Student Project: Modelling Monthly Canadian Automobile Sales

Introduction

It was suggested in the homework assignment to Module 20 that sales for automobiles fluctuate from month to month with a seasonal period of one year. In order to test whether this is correct, this project will fit historical data relating to automobile sales in Canada to a seasonal ARIMA process. The fitted ARIMA process will then be used to forecast automobile sales using the same data as a means of assessing its appropriateness. A well-fitting model with a high degree of forecast accuracy can be employed by the automotive industry to better plan product release cycles and coordinate marketing efforts in order to maximise their sales.

Data

This project used monthly data on automotive sales collected and prepared by Statistics Canada from January 1970 to December 2012. The data was extracted from CANSIM Table 079–0003. Each passenger car sold in a month represents one unit of sales in the data. The data was reviewed for reasonableness; however, no further tests were conducted to assess the data's accuracy. No adjustments were made to the data to account for the different number of days in each calendar month or for leap years.

Figure 1: Raw Data

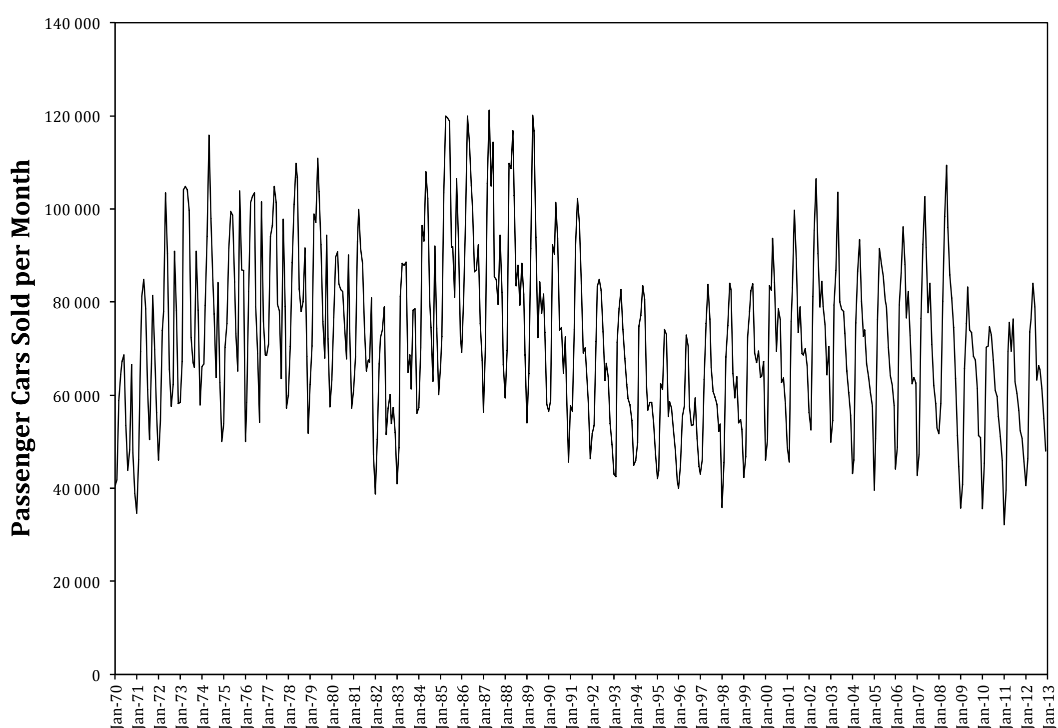
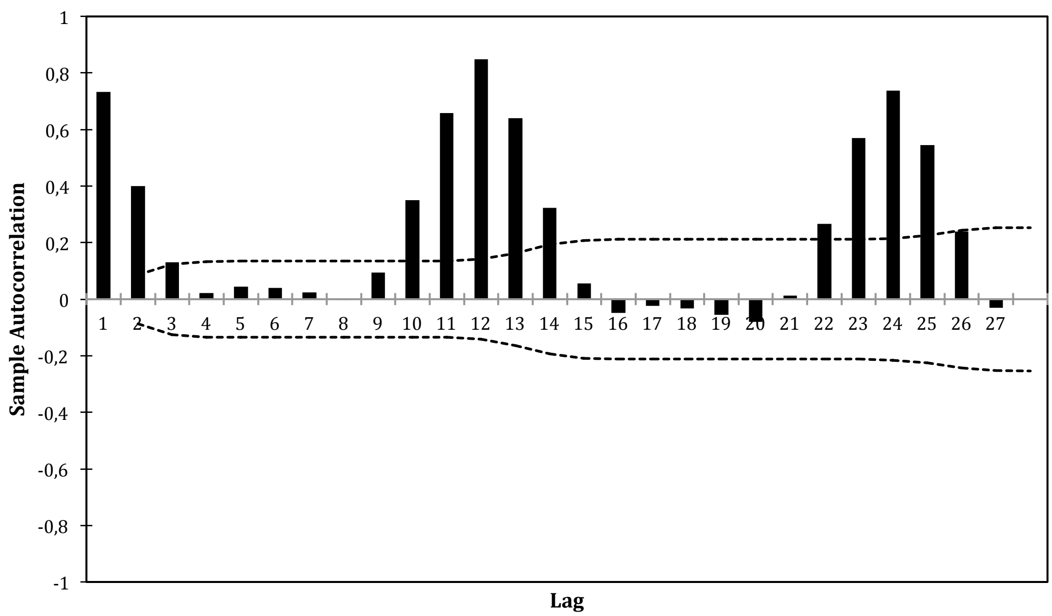


Figure 1 shows units of passenger cars sold by month. One can plainly see noticeable seasonality to automobile sales.

Analysis

There are two ways to remove seasonality in order to arrive at a stationary process: by seasonal differencing or by including a seasonal autoregressive lag term. Seasonal differencing is more reasonable—since one would expect annual trends to be similar (reflected by applying seasonal differencing), but one would not necessary expect a month's sales to be correlated to that of one year ago (reflected by introducing a seasonal autoregressive lag term).

Figure 2: Sample Autocorrelation of Raw Data



From Figure 2, we see a strong cyclical nature to the sample autocorrelations at lags 12 and 24. This is an indication that a seasonal model with lag 12 is appropriate. After taking seasonal difference at lag 12, we see in Figure 3 that the seasonality component of the process has largely been removed. However, the transformed process is still not stationary, since the process shown in Figure 3 does not appear to centre around a mean of zero. Non-stationarity is confirmed by calculating sample autocorrelations—as seen in Figure 4, the sample correlations die out gradually, rather than exponentially (as one would expect for a stationary autoregressive process) or abruptly (as one would expect for a stationary moving average process).

Figure 3: Seasonal Differences, D=1, s=12

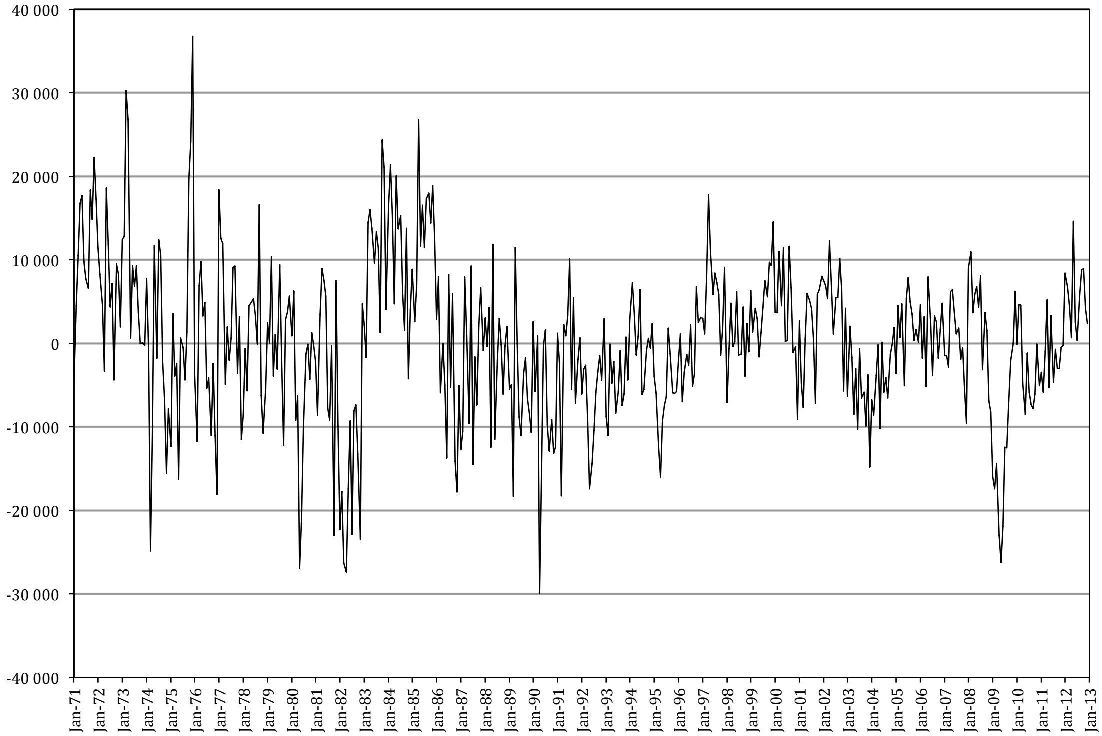
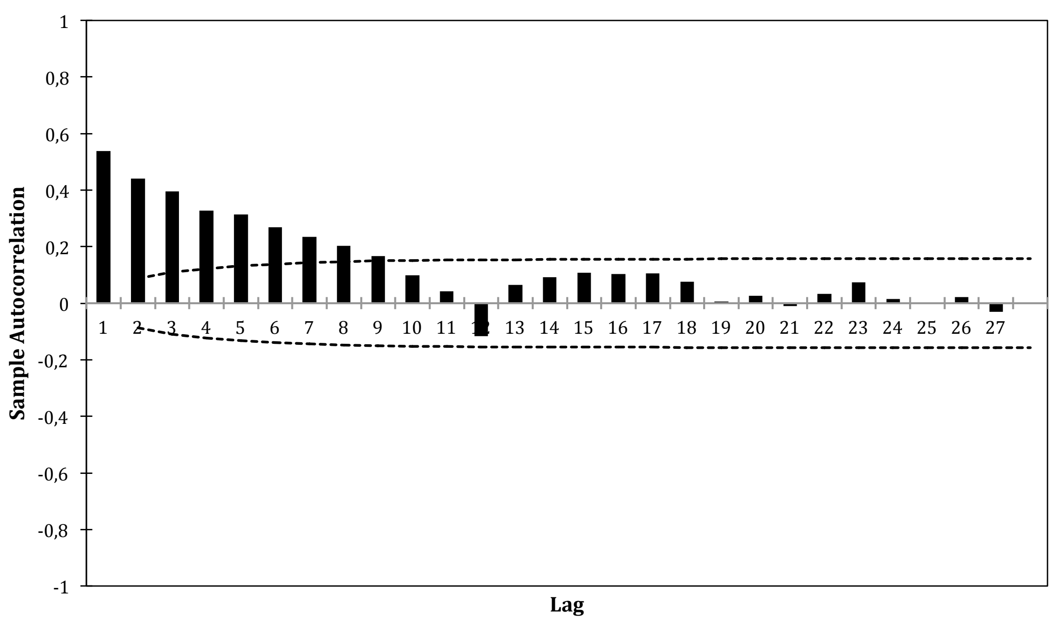
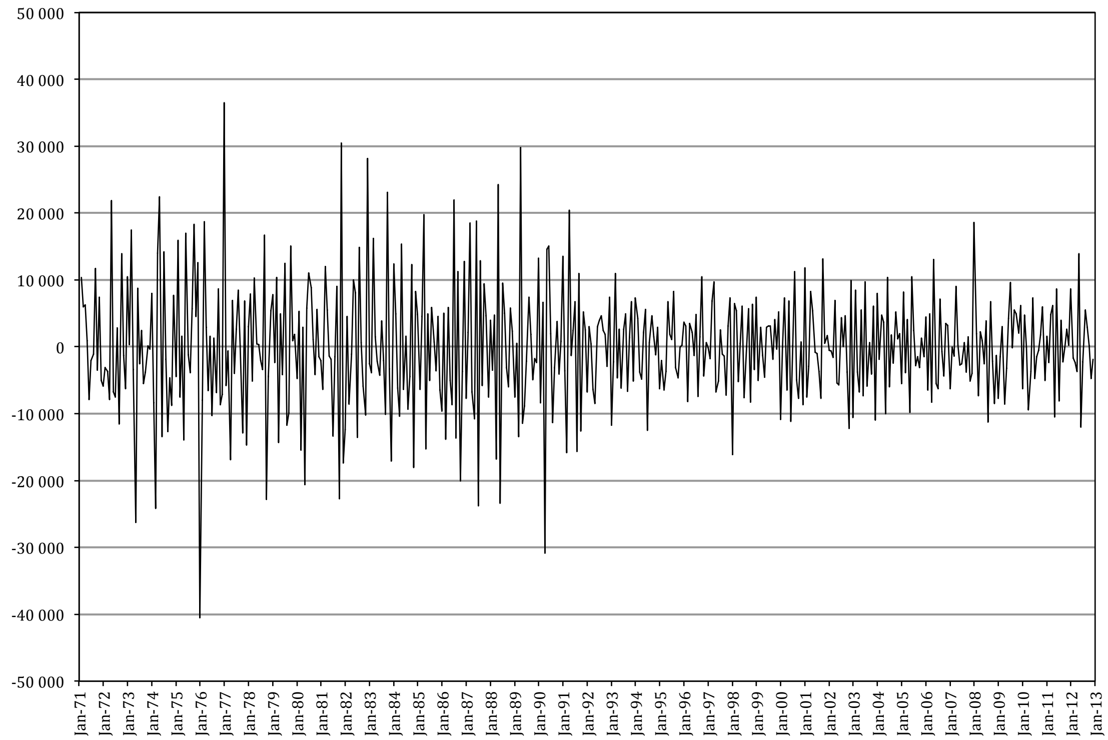


Figure 4: Sample Autocorrelation of Seasonal Differences, D=1, s=12



After taking differences at lag 1, we finally arrive at what appears to be a stationary process, as shown in Figure 5. The process is centred around a mean of zero with no discernable trends.

Figure 5: First-Order Differences of Seasonal Differences, d=1, D=1, s=12



To confirm that the ARIMA process with d=1, D=1 and s=12 is stationary, I created correlograms for sample autocorrelations and sample partial autocorrelations, shown in Figures 6 and 7.

Figure 6: Sample Autocorrelations of First-Order Differences of Seasonal Differences, d=1, D=1, s=12

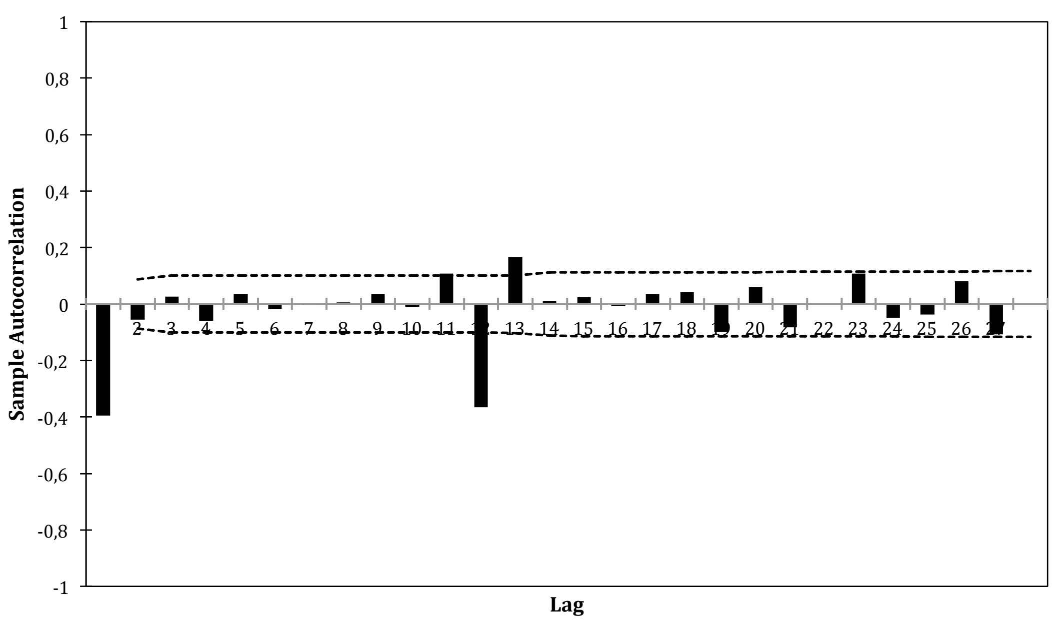


Figure 7: Sample Partial Autocorrelations of First-Order Differences of Seasonal Differences, d=1, D=1, s=12

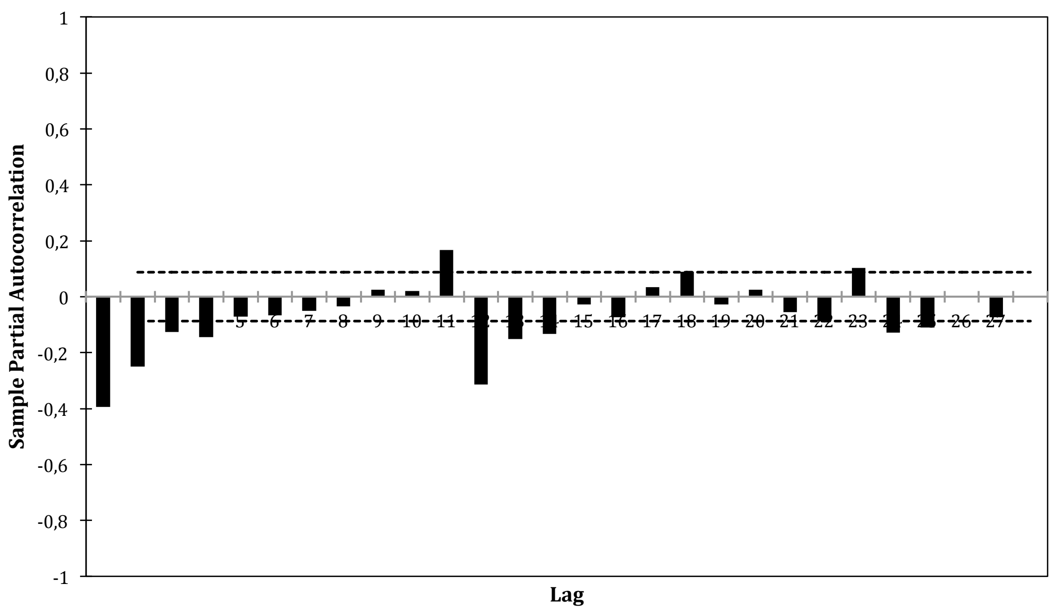


Figure 6 shows significant sample autocorrelation at lags 1 and 12, while Figure 7 shows sample partial autocorrelation that tails off from lags 1 and 12. These suggest that the process is best modelled through a moving average process.

I evaluated two seasonal ARIMA processes, one with a non-zero moving average coefficient at lag 1 only, and another with non-zero coefficients at lags 1 and 12. In Figures 8 and 9, the fitted ARIMA models are plotted with the actual data. It is evident that the ARIMA(0, 1, 1)×(0, 1, 1) model in Figure 9 fits the actual data better than the ARIMA(0, 1, 1)×(0, 1, 0) model. A comparison of the sum of squared errors statistic for the two model confirms that the ARIMA(0, 1, 1)×(0, 1, 1) model is a better fit (~4.85 billion) versus the ARIMA(0, 1, 1)×(0, 1, 0) model (~12.11 billion).

The parameters of the fitted ARIMA(0, 1, 1)×(0, 1, 1) model are: $\theta_0 = -11.085$, $\theta_1 = 1.549$ and $\Theta_1 = 1.378$.

Figure 8: Raw Data vs. Fitted ARIMA(0, 1, 1)×(0, 1, 0), s=12

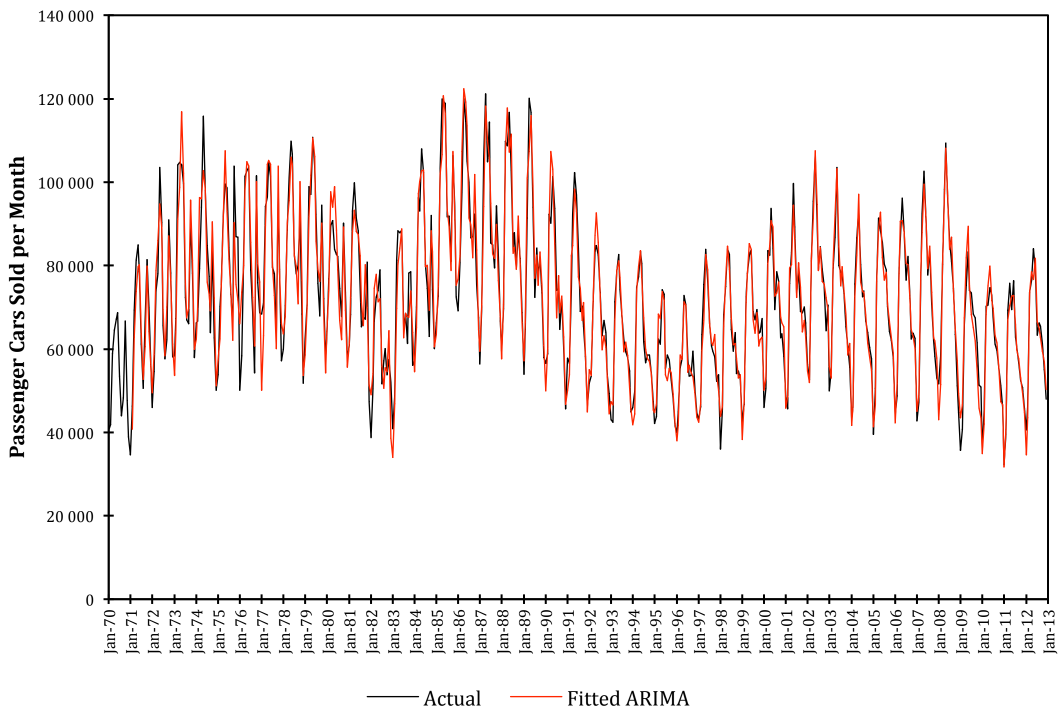
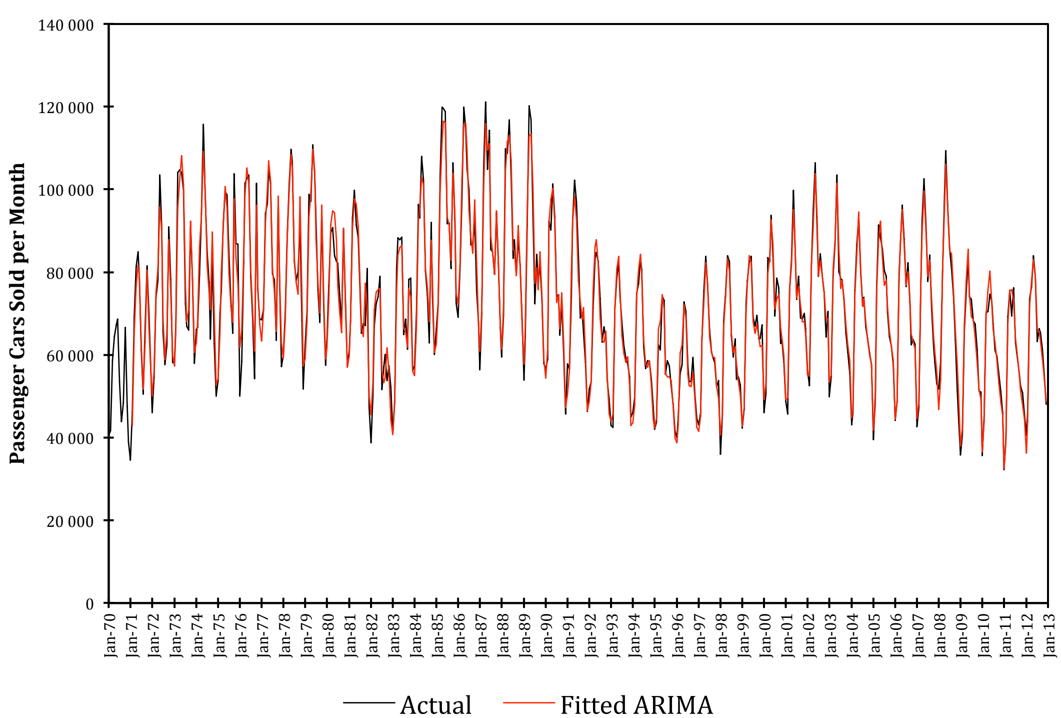


Figure 9: Raw Data vs. Fitted ARIMA(0, 1, 1)×(0, 1, 1), s=12



Conclusion

Based on the preceding analysis, an ARIMA(0, 1, 1)×(0, 1, 1) model with a seasonal period of 12 and parameters $\theta_0 = -11.085$, $\theta_1 = 1.549$ and $\Theta_1 = 1.378$ is an appropriate fit for the data on Canadian automobile sales. By extension, this confirms that Canadian automobile sales indeed exhibit annual seasonality.