OMSBA 5305
Data Translation Challenge
Step 1

Julia Giesbrecht, Aemro Chanie, Valentina Komarova, Jomaica Lei

Company: Walmart

Introduction

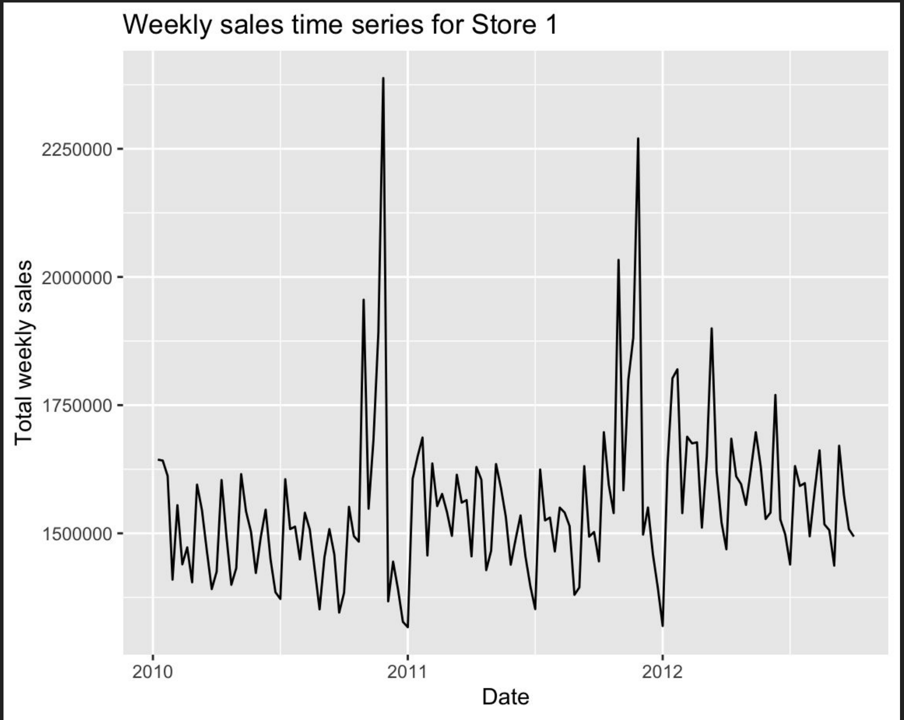
Data selection:

- For this project we have chosen a time series dataset that represents weekly sales amount per store for the year 2010 -2012.
- This dataset contains 421570 observations

Data Preparation:

- We loaded the data to R
- Check the data for missing values
- Convert the date to date format
- Plot the time series
- Upon visual inspection we noticed some level of seasonality on the time series plot
- To remove seasonality and to make the time series stationary, we applied first differencing and seasonal differencing.

Weekly Sales Time Series

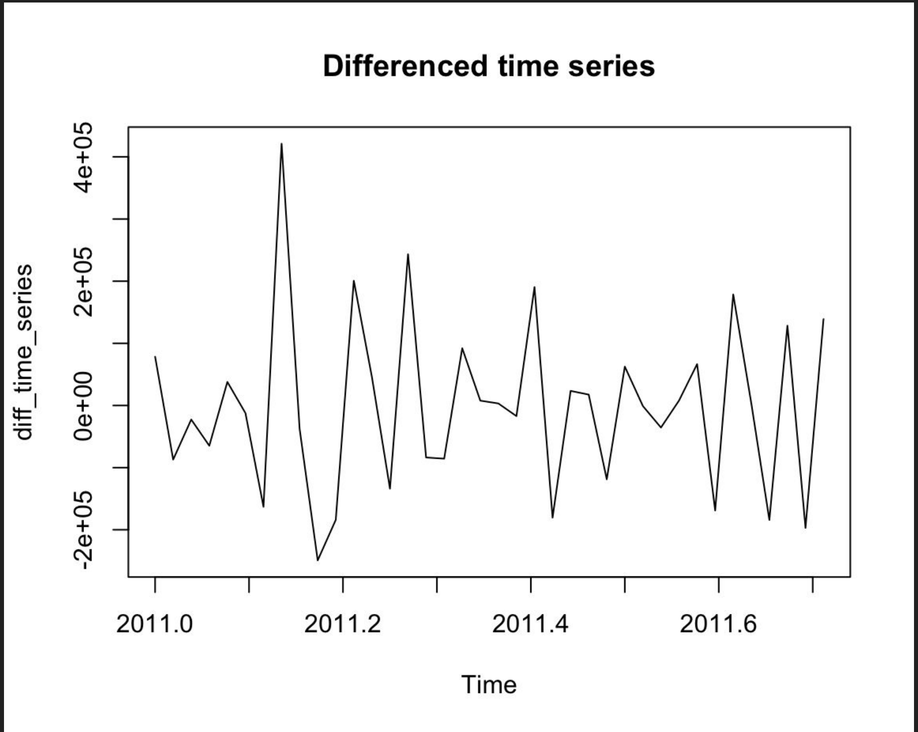


Perform the ADF test- stationary

Perform the KPSS test - not stationary

There is a seasonal trend and a possible slight upward trend

Differenced Time Series



Perform regular differencing - first-order difference to remove the trend

Seasonal differencing- Remove the seasonal pattern that repeats every year.

Perform the ADF test after seasonal differencing - reject the null- stationary

Perform the KPSS test after seasonal differencing - stationary

Models Selection

We chose the MA(1), AR(1), and ARMA(1,1) models based on the characteristics of the ACF and PACF plots of the differenced series.

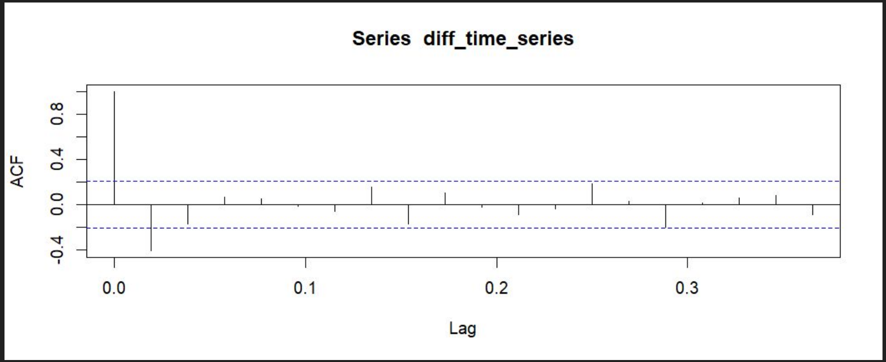
Specifically, the ACF plot showed a sharp drop after one lag, which is typical of a MA(1) process. The PACF plot showed a gradual decrease, which is typical of an AR(1) process.

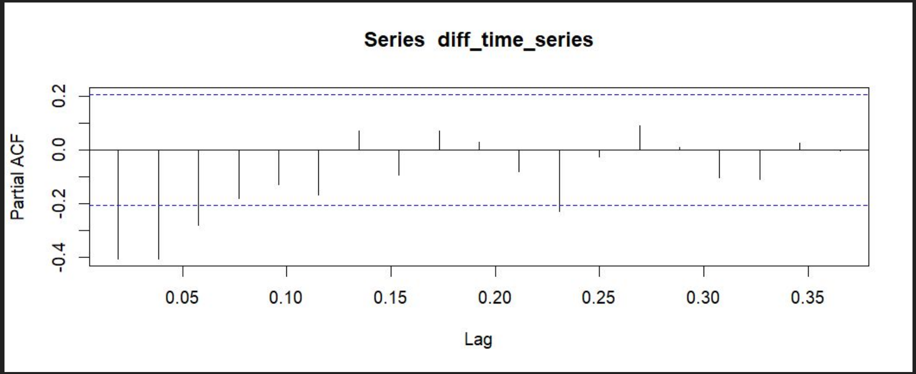
The ARMA(1,1) model was chosen as a combination of the two since it includes both an AR(1) and an MA(1) component.

Overall, these models were chosen as reasonable candidates based on the characteristics of the autocorrelation and partial autocorrelation functions.

Models Selection

- **Plot ACF and PACF for model selection**



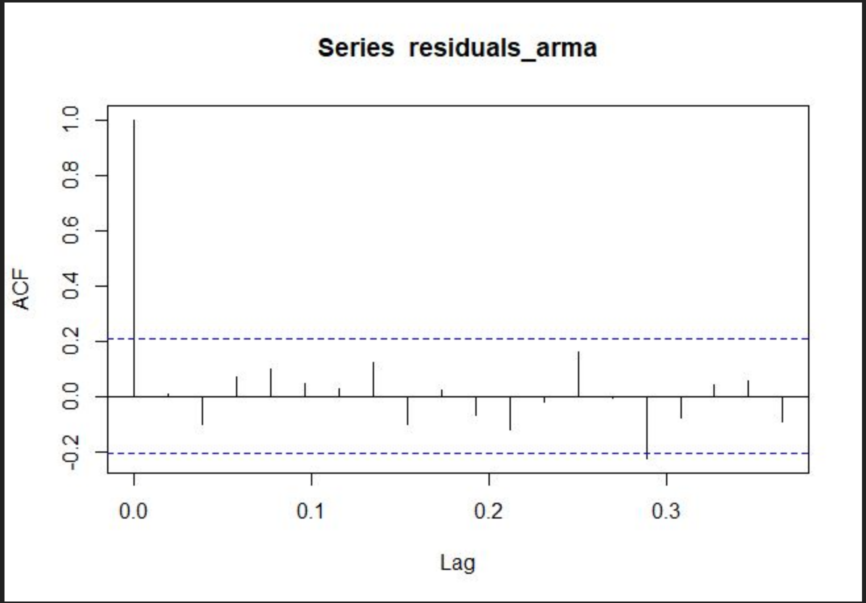


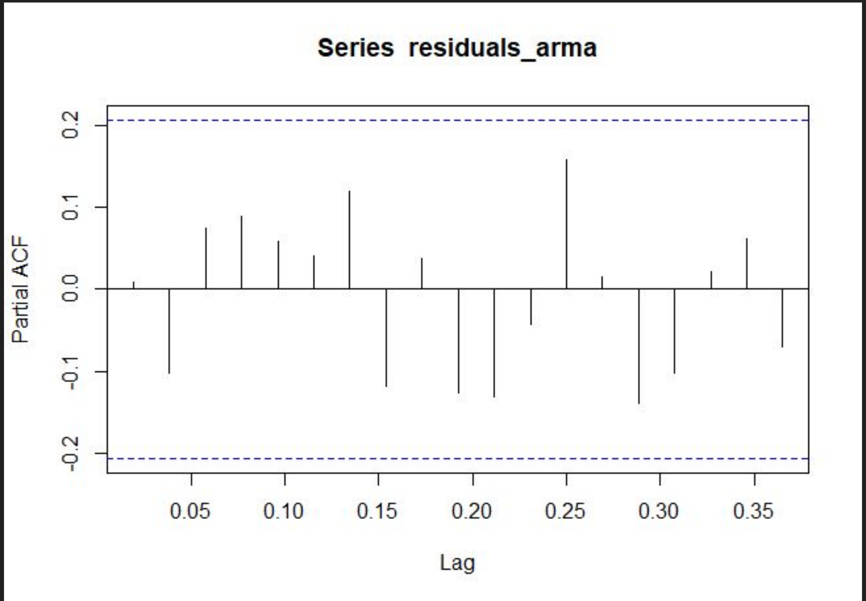
- **ACF plot of the differenced series shows a sharp drop after one lag**
- **The PACF plot shows a gradual decrease**
- **Based on the result from ACF and PACF, we have decided to do MA(1), AR(1) and ARMA(1,1)**

Model Comparison

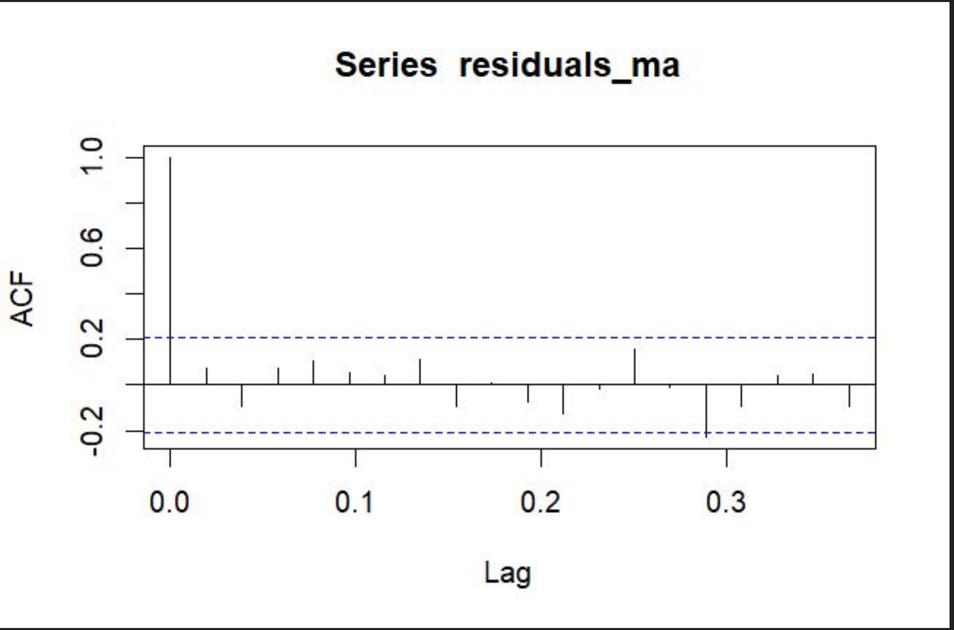
	MA(1)	AR(1)	ARMA(1,1)
Covariance - stationary	yes	yes	yes
Invertibility	yes	yes	yes
White noise residuals	yes	no	yes
Q-statistics	0.086228	0.0020268	1.06E-05
(p-value)	0.769	0.9641	0.9974
Residual variance	4.06E+09	6.29E+09	1.00E+10
Adjusted R-squared	null	null	null
AIC	3590.52	3589.554	3591.5
SIC	3.60E+03	3598.179	3603.005
Coefficients	-0.9195	-0.3839	0.1408 -0.9430
intercept	24071.57	27483.4	25110.99

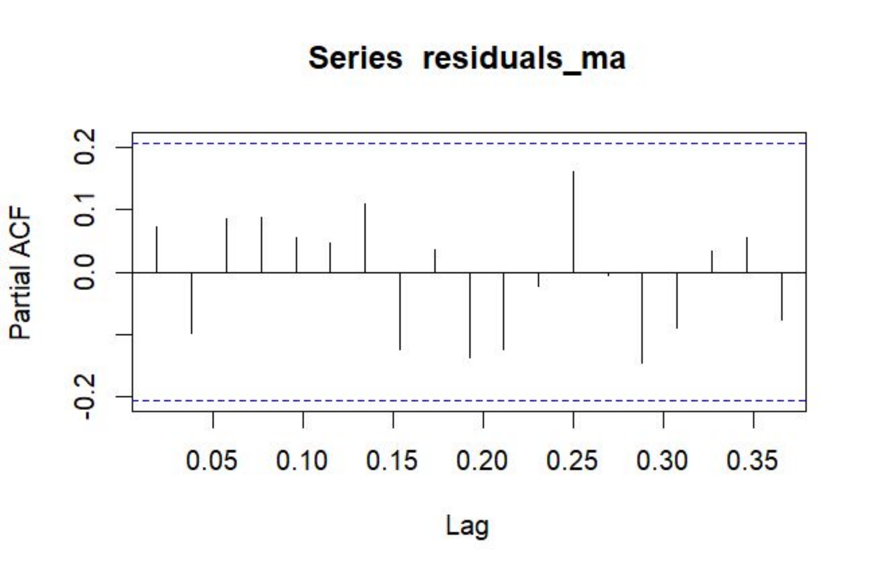
ARMA(1,1) Residual ACF & PACF



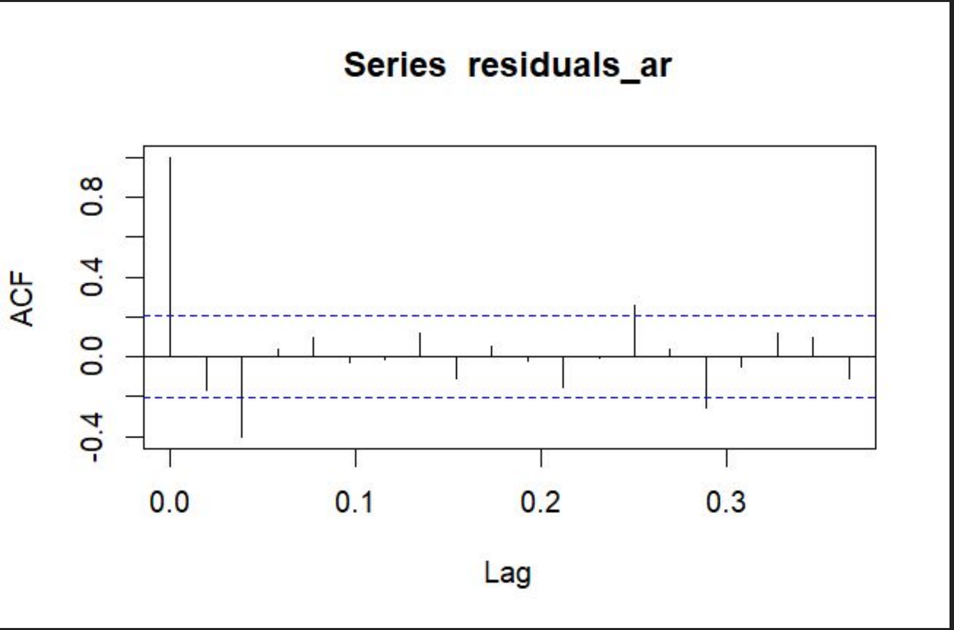


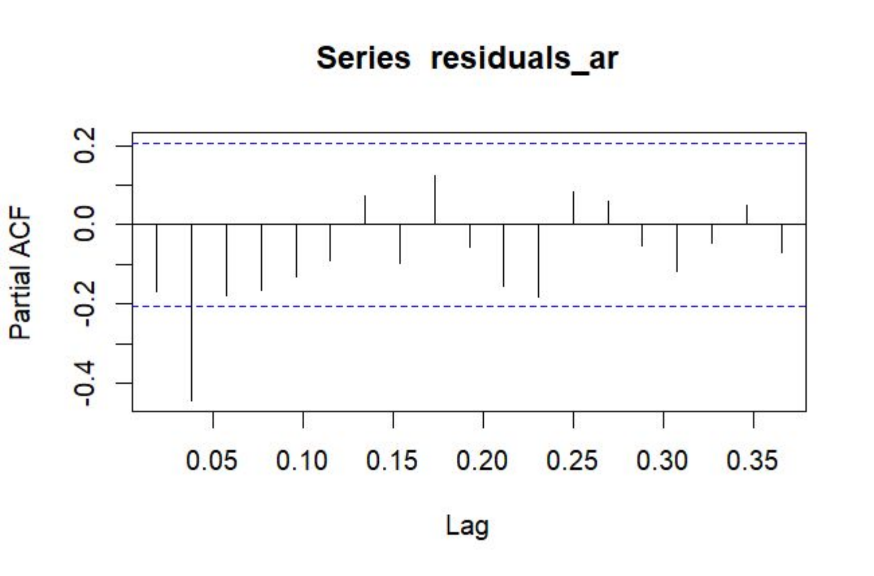
MA(1) Residual ACF & PACF





AR(1) Residual ACF & PACF





Conclusion

Based on our model results, we can conclude the following:

- Covariance-stationarity and invertibility assumptions are satisfied by all three models.
- The MA(1) and ARMA(1,1) models have white noise residuals, whereas the AR(1) model does not.
- The Q-statistics (measuring the goodness of fit of the residuals) for the MA(1) and AR(1) models are relatively high, while for the ARMA(1,1) model it is low.
- The p-values for all three models are high (> 0.05), indicating that the residuals are consistent with white noise.
- The residual variances for the three models are relatively high, with the ARMA(1,1) model having the highest variance.
- The adjusted R-squared values are null for all models since these models are not regression models.
- The AIC is lowest for the AR(1) model, indicating that it is the best-fitting model based on this criterion.
- The AR model has the lowest BIC value, indicating it is the best fit for the data among the three models considered.

Overall, based on the results, the AR(1) model may be the best choice as it has white noise residuals and the lowest AIC value.

Forecasting (AR1)