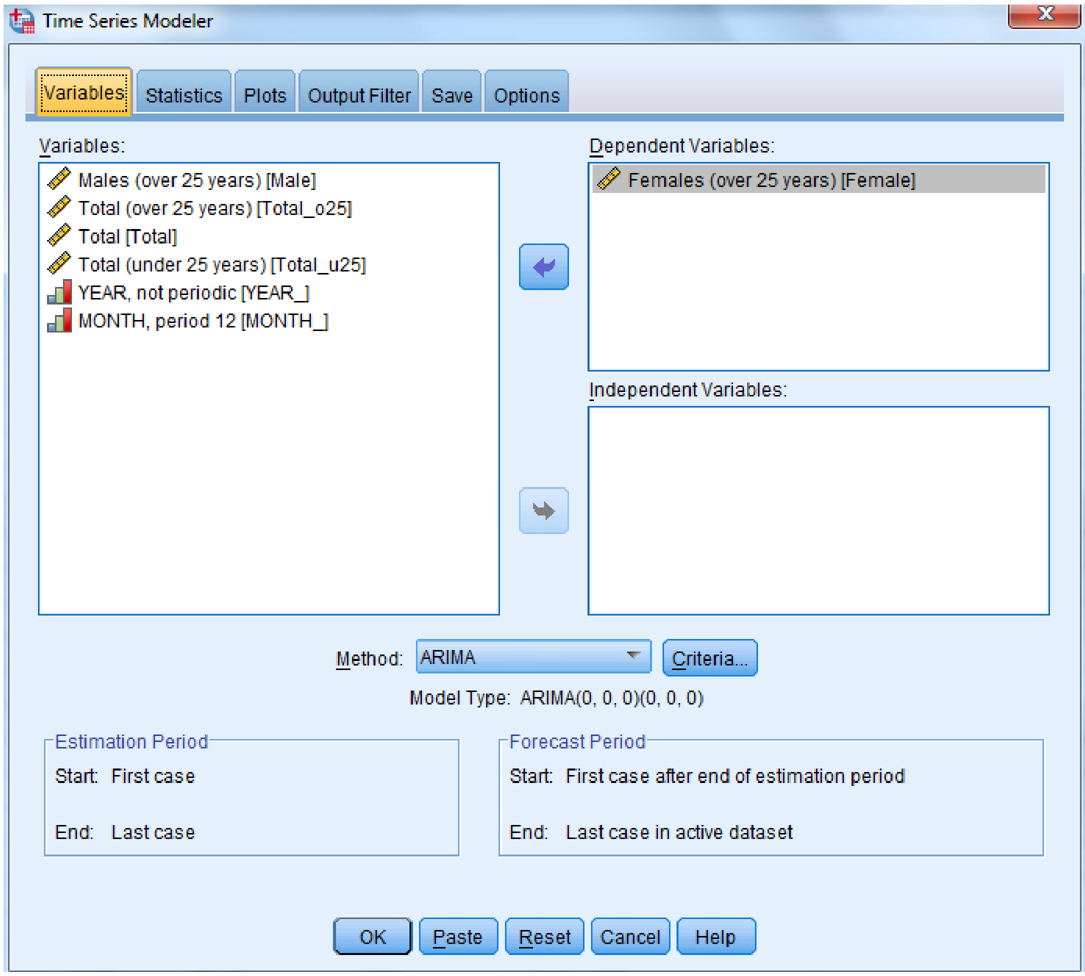
Thus, we will begin with an ARIMA(1,1,1) model in Time Series Modeler. If that seems to be inadequate, we will try other models.

Estimating the ARIMA Model

At the end of the chapter we will request an automatic model with the Expert Modeler. Here, we specify the model ourselves.

Select **Analyze…Forecasting…Create Traditional Models**
Move **Female** into the Dependent list box
Select **ARIMA** from the dropdown for Method

Figure 9.10 Main Time Series Modeler Dialog Box

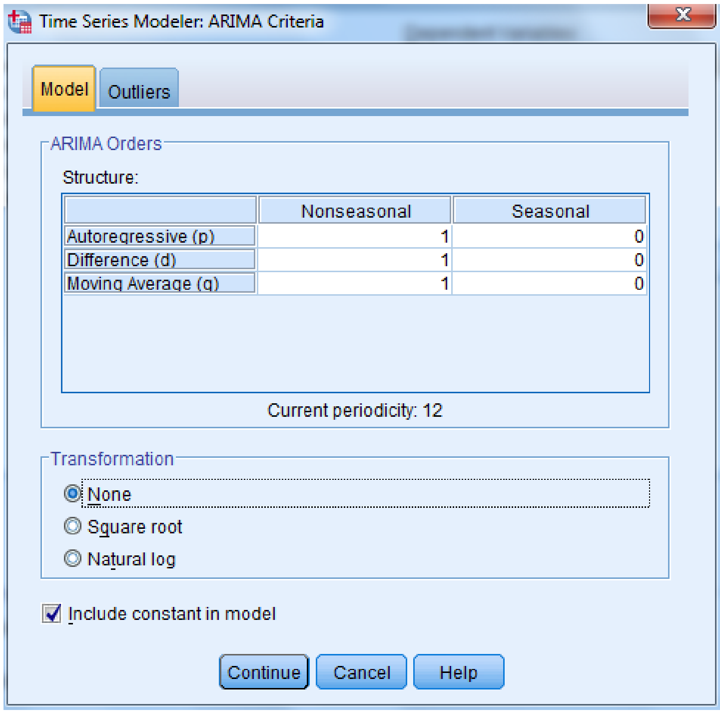


Click the **Criteria** button

The Criteria dialog box changes depending on which method is selected. For ARIMA, you can select values of p, d, and q for both nonseasonal and seasonal terms, request a transformation of the dependent variable, and include a constant in the model (the default choice).

Enter values of 1 for the **Autoregressive (p)**, **Difference (d)**, and **Moving Average (q)**, in the Nonseasonal text boxes

Figure 9.11 ARIMA Criteria Dialog Box

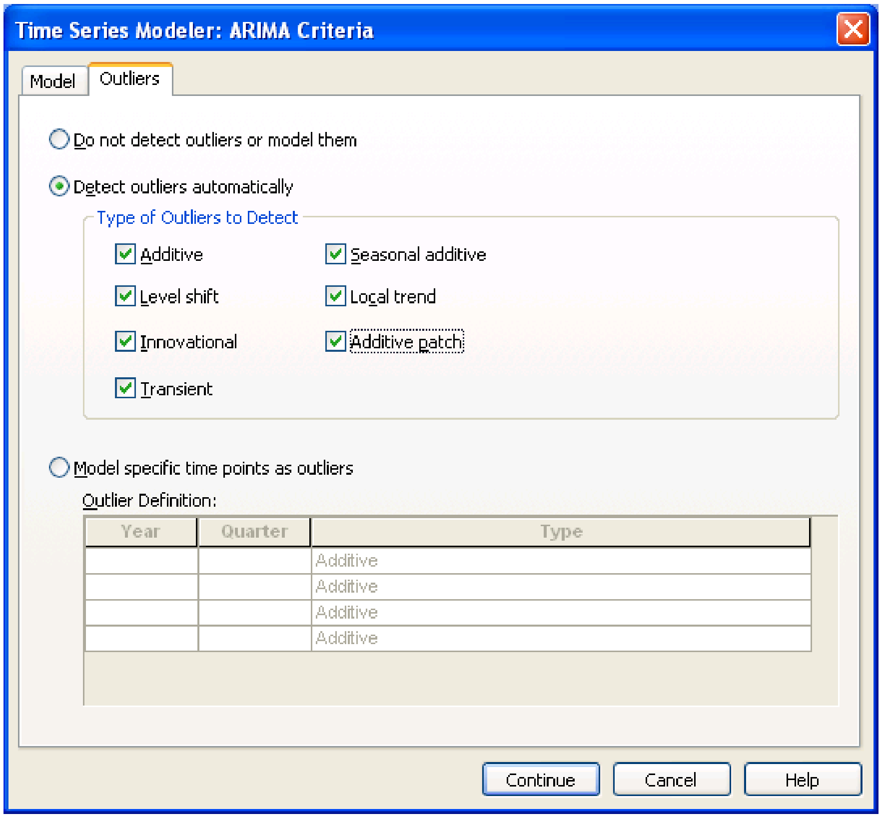


Click the **Outliers** tab

We noticed the large values of the differenced series in the sequence chart (see Figure 9.6). The Modeler can adjust in several ways for outliers, and we will request this correction. In a later lecture, we fully consider the matter of outliers in time series.

Click **Detect outliers automatically** option button
Click **all** check boxes for the **Type of Outliers to Detect**

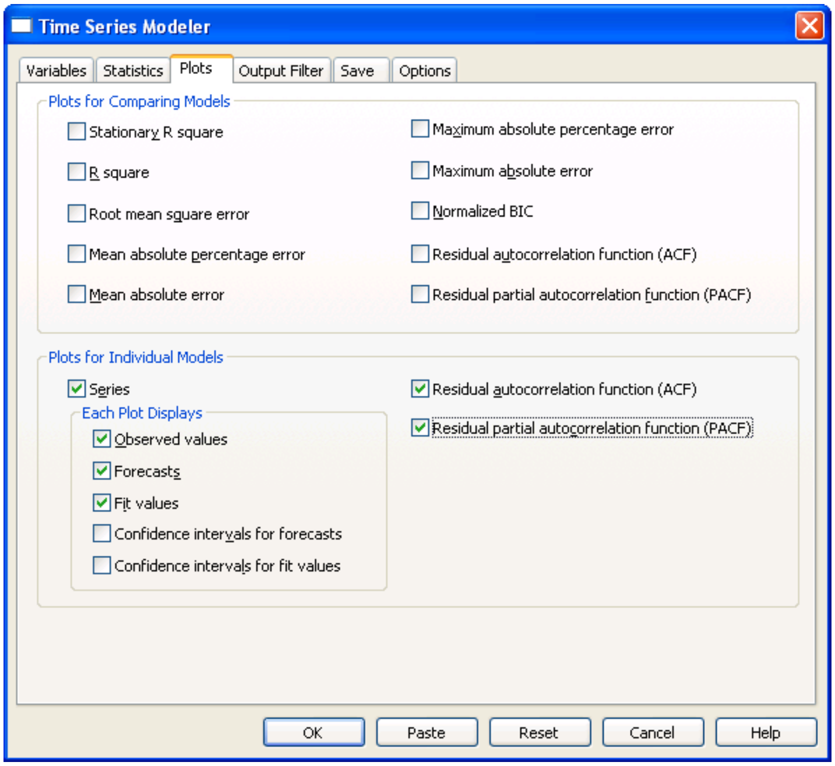
Figure 9.12 Outliers Tab



Click **Continue**
Click the **Statistics tab**
Check the check boxes for **all** the **Fit Measures**
Click **Parameter estimates** check box (not shown)

Click the **Plots tab**
In the Plots for Individual Models section, click **Fit values** and the **two autocorrelation** plots

Figure 9.13 Plots Tab



Now we will save the fit value and its residual.

Click the **Save tab**
Check **Predicted Values** and **Noise Residuals** check boxes (not shown)
Click **OK**

Examining the Results

The Model Description table verifies that we requested an ARIMA(1,1,1)(0,0,0) model. The Modeler always lists the seasonal component, even if zero was specified for its three components.

Figure 9.14 ARIMA Model

Model Description

			Model Type
Model ID	Females (over 25 years)	Model_1	ARIMA(1,1,1)(0,0,0)

Next we will look at the model coefficients.

Figure 9.15 Model Coefficients

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Females (over 25 years)	No Transformation	Constant		-.031	.016	-1.883	.061
		AR	Lag 1	.590	.072	8.159	.000
		Difference		1			
		MA	Lag 1	-.031	.088	-.354	.724

The Constant term is not significant. This is to be expected with a first differenced dependent variable (d=1) that has no trend, as the constant in such a model represents the non-zero average trend. There is none, which was our intent in differencing. The fact that the variable *Female* was differenced once is indicated by the 1 in the *Estimate* column for Difference.

The AR(1) term is significant, while the MA(1) term is insignificant at the .05 level. The value of .590 for the AR(1) term means that the value of unemployment at t–1 is multiplied by that coefficient and added to the other terms in the ARIMA prediction equation to predict unemployment at time t. Larger values of unemployment last month lead to larger values this month, all things being equal.

Conversely, the negative coefficient of –.031 for the MA(1) means that a positive error in the series at time t–1 is followed by a decrease in the value of unemployment at time t. This can be hard to conceptualise, although often we care about the predictive power of the model rather than theoretical understanding.

Figure 9.16 Model Fit Statistics

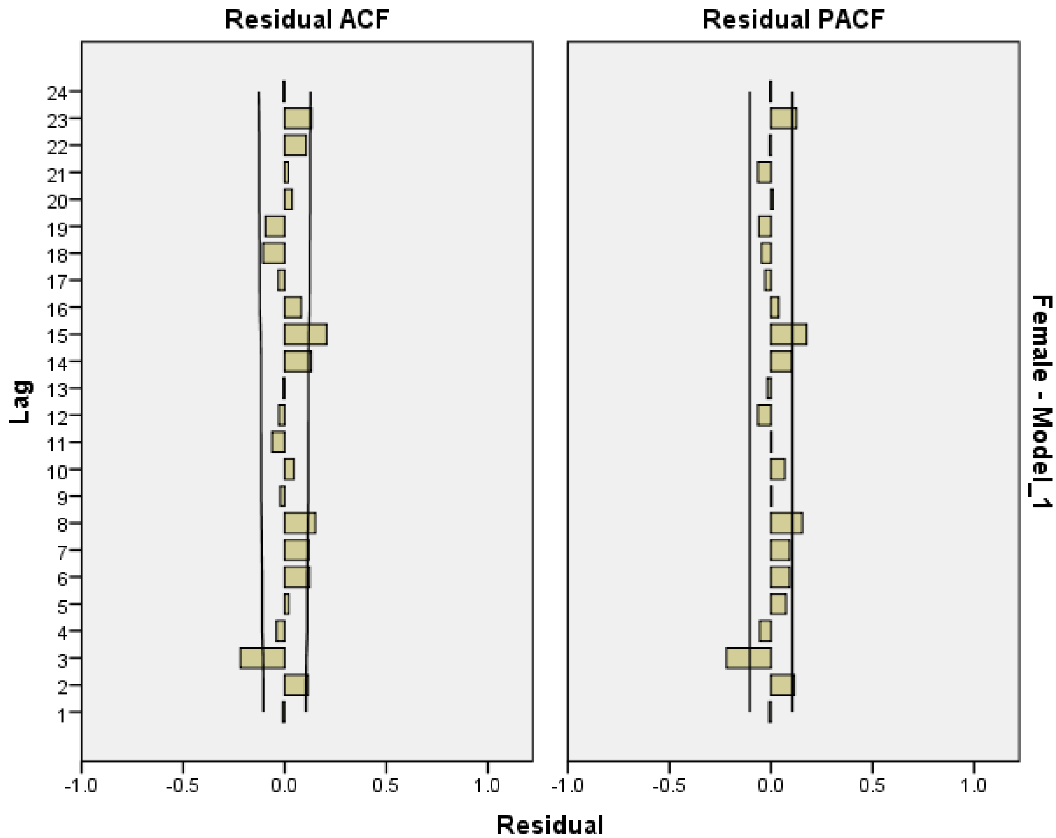
Model Statistics

		Model
		Females (over 25 years)-Model_1
Number of Predictors		0
Model Fit statistics	Stationary R-squared	.478
	R-squared	.999
	RMSE	.114
	MAPE	1.563
	MAE	.087
	MaxAPE	9.330
	MaxAE	.405
	Normalized BIC	-4.233
Ljung-Box Q(18)	Statistics	72.815
	DF	16
	Sig.	.000
Number of Outliers		4

The Stationary R-Squared is .478, which is not terribly high but still probably acceptable. The MAE is .087, meaning that the average error in predicting the unemployment rate is about 0.09%. The MAPE is quite small, about 1.563%. Note that four outliers were identified (all in the late 2000's; see Outliers table).

The Ljung-Box statistic is significant, unfortunately, meaning there is some pattern in the residuals. Next we view the autocorrelation plots.

Figure 9.17 Residual ACF and PACF Plots

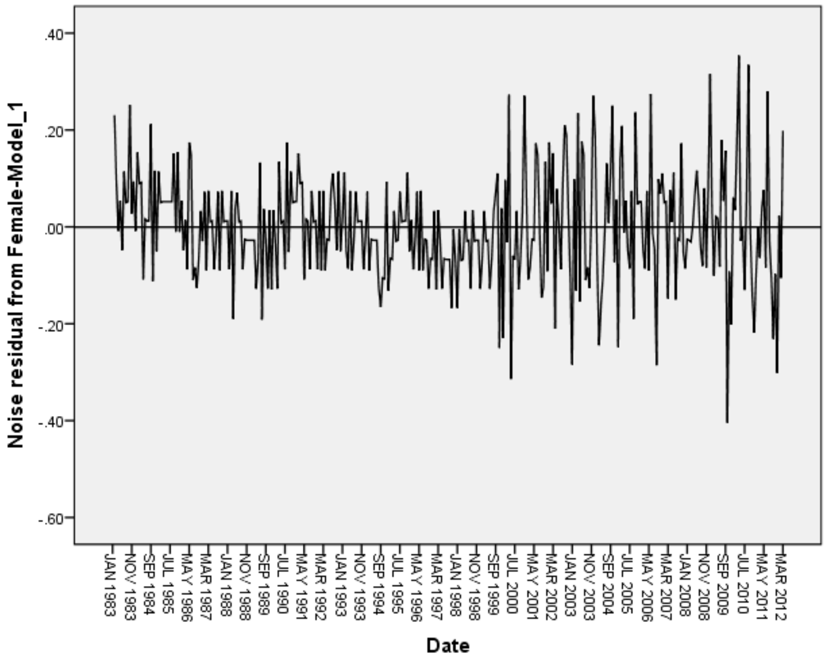


There are significant values of the ACF and PACF at the lower lag 3, this despite the fact the series has been seasonally adjusted.

As a last check, we can view a sequence chart of the residuals.

> Click Dialog Recall Tool, and then click **Sequence Charts**
> Remove **Female** from the Variable list, deselect difference and put in
> > **Nresidual_Female_Model_1_A** (not shown)
>
> Click **OK**

Figure 9.18 Sequence Chart of Residuals for ARIMA(1,1,1) Model



The errors look random enough, although we know from the ACF and PACF plots that they are not. There are still some errors that look like potential outliers, (the model has adjusted for those highlighted earlier).

The model is still not satisfactory, though, because of the significant ACF and PACF values. This means you need to try another ARIMA model. And that process involves deciding on the specific model to try, estimating that model, looking at its fit statistics and errors, and, if that model is still not acceptable, continuing this process until you find an acceptable model. It is important to note that in most cases you will want to find a model that makes conceptual sense. For example, if we found a moving average term of order 14 was significant in an ARIMA model for monthly unemployment, this would be very difficult to understand or justify (errors in predicting unemployment 14/12 years ago affect unemployment now?).

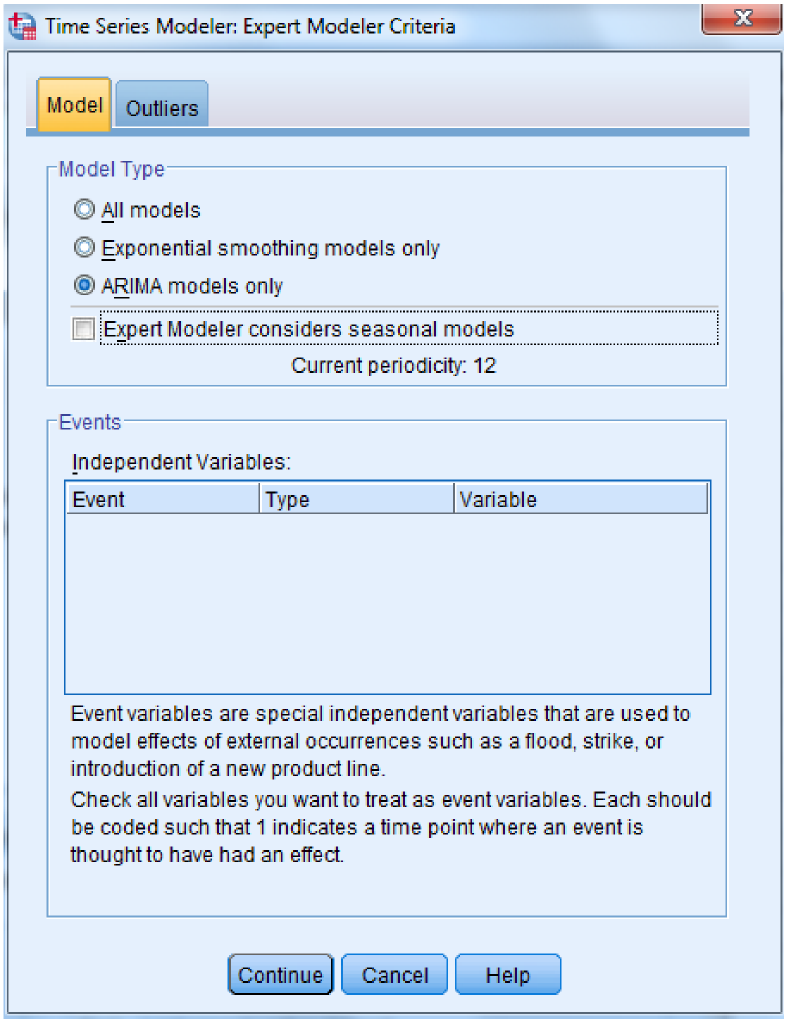
Rather than continue with the intellectually challenging, if also somewhat tedious, job of finding the best ARIMA model, we will invoke the Expert Modeler instead.

Using Expert Modeler to Forecast Female. Unemployment Rate

To start the analysis we return to the Time Series Modeler dialog.

Click **Analyze…Forecasting…Create Traditional Models**
Click the Method dropdown and select **Expert Modeler**
Click **Criteria**
Click **ARIMA models only** option button
Click **Expert Modeler considers seasonal models** check box to deselect it

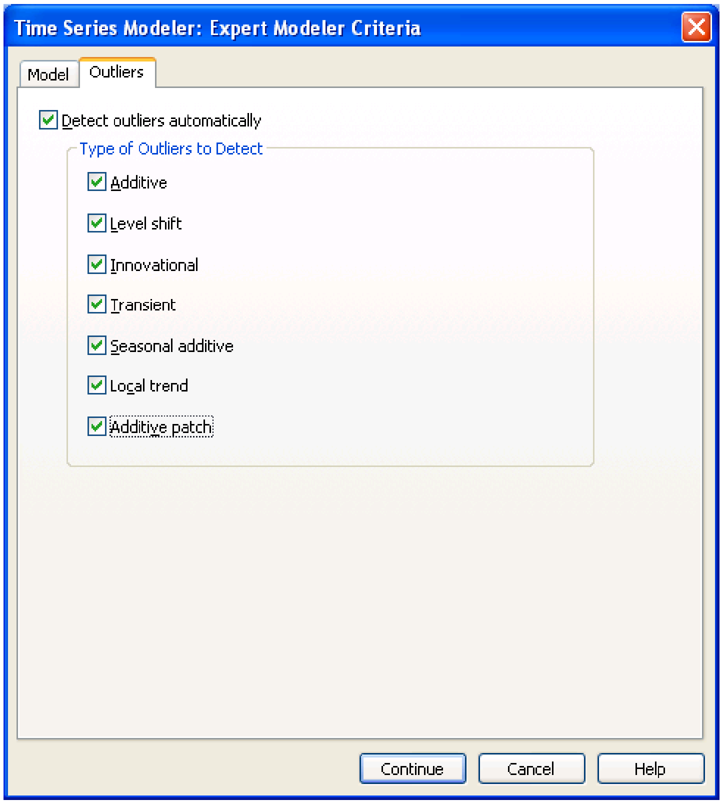
Figure 9.19 Model Tab of Criteria Dialog Box



Although we saw indications of seasonality in the analysis above, we will ignore seasonality for now. We will continue to request adjustment for outliers.

Click **Outliers tab**
Click **Detect outliers automatically**, and then select **all outliers** under Type of Outliers to Detect

Figure 9.20 Outliers Tab of Criteria Dialog Box



All the other settings can remain the same.

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

The Model Description table tells us that the Expert Modeler selected an ARIMA(2,1,1) model. This means we were close to obtaining the correct model on our own, although with much more preliminary work than simply invoking the Expert Modeler.

Figure 9.21 ARIMA Model

Model Description

			Model Type
Model ID	Females (over 25 years)	Model_1	ARIMA(1,1,3)

As before, we now look at the model coefficients.

Figure 9.22 Model Coefficients

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Females (over 25 years)	No Transformation	AR	Lag 1	.981	.017	59.303	.000
		Difference		1			
		MA	Lag 1	.385	.049	7.797	.000
			Lag 3	.445	.048	9.324	.000

There is no constant term now. All the AR and MA terms are significant, but that will always be true because the Expert Modeler only includes significant terms.

Figure 9.23 Model Fit Statistics

Model Statistics

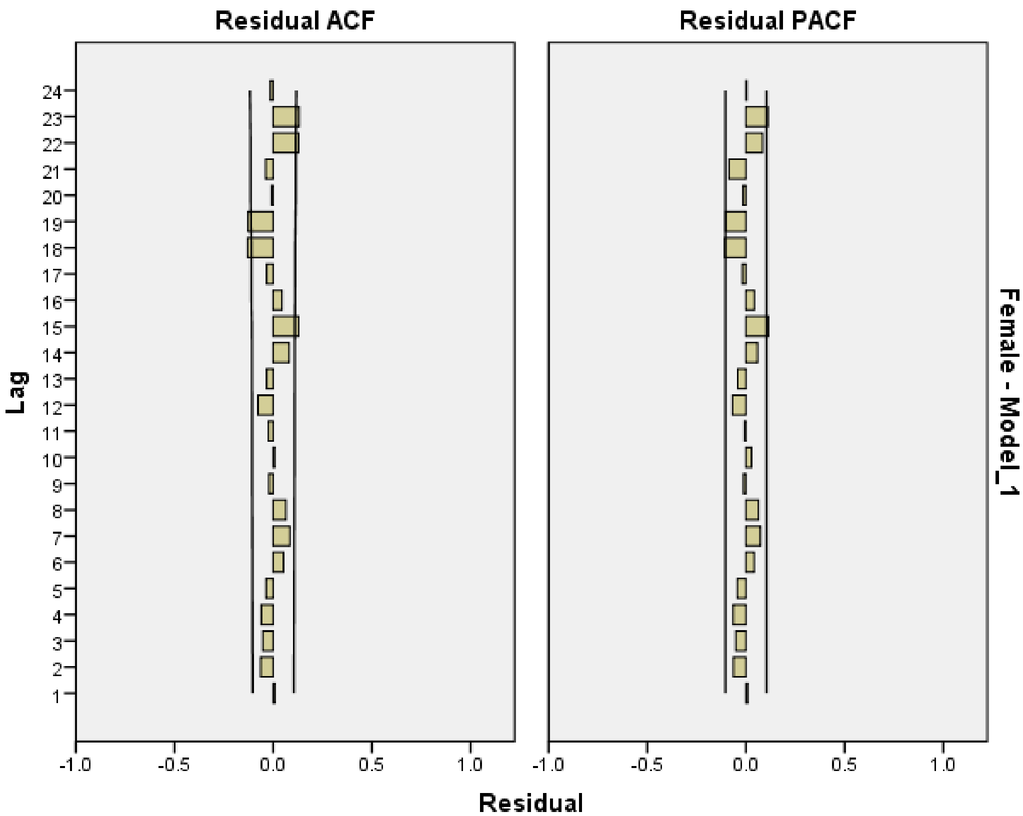
		Model: Females (over 25 years)-Model_1
Number of Predictors		0
Model Fit statistics	Stationary R-squared	.536
	R-squared	1.000
	RMSE	.107
	MAPE	1.492
	MAE	.082
	MaxAPE	9.526
	MaxAE	.376
	Normalized BIC	-4.352
Ljung-Box Q(18)	Statistics	28.065
	DF	15
	Sig.	.021
Number of Outliers		4

The Stationary R-Squared has increased to .536 (compared to .478). The MAE has decreased slightly to .082 (from .087). The MAPE is also smaller at 1.492% (to 1.563%. in our first model). Still 4 outliers were identified.

We can now use Normalized BIC to compare the two models. For our first model, its value was –4.233. For this model it is lower, at –4.352. This also suggests that the ARIMA(1,1,3) model is to be preferred.

The Ljung-Box statistic is still significant (sig < 0.05). It is possible that there is no ARIMA model that can remove all the autocorrelation from the errors for a series, so you simply try to reduce autocorrelation as much as possible.

Figure 9.23 Residual ACF and PACF Plots

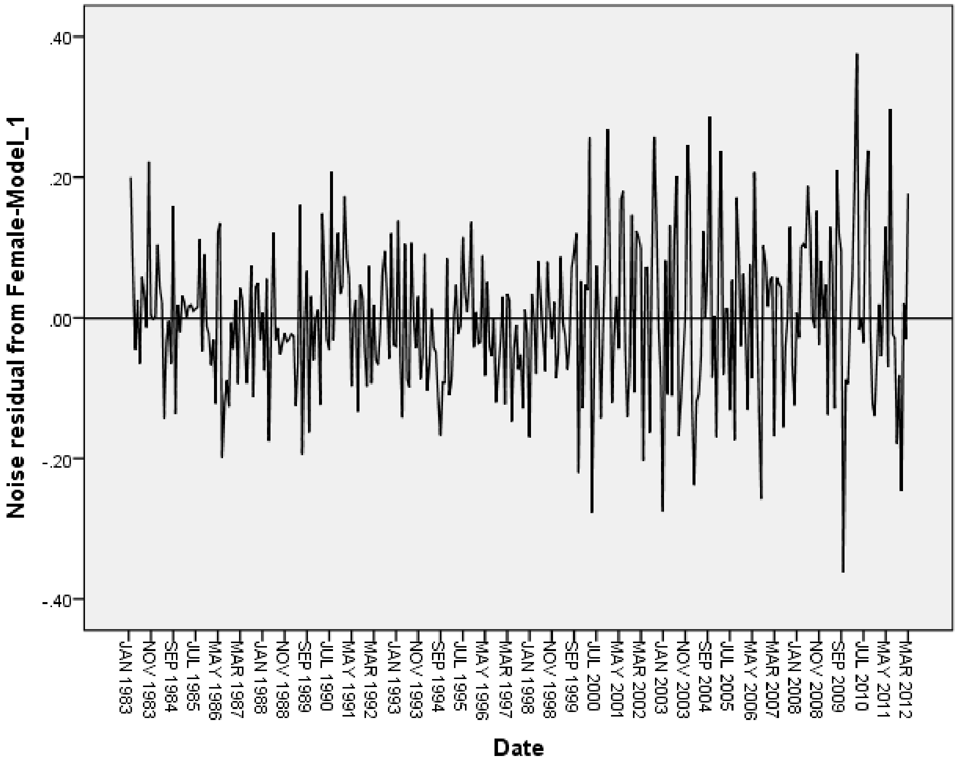


The good news is that the lower order lags don’t show autocorrelation, and the lag 12 values are now insignificant, which is encouraging.

As a last check, we can view a sequence chart of the residuals.

Click Dialog Recall Tool, and then click **Sequence Charts**
Remove **Nresidual_Female_Model_1** from the Variable list and put in
Nresidual_Female_Model_1_B (not shown)
Click **OK**

Figure 9.24 Sequence Chart of Residuals for ARIMA(1,1,3) Model



The errors look random enough.

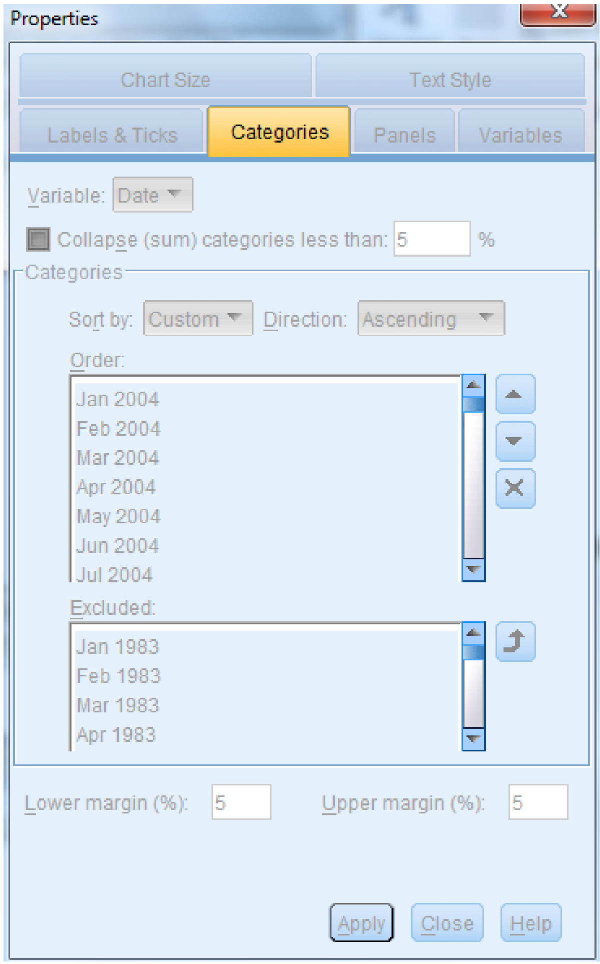
Let's see how well the predictions fit the historical values. The fit is so close that it is difficult to separate the two series on the sequence chart produced by Expert Modeler. Also, you typically are interested in the last few years of a series. So we will edit the sequence chart to focus on the latter portion of the unemployment rate series.

> Double-click on the chart of Observed and Fit values to edit it
> Double-click on the **date labels** on the X axis

This opens the Properties dialog box. Open the **Categories** tab. We need to move most of the series dates to the Excluded box.

> Click on **Jan 1983**
> Shift-click on **Dec 2003**
> Click on the **Exclude** button 

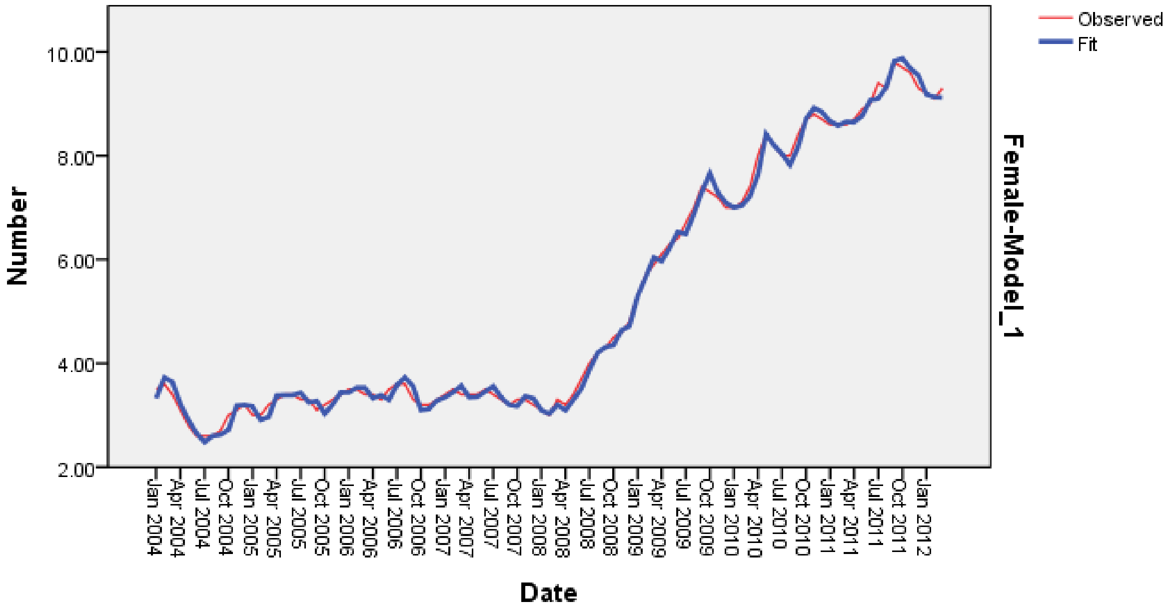
Figure 9.25 Excluding a Range of Dates in Sequence Chart



Click **Apply**
Click **Close**
Close the Chart Editor window

The revised chart is displayed in Figure 9.26, where we have also increased the thickness of the fit series (the predicted values). The fit values generally follow the observed series values closely.

Figure 9.26 Plot of Observed and Fit Values for Unemployment Rate (2004 to 2004)



At this stage we may decide that the ARIMA(1,1,3) model is satisfactory. Nonetheless, you should never accept a model from the Expert Modeler blindly, and you can certainly try other models after this as comparison. But when you have chosen a model, you can use it to make forecasts in the usual manner from the Time Series Modeler Options tab. You can also use a holdout sample with ARIMA modeling, as we discussed in an earlier lab, so a model can be applied to the holdout sample as a further test.