Analysis of Package Deliveries in Seattle

Step 1: Data Translation Challenge

Team Members: Anjali, Saad, and Tavnit

Introduction

Dataset
Project Workflow
Exploratory Data Analysis
Decomposition Analysis
ACF and PACF Analysis – Part 1 Original Series
ACF and PACF Analysis – Part 2 Weekly Cyclicity

Project Introduction

Forecast the number of daily deliveries in the Seattle Area for Company A.

Logistics: Transportation, Inventory, Order Processing Warehousing, Inventory Management

Purpose: To forecast delivery plans in order to allocate drivers, do demand planning & save money

Dataset

Real Data from Company A

Total Number of package deliveries in Seattle Area (Tacoma – Everett)

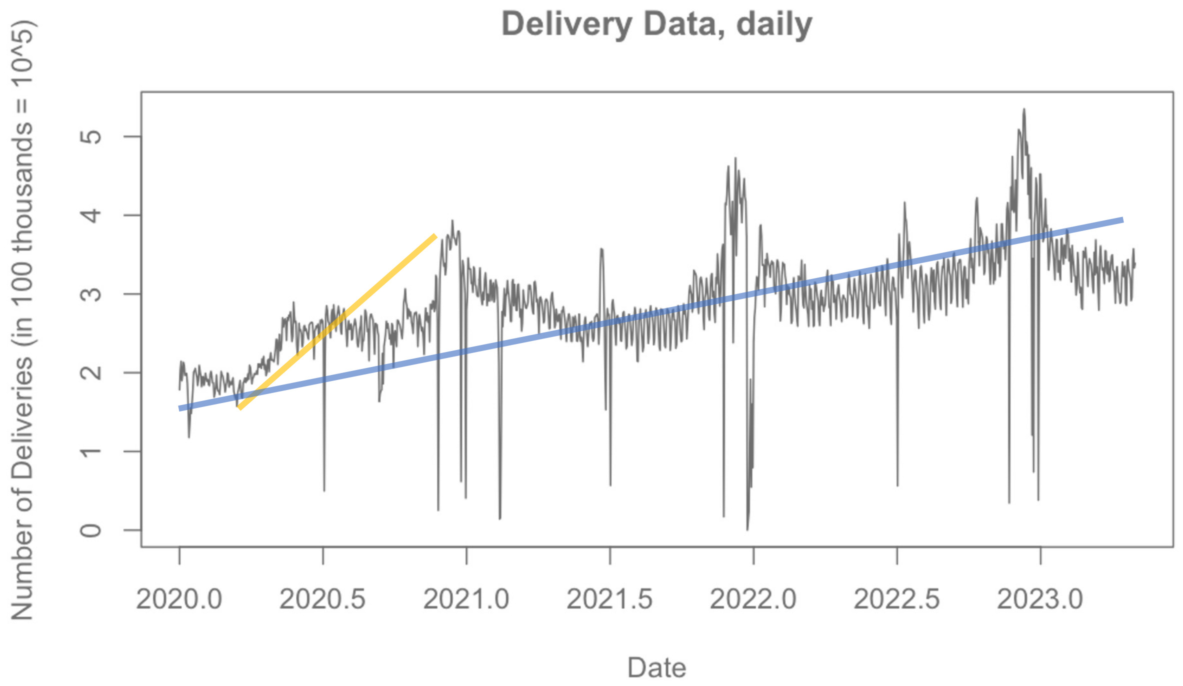
Frequency = Daily

January 2nd 2020 to March 31st 2023

Scale down by 10^5

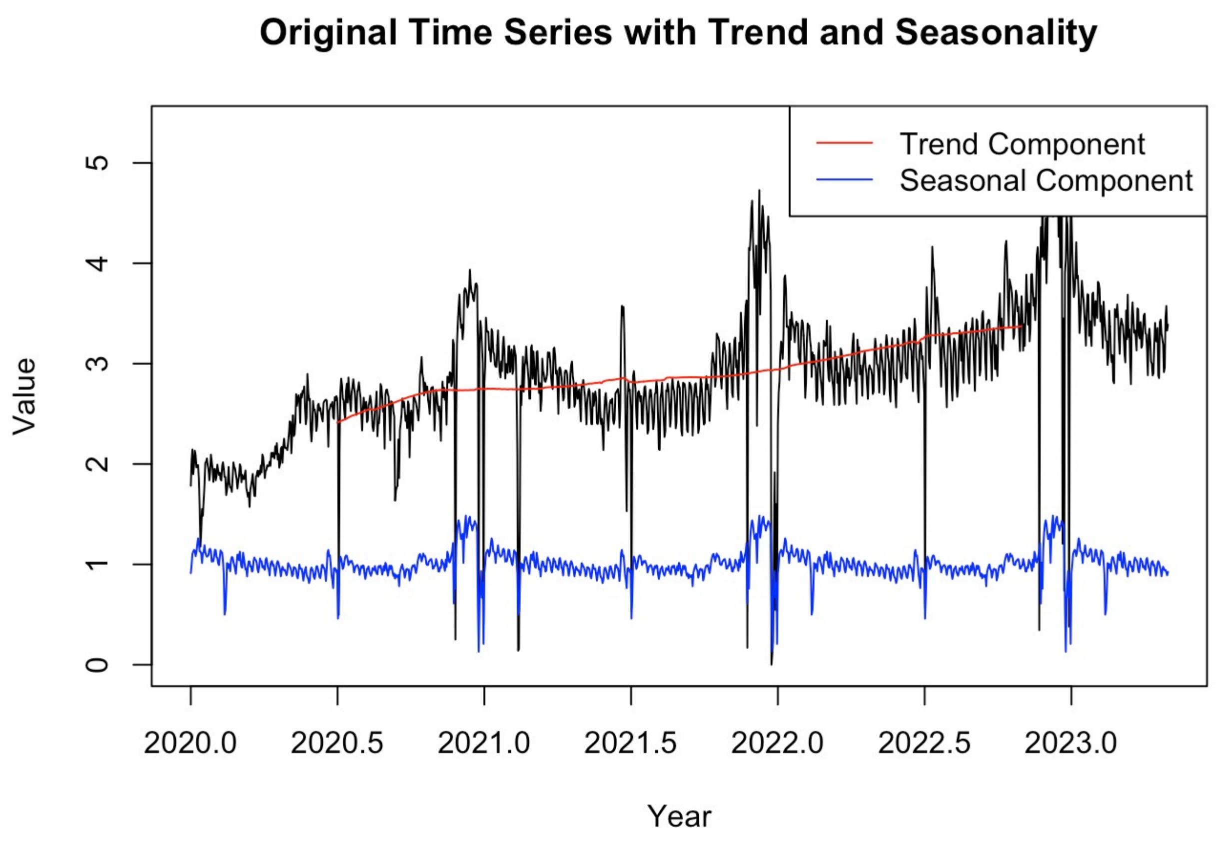
(in 100 thousands)

Estimation – Prediction Sample Split 90 % - 10 %



Exploratory Data Analysis

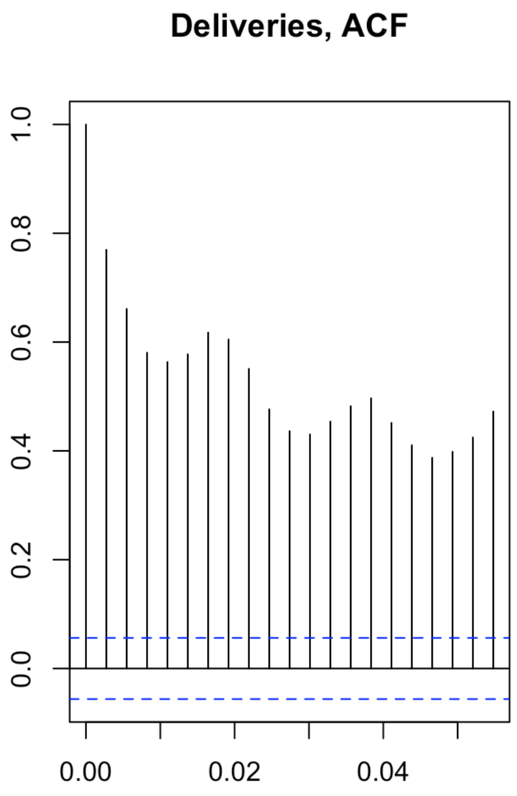
- Median = 2.85, Mean = 2.88, Std = 0.679
- Slight upward trend
- Clear annual seasonality. Holiday shopping.
- Weekly cyclical
- COVID-19 buying behavior - March 2020 onwards
- **Dickey Fuller Test indicates Deterministic Trend.** Test statistic is very low (-12.582) -> **no unit root**

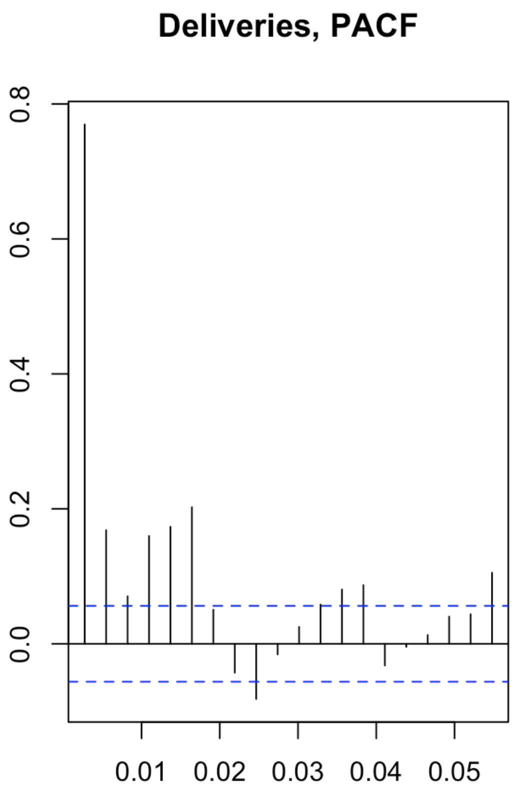


Decomposition Analysis

- There is a seasonal trend.
- Not enough years to predict annual seasonality
- Weekly Cyclicity can be Analyzed.
- Analysis with lag 7 and seasonal models with period = 7

ACF & PACF
PART 1 – Original Series





Features of the ACF

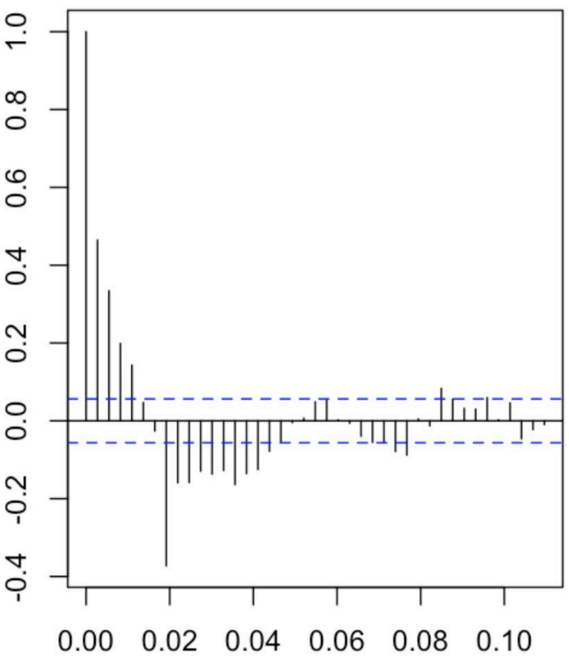
- Gradually falling -> indicates AR models
- Bumps (7 lags width) weekly cyclical pattern -> cyclical models

Features of PACF

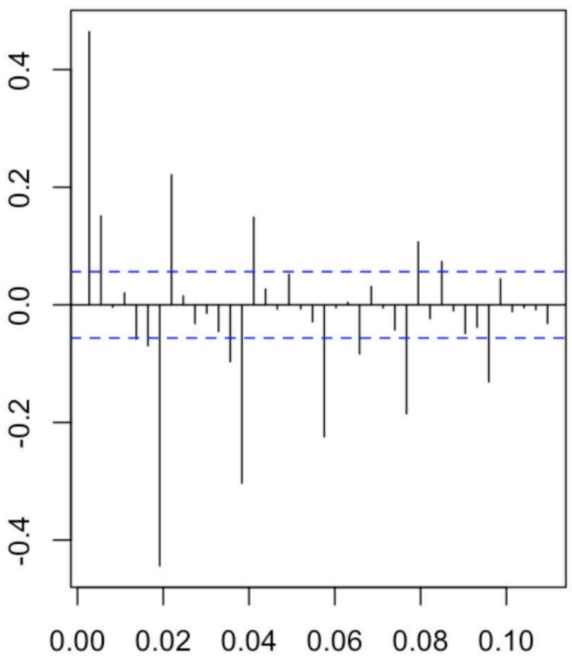
- No smooth decay towards zero – MA models not indicated.
- Large Spike at lag = 1 -> AR(1)
- Spikes at lag = 2,4,5,6 -> AR(2),AR(6)
- There are recurring spikes -> ARMA models of ARMA(1,1), ARMA(2,2), ARMA(6,1), ARMA(6,2), ARMA(6,6)

ACF & PACF
Part 2 - Weekly Cycles

Deliveries with Difference = 7, ACF



Deliveries with Difference = 7, PACF



Seasonal Models

- Identify order using PACF, ACF of Difference = 7
- Try Simpler models before complex.

Features of the ACF:

- Gradually falling spikes indicate AR behavior.
- Spikes at lag = 1,2,3,4 and 7 -> MA 1 ... MA 4, MA7

Features of the PACF:

- PACF has rapidly declining autocorrelation indicates we could try MA models
- Large spike at lag = 1 indicates AR(1).
- Spikes at lag = 2 and 7 indicate AR(2) or AR(7).
- There are recurring spikes. No need to considered the spikes after the 7th lag as these are in the cyclical pattern

Tentative Models first tried

- SARMA models with AR2 and AR7 along with MA1 and MA4.
- We try the simplest form (p,0,0)(0,0,Q)[7]

Model Search
(In-Sample Evaluations)

Tentative Models

Model Evaluation Criteria

Model Selection

Summary of Top 5 Models

Multistep Forecast

In-sample Evaluation

Models Evaluated = 25

✓Check Stationarity

✓Check Invertibility

We consider only if both models passes upper two tests

✓Explanatory Power – R^Squared, AIC, BIC

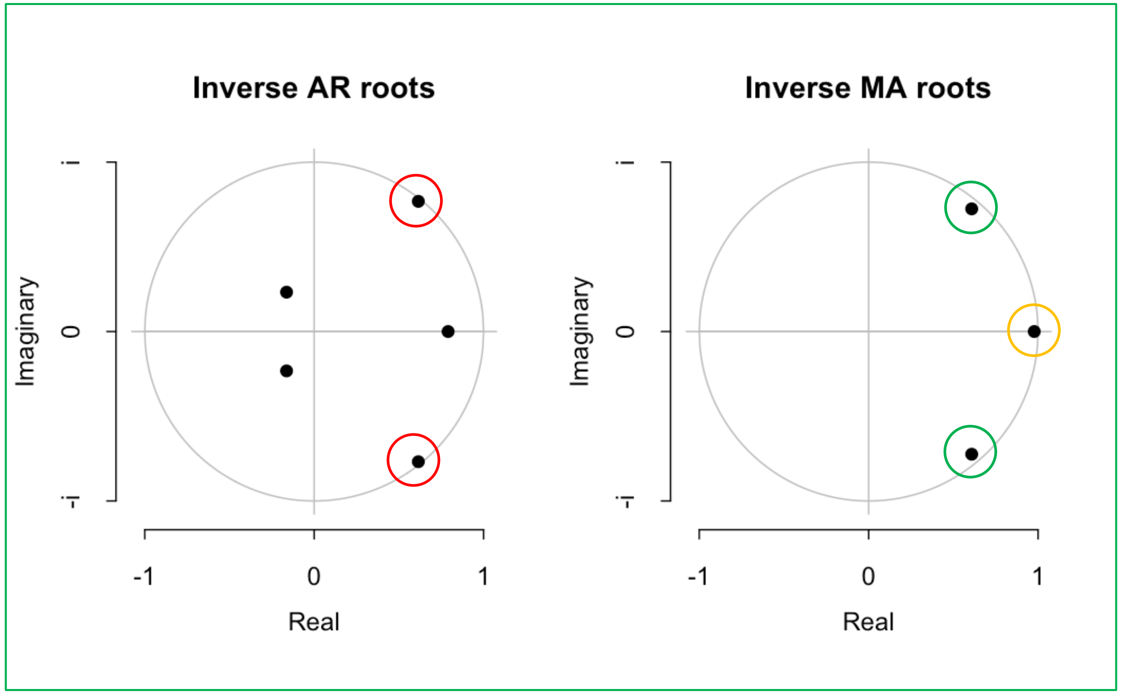
✓Check Residuals – Plot, Box Ljung Test

AR Models MA models (for completeness)	ARMA Models	Seasonal and Auto
1. AR 1 2. AR 2 3. AR 6 4. AR 6, d=1 5. MA 1 6. MA 2	1. ARMA (1,1) 2. ARMA (1,2) 3. ARMA (2,2) 4. ARMA (2,2), d = 1 5. ARMA (2,1) 6. ARMA (2,1), d=1 7. ARMA (6,1) 8. ARMA (6,1), d=1 9. ARMA (6,2) 10. ARMA (6,2), d=1 11. ARMA(6,4) 12. ARMA(6,4), d=1 13. ARMA(6,6) 14. ARMA(6,6), d=1	1.Auto ARIMA: ARIMA(5,1,3) 2.SARIMA Model 1: • order =(2,0,0) • seasonal = (0,0,1) 3.SARIMA Model 2: • order = (2,0,0) • seasonal = (0,0,4) 4.SARIMA Model 3: • order = (7,0,0) • seasonal = (0,0,1) 5.SARIMA Model 4: • order = (7,0,0) • seasonal = (0,0,4)

Check Stationarity and Invertibility

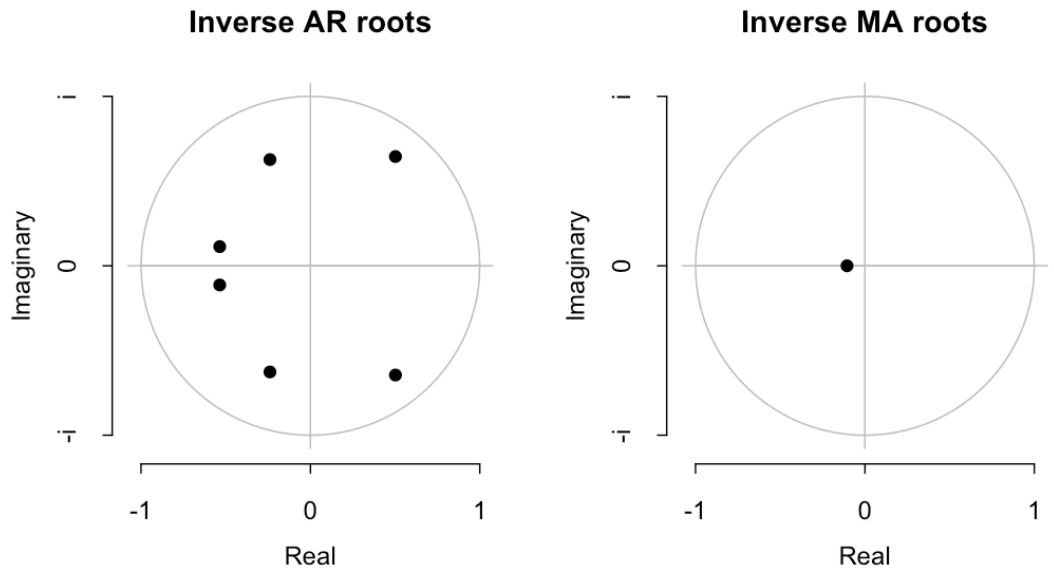
Auto Arima Model - ARMA (5,1,3)

- Some roots are on the circle
- Fail Test



ARMA (6,1,1)

- No roots are on the circle
- Pass Test



Model Selection

Model Name	AIC	BIC	R-Squared	Adjusted R-Squared	Box Ljung (p-value)	AR Roots	MA Roots
Sarima_ar7_ma4	1246.770	1313.113	0.6540682	0.6537832	9.532996e-01	TRUE	TRUE
Sarima_a7_ma1	1267.809	1318.842	0.6461570	0.6458655	9.886236e-01	TRUE	TRUE
ar6_d1	1279.758	1315.475	0.6463945	0.6461032	9.618456e-01	TRUE	TRUE
ar6_ma1_d1	1281.515	1322.335	0.6465330	0.6462418	9.885346e-01	TRUE	TRUE
ar2_ma2_d1	1292.706	1318.218	0.6370250	0.6367260	9.580129e-01	TRUE	TRUE
ar1_ma1_d1	1295.486	1315.893	0.6355851	0.6352849	9.796896e-01	TRUE	TRUE
Sarima_a2_ma4	1305.201	1346.028	0.6342961	0.6339949	6.837093e-01	TRUE	TRUE
a1_ma2	1325.918	1351.434	0.6256931	0.6253848	2.819411e-01	TRUE	TRUE
Sarima_a2_ma1	1356.159	1381.675	0.6162974	0.6159814	6.479105e-01	TRUE	TRUE
ar1_ma1	1373.259	1393.672	0.6100961	0.6097750	8.252470e-02	TRUE	TRUE

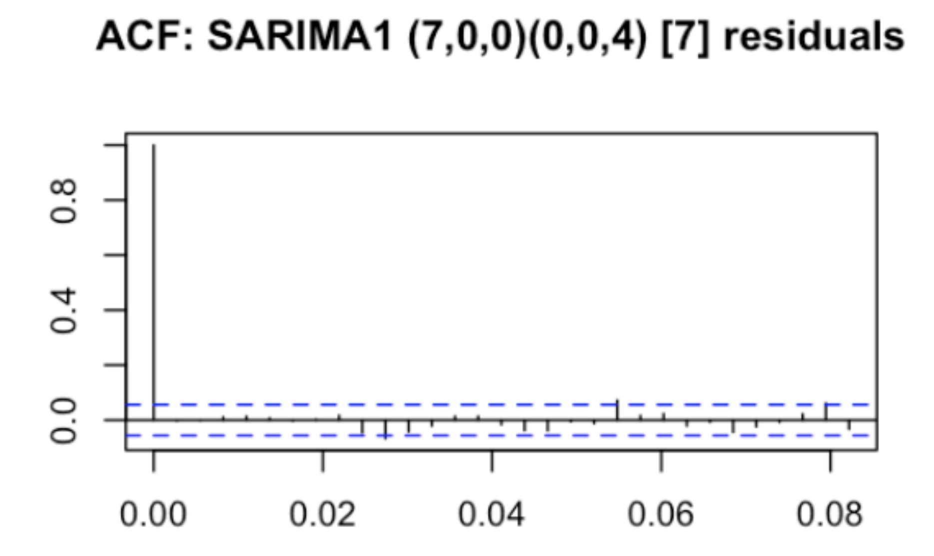
Summary of Top 5 Models

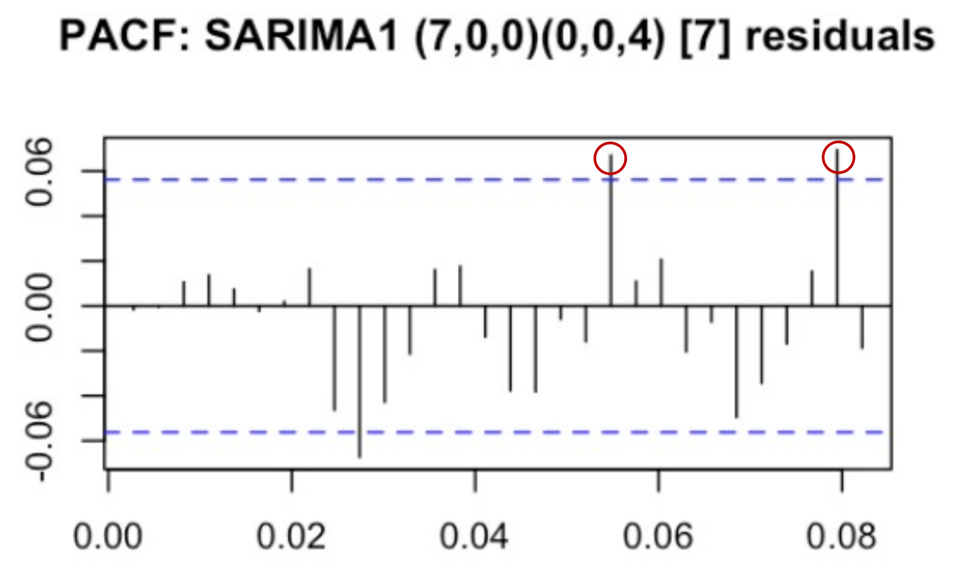
- The top 5 models are selected using AIC, BIC, and Adjusted R2, which balance model fit and complexity. Lower AIC and BIC indicates superior performance.
- Higher Adjusted R-Squared or R-Squared indicates better explanatory power for the model.
- Ljung-Box Test assesses the independence of residuals and check if residuals are white noise.
- All models demonstrate both covariance stationarity and invertibility, which are crucial for model stability and reliability.
- The in-sample evaluation enables us to find models with the most optimal fit.

	Model 1:	**Model 2:**	**Model 3:**	**Model 4:**	**Model 5:**
Name in Code	sarima_ar7_ma4	sarima_ar7_ma1	ar6_d1	ar6_ma1_d1	ar2_ma2_d1
Order	SARMA (7,0,0)(0,0,4)[7]	SARMA (7,0,0)(0,0,1)[7]	ARMA(6,1,0)	ARMA(6,1,1)	ARMA(2,1,2)
Coefficients (Theta and Phi)	ar1=0.5330 ar2=0.1192 ar3=-0.0327 ar4=0.0450 ar5=0.0553 ar6=0.1641 ar7=-0.0181 sma1=0.0628 sma2=0.0605 sma3 =0.1325 sma4=0.0276	ar1=0.5422 ar2=0.0963 ar3=-0.0394 ar4=0.0347 ar5=0.0500 ar6=0.1752 ar7=0.0485 sma1=0.0066	ar1=-0.4408 ar2=-0.3343 ar3=-0.3636 ar4=-0.3172 ar5=-0.2566 ar6=-0.0705	ar1=-0.5618 ar2=-0.3859 ar3=-0.4018 ar4=-0.3585 ar5=-0.2926 ar6=-0.0987 ma1=0.1214	ar1=-0.3187 ar2=0.4709 ma1=-0.1251 ma2=-0.7833
AIC	1246.77	1267.81	1279.76	1281.52	1292.71
BIC / SIC	1313.113	1318.842	1315.475	1322.335	1318.218
R- Squared	0.6541	0.6462	0.6464	0.6465	0.6370
Adj R-Squared	0.6538	0.6459	0.6461	0.6462	0.6367
White Noise Residuals (p-value)	0.9533 PASS	0.9886 PASS	0.9618 PASS	0.9885 PASS	0.958 PASS
Invertibility	Yes	Yes	Yes	Yes	Yes

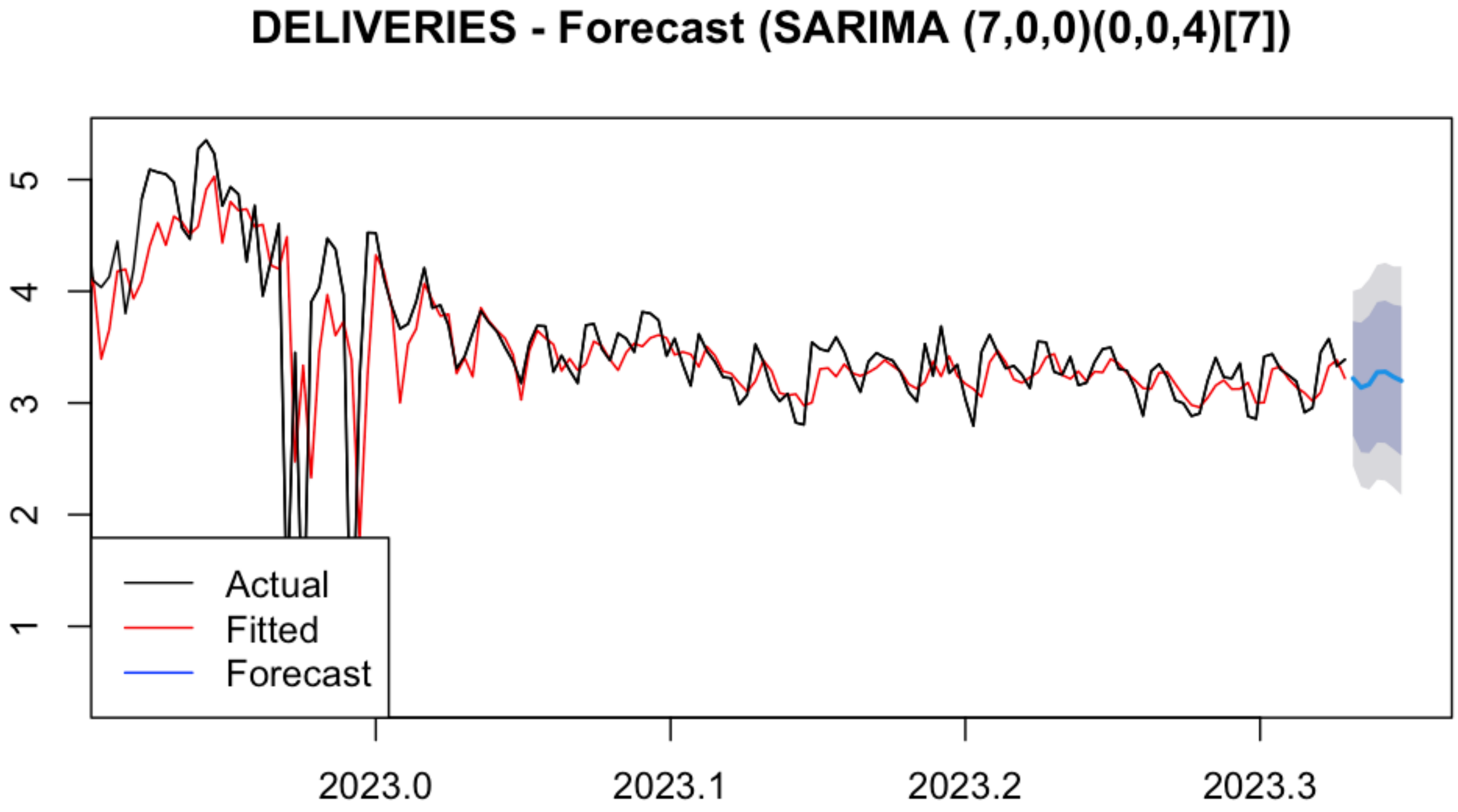
In sample evaluation Summary

- Best Model: ARIMA (7,0,0) $(0,0,4)_7$
- All the top 5 models have residuals that pass Box-Test
- But the residual ACF and PACF plots show some spikes
- Generate in-sample multistep forecast for 7-days
- Appropriate period for usefulness
- The top 5 models will be considered for out of sample evaluation.





Multistep Forecast – Best Model – 7 Days Ahead



Annexure 3 has multistep forecast for the other top 5 models

Out of Sample Evaluations

Recursive Forecasting Environment
Forecast Optimality Tests
Assessment of Forecasts
Linear Combination models
Forecast Visualization

Out of Sample Evaluation

Model 1	Seasonal Model	
	SARIMA (7,0,0)(0,0,4)[7]	
Model 2	Seasonal Model	
	SARIMA (7,0,0)(0,0,1)[7]	
Model 3	Non-Seasonal Model	
	AR6 = ARMA(6,1,0)	
Model 4	Non-Seasonal Model	
	ARMA(6,1,1)	
Model 5	Non-Seasonal Model	
	ARMA(2,1,2)	
Naïve Model	last period value	
Simple Average Model	Average of last 4 periods	

Split Data: Estimation – Prediction (90% -10%)

Forecast Recursively by updating the estimation sample

Each forecast is for 1 period (h = 1)

Check Forecast Optimality

- MPE Test (Mean Prediction Error Test)
- IET Test (Informational Efficiency Test)

Assess the Forecast

- MSE (Mean Squared Error)
- Other Metrics – RMSE, MAE, MAPE

Select the 3 best models

Construct Combination Models from 3 best models

- Equal Weighted
- Inverse MSE weighted
- OLS Linear Combination

Visualize Forecast

- Use best model and visualize the forecast for prediction sample data

Predictive Ability Assessment Outputs

Model	MSE	ME	RMSE	MAE	MPE	MAPE
Model 1 SARMA (7,0,0)(0,0,4)[7]	0.02670103	0.05941091	0.1634045	0.1264254	1.499267	3.675692
Model 2 SARMA (7,0,0)(0,0,1)[7]	0.01878746	0.04820287	0.1370673	0.1071776	1.19126	3.120265
Model 3 = ARMA (6,1,0)	0.01605916	0.001266935	0.1267247	0.09687102	-0.1688755	2.883118
Model 4 = ARMA(6,1,1)	0.01643544	0.001367634	0.1282008	0.09796088	-0.1673004	2.914466
Model 5 = ARMA(2,1,2)	0.013016	-0.03837299	0.1140877	0.08870377	-1.336391	2.6858
Naïve Model	0.06972841	0.001369918	0.2640614	0.1928634	-0.2490005	5.621549
Simple Average Model	0.1097014	0.01598162	0.331212	0.2437092	-0.1190842	7.071301

Annexure 2a has details on all the metrics presented here

Out of Sample Evaluation Results

Model Name	MPE Test		IET Test		MSE
	P-value	Result	P-value	Result	
Model 1 SARMA (7,0,0)(0,0,4)[7]	3.57e-05	FAIL	0.0287	FAIL	0.02670103
Model 2 SARMA (7,0,0)(0,0,1)[7]	3.57e-05	FAIL	3.532e-05	FAIL	0.01878746
Model 3 = ARMA (6,1,0)	0.913	PASS	0.011	FAIL	0.01605916
Model 4 = ARMA(6,1,1)	0.907	PASS	0.01489	FAIL	0.01643544
Model 5 = ARMA(2,1,2)	0.000143	FAIL	7.34e-07	FAIL	0.013016
Naïve Model	0.955	PASS	9.021e-07	FAIL	0.06972841
Simple Average Model	0.596	PASS	1.871e-07	FAIL	0.1097014

- All individual models fail at least 1 test.
- The 5 models significantly better (much lower) MSE that the Naïve Model and Simple Average Model indicating they have better explanatory power

Combination Models

- Based on MSE we consider the top 3 models
  - Model 3 = ARMA (6,1,0)
  - Model 4 = ARMA(6,1,1)
  - Model 5 = ARMA (2,1,2)
- Seasonal Models perform better in sample as compared to out of sample.
- 3 combination models

Combination Model 1:

- Equally Weighted
- 1/3*Model 3 + 1/3*Model 4 + 1/3*Model 5

Combination Model 2:

- Inverse of MSE
- 1/MSE3*Model 3
  + 1/MSE4 *Model4 + 1/MSE5*Model 5

Combination Model 3:

- OLS Weighted

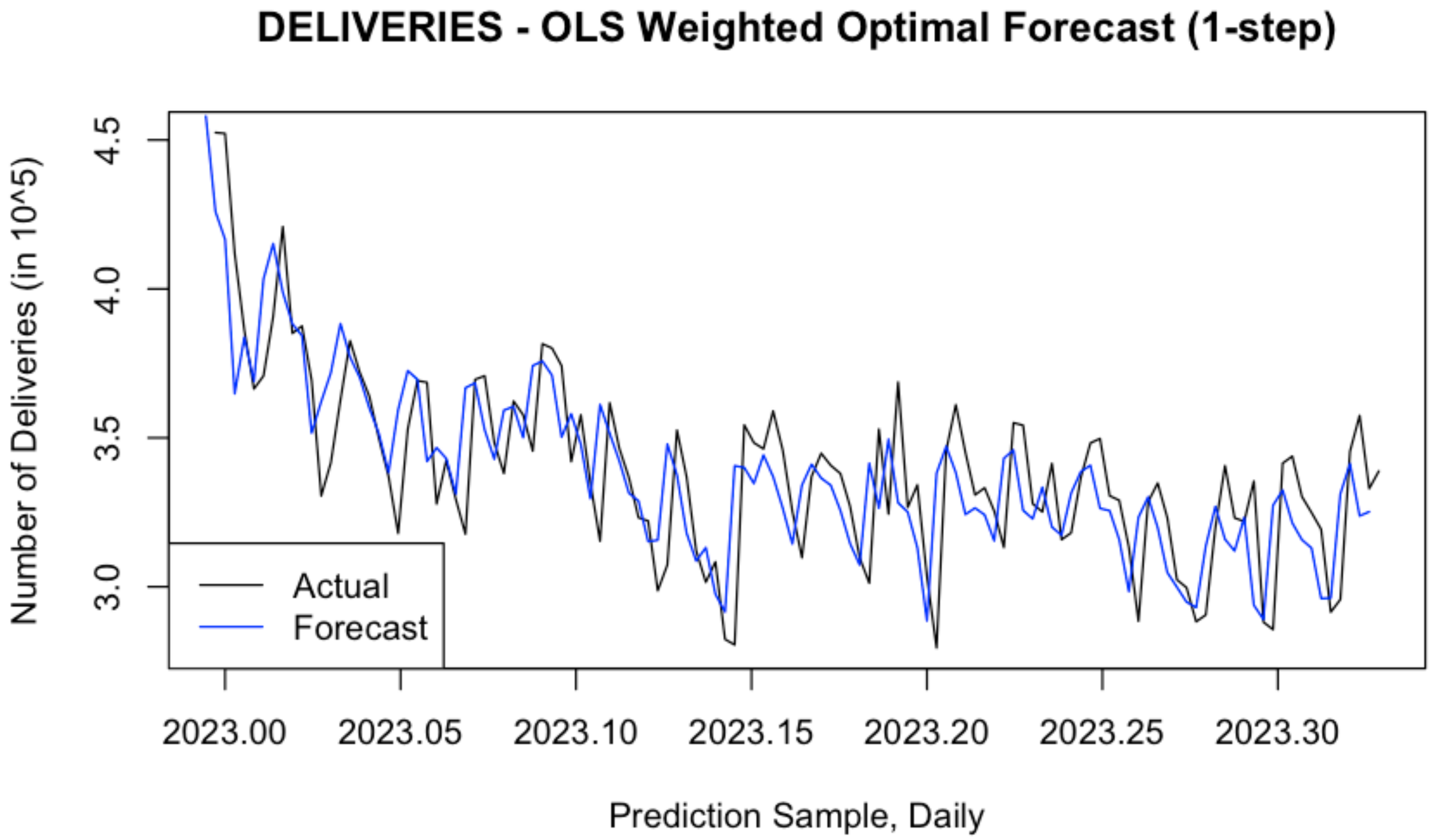
Combination of Forecast

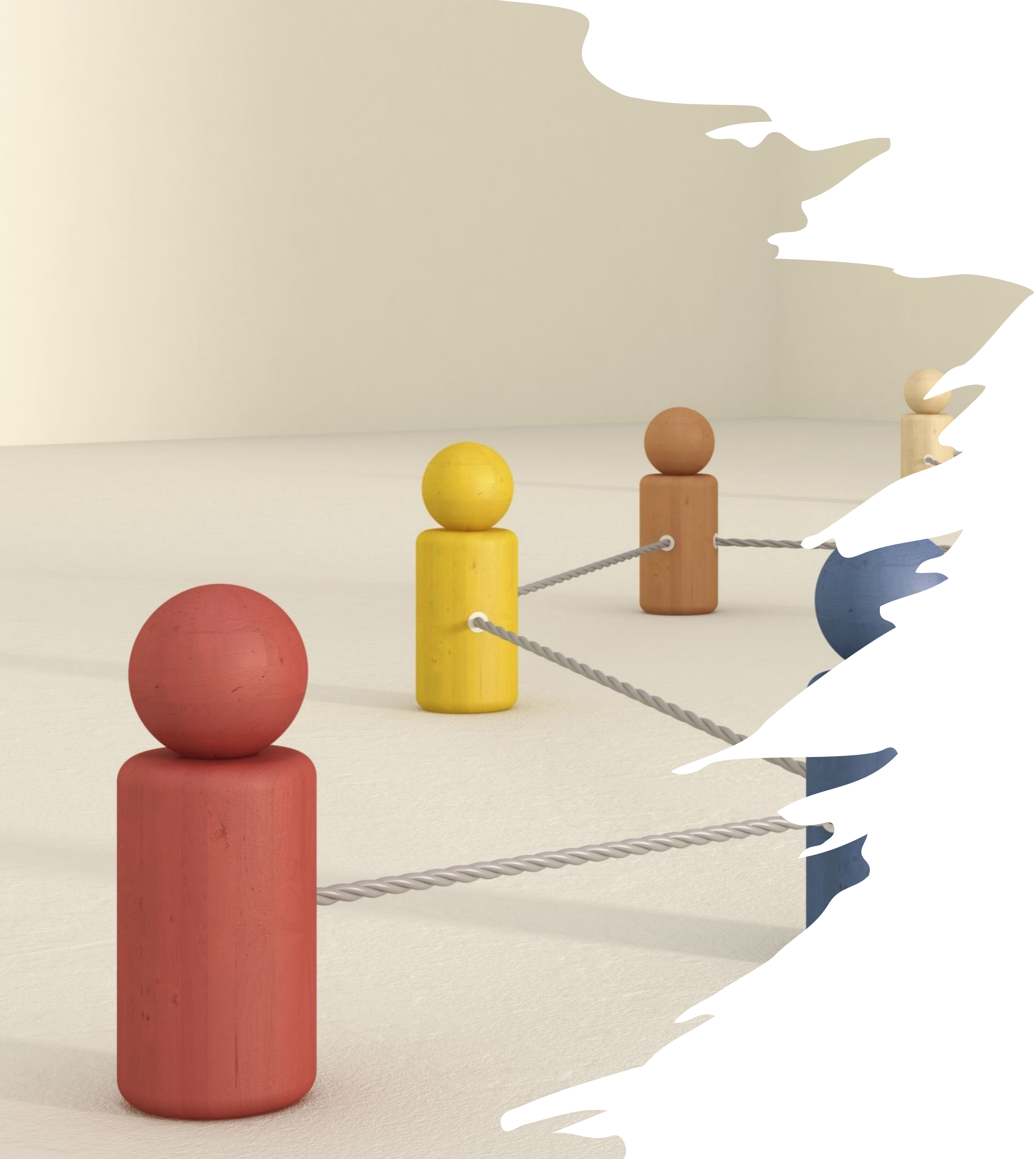
	Combination Model 1 Equal Weighted	Combination Model 2 Inverse MSE Weighted	Combination Model 3 OLS weighted
Model 3 = ARMA (6,1,0) MSE = 0.01605916	1/3	0.3114385	3.8382
Model 4 = ARMA (6,1,1) MSE = 0.01643544	1/3	0.3043083	-3.5130
Model 5 = ARMA (2,1,2) MSE = 0.013016	1/3	0.3842533	0.8644
MSE of combined forecast	0.01310528	0.01280656	0.008805348
MPE Test	P-value = 0.252, PASS	P-value = 0.175, PASS	P-value = 1, PASS
IET Test	P-value = 3.243e-05, FAIL	P-value = 1.239e-05, FAIL	P-value = 1, PASS

- Combination Model 3 passes all tests and has significant improvement in MSE.
- The OLS weighted forecast optimal forecast is the best forecast
- Visualize the forecast

OLS Weighted Combination Model

<u>Forecast Equation</u> = -0.6762
+ 3.8382*(forecast from ARMA(6,1,0))
+ (-3.5130)*(forecast from ARMA(6,1,1))
+ 0.8644*(forecast from ARMA (2,1,2))





Conclusion & Future Work

- Analyzed a real time series data and generated forecast that Company A can use to predict delivery volumes.
- Enable optimally allocate resources to ensure correct delivery planning
- Multivariate analysis
  - personal income /consumption data
  - interest rate data
  - other company specific data (discounts, deals)
- Analyze deeper why many of our models failed the forecast optimality test.

Thank you!

Annexure 1: References

- https://stats.stackexchange.com/questions/194453/interpreting-accuracy-results-for-an-arima-model-fit
- https://online.stat.psu.edu/stat510/lesson/4/4.2
- https://otexts.com/fpp2/seasonal-arima.html
- https://otexts.com/fpp2/arima-r.html
- https://otexts.com/fpp2/non-seasonal-arima.html
- https://online.stat.psu.edu/stat510/lesson/4/4.1
- https://timeseriesreasoning.com/contents/regression-with-arima-errors-model/
- https://stackoverflow.com/questions/59918142/creating-arima-model-with-a-minimum-value-of-forecast-error-in-r
- https://stackoverflow.com/questions/54232180/appending-time-series

Annexure 2a – Glossary of Metrics

- R-Squared: Proportion of the sample variation of the dependent variable in a regression model explained by the model's regressors. Adjusted R-square penalizes additional irrelevant regressors. Higher value is better.
- AIC (Akaike Information Criterion): Measures to select the best time series models by minimizing the residual variances. Penalizes additional irrelevant regressors. Lower value is better.
- MSE (Mean Squared Error): The sum of squared difference between the values that are fitted by the model, and observed values that are divided by the number of historical points, minus the number of parameters in the model. The number of parameters in the model is subtracted from the number of historical points to be consistent with an unbiased model variance estimate.
- RMSE (Root Mean Squared Error): The square root of the MSE.
- MAE (Mean Absolute Error):Computed as the average absolute difference between the values fitted by the model (one-step ahead in-sample forecast), and the observed historical data.
- MPE (Mean Percentage Error): showing the average percentage difference between the predicted and actual values.
- MAPE (Mean Absolute Percentage Error): The average absolute percent difference between the values that are fitted by the model and the observed data values.

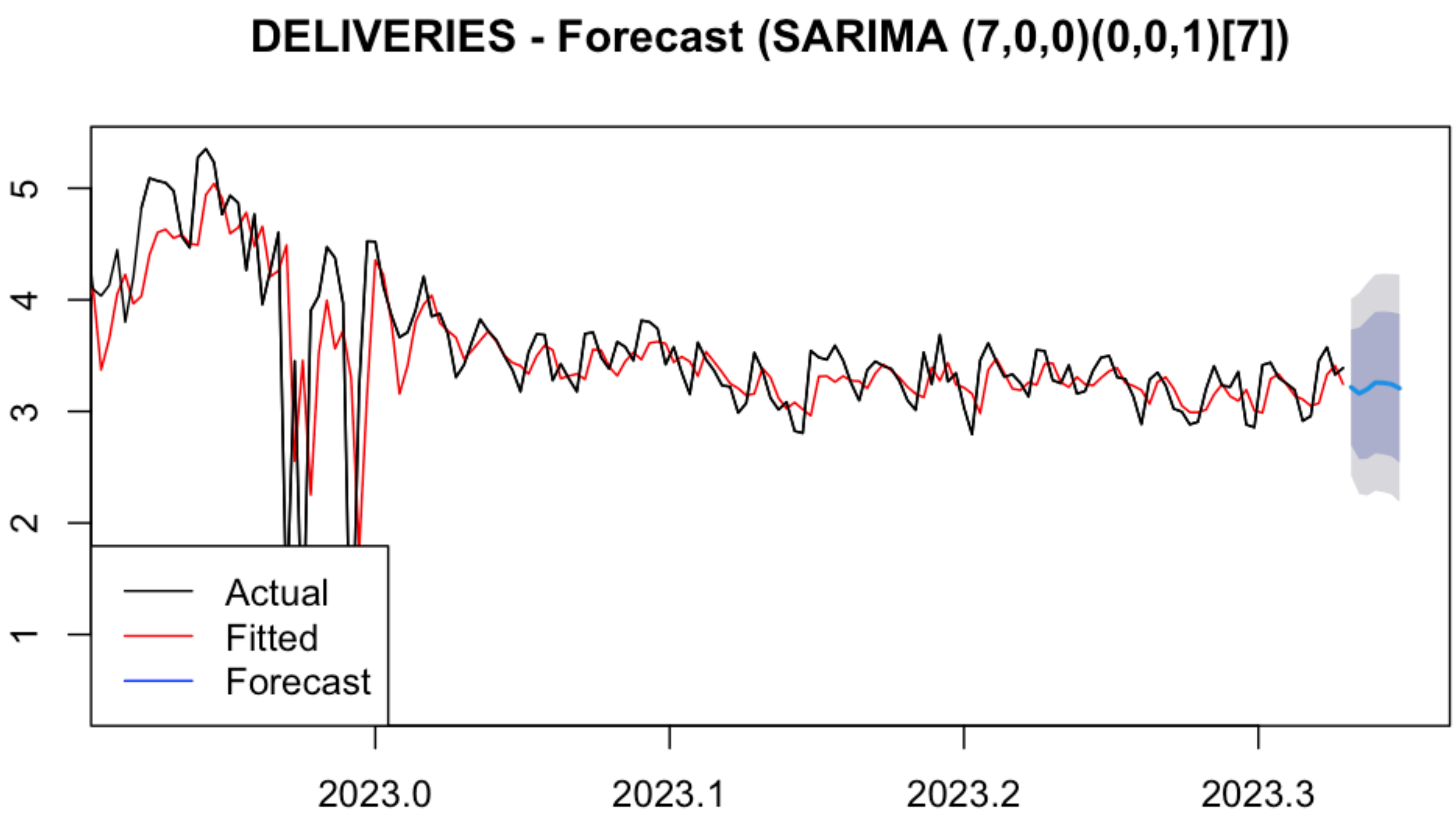
Annexure 2b – Glossary of Tests

- DF Test / Dickey Fuller Test: Statistic to assess the presence of unit root. If the null hypothesis can be rejected there is no unit root.
- Box LJung Test for White Noise: Test used to check if the residuals are white noise i.e. lack of autocorrelation. Rejecting the null hypothesis indicates that the residual is not white noise.
- MPE Test (mean prediction error test): Statistical assessment of whether the expected value of the forecast is zero in the case of symmetric loss function. Rejecting the null hypothesis indicates that the expected value of the forecast is not zero for symmetric loss function.
- IET (information efficiency test): Statistical assessment of whether the optimal forecast error in uncorrelated with any variables(s) in the information set. Rejecting the null hypothesis indicates that the optimal forecast is not completely uncorrelated with any variables in the information set.

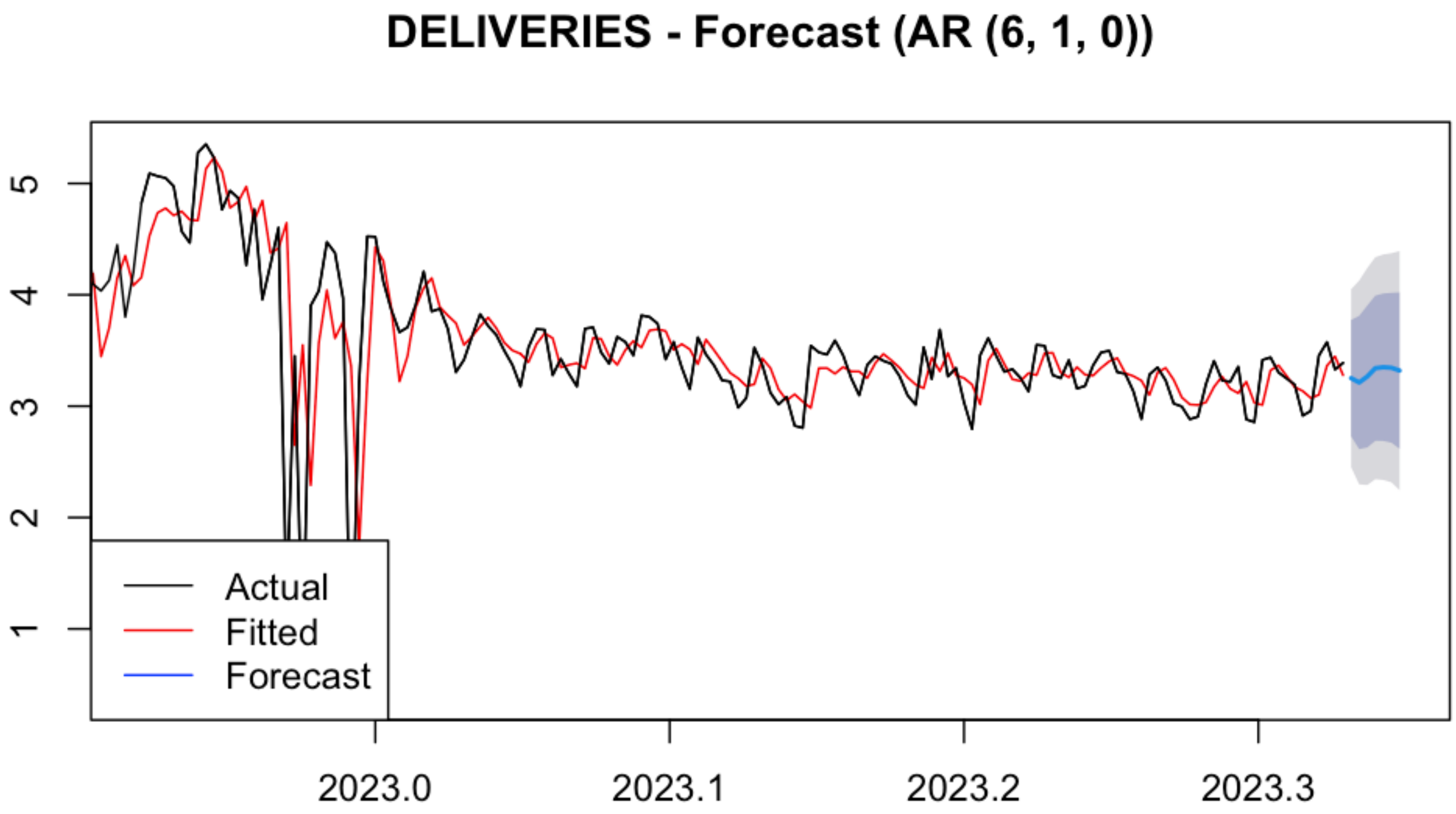
Annexure 2c – Glossary of Terms

- ACF (Auto-correlation Function): Collection of correlation coefficients between any two random variables in the stochastic process that are k periods apart for k = 1,2,3 ...
- PACF (Partial Auto-correlation Function): Controls for the information that runs in between the periods
- ARMA (auto regressive moving average model): Equation = (p, d, q) where p = order AR process, d = order of the difference, q = order of the MA process.
- SARMA (Seasonal auto regressive model) : Equation = (p, d, q)(P, D, Q)[period] where (p, d, q) is same as ARMA model. (P, D, Q) represent the order of the seasonal term and period denotes the frequency.
- OLS (ordinary least squares): Estimation technique to obtain the regression of coefficients that consists of minimizing the sum of squared residuals of the model.

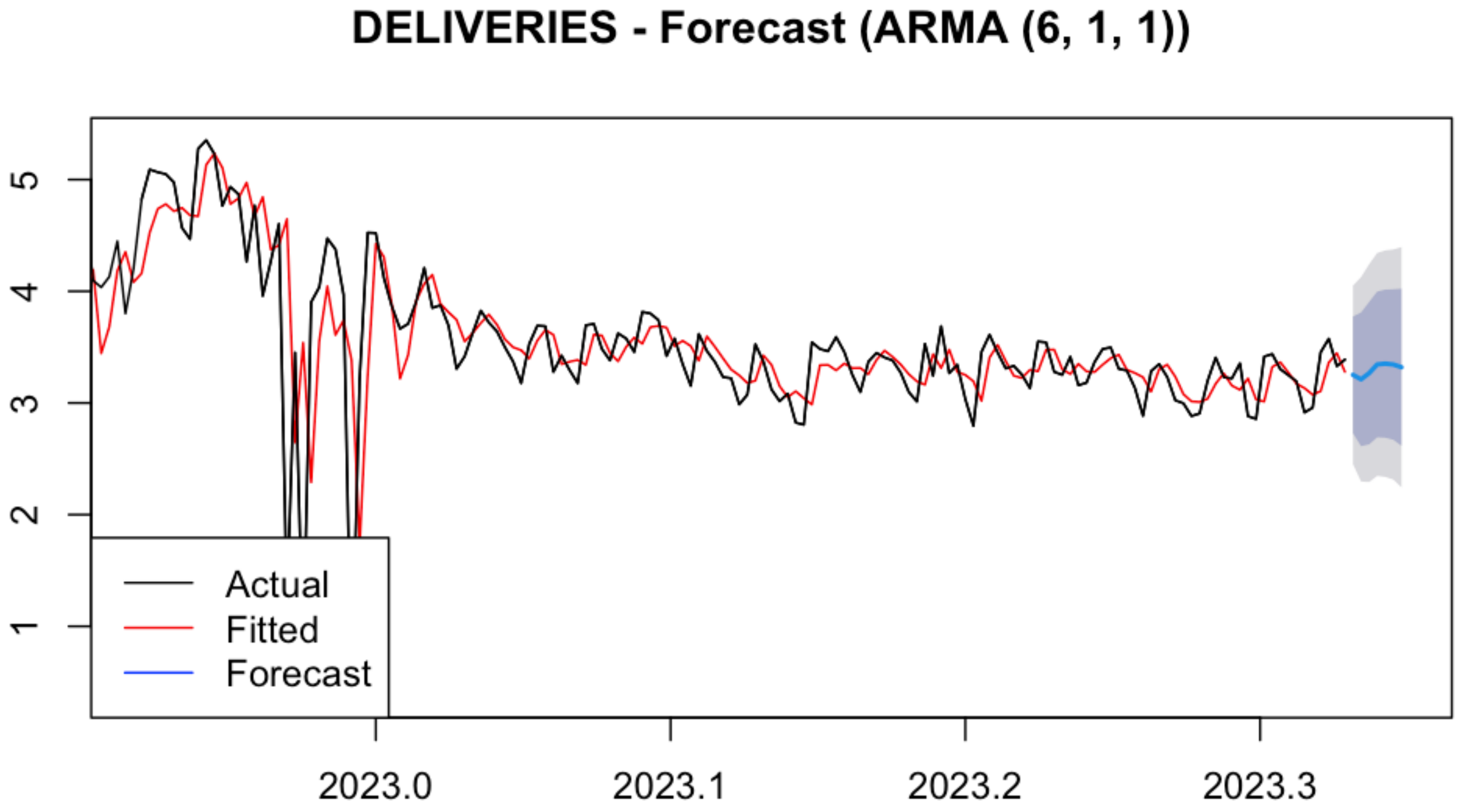
Annexure 3a: Multistep Forecast Model 2 – 7 Days Ahead



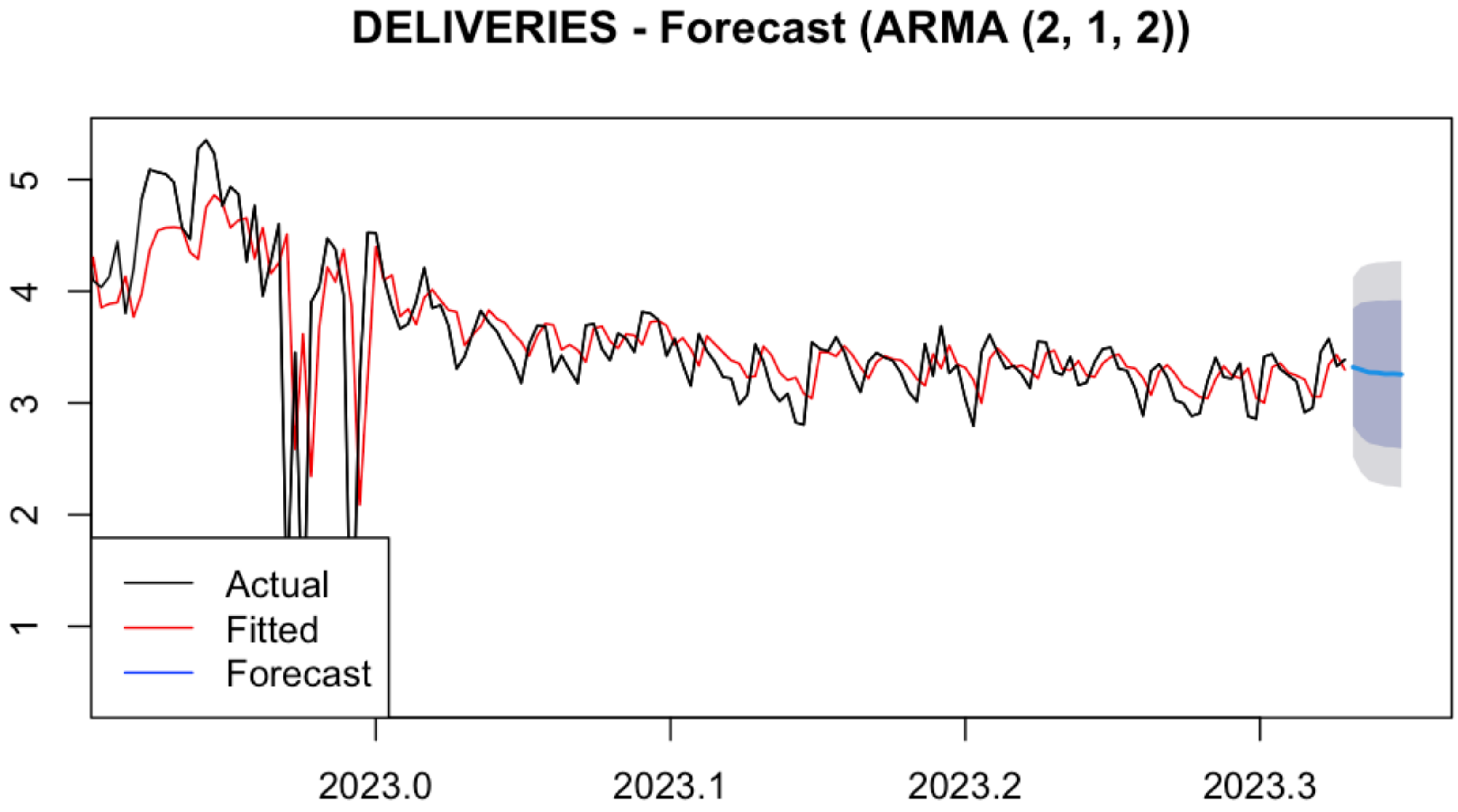
Annexure 3b: Multistep Forecast Model 3 – 7 Days Ahead



Annexure 3c: Multistep Forecast Model 4 – 7 Days Ahead



Annexure 3d: Multistep Forecast Model 5 – 7 Days Ahead



Annexure 4: Forecast Optimality Test Code

```
# Model 2 - ARMA (6,1,1)
fcast2<-numeric(prediction_size)
ferror2<-numeric(prediction_size)
loss2<-numeric(prediction_size)

for (i in  1: prediction_size) {
refit_ar6_ma1_d1 <- Arima(DELIVERIES[1:estimation_size + i], model=ar6_ma1_d1)
fcast2[i]<-forecast(refit_ar6_ma1_d1, h=1)$mean
ferror2[i] <- ps[i] - fcast2[i]
loss2[i] <- ferror2[i]^2
}

mpetest2 <- lm(ferror2 ~ 1)
summary(mpetest2)

IETest2 <- lm(ferror2 ~ fcast2)
summary(IETest2)
```