Forecasting Catalog Sales of Men's Clothing Using ARIMA Time-Series Analysis

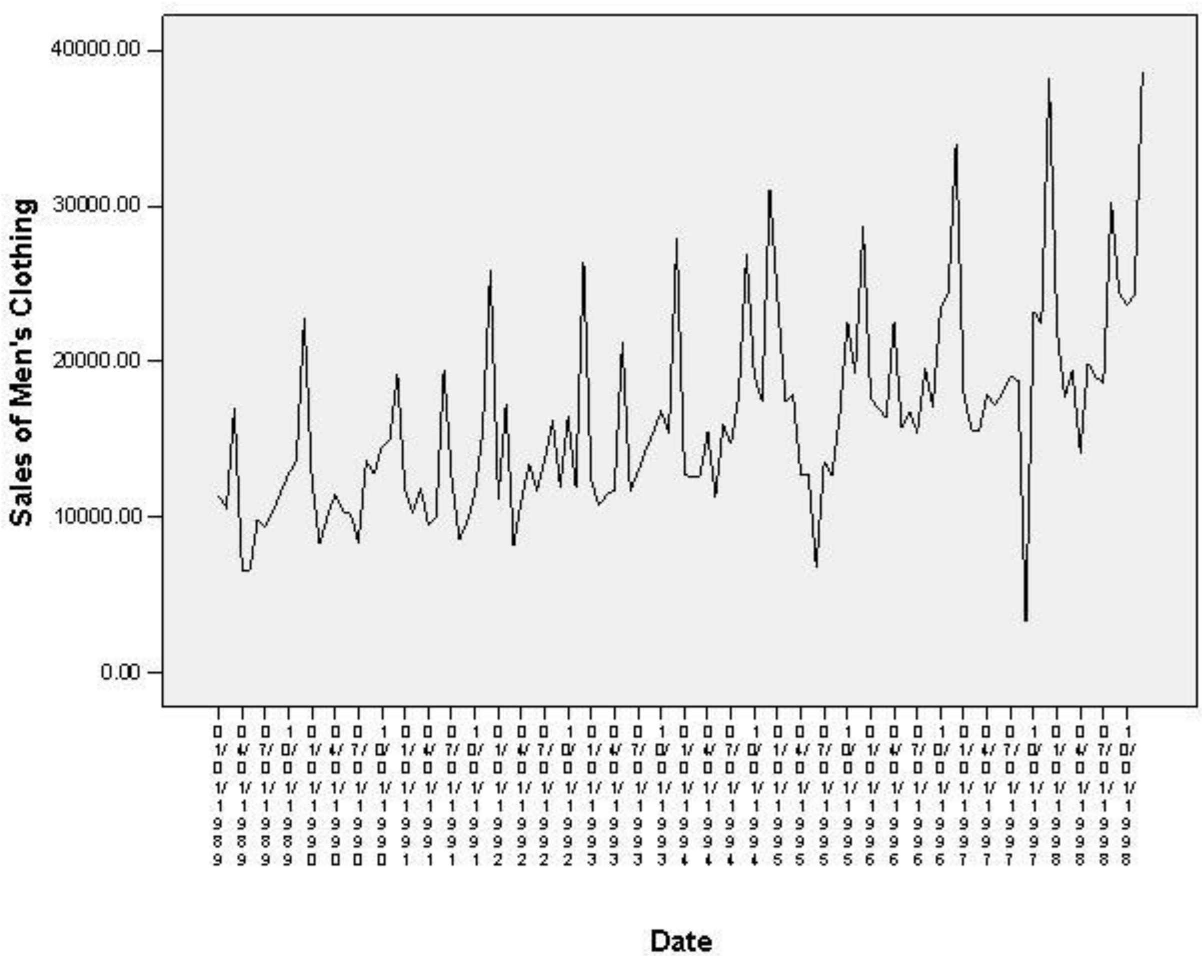
Introduction

A catalog company, interested in developing a forecasting model, 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; and the amount spent on print advertising.

Are any of the predictors useful for forecasting? Is a model with predictors really better than one without? We will use an **ARIMA (AutoRegressive Integrated Moving-Average time-series)** analysis to create forecasting models with and without predictors, and see if there is a significant difference in predictive ability.

ARIMA Analysis

The first step in the model-building process is to plot the series and look for any evidence that the mean or variance is not stationary. (The ARIMA procedure assumes that the original series is stationary.)

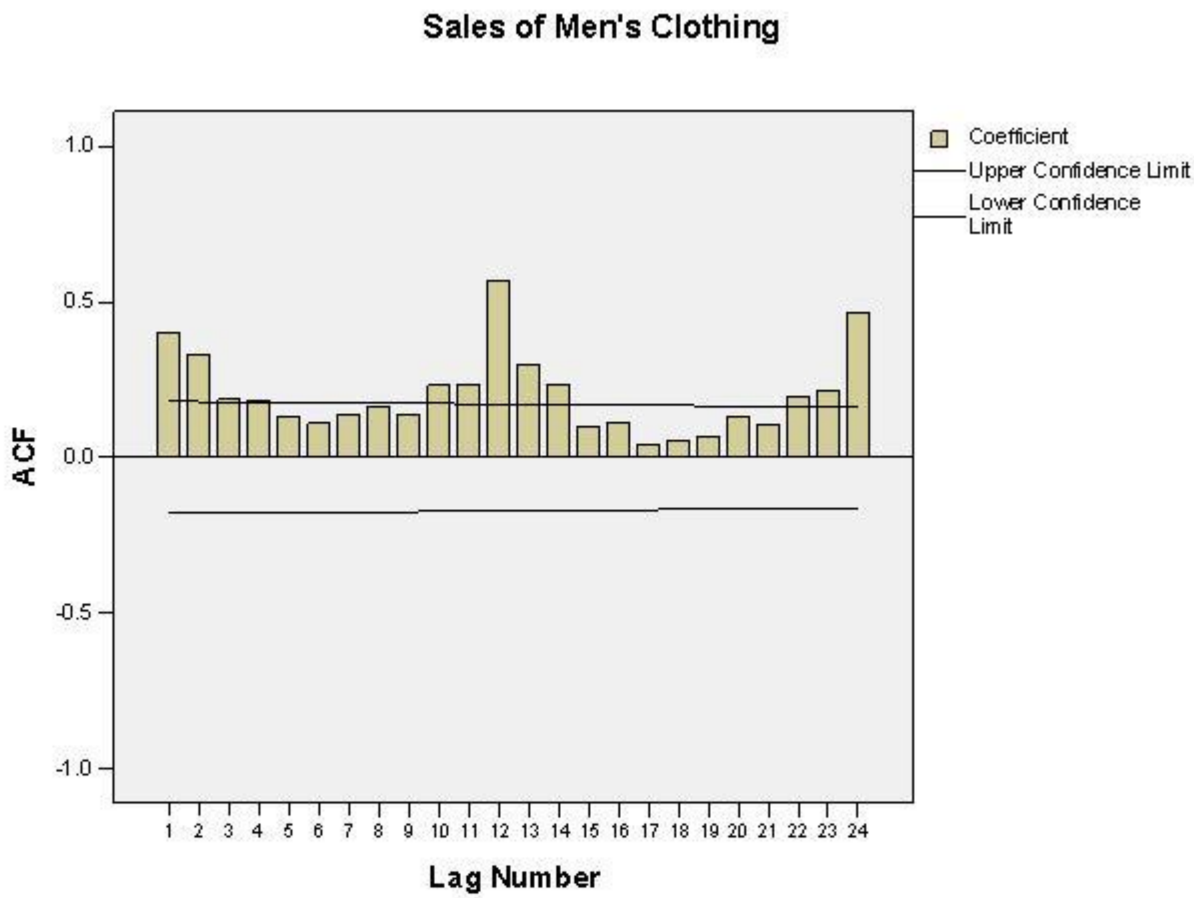


The series shows a global upward trend, making it clear that the level of the series is not stationary. Some degree of differencing will be necessary to stabilize the series level. The variance of the series appears stationary.

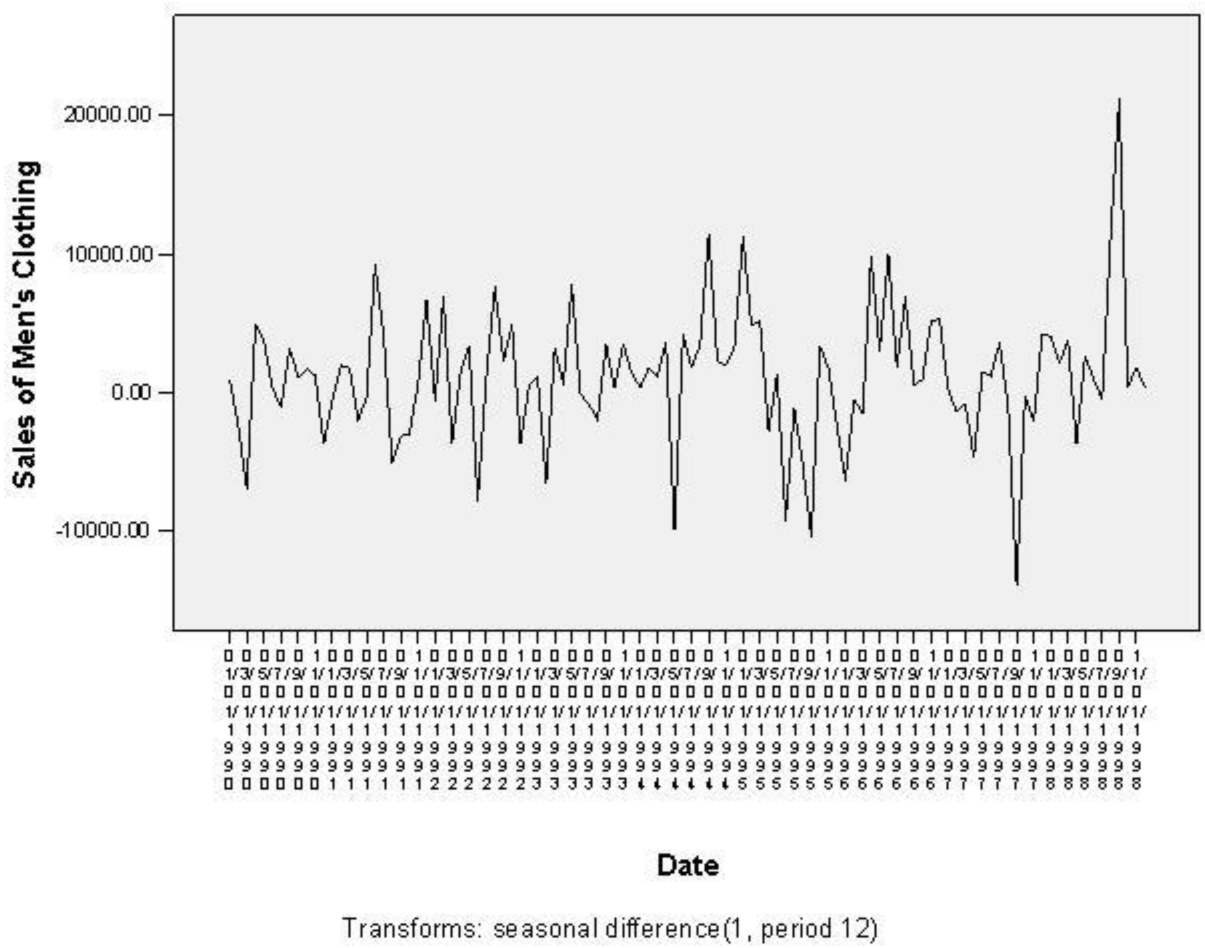
The series also exhibits numerous peaks, many of which appear to be equally spaced. This suggests the presence of a periodic component to the time series. Given the seasonal nature of sales, with highs typically occurring during the holiday season, we shouldn't be surprised to find an annual seasonal component to the data.

We've established that the series has a trend, so some amount of differencing will be required to obtain a stationary series. The likely presence of a seasonal component means that seasonal differencing may be needed. A plot of the autocorrelation function will tell if seasonal differencing is required. If there is a slow decrease in autocorrelations separated by the seasonal interval—for example, a separation of 12 for annual seasonality—then seasonal differencing is necessary to stabilize the series.

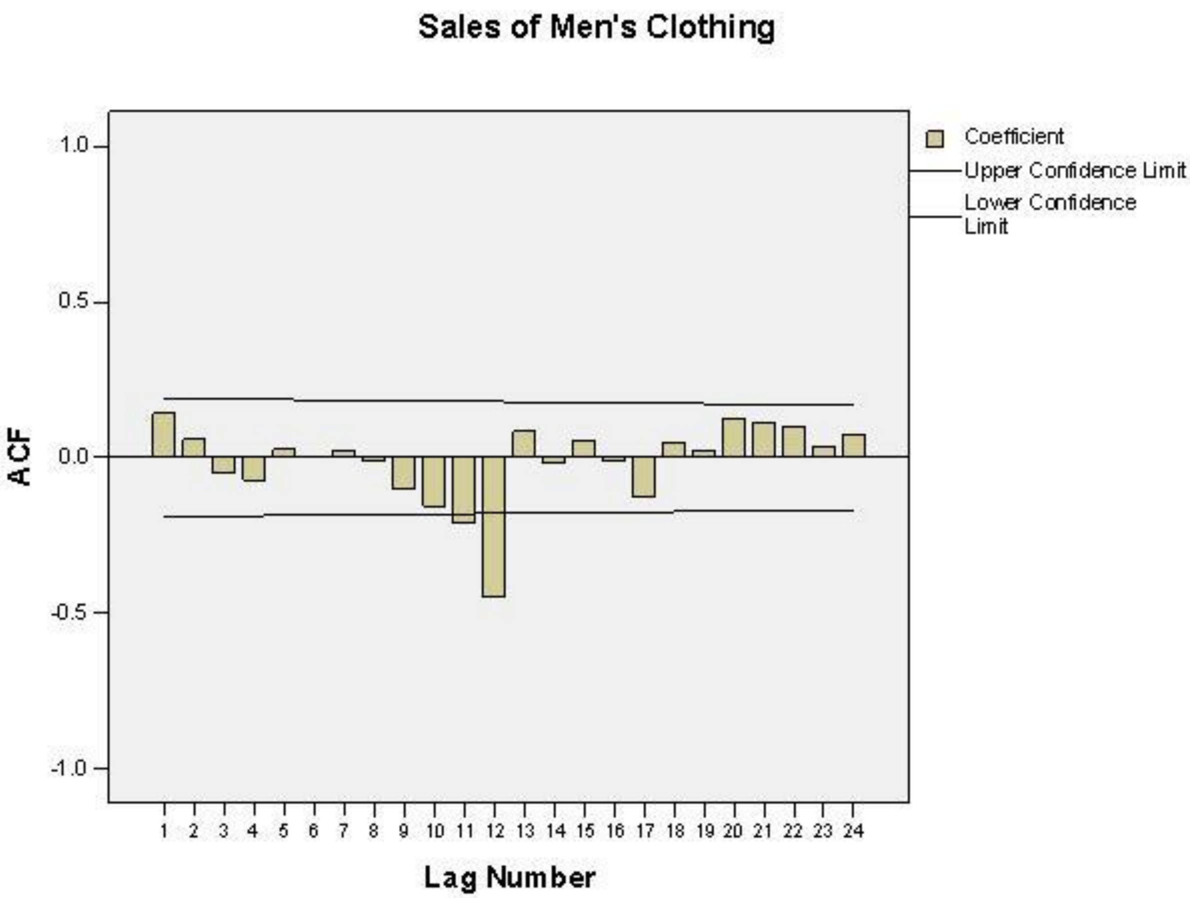
To allow for an investigation of the need for seasonal differencing, the scope of the autocorrelation function (ACF) plot has to be extended beyond the default of 16 lags.



The **autocorrelation** function exhibits significant peaks at lags 1 and 2 as well as significant peaks at lags 12 and 24. Since each data point represents one month, the lag 12 and 24 peaks confirm the presence of an annual seasonal component. The small drop in the ACF at lag 24 relative to the value at lag 12 reflects the fact that the series level is not stationary and indicates that seasonal differencing is necessary. Nonseasonal differencing may also be necessary but will be easier to detect once the series has been seasonally differenced.



Seasonally differencing the data once stabilizes the series level. Notice that the mean of the differenced series appears to be 0. The global upward trend, present in the original series, has been removed. The ACF plot of the seasonally differenced series will show if additional differencing is required.



Seasonal differencing has removed the slow decay of the ACF over seasonal lags. And there is no evidence that further differencing, either seasonal or nonseasonal, is required. The conclusion is that one order of seasonal differencing is sufficient for stabilizing the series.

Next we determine any autoregressive and/or moving-average orders needed to model the series. The strong seasonality of the data suggests that seasonal ARIMA orders are present. An effective approach for isolating seasonal orders is to examine the ACF and PACF plots at the seasonal lags, ignoring, for the moment, the correlations at nonseasonal lags.

The PACF plot shows a significant peak at a lag of 12, followed by evidence of a tail extending beyond lag 48.

The ACF plot shows a significant peak at a lag of 12 without strong evidence of a substantial tail.

The characteristic ACF and PACF patterns produced by seasonal processes are the same as those shown for nonseasonal processes, except that the patterns occur in the first few seasonal lags rather than the first few lags.

The spikes in the ACF/PACF plots at the first seasonal lag (lag 12), coupled with a tail in the PACF plot, indicate a seasonal moving-average ARIMA component of order 1.

Given that we've already identified a seasonal differencing component of order 1, this suggests that an ARIMA(0,0,0)(0,1,1) model may be most appropriate for this series.

The general ARIMA model includes a constant term, whose interpretation depends on the model we are using:

In MA models, the constant is the mean level of the series.

In AR(1) models, the constant is a trend parameter.

When a series has been differenced, the above interpretations apply to the differences.

We've determined that a candidate model is ARIMA(0,0,0)(0,1,1), which is an MA model of a differenced series. Therefore, the constant term will represent the mean level of the differences. Since we know that the mean level of the differences is about 0 for the series of men's clothing sales, the constant term in the ARIMA model should be 0. Therefore we can suppress the estimation of the constant term. This speeds up the computation, simplifies the model, and yields slightly smaller standard errors of the other estimates.

Diagnosing an ARIMA model is a crucial part of the model-building process and involves verifying that the residuals are random. The most direct evidence of random residuals is the absence of significant values of the Box-Ljung Q statistic at lags of about one quarter of the sample size. Since the current sample size is 120, we analyze values in the region of the lag 30 statistic.

Lag	Autocorrelation	Std.Error(a)	Box-Ljung Statistic		
			Value	df	Sig.(b)
1	.145	.095	2.342	1	.126
2	.032	.094	2.455	2	.293
3	-.059	.094	2.855	3	.415
4	-.081	.094	3.597	4	.463
5	.011	.093	3.610	5	.607
6	.013	.093	3.631	6	.727
7	.060	.092	4.061	7	.773
8	.001	.092	4.061	8	.852
9	-.098	.091	5.205	9	.816
10	-.131	.091	7.301	10	.697
11	-.219	.090	13.162	11	.283
12	-.237	.090	20.120	12	.065
13	.132	.089	22.299	13	.051
14	.027	.089	22.389	14	.071
15	.060	.088	22.845	15	.087
16	.009	.088	22.854	16	.118
17	-.111	.088	24.474	17	.107
18	.028	.087	24.575	18	.137
19	.028	.087	24.680	19	.171
20	.092	.086	25.817	20	.172
21	.099	.086	27.165	21	.165

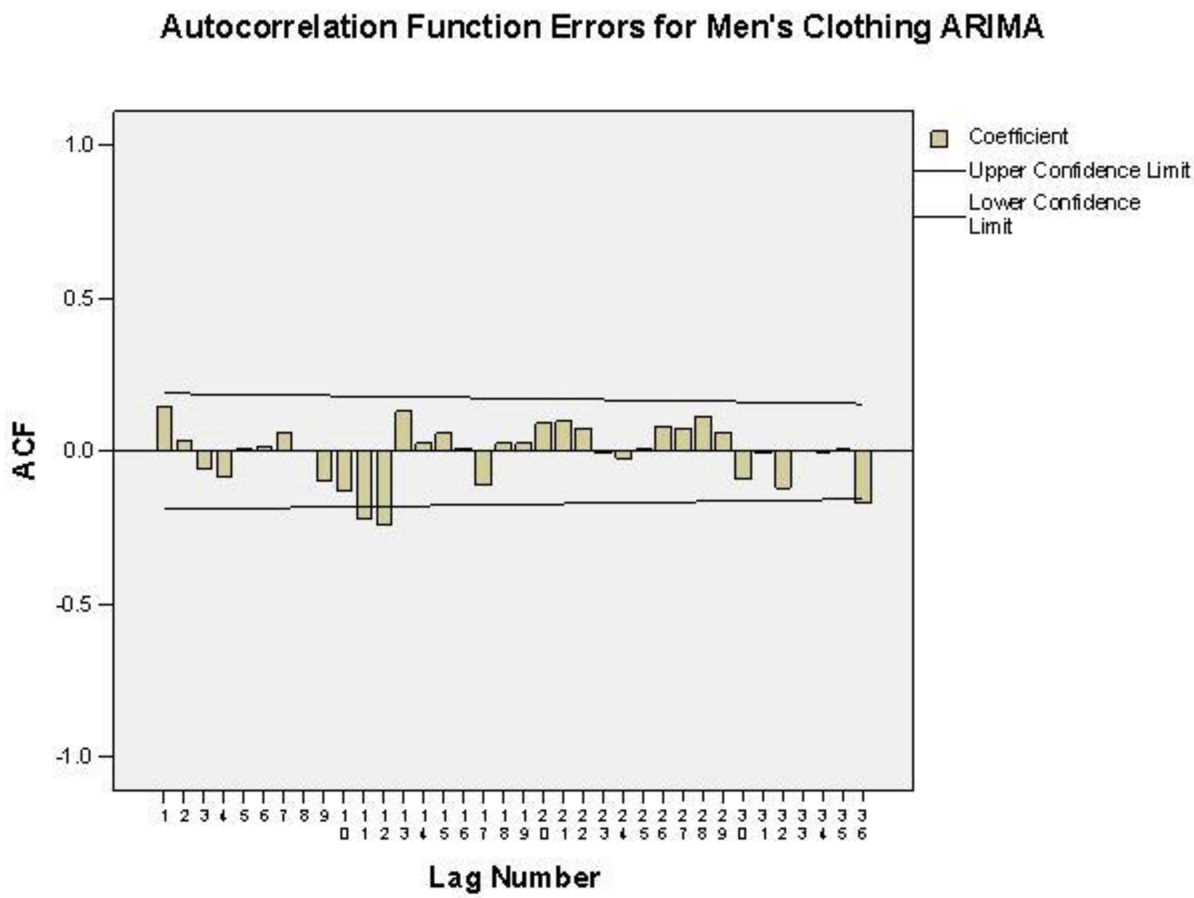
22	.076	.085	27.962	22	.177
23	-.006	.085	27.967	23	.217
24	-.022	.084	28.038	24	.258
25	.007	.084	28.046	25	.306
26	.083	.083	29.036	26	.309
27	.077	.083	29.896	27	.319
28	.114	.082	31.842	28	.281
29	.063	.082	32.431	29	.301
30	-.091	.081	33.681	30	.294
31	-.005	.081	33.685	31	.339
32	-.119	.080	35.895	32	.291
33	.001	.079	35.895	33	.334
34	-.002	.079	35.895	34	.380
35	.005	.078	35.900	35	.426
36	-.168	.078	40.540	36	.277

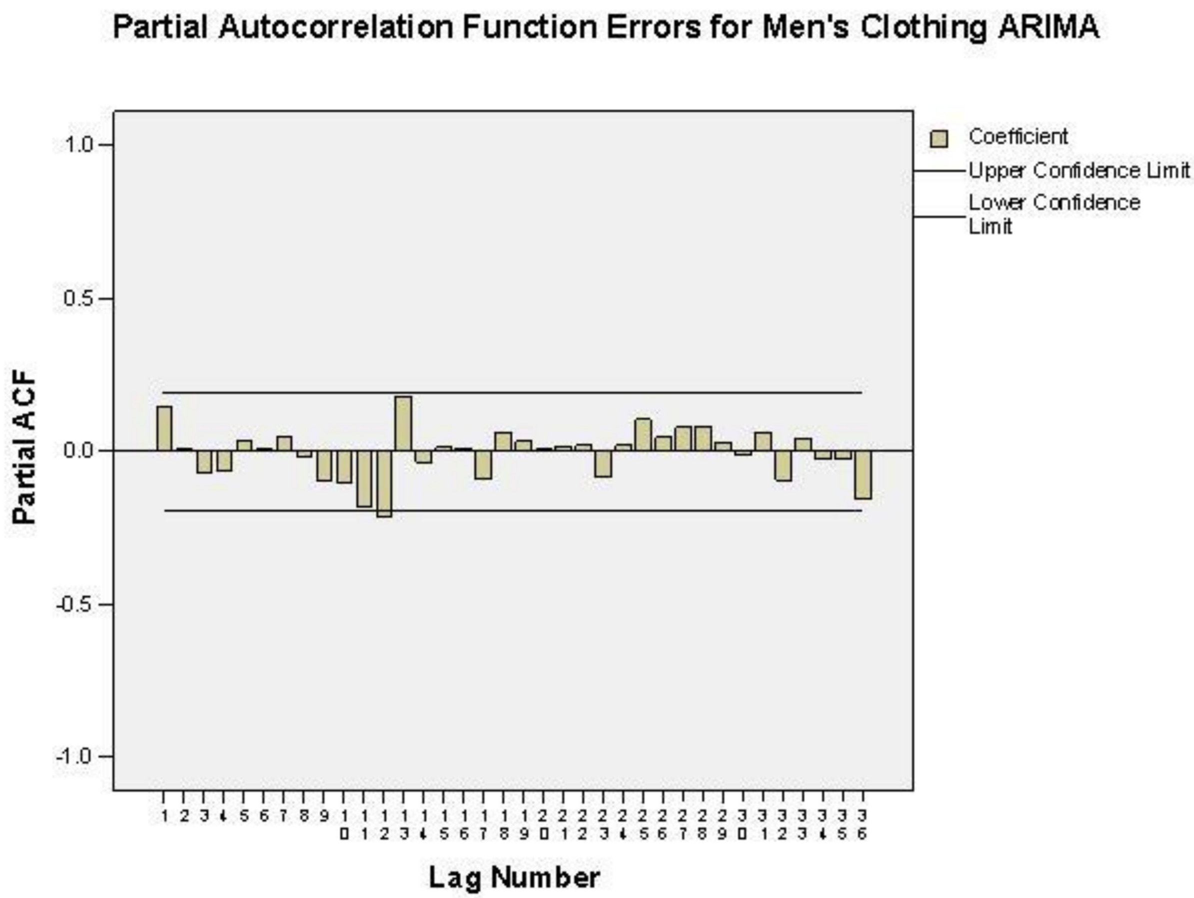
a The underlying process assumed is independence (white noise).

b Based on the asymptotic chi-square approximation.

None of the Box-Ljung values in the vicinity of lag 30 is significant. This confirms that the residuals for the ARIMA(0,0,0)(0,1,1) model are random, which also means that no essential components have been omitted from the model.

In addition, the autocorrelation function errors and partial autocorrelation function errors are within acceptable limits, as shown below.


Autocorrelation Function Errors for Men's Clothing ARIMA
Coefficient
Upper Confidence Limit
Lower Confidence Limit
1.0
0.5
0.0
-0.5
-1.0
ACF
Lag Number



We've determined that an ARIMA(0,0,0)(0,1,1) model does a good job of capturing the structure of the time series; however, the model is based only on the series itself and doesn't incorporate information about the possible predictor series included with the original data set.

Can we build a better forecasting model by treating sales of men's clothing as a dependent variable and treating variables, such as the number of catalogs mailed and the number of phone lines open for ordering, as independent variables? ARIMA treats these predictor, or independent, variables much like predictor variables in regression analysis - it estimates the coefficients for them that best fit the data.

The parameter estimates table provides estimates of the model parameters and associated significance values, including both the AR and MA orders as well as any predictors.

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Seasonal Lags	Seasonal MA1	.595	.105	5.652	.000
Regression Coefficients	Number of Catalogs Mailed	1.054	.183	5.746	.000
	Number of Pages in Catalog	4.633	17.334	.267	.790
	Number of Phone Lines Open for Ordering	313.237	32.335	9.687	.000
	Print Advertising	.368	.056	6.532	.000

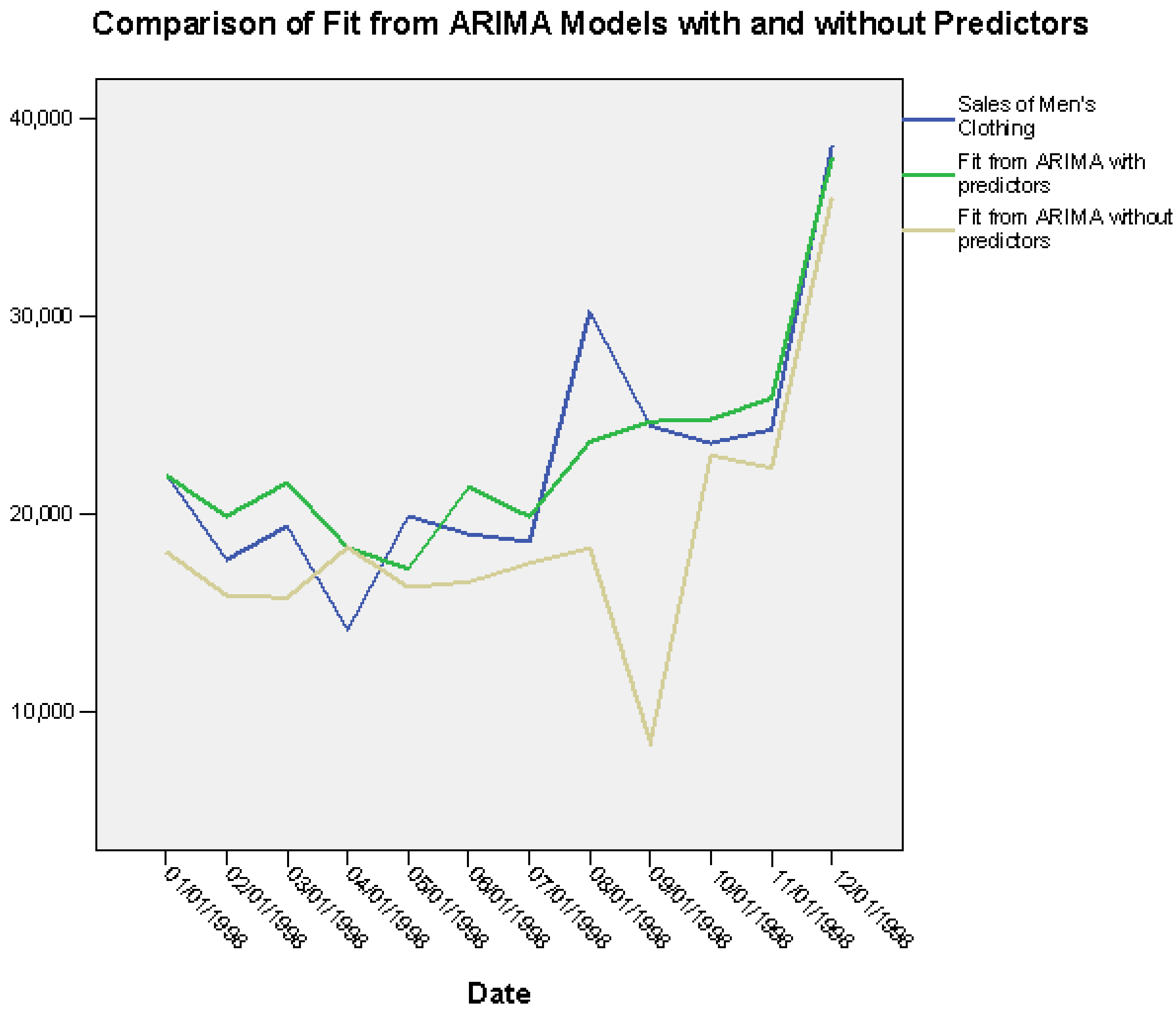
Melard's algorithm was used for estimation.

Notice that the parameter representing the seasonal moving-average component (labeled Seasonal MA1) is significant. This is expected, since we've already determined that it should be part of the model. Note also that the variable representing the number of pages in a catalog is not significant. However, the number of catalogs mailed, the number of phone lines open for ordering, and the coordinated print advertising campaign are all statistically significant influences on catalog sales of men's clothing.

Is the model with predictors really better than the one without predictors? We can test the predictive ability of a model by using holdouts. A holdout is a historical series point that is not used in the computation of the model parameters, thus removing its effect on the computation of forecasts. By forcing the model to predict values we actually know, we can get an idea of how well the model forecasts. This method can be illustrated by holding out the data from January 1998 through December 1998. The data prior to January, 1998, are used to build the model, and the model is then used to forecast sales in 1998.

So we first rerun the ARIMA procedure, with only the significant predictors, using the data from 01/1989 to 12/1997 to determine the best-fit parameters. The analysis also includes predictions of sales of men's clothing during the holdout period (01/1998 to 12/1998), using the parameters from the best-fit model.

Then we also rerun the ARIMA procedure, this time with no predictors, using the data from 01/1989 to 12/1997 to determine the best-fit parameters. Comparison of the model predictions for the holdout period with the actual data is best done by limiting the cases to the holdout period itself, as shown in the graph below.



It is clear from the plot that the ARIMA model with predictors fits the actual data much better than the model without predictors.

Conclusions

We have demonstrated how to build a seasonal **ARIMA** model using the **autocorrelation and partial autocorrelation functions** to identify the ARIMA orders. A number of candidate predictor variables were added to the model and evaluated based on their statistical significance. The final model, keeping only significant predictors, was compared to the model with no predictors. Results clearly showed that the model with predictors did a better job of explaining the variance of the data.

The foregoing case study is an edited version of one originally furnished by SPSS, and is used with their permission.