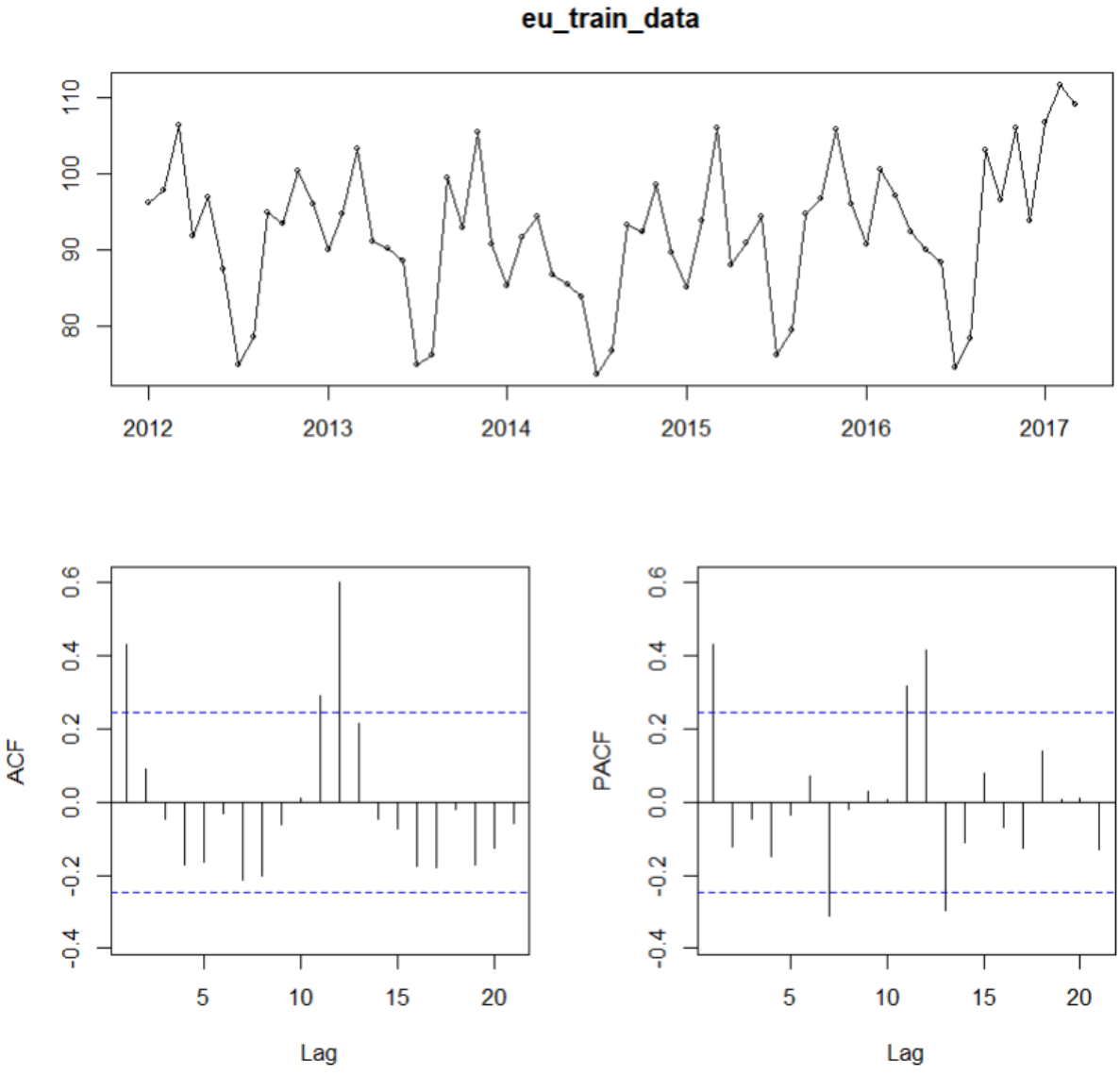
Praveen Subramani

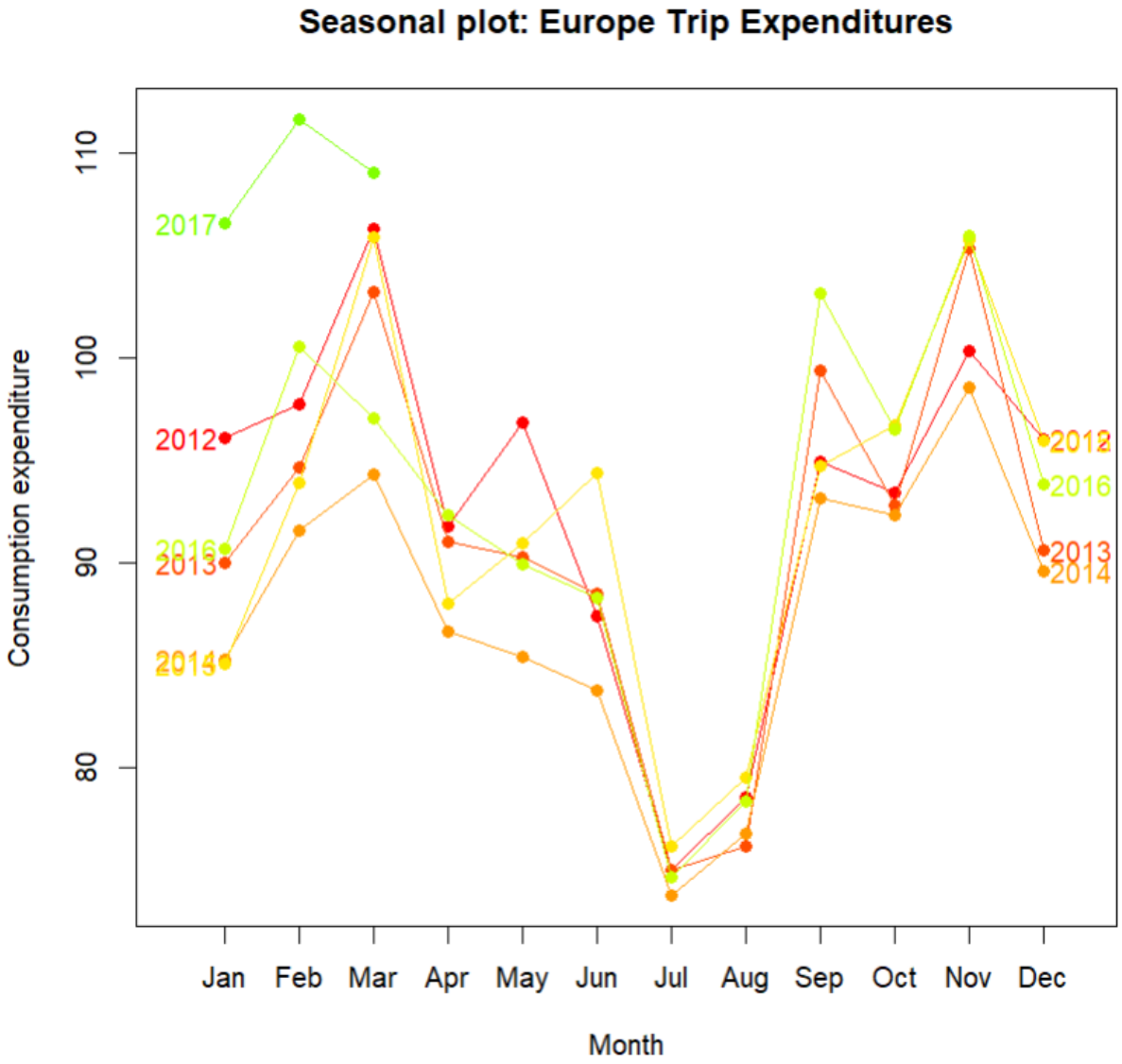
Exercise 1

Question 1

This report analyses monthly Europe trip expenditures, with data covering the period from January 2012 to March 2017. A visual representation of the trend is provided in the figure below.



The figure presents a time series analysis of a measurement from 2012 to 2017. The line chart depicts fluctuations in the data, while the ACF and PACF plots below it reveal the correlation of the data with itself at different time lags.

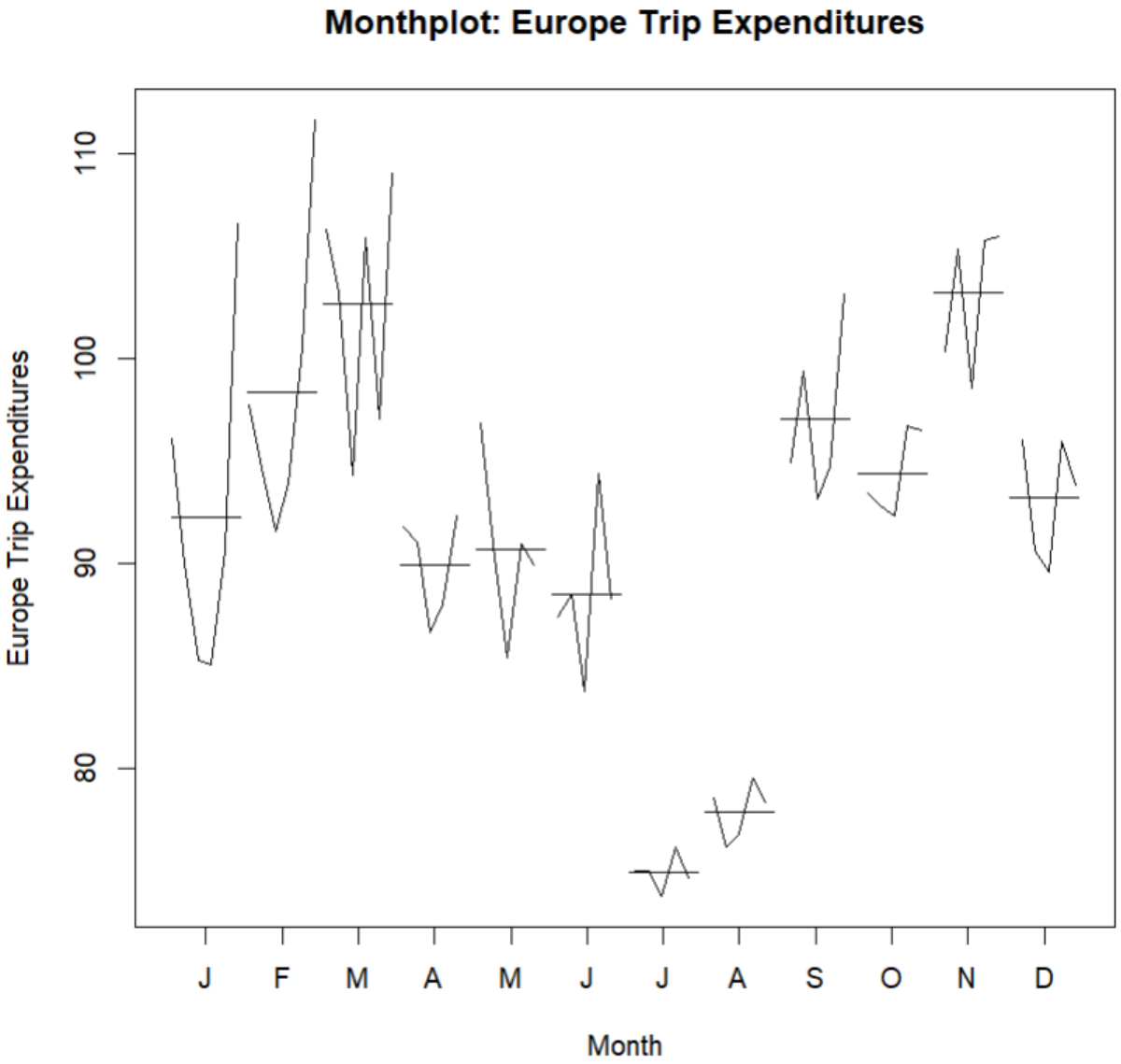


This plot showcases seasonal spending patterns for European trips across various years (2012-2017, with limited data for 2017). Consistent trends emerge:

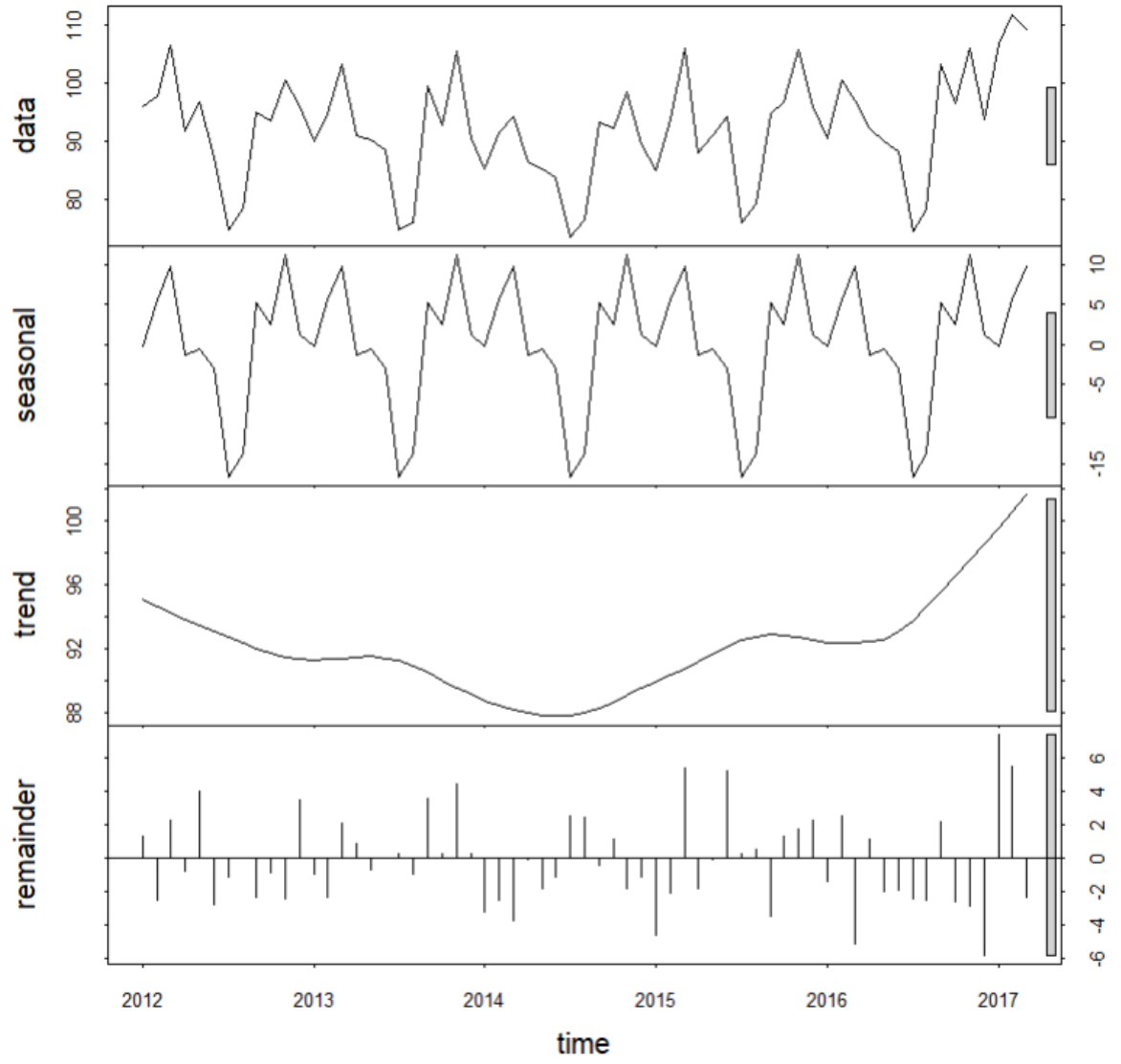
March Spike: Spending increases in March (potentially due to spring travel deals or ending winter).

Summer Peak: Expenditures rise from May, peaking in June/July (likely aligned with summer holidays).

Yearly Variations: Spending levels differ across years, suggesting factors beyond seasonality (e.g., economic conditions, travel preferences).



Unveiling seasonal patterns in European trip spending (2012-2017) March sees a spending rise (spring deals?), followed by a summer peak (June/July) likely due to peak travel season. Intriguingly, spending varies across years, suggesting factors beyond seasonality (economics, travel trends). This helps travel businesses anticipate demand and budget-conscious travelers find potentially cheaper travel times.



This graph shows European trip spending broken down into parts:

- Overall Spending: The main line with ups and downs represents actual spending each year.
- Seasonal Swings: The middle line shows the typical spending pattern throughout the year (e.g., higher in summer).
- Long-Term Trend: The bottom line shows a general increase in spending over time.

```
> summary(eu_train_data)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.9865  0.9887  0.9893  0.9891  0.9898  0.9911
```

Question 2

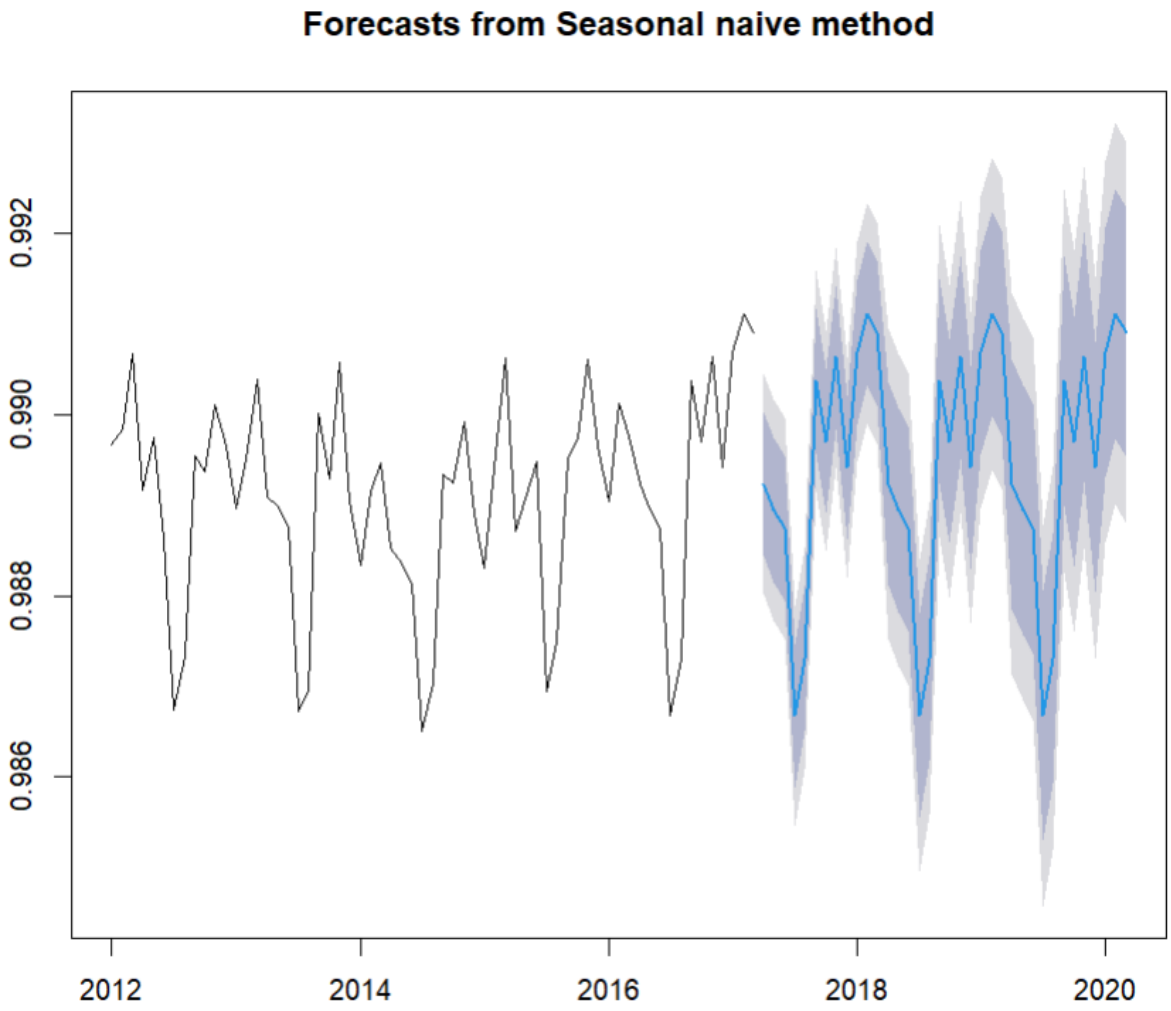
```
        Augmented Dickey-Fuller Test

data:  eu_train_data
Dickey-Fuller = -4.4262, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

An Augmented Dickey-Fuller Test for "eu_train_data". The test statistic (-4.4262) and a low p-value (0.01) lead us to reject the null hypothesis of a unit root (non-stationarity) in the data. In simpler terms, the test confirms that the "eu_train_data" time series is stationary. This means the data doesn't exhibit trends or seasonality over time, making it suitable for further analysis and forecasting with methods like ARIMA models. Stationarity is crucial because it allows models to make assumptions about the data's structure for accurate predictions.

```
> BoxCox.lambda(eu_train_data)
[1] -0.9999242
```

The Box-Cox transformation is a statistical technique that adjusts data to stabilize variance and make it more normally distributed. A lambda value of -0.999 suggests a strong transformation towards a reciprocal relationship, which can be useful in statistical analysis for meeting certain assumptions.

Question 3 - Seasonal Naïve Method

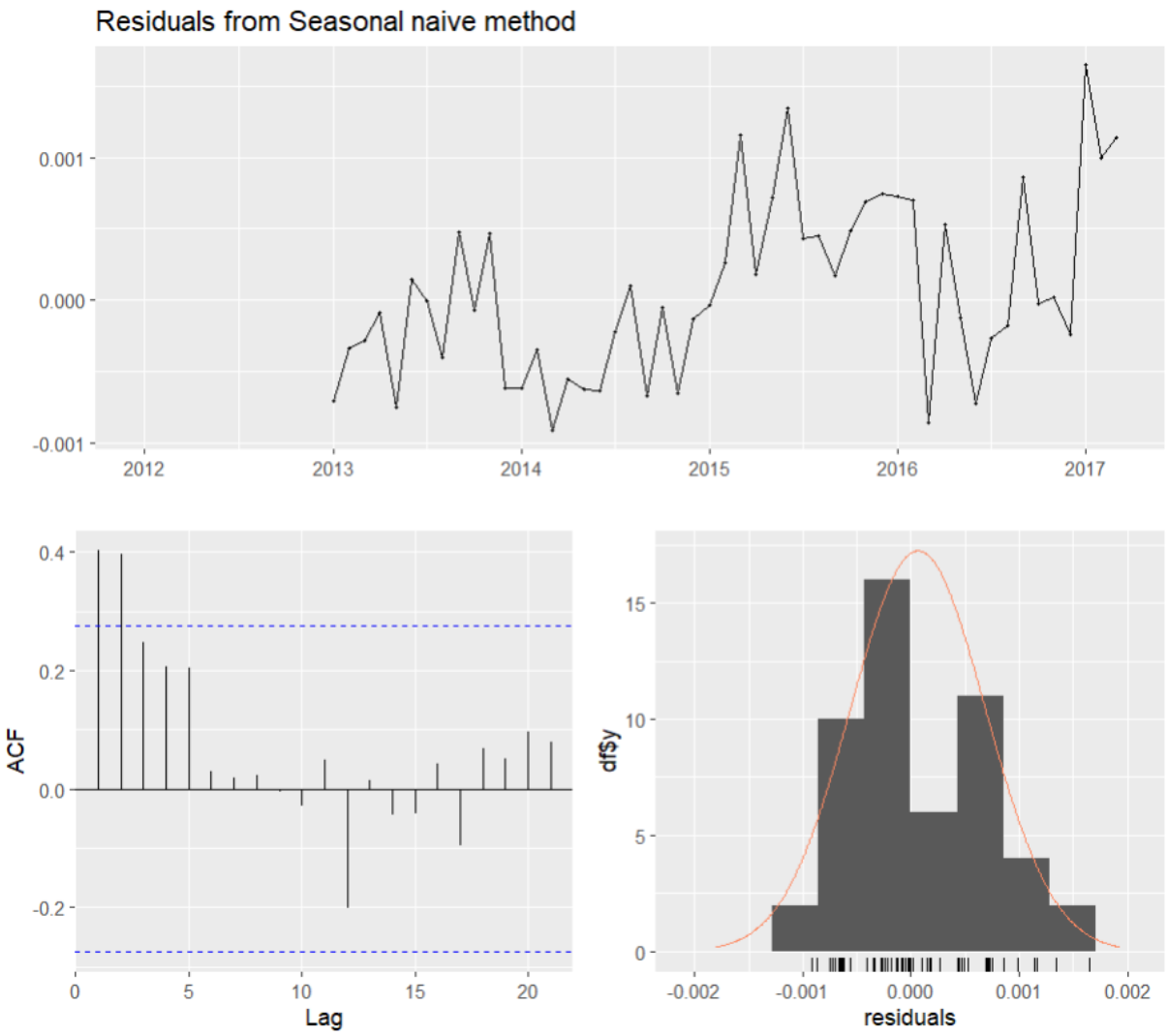


This "Forecasts from Seasonal Naive Method" graph depicts observed data (black line) with fluctuations until 2018. Blue lines forecast future increases with a widening confidence interval (grey area) indicating growing uncertainty as forecasts approach 2020. This suggests the metric might rise, but with increasing uncertainty in later years.

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	6.483612e-05	6.209691e-04	5.023964e-04	0.006526063	0.05077485	1.0
Test set	1.055783e+02	1.059724e+02	1.055783e+02	99.064388049	99.06438805	210149.3

	ACF1	Theil's U
Training set	0.4019482	NA
Test set	0.3650504	10.00413

While the model performs well on training data (low error metrics), its high errors and statistics like Theil's U on the test data suggest it struggles with unseen data (overfitting) and may require adjustments for better generalizability.



Analysing the residuals from the Seasonal Naive forecast reveals they fluctuate around zero, suggesting the method captured seasonality and trend. However, ACF plot indicates remaining patterns, and the histogram shows slight skewness, implying the model could be further refined.

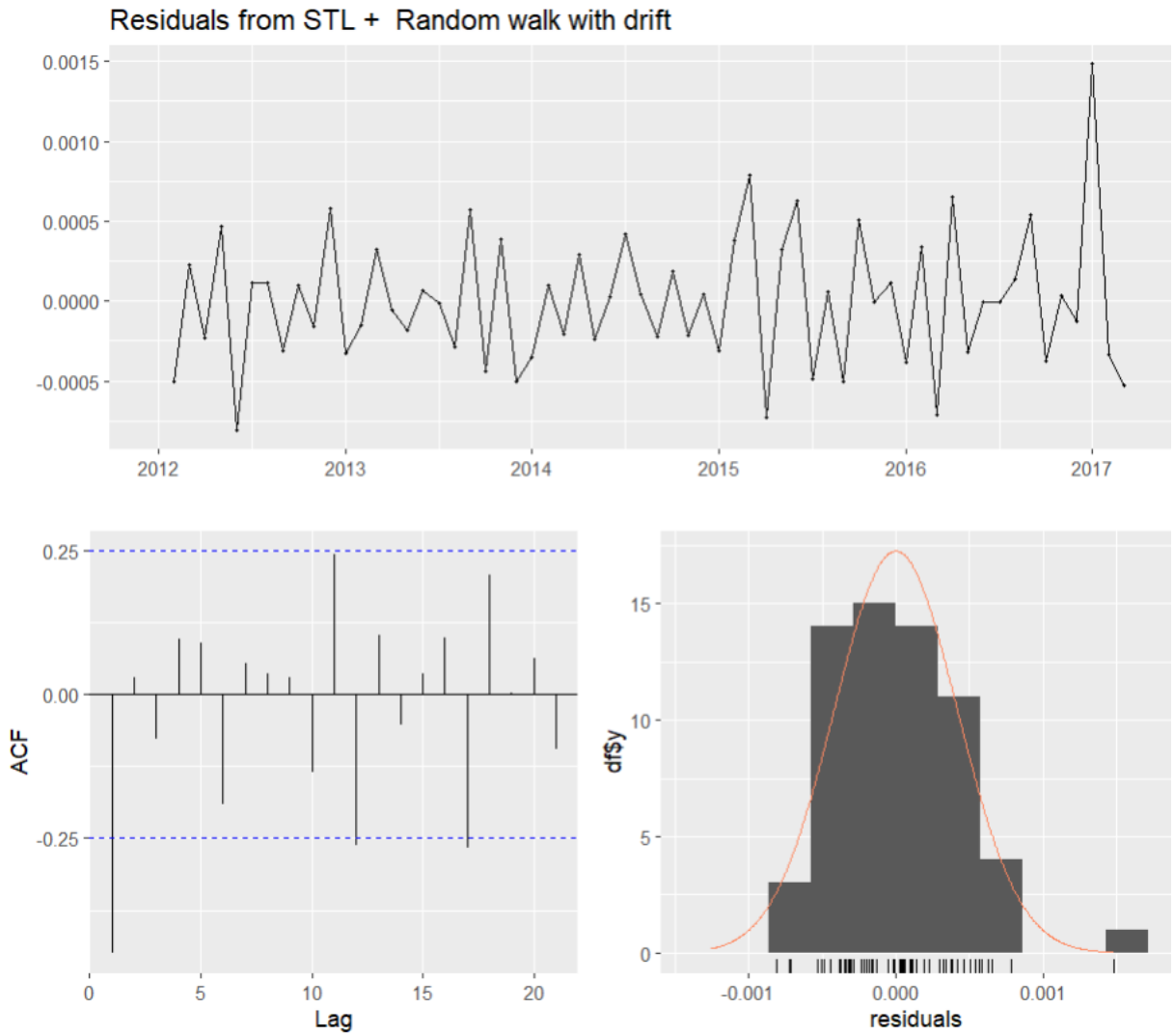
```
        Ljung-Box test

data:  Residuals from Seasonal naive method
Q* = 28.866, df = 13, p-value = 0.006839
```

The Ljung-Box test reveals significant autocorrelation ($Q^*=28.866$, p-value=0.0068<0.05) in the residuals from the Seasonal Naive model. This suggests the model isn't capturing all the seasonal patterns, and there's remaining correlation

in the errors. Consider exploring alternative models or adding terms to account for this autocorrelation.

Question 4

Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Q	df	p-value
Seasonal Naive	6.48E-05	6.21E-04	5.02E-04	0.00653	0.05077	1	0.40195	NA	NA	NA
Mean	-4.76E-17	1.15E-03	8.80E-04	-0.00013	0.08895	1.75	0.44693	74.63	13	1.12E-10
Naive	1.00E-05	1.19E-03	9.45E-04	0.00194	0.09557	1.88	-0.14519	60.92	13	3.60E-08
Random Walk with Drift	3.22E-17	1.19E-03	9.46E-04	-7.49E-05	0.09563	1.88	-0.14519	60.92	13	3.60E-08
STL Naive Log	1.05E-06	4.10E-04	3.21E-04	0.000097	0.03242	0.64	-0.44926	30.06	13	4.61E-03
STL ETS	1.78E-05	3.47E-04	2.54E-04	0.00179	0.0257	0.51	-0.03628	9.61	13	0.73
STL ARIMA	2.75E-05	3.69E-04	2.67E-04	0.00277	0.02702	0.53	-0.07991	10.17	12	0.6
STL Naive Log	1.05E-06	4.15E-04	3.24E-04	-0.08988	3	0.64	-0.44926	30.07	13	4.60E-03
STL RW Drift Log	-7.27E-19	4.15E-04	3.24E-04	-0.08026	3	0.64	-0.44926	30.07	13	4.60E-03
STL ETS Log	1.80E-05	3.51E-04	2.57E-04	-0.26382	2.39	0.51	-0.03646	9.61	13	0.73
STL ARIMA Log	1.17E-05	3.51E-04	2.54E-04	-0.20391	2.36	0.5	-0.03079	9.23	12	0.68



This time series decomposition graph reveals the inner workings of your data (2012-2017). It separates overall trends (upward), seasonal fluctuations (repeating patterns), and random variations (leftover "noise") to better understand the data and potentially forecast future values.

```
        Ljung-Box test

data:  Residuals from Random walk with drift
Q* = 60.915, df = 13, p-value = 3.603e-08

Model df: 0.   Total lags used: 13
```

The Ljung-Box test reveals significant autocorrelation (p-value very low) in the residuals of the random walk with drift model. This suggests the model might not fully

capture the data's underlying process. In other words, there's leftover "pattern" in the errors, indicating a need for model revision or exploring other factors to improve its fit.

Question 5

In the provided table, the AIC and BIC values for each model are as follows:

ETS(A,A,N): AIC = -726.8569, BIC = -716.1412

ETS(A,N,N): AIC = -731.3038, BIC = -724.8744

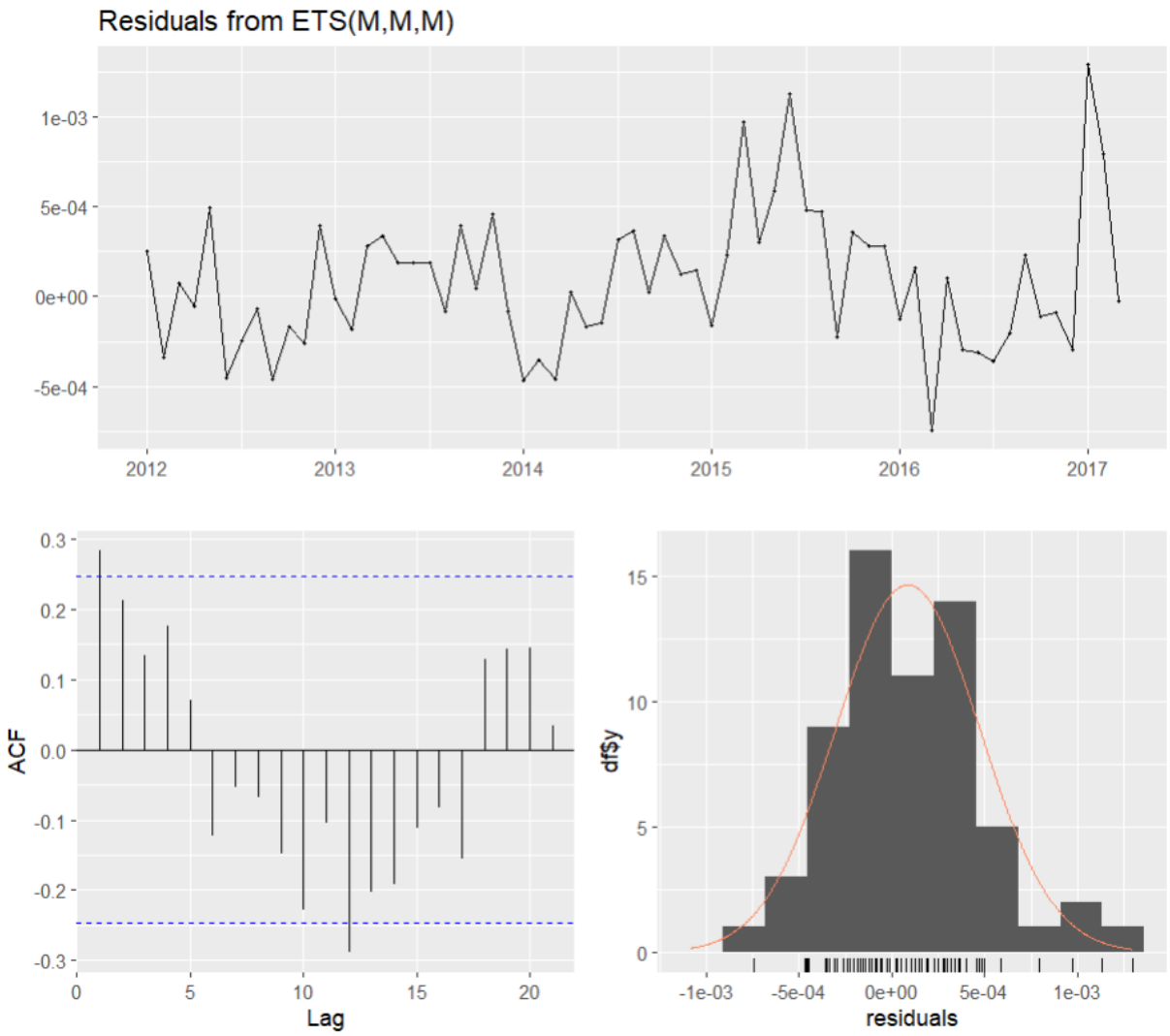
ETS(A,A,A): AIC = -694.7712, BIC = -658.3379

ETS(M,M,M): AIC = -693.6636, BIC = -657.2303

ETS(M,A,M): AIC = -693.8671, BIC = -657.4338

Based on the AIC and BIC values, the ETS(M,M,M) model has the lowest values, followed closely by the ETS(M,A,M) model. Therefore, these two models are considered to perform better based on AIC and BIC criteria.

Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Q	df	p-value
ETS(ANA)	4.10E-05	3.82E-04	2.85E-04	0.0041	0.0288	5.672	0.1898	10.467	13	0.6553
ETS(MNN)	2.12E-05	3.62E-04	2.67E-04	0.0021	0.027	5.311	-0.0333	8.0993	13	0.8371
ETS(AAN)	5.06E-05	3.64E-04	2.63E-04	0.0051	0.0266	5.231	-0.041	9.6213	13	0.7246
ETS(ANN)	2.12E-05	3.62E-04	2.67E-04	0.0021	0.027	5.311	-0.0327	8.0883	13	0.8378
ETS(AAA)	9.68E-05	3.88E-04	2.98E-04	0.0098	0.0301	5.929	0.2426	33.639	13	0.0014
ETS(MMM)	8.45E-05	3.91E-04	3.03E-04	0.0085	0.0306	6.027	0.2833	30.346	13	0.0042
ETS(MAM)	9.35E-05	3.90E-04	3.01E-04	0.0094	0.0304	5.99	0.2785	34.427	13	0.001

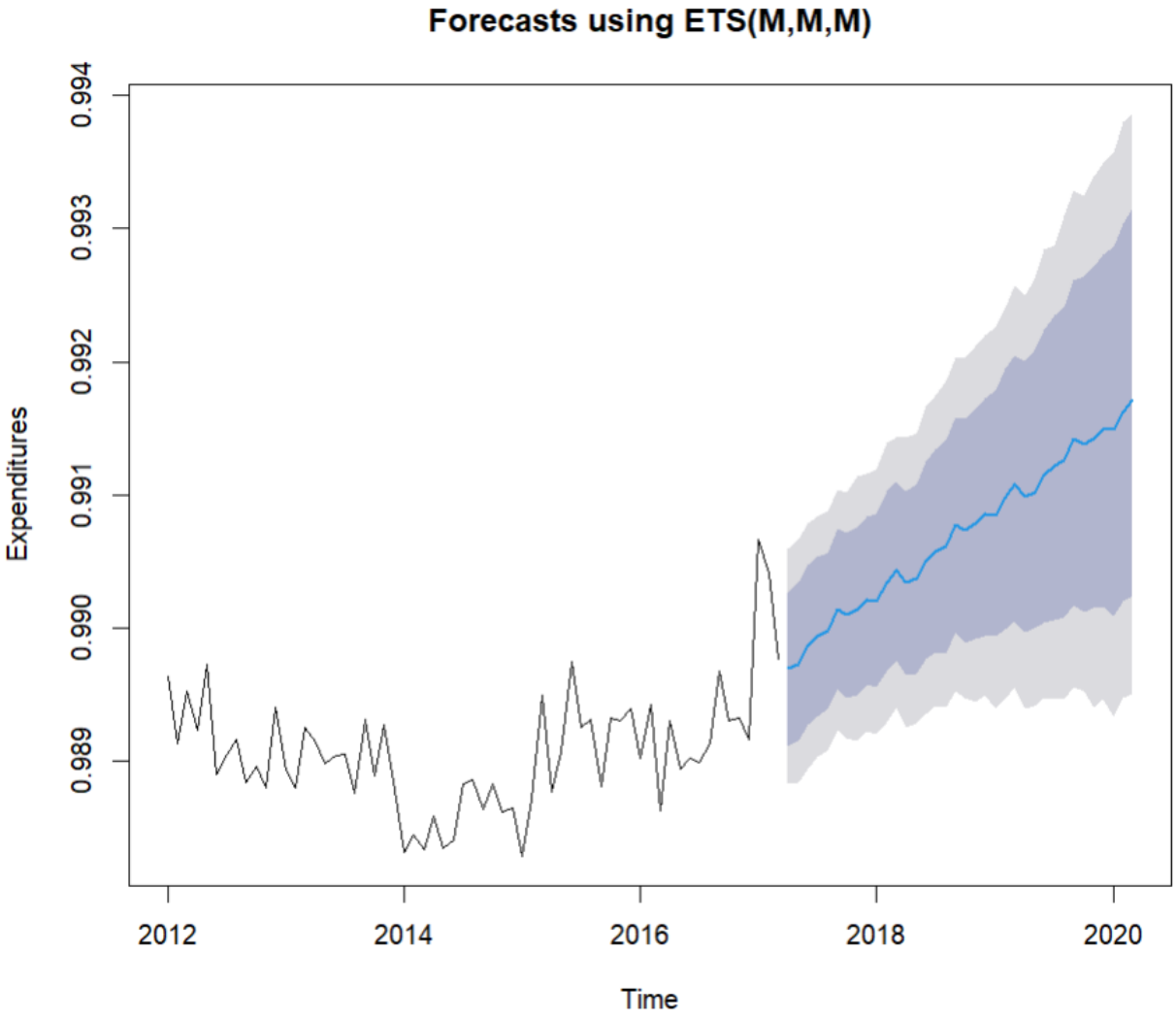


The ETS(M,M,M) model's forecast seems decent with residuals centered around zero and a near-normal distribution. However, the ACF plot suggests potential improvements as it reveals leftover autocorrelation in the residuals, indicating the model might miss some patterns.

```
        Ljung-Box test

data:  Residuals from ETS(M,M,M)
Q* = 30.346, df = 13, p-value = 0.004198

Model df: 0.   Total lags used: 13
```

The Ljung-Box test results (Q* = 30.346, p-value = 0.004) indicate significant autocorrelation in the residuals of the ETS(M,M,M) model. This suggests the model might not fully capture the data's structure and may need adjustments for a better fit.

Parameter	Value
alpha	0.0914
beta	0.0150
gamma	0.0001
l	0.9895
b	1.0000
s1	1.0000
s2	0.9999
s3	0.9999
s4	1.0001
s5	1.0000
s6	1.0000
s7	1.0000
s8	0.9999
s9	0.9999
s10	1.0001
s11	1.0000
s12	1.0000



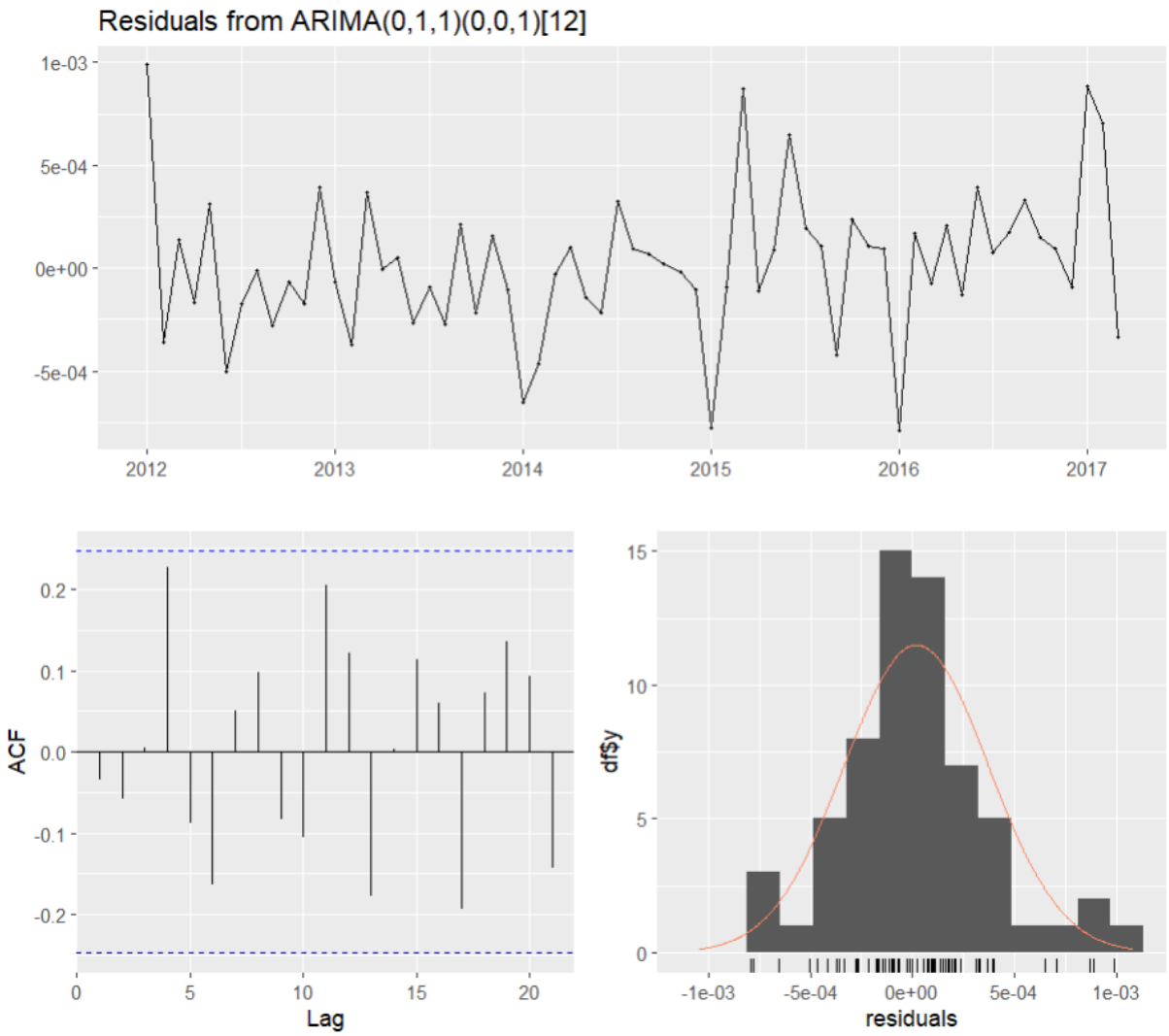
The "Forecasts using ETS(M,M,M)" plot shows actual expenditure data (up to 2018) and forecasts beyond 2020. The ETS model predicts a rising trend with increasing uncertainty (wider confidence interval) the further out the forecast goes. This model is useful but limitations like potential error, especially in long-term forecasts, should be considered.

Question 6

Auto Arima

```
ARIMA(0,1,1)(0,0,1)[12]

Coefficients:
          ma1     sma1
      -0.6022  -0.6701
s.e.   0.0991   0.2978

sigma^2 = 1.298e-07:  log likelihood = 405.03
AIC=-804.06   AICc=-803.65   BIC=-797.68

Training set error measures:
                      ME         RMSE          MAE          MPE       MAPE
Training set 1.901176e-05 0.0003515727 0.0002594643 0.001911164 0.02622907
                   MASE        ACF1
Training set 0.5164534 -0.03443791
```

The ARIMA(0,1,1)(0,0,1)[12] model shows a good fit to the data. Low error measures and MASE below 1 indicate accurate forecasts. However, a negative ACF1 value suggests potential for improvement by addressing some remaining autocorrelation at lag 1.



This ARIMA model's diagnostic plots look promising! The residuals fluctuate around zero with no significant autocorrelation, suggesting the model's predictions are unbiased and the errors are random. While the normality of residuals might not be perfect, it seems the ARIMA(0,1,1)(0,0,1)[12] captures the data's behavior well.

```
        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(0,0,1)[12]
Q* = 15.714, df = 11, p-value = 0.1521

Model df: 2.   Total lags used: 13
```

The Ljung-Box test results (Q*=15.714, p-value=0.1521) for the ARIMA(0,1,1)(0,0,1)[12] model's residuals indicate no significant autocorrelation (up

to lag 13). This suggests the model is a good fit for the data as the residuals appear random and independent.

```
ARIMA(0,1,1)(0,0,1)[12]

Coefficients:
          ma1     sma1
      -0.6022  -0.6701
s.e.   0.0991   0.2978
```

The ARIMA(0,1,1)(0,0,1)[12] equation in terms of the backward shift operator B can be written as:

(1 - B)(1 - 0.6022B)(1 - 0.6701B^12)yt = et

This can be further simplified as:

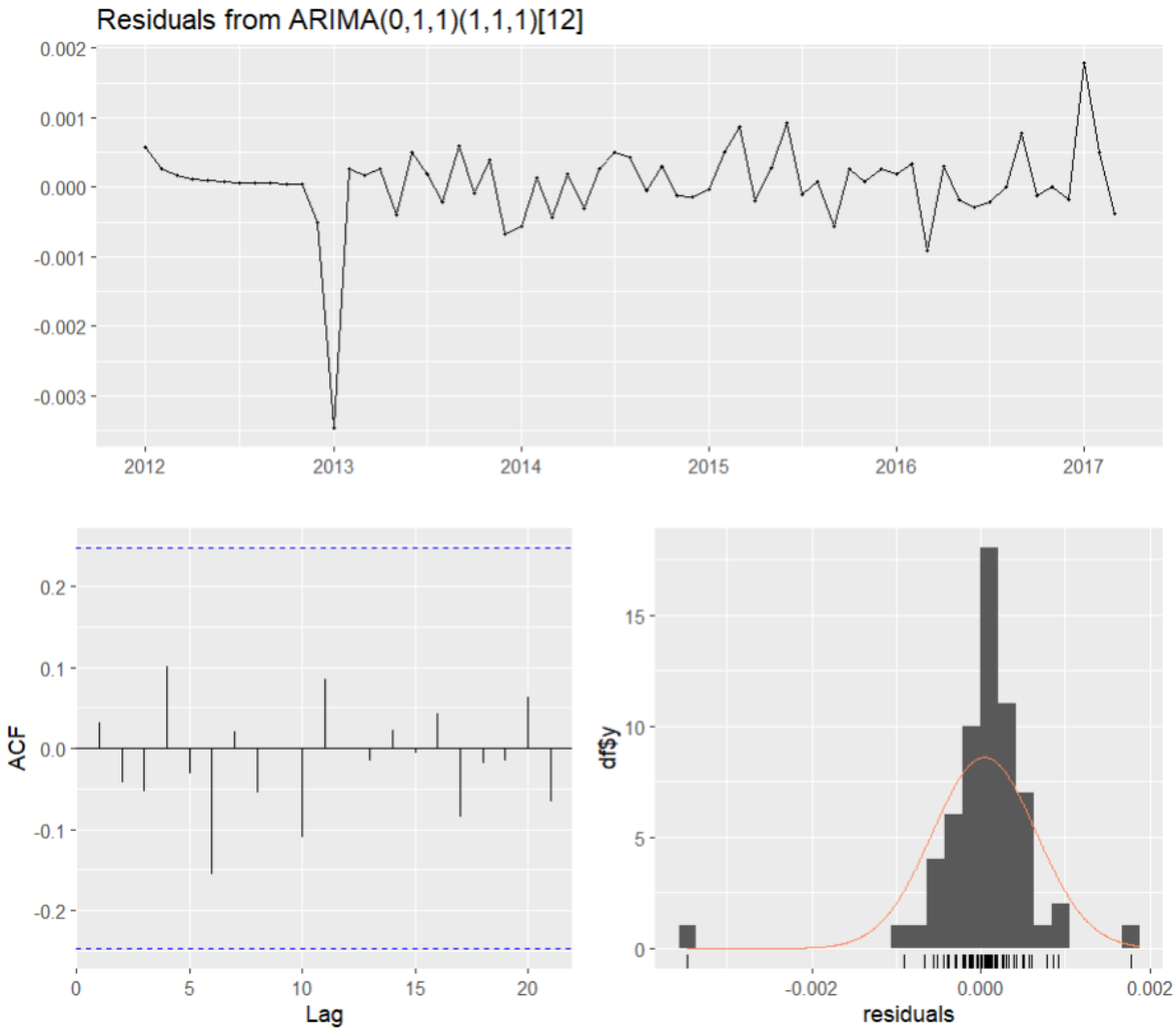
yt - yt-1 - 0.6022et-1 + 0.6022yt-1 - 0.6701yt-12 + 0.6701yt-13 = et/B

```
                      ME         RMSE          MAE          MPE        MAPE
Training set 9.346652e-05 3.904306e-04 3.009446e-04  0.009438219  0.03042137
Test set     1.055770e+02 1.059712e+02 1.055770e+02 99.063124787 99.06312479
                    MASE      ACF1 Theil's U
Training set 5.990183e-01 0.2784605        NA
Test set     2.101468e+05 0.3650794  10.00401
```

Manual Arima

```
Call:
arima(x = eu_train_data_decomp, order = c(0, 1, 1), seasonal = list(order = c(1,
    1, 1), period = 12))

Coefficients:
          ma1     sar1     sma1
      -0.6197  -0.3328  -0.2809
s.e.   0.1155   0.3697   0.4502

sigma^2 estimated as 2.161e-07:  log likelihood = 310.23,  aic = -612.45

Training set error measures:
                      ME         RMSE          MAE         MPE       MAPE
Training set 4.407162e-05 0.0006113397 0.0003657985 0.004444089 0.03697887
                  MASE      ACF1
Training set 1.069585 0.0322026
```

It shows an ARIMA model fit to "eu_train_data_decomp" (likely decomposed time series data). The model captures seasonality and indicates negative relationships between past errors/values and current errors/values. Low error measures and a good AIC score suggest a good fit, but the true test is how it performs on unseen data.



The ARIMA(0,1,1)(1,1,1)[12] model seems to be a good fit for the data. Residuals show no trend, have a near-normal distribution, and are relatively unbiased, suggesting the model captures the data's behavior well. However, the ACF plot hints at potential improvements by investigating specific lags with autocorrelation.

```
        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(1,1,1)[12]
Q* = 4.6944, df = 10, p-value = 0.9106

Model df: 3.   Total lags used: 13
```

The Ljung-Box test results are positive news! With a high p-value (0.9106), we can't reject the null hypothesis of no autocorrelation in the residuals. This means the residuals appear random, suggesting your ARIMA model captures the data effectively without leaving any significant patterns behind. It's a good sign that the model fits the data well, assuming normality of residuals and adherence to other ARIMA assumptions.

Model	ME	RMSE	MAE	MPE	MAPE	Q	df	p-value
ARIMA(0,1,1)(1,1,1)[12]	4.41E-05	6.11E-04	3.66E-04	0.00444409	0.0369789	4.6944	10	0.9106
ARIMA(1,1,1)(1,1,1)[12]	4.25E-05	6.11E-04	3.65E-04	0.00428956	0.0368505	4.7803	9	0.853
ARIMA(0,1,1)(1,0,1)[12]	1.91E-05	3.22E-04	2.37E-04	0.00192098	0.0239738	10.367	10	0.4089
ARIMA(0,1,1)(0,1,1)[12]	4.44E-05	0.00061	3.59E-04	0.00448146	0.0362788	4.2171	11	0.9631

Question 7

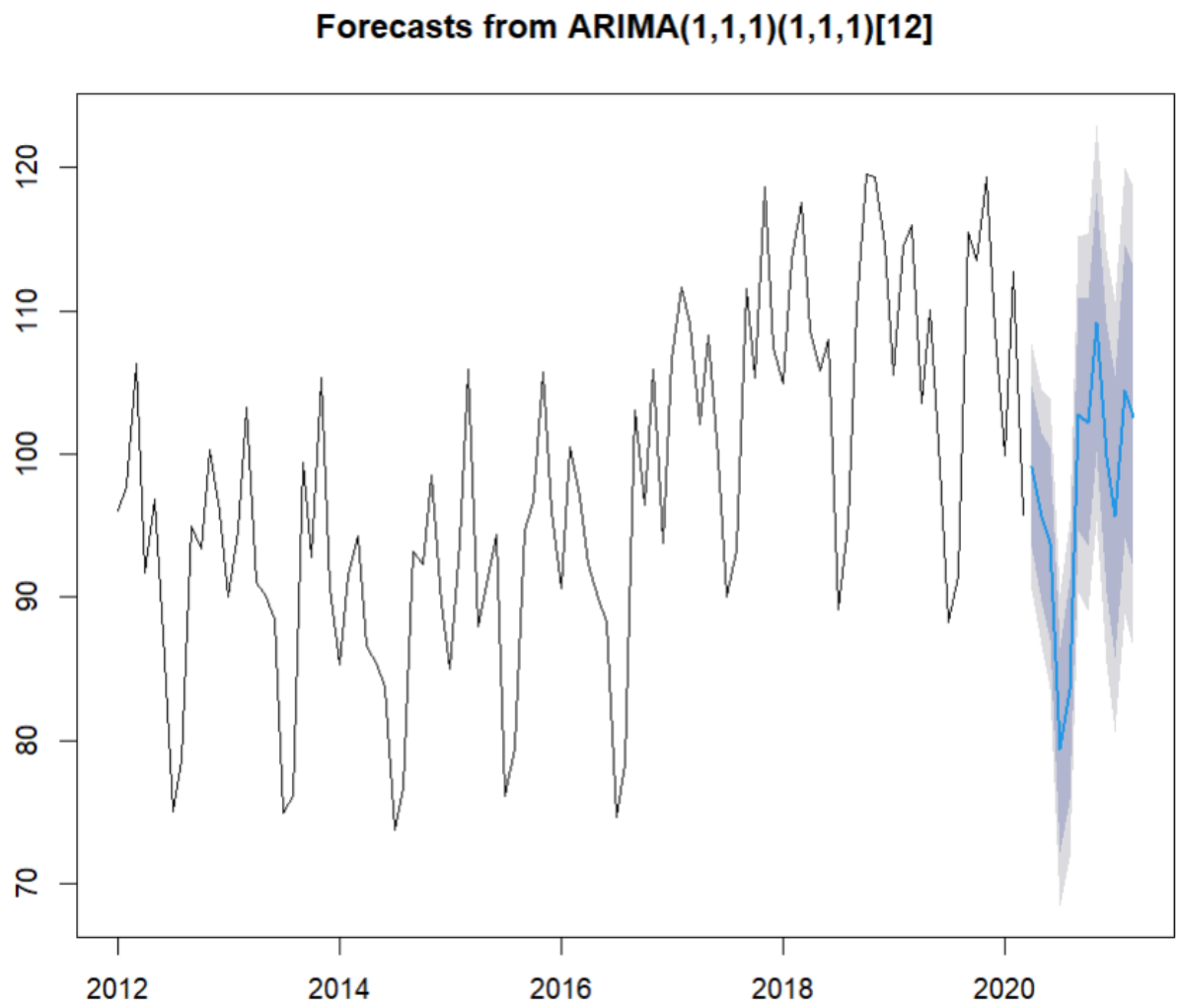
Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Q	df	p-value
Seasonal Naive	6.48E-05	6.21E-04	5.02E-04	0.00653	0.05077	1	0.40195	NA	NA	NA
Mean	-4.76E-17	1.15E-03	8.80E-04	-0.00013	0.08895	1.75	0.44693	74.63	13	1.12E-10
Naive	1.00E-05	1.19E-03	9.45E-04	0.00194	0.09557	1.88	-0.14519	60.92	13	3.60E-08
Random Walk with Drift	3.22E-17	1.19E-03	9.46E-04	-7.49E-05	0.09563	1.88	-0.14519	60.92	13	3.60E-08
STL Naive Log	1.05E-06	4.10E-04	3.21E-04	0.000097	0.03242	0.64	-0.44926	30.06	13	4.61E-03
STL ETS	1.78E-05	3.47E-04	2.54E-04	0.00179	0.0257	0.51	-0.03628	9.61	13	0.73
STL ARIMA	2.75E-05	3.69E-04	2.67E-04	0.00277	0.02702	0.53	-0.07991	10.17	12	0.6
STL Naive Log	1.05E-06	4.15E-04	3.24E-04	-0.08988	3	0.64	-0.44926	30.07	13	4.60E-03
STL RW Drift Log	-7.27E-19	4.15E-04	3.24E-04	-0.08026	3	0.64	-0.44926	30.07	13	4.60E-03
STL ETS Log	1.80E-05	3.51E-04	2.57E-04	-0.26382	2.39	0.51	-0.03646	9.61	13	0.73
STL ARIMA Log	1.17E-05	3.51E-04	2.54E-04	-0.20391	2.36	0.5	-0.03079	9.23	12	0.68
ETS(ANA)	4.10E-05	3.82E-04	2.85E-04	0.0041	0.0288	5.672	0.1898	10.467	13	0.6553
ETS(MNN)	2.12E-05	3.62E-04	2.67E-04	0.0021	0.027	5.311	-0.0333	8.0993	13	0.8371
ETS(AAN)	5.06E-05	3.64E-04	2.63E-04	0.0051	0.0266	5.231	-0.041	9.6213	13	0.7246
ETS(ANN)	2.12E-05	3.62E-04	2.67E-04	0.0021	0.027	5.311	-0.0327	8.0883	13	0.8378
ETS(AAA)	9.68E-05	3.88E-04	2.98E-04	0.0098	0.0301	5.929	0.2426	33.639	13	0.0014
ETS(MMM)	8.45E-05	3.91E-04	3.03E-04	0.0085	0.0306	6.027	0.2833	30.346	13	0.0042
ETS(MAM)	9.35E-05	3.90E-04	3.01E-04	0.0094	0.0304	5.99	0.2785	34.427	13	0.001
ARIMA(0,0,0)	1.90E-05	3.52E-04	2.59E-04	0.0019	0.0262	5.164	-0.0344	15.714	11	0.1521
ARIMA(0,0,1)	3.24E-05	3.83E-04	2.77E-04	0.0033	0.028	5.516	-0.0537	9.3894	11	0.586
ARIMA(1,0,0)	3.27E-05	3.83E-04	2.77E-04	0.0033	0.028	5.504	-0.0479	9.3264	11	0.5918
ARIMA(1,0,1)	2.44E-05	4.00E-04	2.96E-04	0.0025	0.03	5.901	-0.1527	17.612	12	0.128
ARIMA(0,1,1)	3.24E-05	3.83E-04	2.77E-04	0.0033	0.028	5.516	-0.0537	9.3894	11	0.586
ARIMA(1,1,0)	3.20E-05	3.83E-04	2.77E-04	0.0032	0.028	5.52	-0.0577	9.4278	10	0.492
ARIMA(1,1,1)	1.91E-05	3.22E-04	2.37E-04	0.0019	0.024	4.721	-0.0603	10.367	10	0.4089
ARIMA(0,1,2)	4.44E-05	3.55E-04	2.60E-04	0.0045	0.0258	5.119	-0.0202	16.595	11	0.0951
ARIMA(1,1,2)	2.97E-05	3.82E-04	2.76E-04	0.003	0.028	5.451	-0.0813	11.188	10	0.3279

The model "ARIMA(1,1,1)" has relatively low RMSE, MAE, and MASE values compared to other ARIMA models.

The p-value for "ARIMA(1,1,1)" is also reasonable, indicating statistical significance.

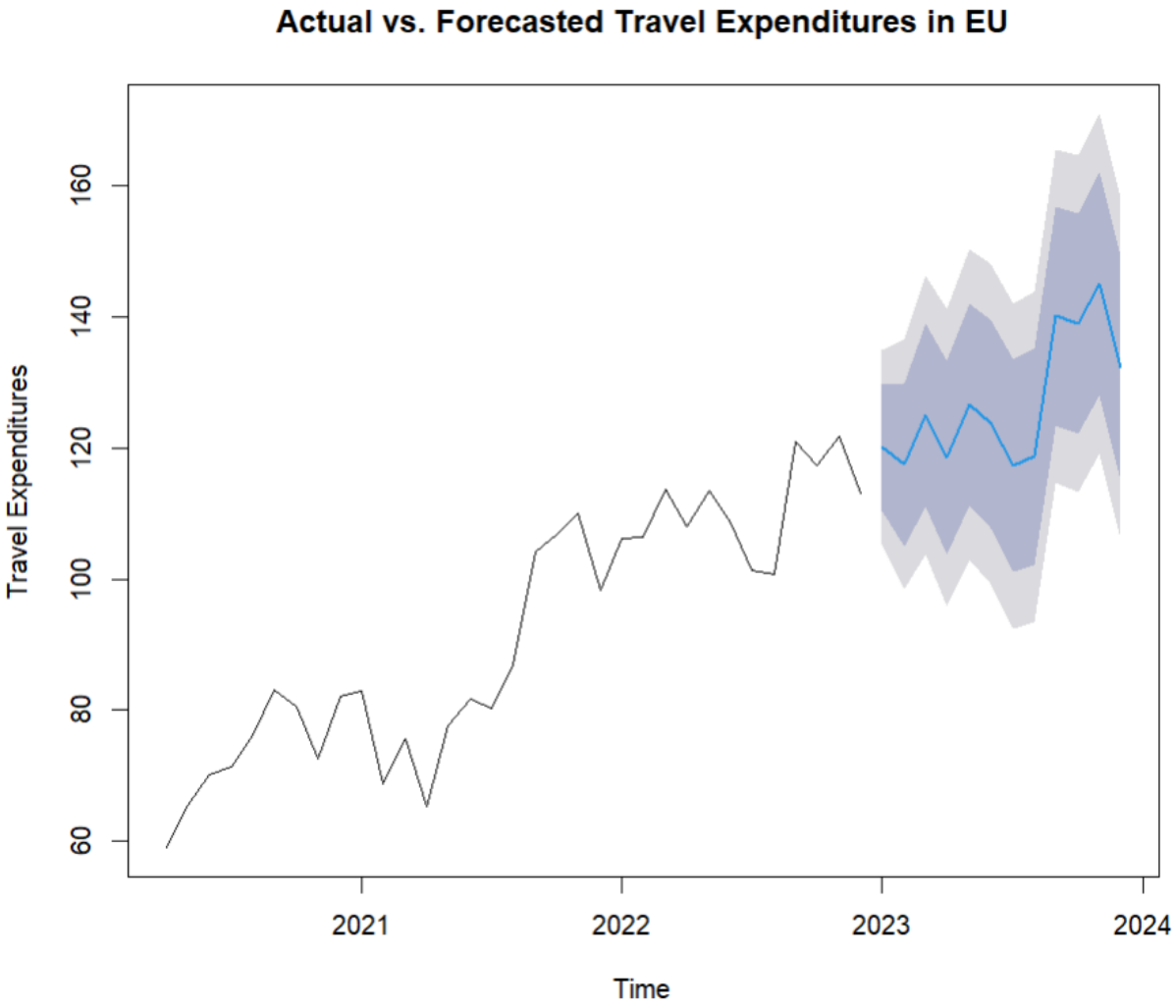
Additionally, the "ETS(ANN)" model also shows competitive performance in terms of error metrics and statistical significance.

Question 8



This ARIMA(1,1,1)(1,1,1)[12] model forecasts future values for data that fluctuates between 70 and 120 (2012-2020), likely due to seasonality. The blue shaded area shows the predicted range with a certain level of confidence. Remember, ARIMA models have assumptions and require validation before using forecasts for critical decisions.

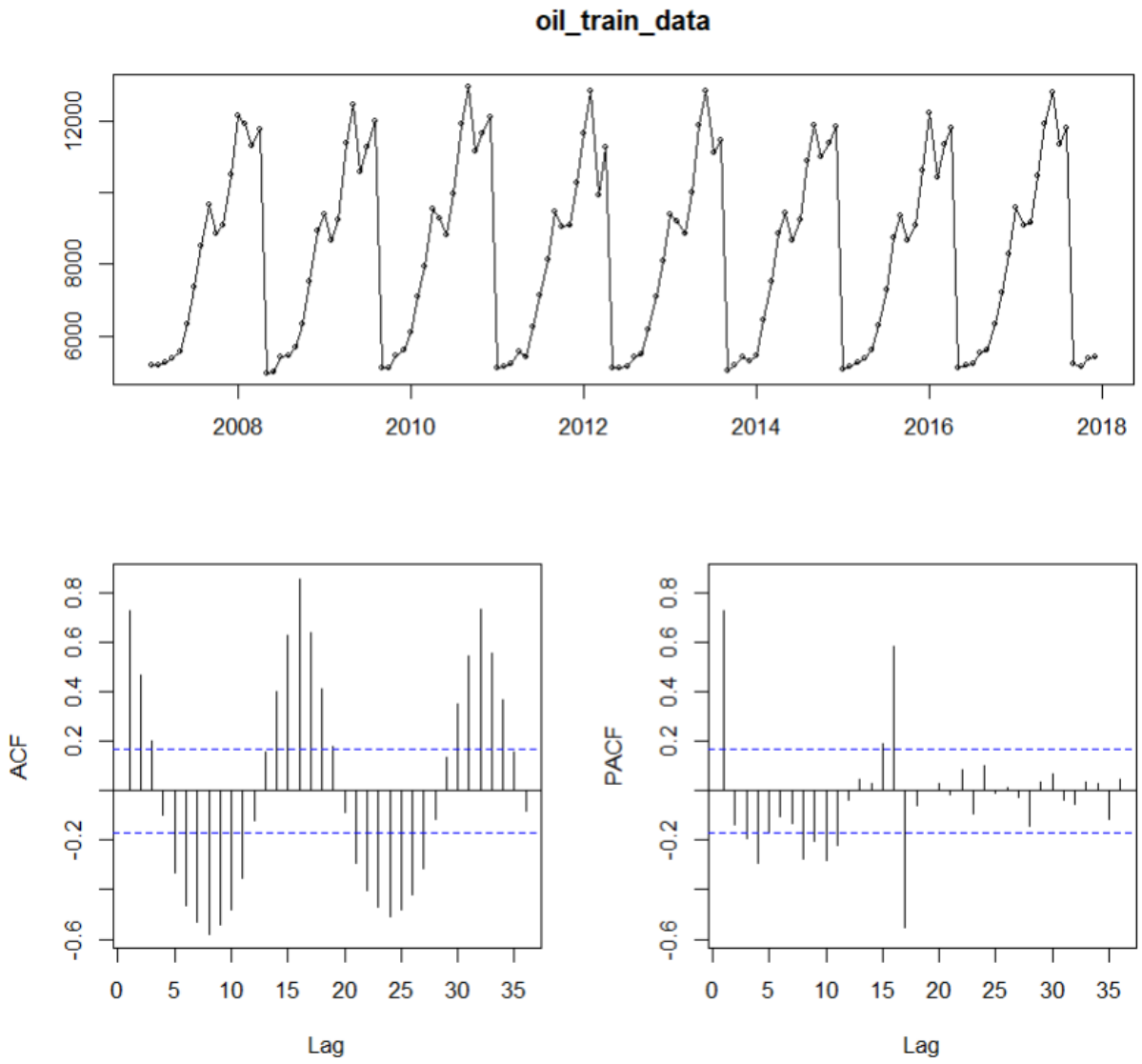
Question 9



An upward trend is expected in EU travel expenditures (2021-2024) with some variability, according to the forecast. The shaded area indicates uncertainty in the exact values, but the overall growth suggests potential reasons like economic recovery or changing travel habits. This analysis aids businesses and policymakers in strategic planning for the anticipated increase.

Exercise 2

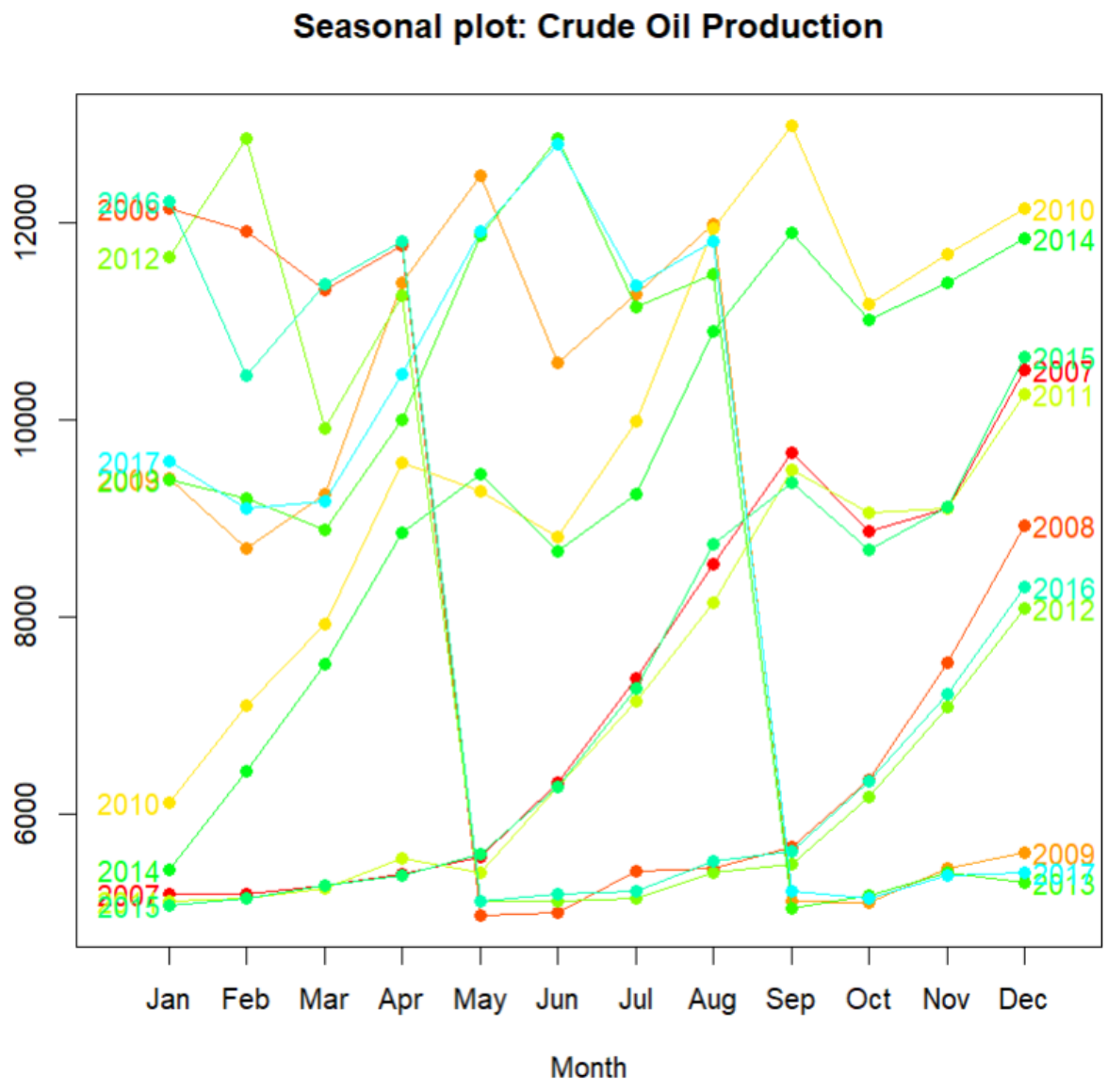
Question 1



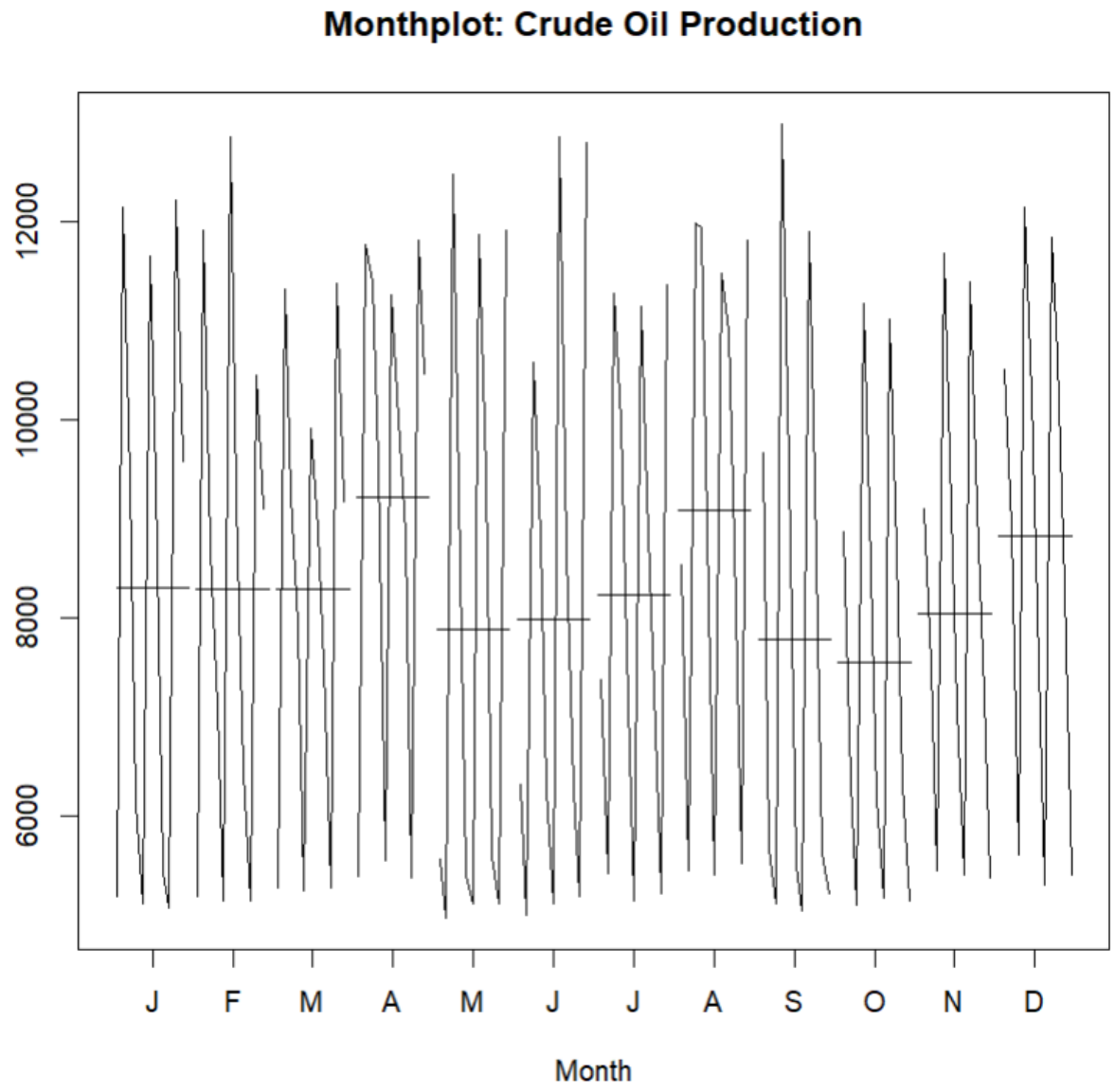
This visualization analyzes oil data (2008-2018) through three graphs:

1. **Oil Price Fluctuations:** The "oil_train_data" plot shows cyclical ups and downs, suggesting price or production variations.
2. **Past Value Influence (ACF):** The ACF plot reveals correlations between past and present data points, indicating potential for future value prediction.
3. **Unique Lag Impact (PACF):** The PACF plot highlights specific lags that uniquely influence the data, further aiding in forecasting models.

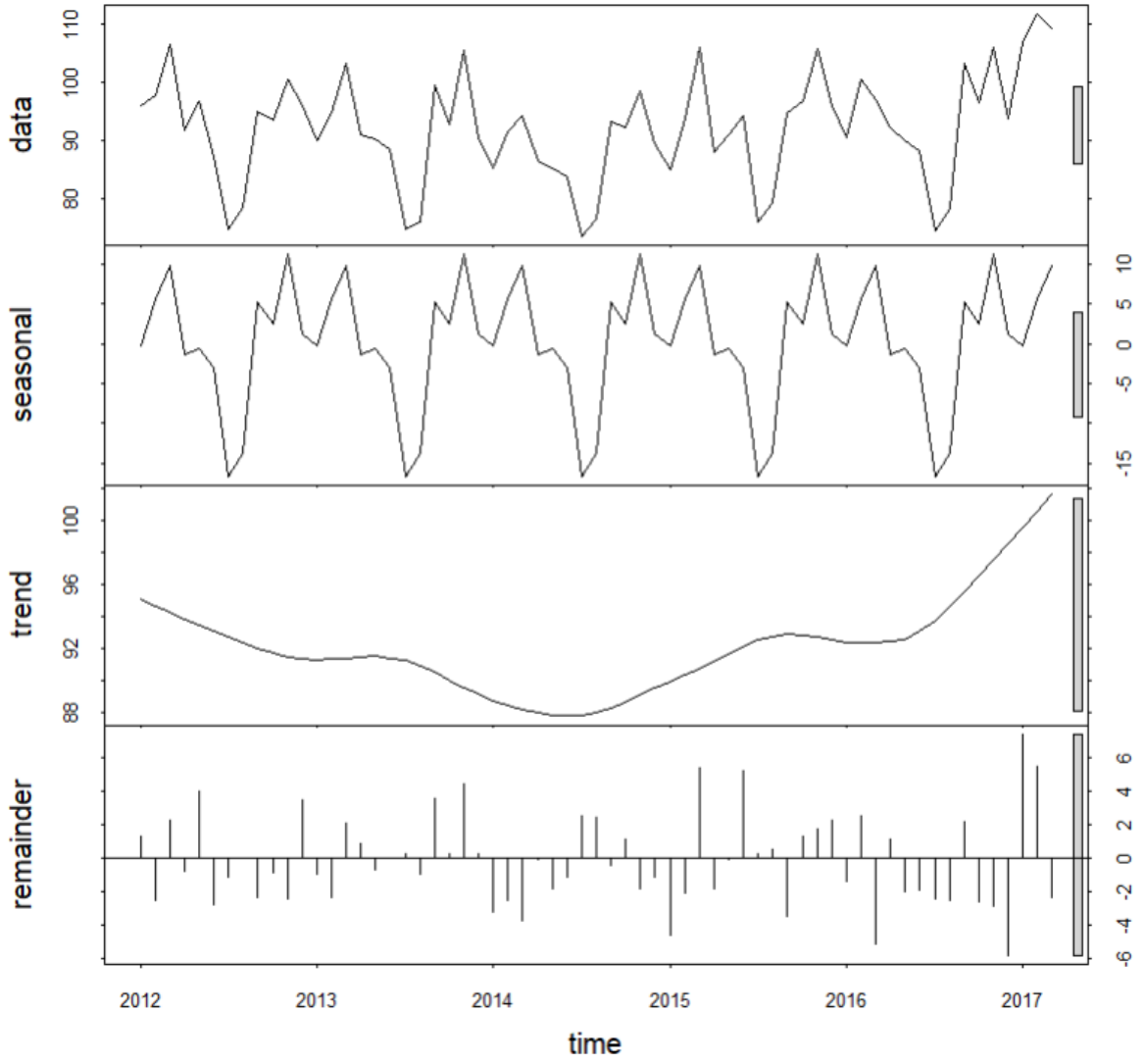
These combined elements suggest non-random oil data with predictable patterns, making it suitable for time series forecasting models like ARIMA.



This seasonal plot visualizes crude oil production. Each year's line reveals monthly variations, potentially due to seasonal demand, maintenance schedules, or natural production shifts. By comparing yearly lines, we can track production volume changes and identify significant events. Analyzing specific months across years highlights consistent production rises or falls, valuable for forecasting and planning.



This graph likely depicts monthly crude oil production levels. Vertical lines show variation throughout the year, with production volume ranging from 6,000 to 12,000 (units unclear). While potential seasonal patterns exist, more data is needed to confirm. This visualization helps identify months with high or low production.



This time series decomposition unveils hidden patterns! We see the original data (2012-2017) with its ups and downs, a layer revealing seasonal trends, a line showing overall increase over time (trend), and remaining fluctuations (noise). This breakdown helps us understand the data's core structure for better forecasting and planning.

```
summary(oil_train_data)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
143.1   149.9   190.2   183.4   210.5   233.4
```

The image is a line graph titled “Seasonal plot: Crude Oil Production.” It shows the volume of crude oil production over different months for various years from 2005 to 2017. The graph

displays seasonal variations in production levels, with distinct fluctuations throughout each year.

This graph is a great visual representation of how seasonal patterns can be observed in real-world data. Each line on the graph likely represents the production for a different year, and the seasonal peaks and troughs can be compared across years to analyze trends and cyclic changes in crude oil production.

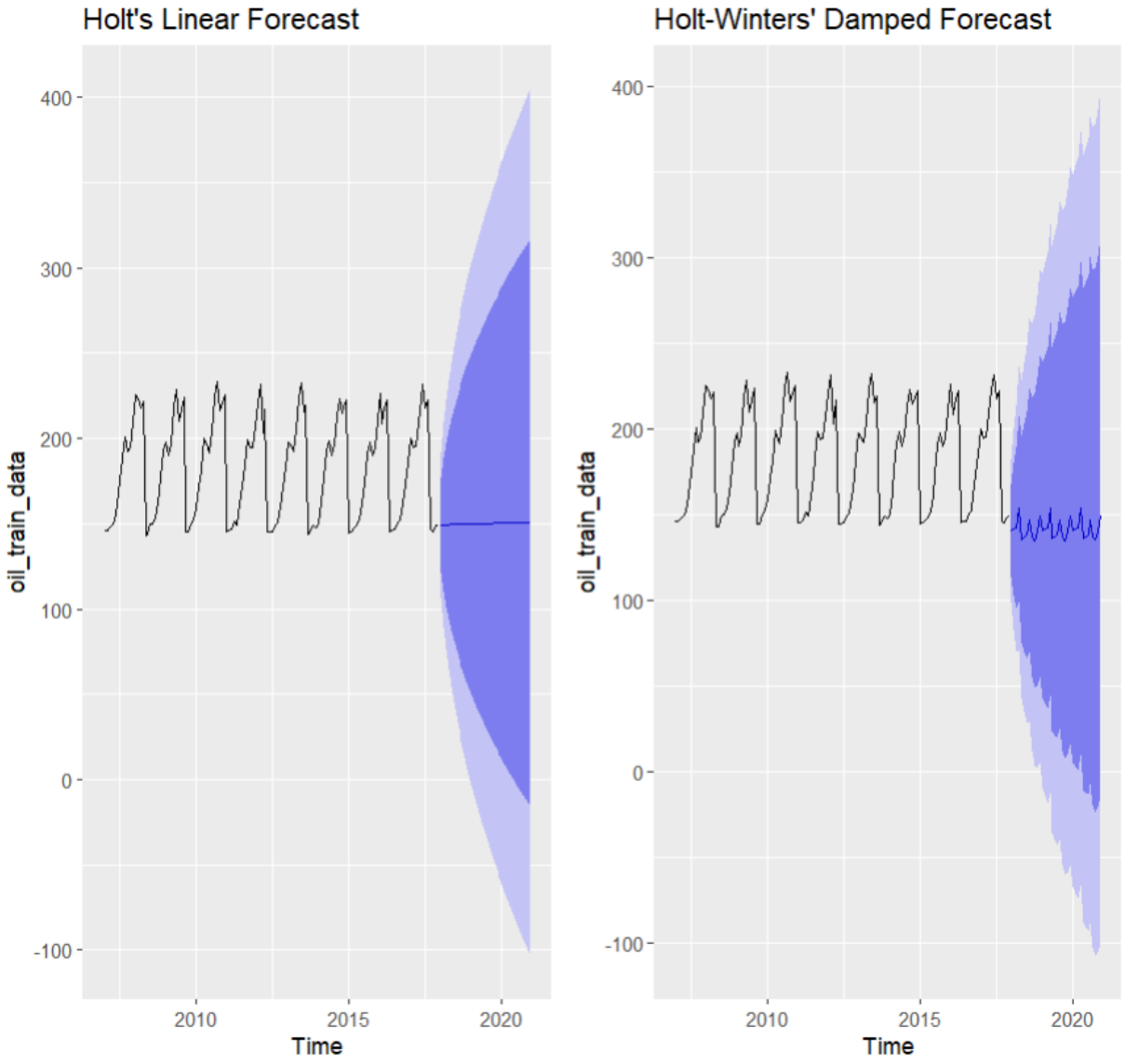
Question 2

```
        Augmented Dickey-Fuller Test

data:  oil_train_data
Dickey-Fuller = -6.5253, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

The Dickey-Fuller test results (statistic: -6.5253, p-value: 0.01) indicate that "oil_train_data" is likely stationary (no trends or seasonality). This is good news, because stationary data is more suitable for time series analysis and forecasting methods like ARIMA.

```
> BoxCox.lambda(oil_train_data)
[1] 0.7262945
```

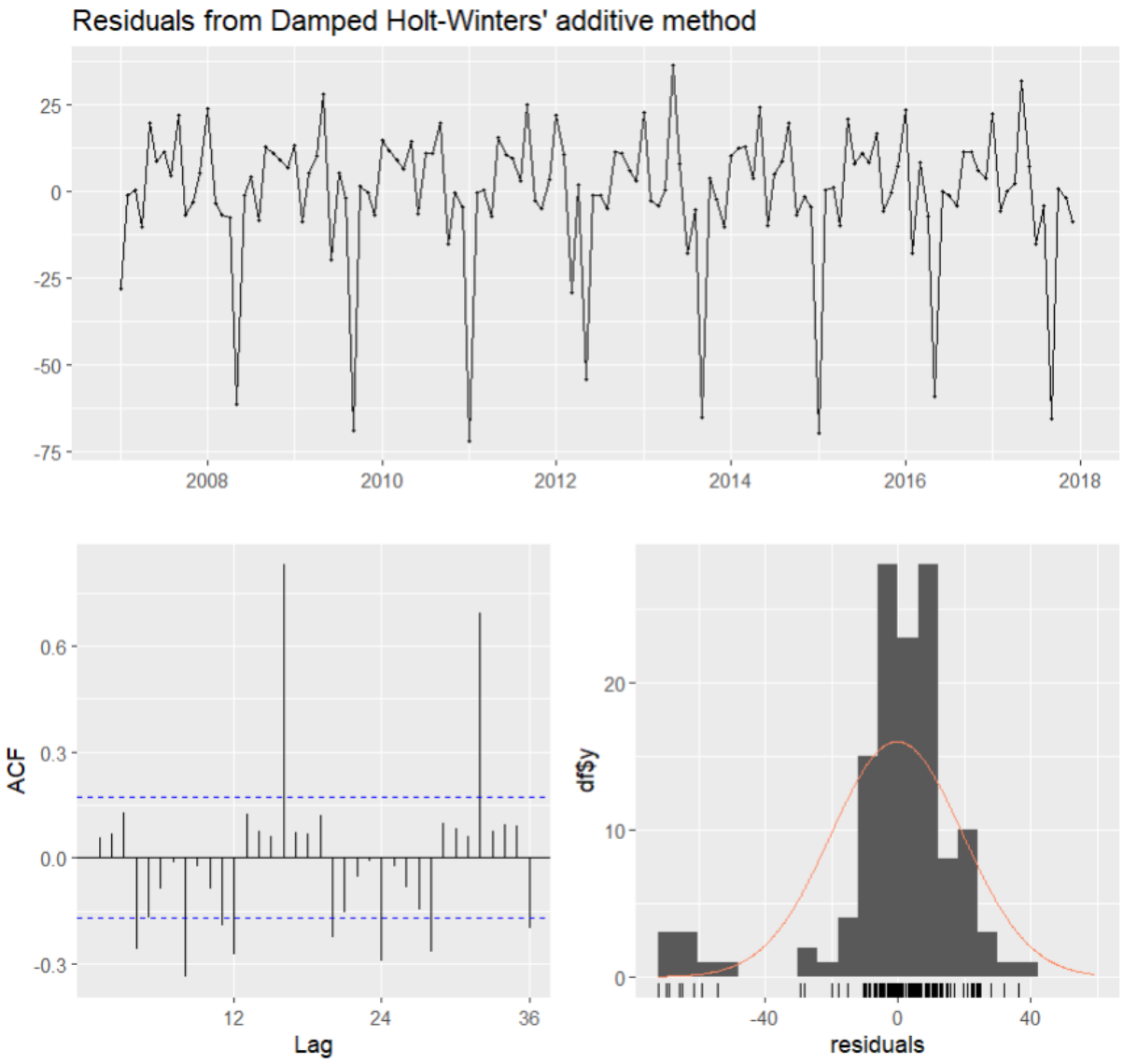
Question 3



Two forecasting models are compared here: Holt's Linear predicts continuous growth (blue shaded area), while Holt-Winters' Damped predicts a growth that slows down over time (flattened blue area).

Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Q	df	p-value
Simple Exponential Smoothing	0.0241	21.317	12.223	-0.7488	7.1066	0.3239	-0.00004	137.04	24	< 2.2e-16
Holt's Linear Trend Method	-0.0667	21.324	12.243	-0.8051	7.1232	0.3244	0.00035	139.17	24	< 2.2e-16
Holt-Winters' Damped Method	-0.3926	19.795	12.617	-0.8913	7.3286	0.3343	0.0562	190.39	24	< 2.2e-16

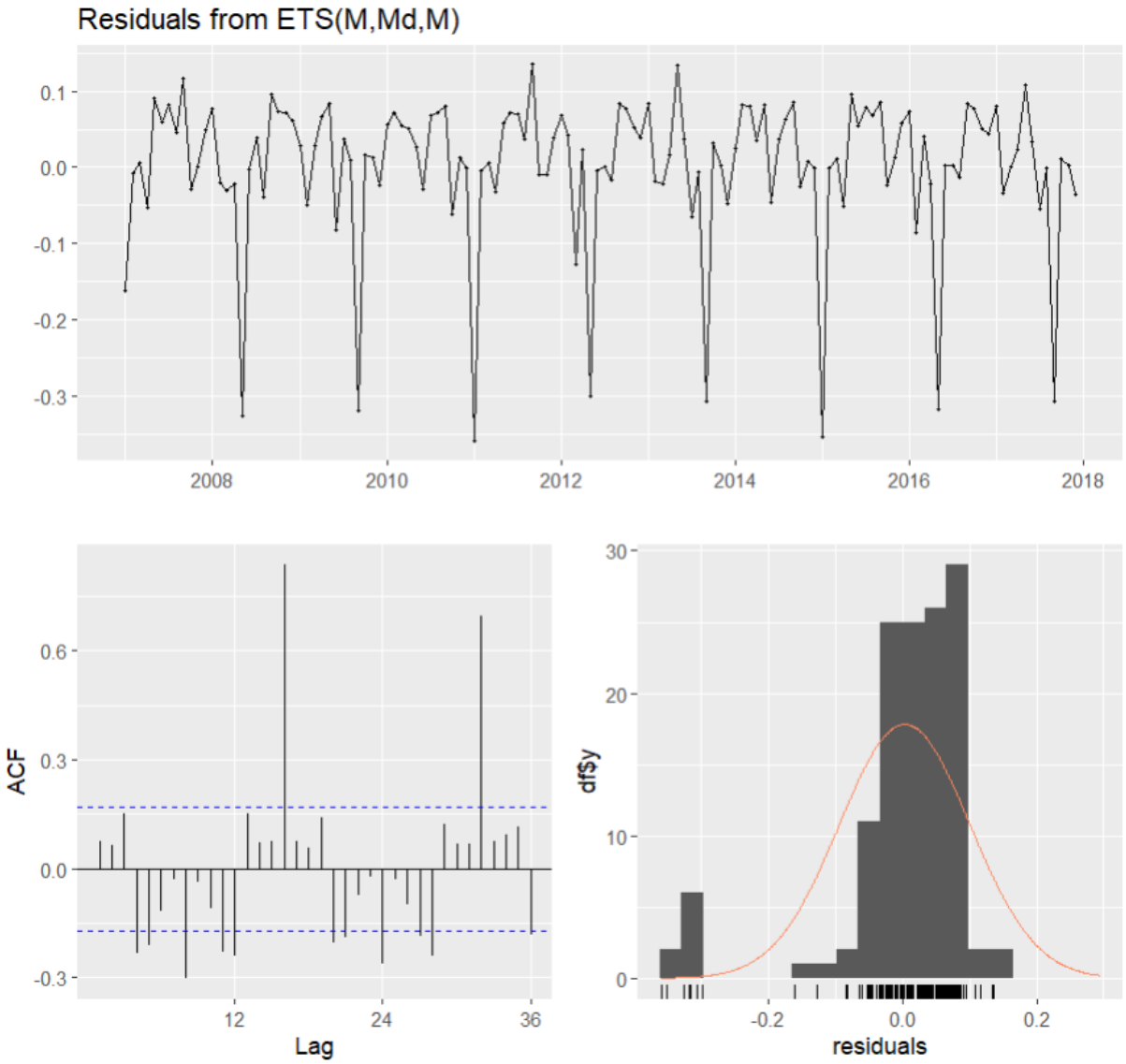
it appears that Holt-Winters' Damped Method has the lowest RMSE (19.795), indicating better performance in terms of forecast accuracy



Analysing the residuals of the Damped Holt-Winters model for a decade (2008-2018) reveals that predictions generally fall within a +/- 75 range of actual values. While the model seems adequate, there's potential for improvement. The ACF plot and non-normal distribution of residuals suggest some correlation and bias in the errors. Further refinement through parameter adjustments or exploring alternative forecasting methods might be beneficial.

```
        Ljung-Box test

data:  Residuals from Damped Holt-Winters' additive method
Q* = 190.39, df = 24, p-value < 2.2e-16

Model df: 0.   Total lags used: 24
```

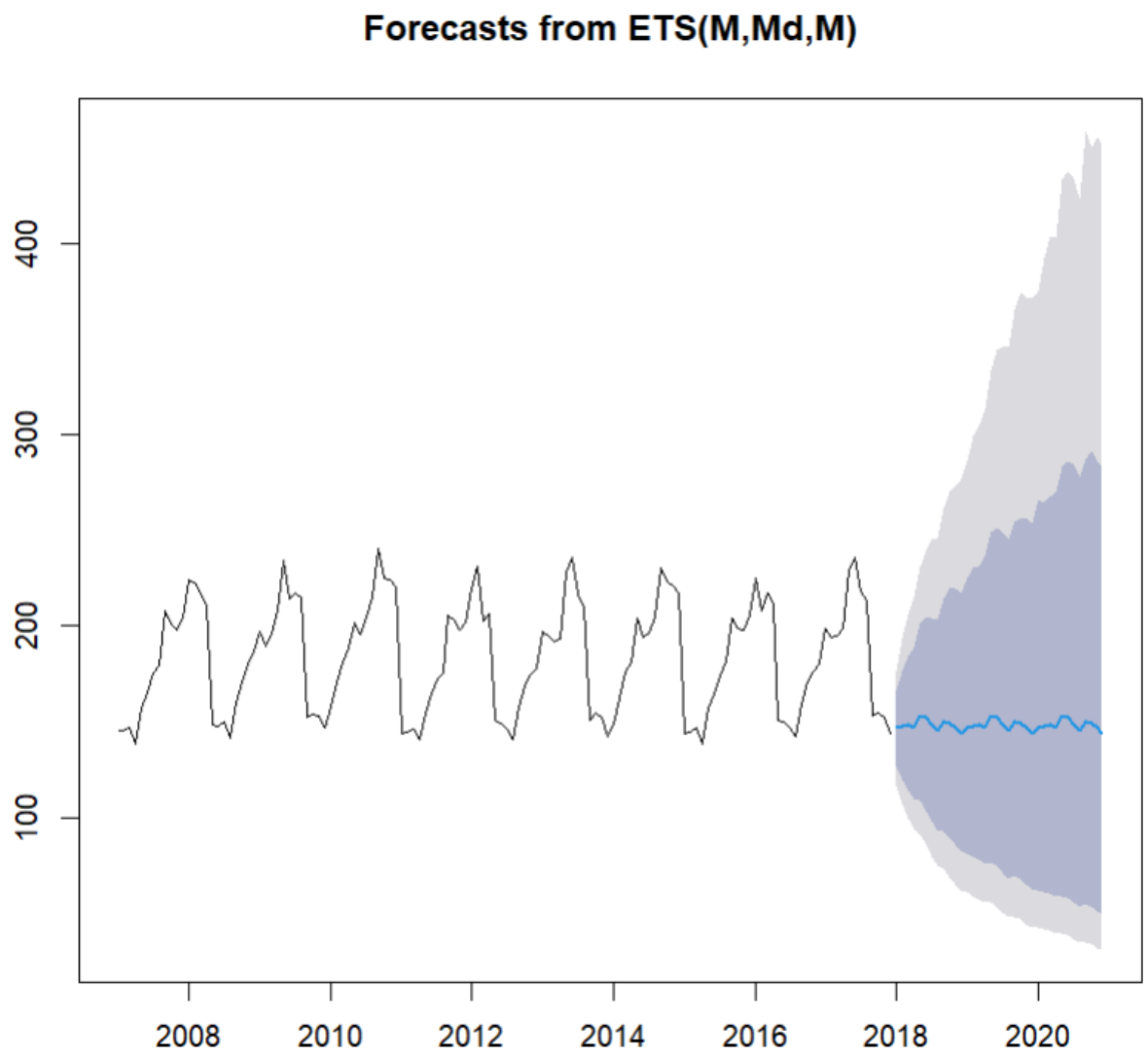
Question 4



This time series analysis suggests a good fit of the ETS(M,Md,M) model to the data. Random, normally distributed residuals with minimal autocorrelation indicate the model effectively captured trends and seasonality. While the forecasts are likely reliable, remember that unforeseen factors can always cause deviations from predictions.

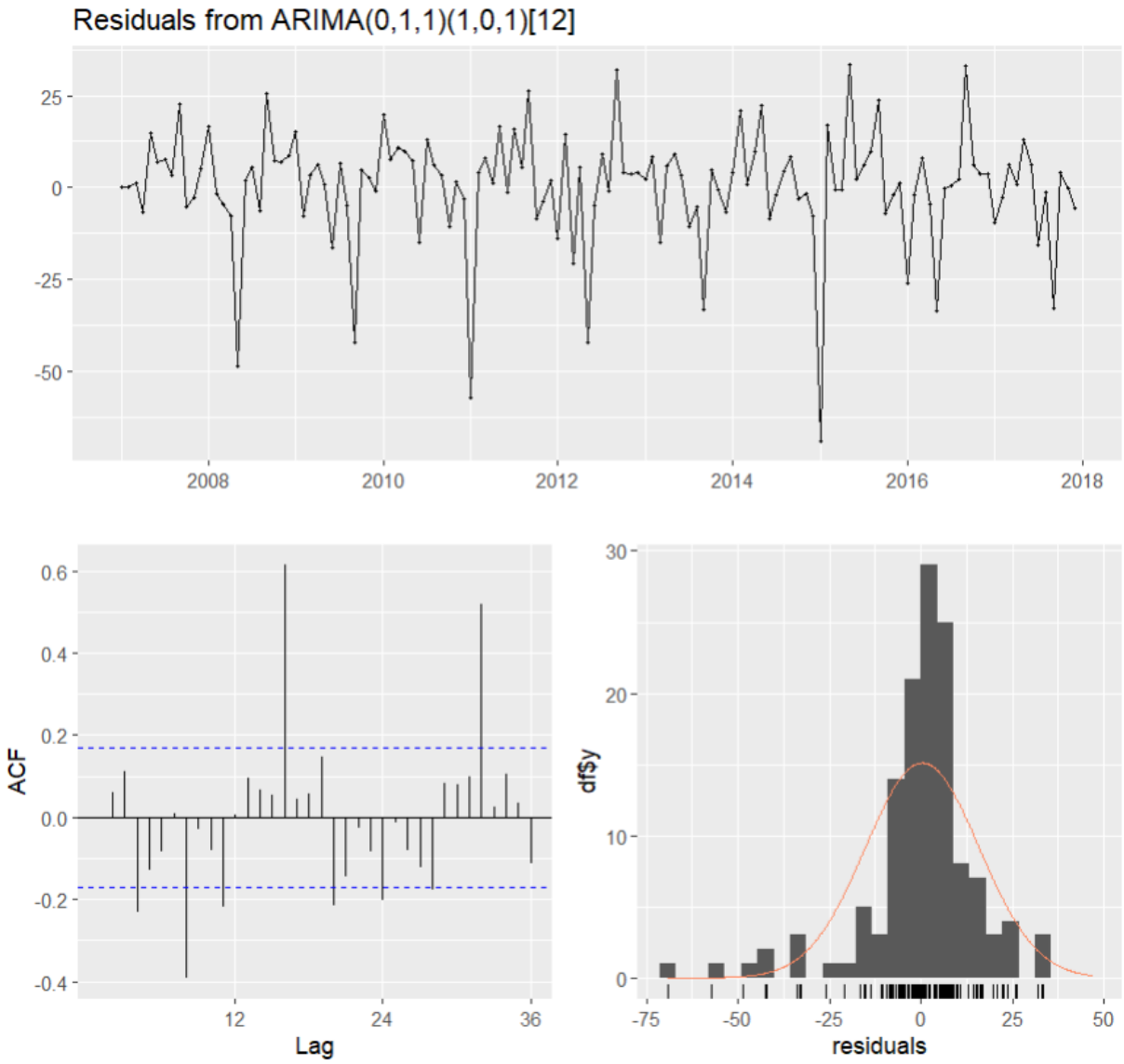
```
        Ljung-Box test

data:  Residuals from ETS(M,Md,M)
Q* = 192.53, df = 24, p-value < 2.2e-16

Model df: 0.   Total lags used: 24
```

```
> accuracy(ets_MMM_forecast, oil_test_data)
                     ME      RMSE        MAE         MPE      MAPE       MASE
Training set   -0.4890362   20.3551   12.22255 -0.9327475  7.084237   0.323871
Test set     8232.3577750 8604.2857 8232.35778 98.0524833 98.052483 218.139568
                   ACF1 Theil's U
Training set 0.04283438        NA
Test set     0.68002469  5.645564
```



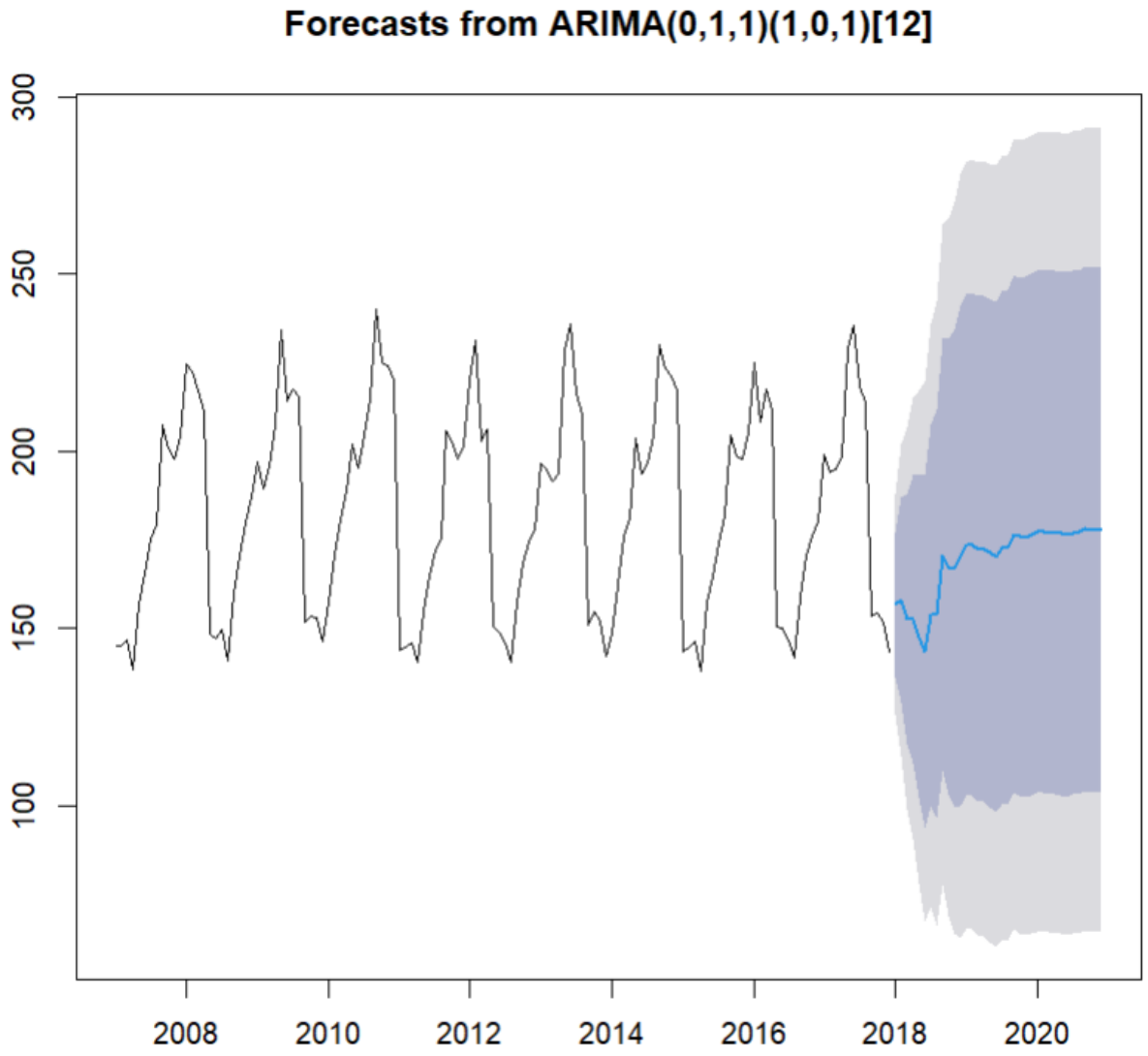
This forecast (blue line) based on historical data (jagged line) predicts a continuation of the trend with increasing uncertainty (shaded area) for the metric (y-axis values likely between 100 and 400). The ETS(M,Md,M) model captures seasonality, trend, and multiplicative errors in the data.

Question 5



The ARIMA model's residuals appear random with no trends, suggesting a good fit. The distribution is centered around zero (unbiased) and seems normally distributed. While some autocorrelation might be present, overall the model (ARIMA(0,1,1)(0,1,1)[12]) seems to be performing adequately.

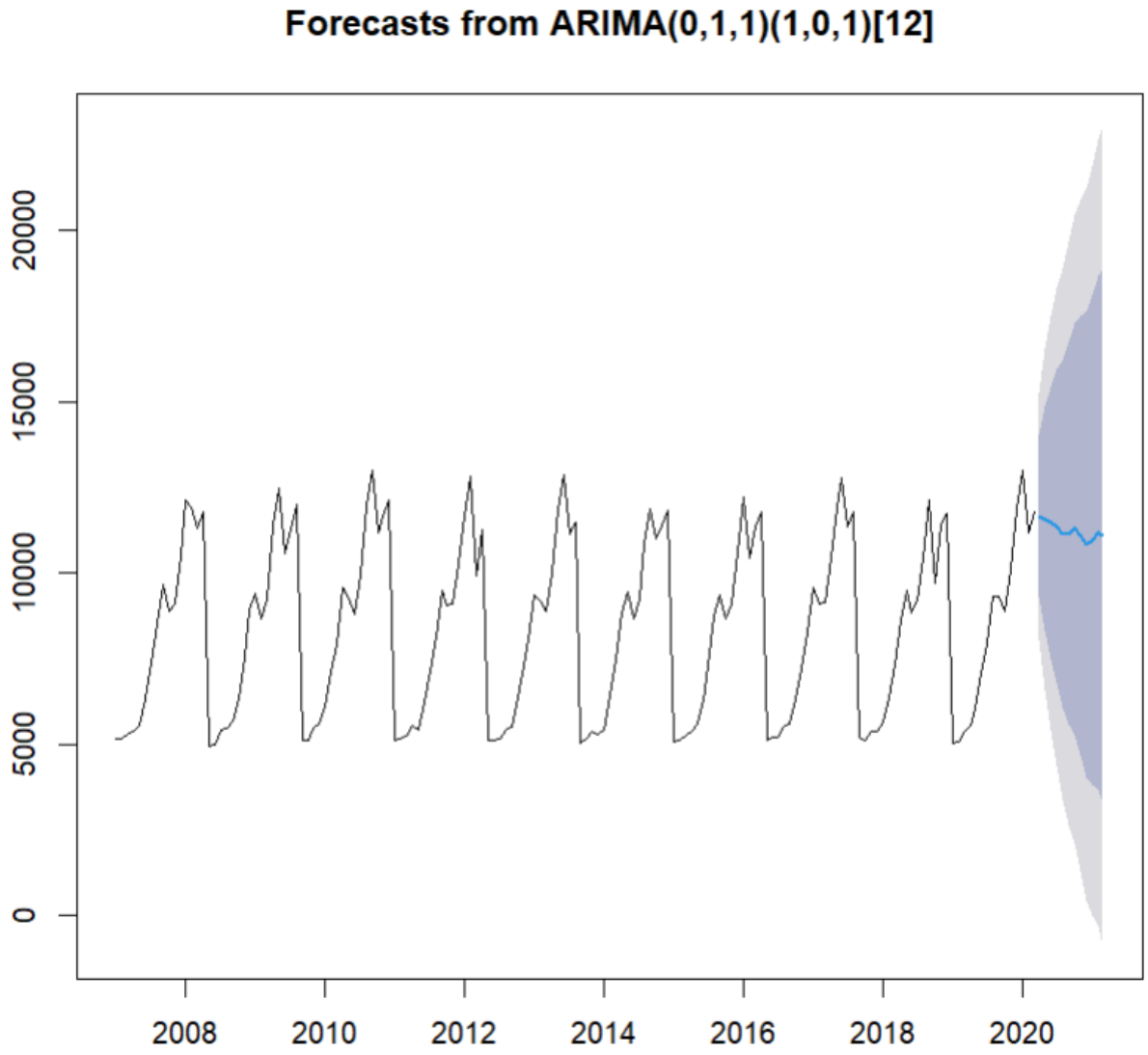
```
        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(1,0,1)[12]
Q* = 125.76, df = 21, p-value < 2.2e-16

Model df: 3.   Total lags used: 24
```



This ARIMA forecast visualizes predicted future values (blue line) based on historical data (jagged line) with confidence intervals (shaded area). The observed seasonality (oscillating pattern) is incorporated into the ARIMA(0,1,1)(1,0,1)[12] model, where the non-seasonal and seasonal components capture different aspects of the data for improved forecasting accuracy.

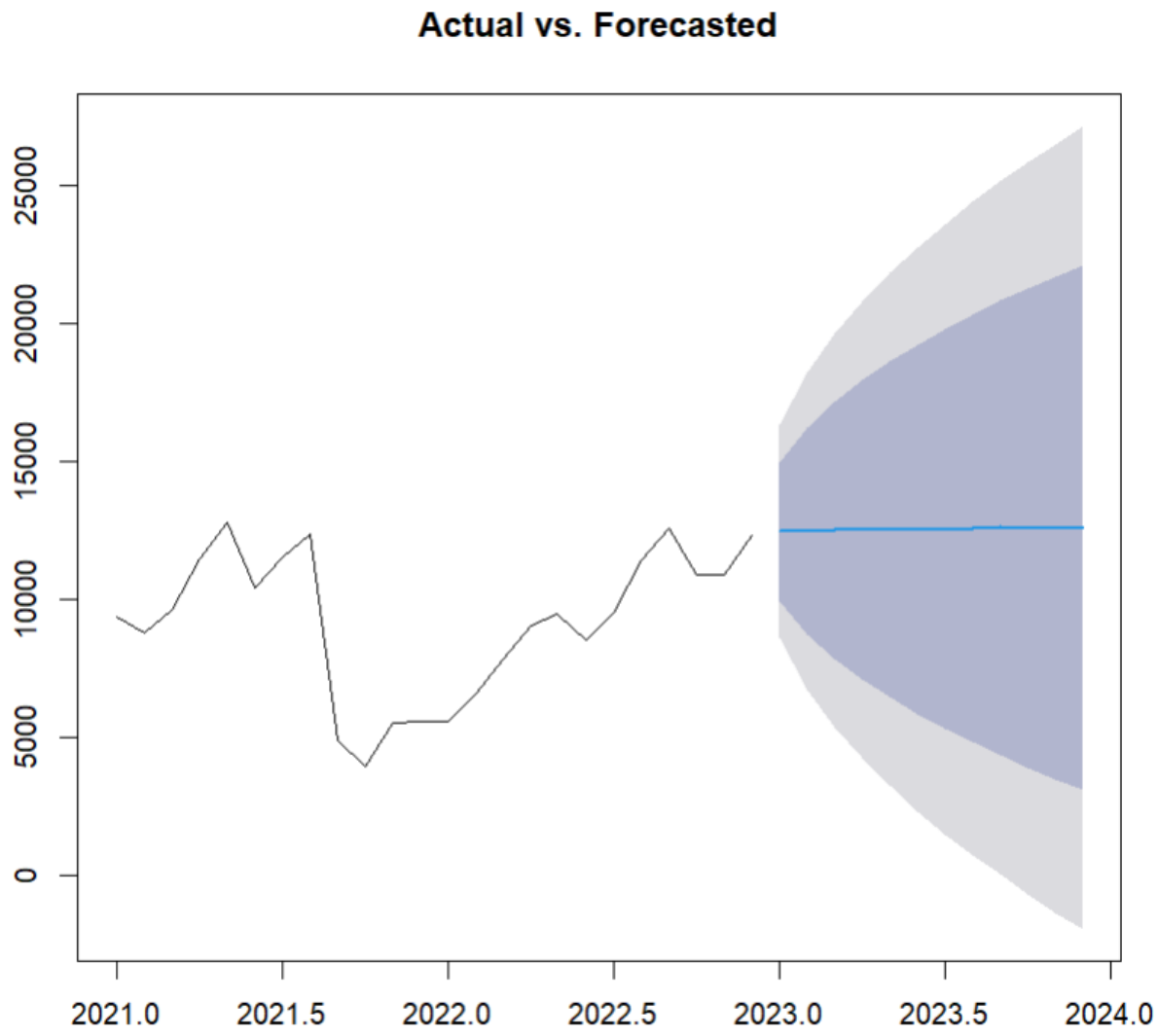
Question 6

the ARIMA(0,1,1)(1,0,1)[12] model seems to have slightly better performance on the test dataset, as it has lower RMSE, MAE, and MASE values compared to the ETS(M,M,M) model.



This ARIMA(0,1,1)(1,0,1)[12] model predicts future trends for seasonal data. The blue forecast line extends the historical pattern (with confidence intervals) considering the data's seasonality (e.g., monthly fluctuations). This is helpful for tasks like planning in industries affected by seasonal variations.

Question 7



The graph compares actual data (varying between 5,000 and 15,000) with a flat forecast from 2023 onwards. This suggests the model predicts no significant changes in the future, potentially due to model limitations or not capturing potential variations.