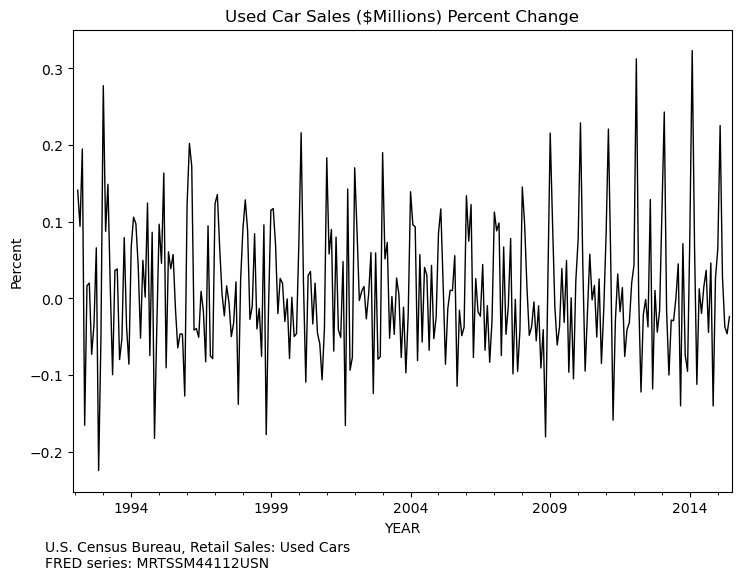
**Used Car Sales Writeup**

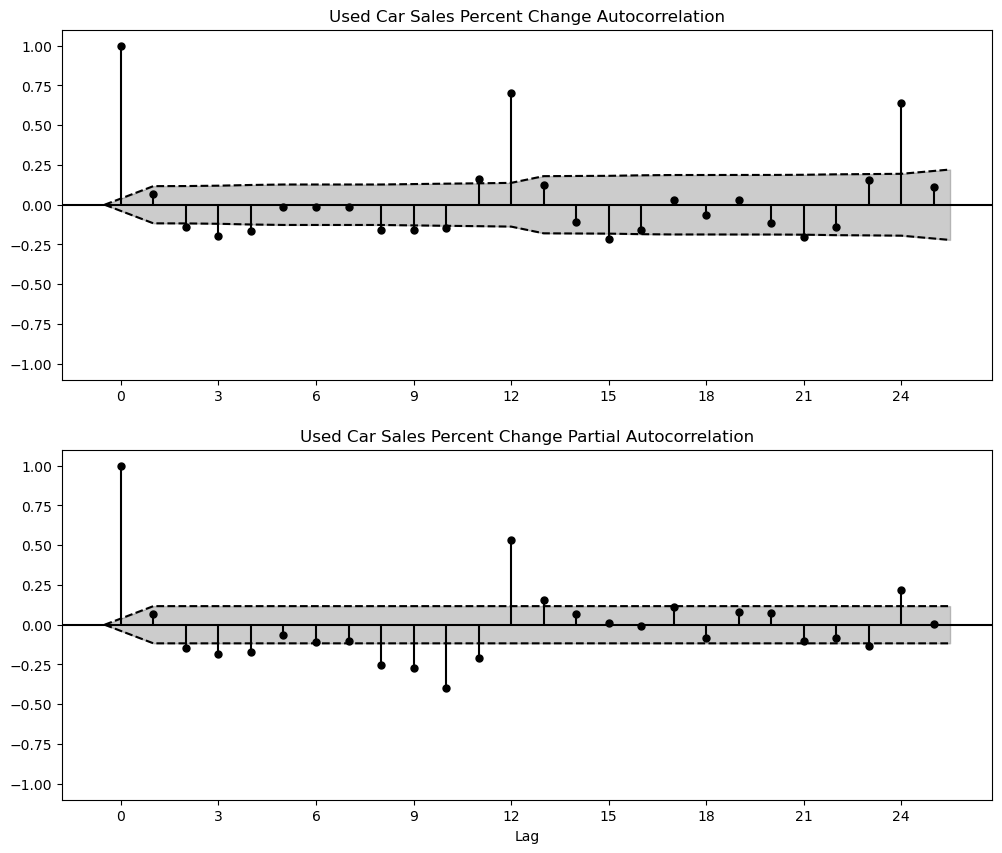
**(To accompany “Used\_Car\_Sales\_Modeling\_Forecasting.ipynb”)**

a. Plot of Car Sales, in Percent Change



b.

Correlograms



Analysis:

The most prominent feature of the ACF plot is the large autocorrelation at lag 12 and lag 24. This indicates some monthly seasonality, which is not surprising as we are working with retail sales data, and retail data typically has strong seasonality. There is also a large autocorrelation at lag 12 in the PACF plot, further supporting the hypothesis of strong seasonality. Otherwise, in both plots we also see some slowly dampening correlations up to lag 11. This tells us that we will have to test many lags with AR, MA, and ARMA models to find a process that explains the patterns we see in the data.

d. Finding best model by iteratively finding best ARMA(p,q) specification

Our model includes a constant term as well as 11 monthly dummies, as strong monthly seasonality is suspected. With that base model, we now explore the model fit of orders of ARMA(p,q) models.

Table of AIC values

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| AR order (p) | MA order (q) | | | | |
|  | 0 | 1 | 2 | 3 |
| **0** |  | -870.74 | -871.85 | -869.88 |
| **1** | -859.27 | -871.83 | -869.85 | -867.97 |
| **2** | -870.12 | -868.16 | -867.93 | -887.23 |
| **3** | -868.14 | -869.13 | -868.21 | **-891.31** |

Table of SIC values

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| AR order (p) | MA order (q) | | | | |
|  | 0 | 1 | 2 | 3 |
| **0** |  | -819.76 | -817.22 | -811.61 |
| **1** | -808.29 | -817.20 | -811.58 | -806.06 |
| **2** | -815.50 | -809.89 | -806.02 | -821.68 |
| **3** | -809.87 | -807.22 | -802.66 | **-822.11** |

Notice that the AIC and SIC values in this case are all negative. Thus, when we are looking for the smallest AIC/SIC, we are actually looking for the largest absolute value. We started with ARMA models with AR(p) and MA(q) term combinations 0 to 3. The best fitting model according to both AIC and SIC is ARMA(3,3).

However, after fitting that model and examining the correlogram of the model residuals, we still see significant autocorrelation at lag 12. Therefore, we explored a few more models with MA term remaining low (0 to 3), but with a larger AR term (10, 11, or 12). The AIC and SIC for those models are below.

AIC Values

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | MA order (q) | | | | |
| AR order (p) |  | **0** | **1** | **2** | **3** |
| **10** | -881.60 | -879.71 | -882.98 | -880.36 |
| **11** | -880.19 | -877.60 | **-911.75** | -884.71 |
| **12** | -890.05 | -890.79 | -910.97 | -910.65 |

SIC values

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | MA order (q) | | | | |
| AR order (p) |  | **0** | **1** | **2** | **3** |
| **10** | -797.83 | -792.31 | -791.93 | -785.67 |
| **11** | -792.79 | -786.55 | **-817.06** | -786.37 |
| **12** | -799.00 | -796.10 | -812.64 | -808.67 |

e. Best model?

The model with lowest overall AIC among *all models* is ARMA(11,2) [AIC = -911]. Notice that if going by the SIC, ARMA(11,2) is not the absolute best model [SIC = -817 for ARMA(11,2) vs SIC= -822 for ARMA(3,3)], although it is the best among the models with higher order AR terms.

We had already rejected ARMA(3,3) as a poor fit due to autocorrelation in the model residuals, so we choose the ARMA(11,2) model based on the absolute best AIC value and a relatively good SIC value. After fitting the model examining the correlogram of the ARMA(11,2) residuals, they appear to be closer to white noise, indicating that the model fits the data well.

f. Estimation output:

