ARIMA

Data 624
March 5, 2024

Our team

Moiya Josephs
Mohamed Hassan
Nick Climaco
Umer Farooq
Mohammed Rahman

Table of contents

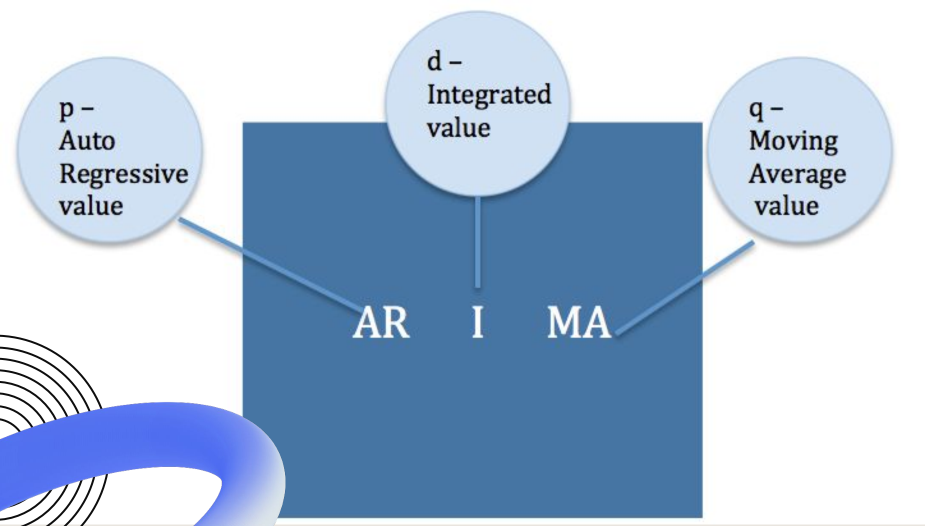
Table of contents

Table of contents

01

Introduction

Introduction



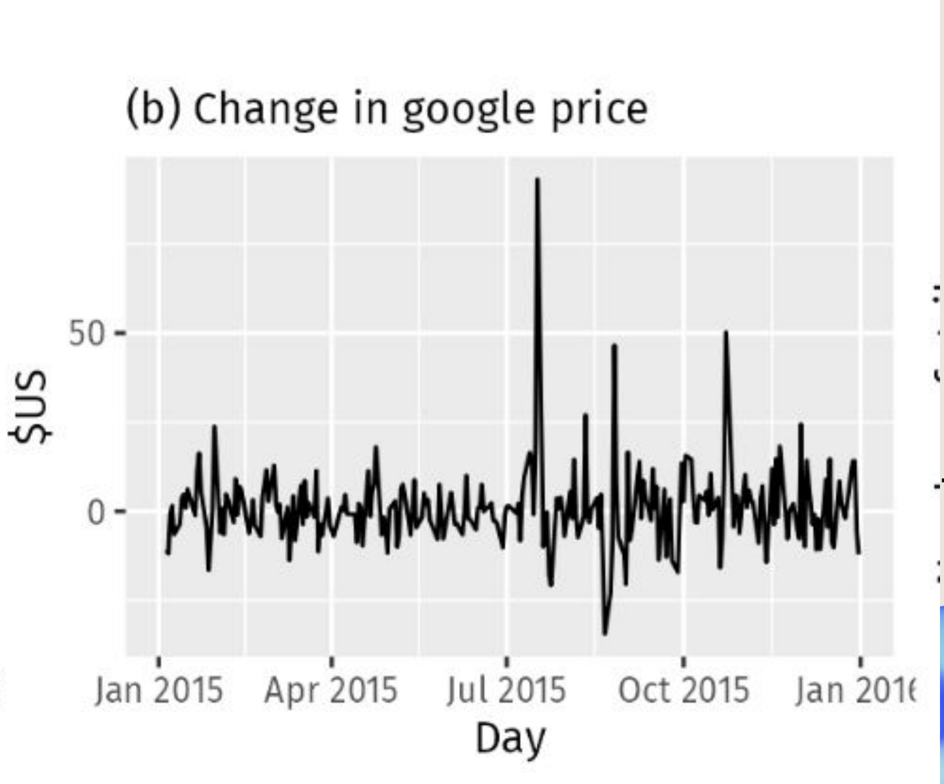
- ARIMA is an acronym for AutoRegressive Integrated Moving Average
- ARIMA models are another way to forecast a time series. It is one of the most widely used approaches.
- ARIMA models aim to describe the autocorrelations in the data.
- There are two main concepts to discuss before we discuss Arima further:
    - Stationary time series
    - Differencing time series

02

Components of ARIMA Models

Stationary

- Time Series whose statistical properties do not depend on the time at which the series was observed
- Does not have seasonality
- For example, a white noise time series is stationary.

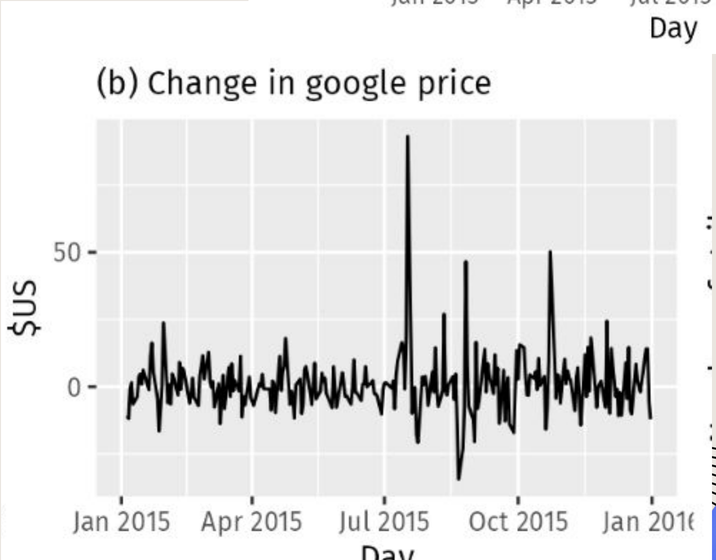


.....

Differencing

- Differencing allows us to take non stationary time series to stationary.
- Transformations such as logarithms can help to stabilise the variance of a time series.
- Differencing can help stabilise the mean of a time series and therefore eliminating (or reducing) trend and seasonality.

.....



ARIMA
Autoregressive

- Autoregressive model is regression of a time series against the lagged values of the series.
- The last p observations are used as predictors in the regression equation.
- Autoregressive models are remarkably flexible at handling a wide range of different time series patterns.
- We refer to this as an AR(p) model, an autoregressive model of order p.
- Changing the phi of the formula results in different patterns and the variance in the error terms changes the scale.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

.....

ARIMA
Moving Average

- Moving Average can also be considered a regression but it regresses against lagged errors.
- The last q errors are used as predictors.
- We refer to this as an MA(q) model, a moving average model of order q.
- Changing the theta changes the time series pattern.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

.....

ARMA

- Putting these together you get ARMA (autoregressive moving average model)
- Multiple regression with lagged observations and errors as predictors.
- This can only work with stationary data, so we have to difference the data.

ARIMA
Integrated

- The I in Arima stands for integrated - it combines ARMA model with differencing
- If the time series needs to be differenced a number of times to make it stationary it results in an **ARIMA(p,d,q)** model.

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

03

Parameter Selection

Parameter Selection – Nonseasonal

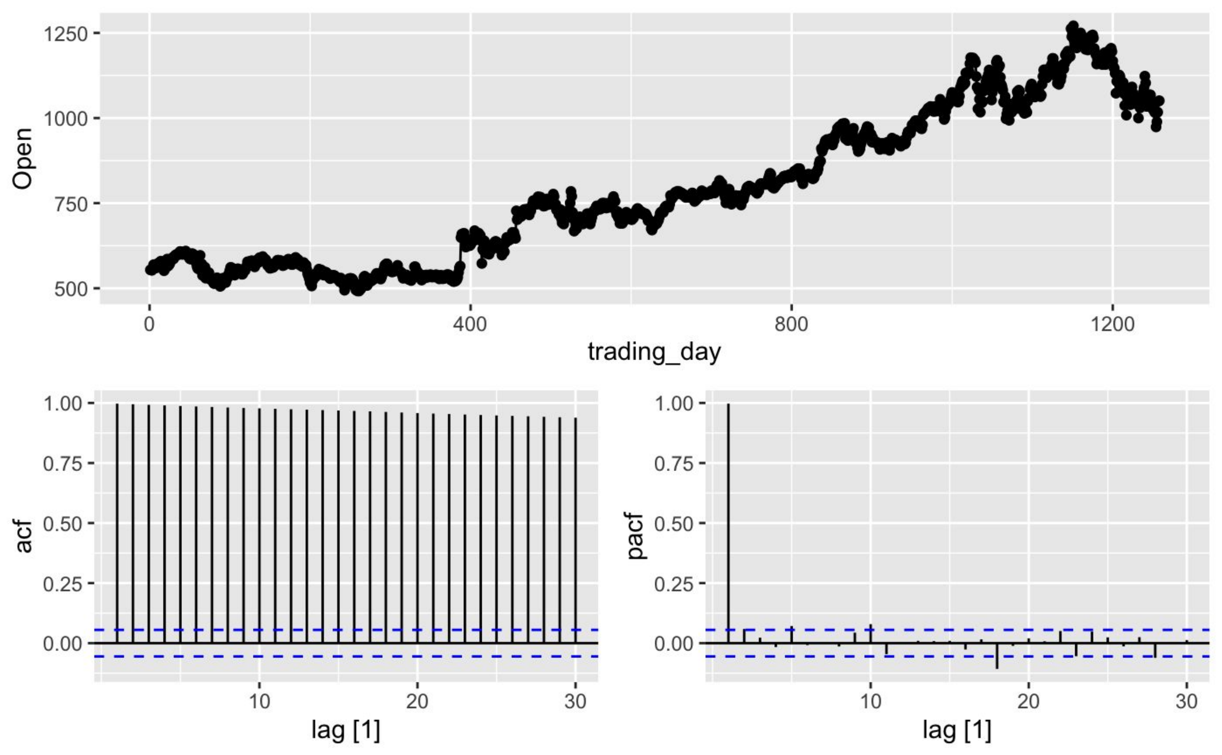
- Three parameters: *p, d, q*
- AR - *p,* I - *d,* MA - *q*
- *p:* Number of lag observations in the model; order of the autoregressive part; measures how much the current values depends on the previous values.
- *d*: Degrees of differencing; value effects prediction interval - higher the value, the faster it increases in size.
- *q*: Order of Moving Average part; number of lag errors in the model; measures how much current value depends on past errors.
- If time series is non-stationary, use differencing *d*.
- *p* can be determined by the PACF plot - last significant spike.
- *q* can be determined by the ACF plot - last significant spike.

Google Stock Price (Daily Open)

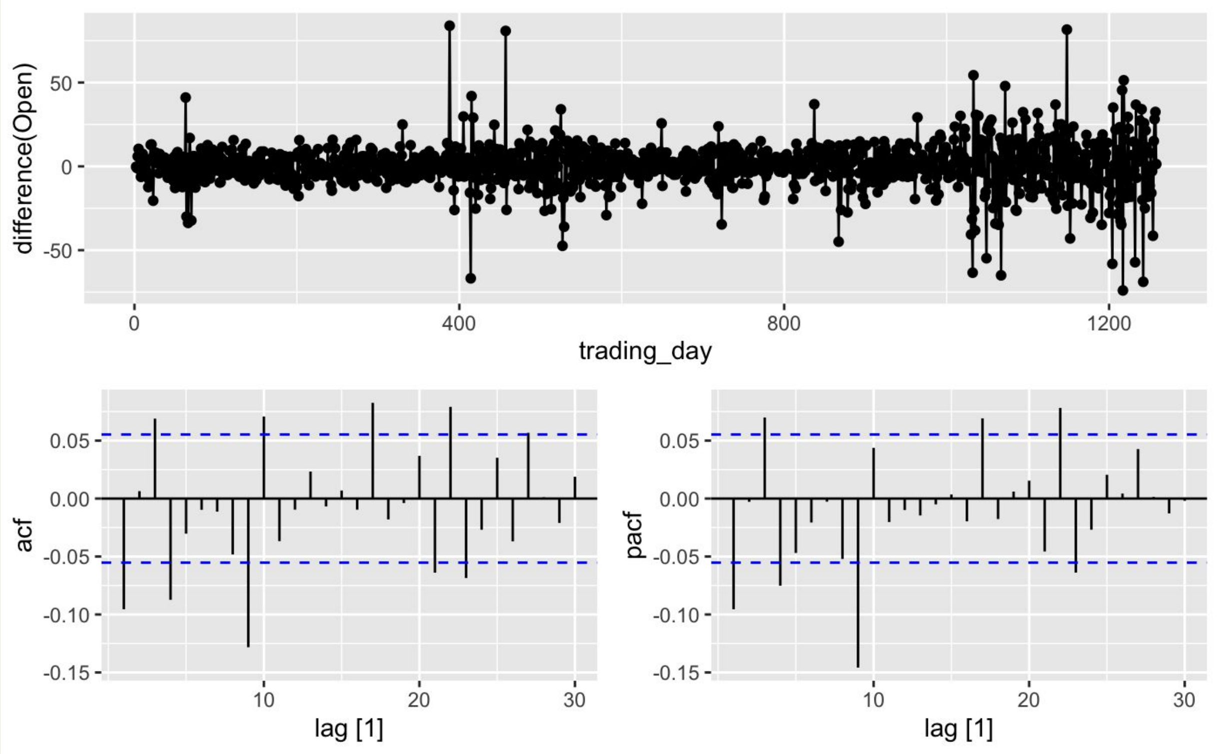
```
fit <- google_stock |>
  filter(Symbol == "GOOG") |>
  model(ARIMA(Open))
report(fit)
```

```
Series: Open
Model: ARIMA(2,1,2)

Coefficients:
          ar1      ar2     ma1     ma2
      -0.2537  -0.8339  0.170  0.8555
s.e.   0.0549   0.0581  0.049  0.0578

sigma^2 estimated as 163.3:  log likelihood=-4984.33
AIC=9978.67   AICc=9978.71   BIC=10004.35
```

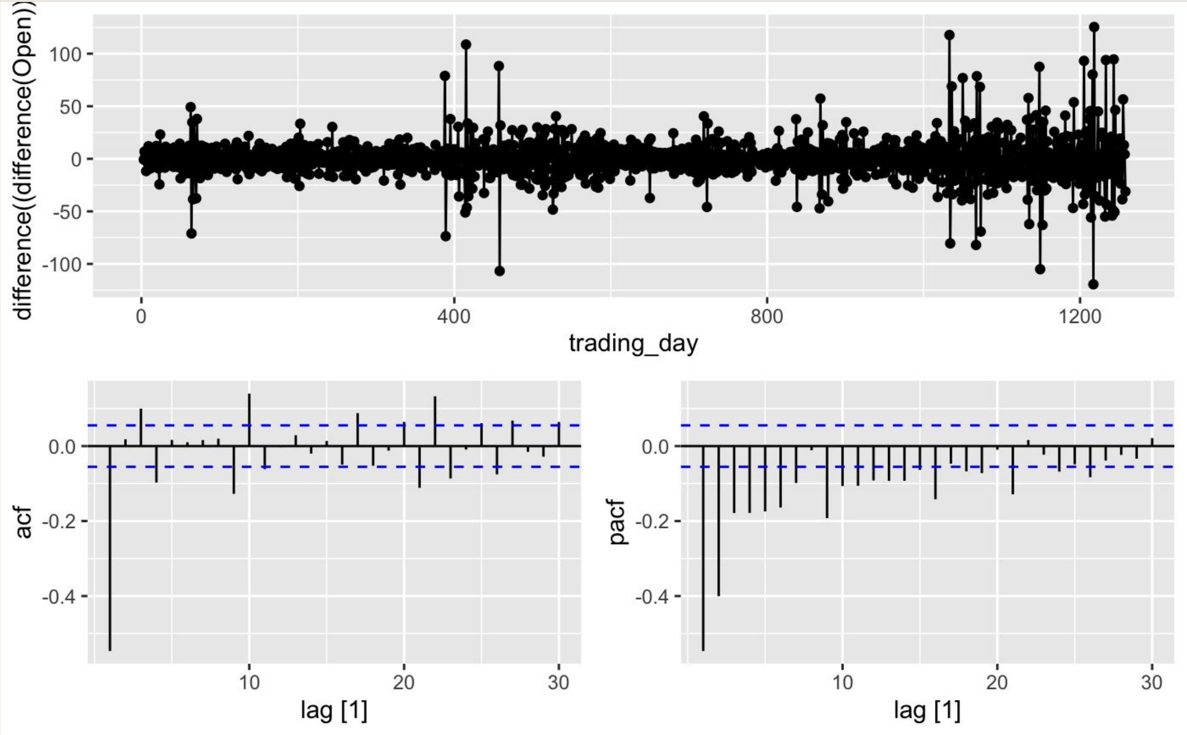
Google Stock Price (Daily Open)



Differenced



Differenced Twice



Parameter Selection – Seasonal

- Incorporates both seasonal and nonseasonal factors in a multiplicative model
- ARIMA(p,d,q) x (P,D,Q)S
- p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern.

** Source: https://online.stat.psu.edu/stat510/lesson/4/4.1*

04

Model Estimation

Maximum Likelihood Estimation (MLE)

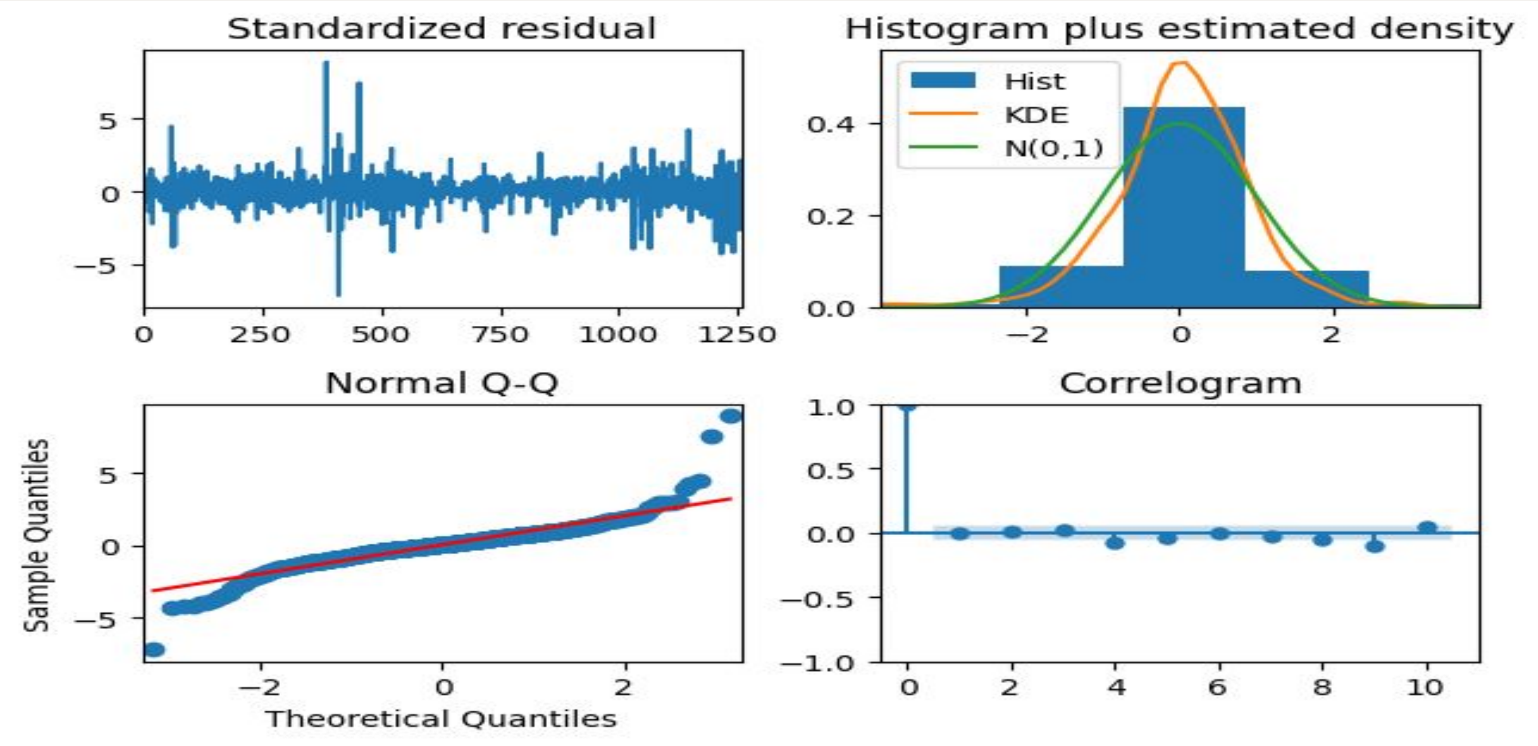
- Maximum likelihood estimation (MLE) is a commonly used method of estimating the parameters of a statistical model given a set of observations.
- It is based on the premise that the best choice of the parameter values should maximize the likelihood of making the observations given these parameters.
- Estimates coefficients

** Source: https://mfe.baruch.cuny.edu/wp-content/uploads/2014/12/TS_Lecture1_2019.pdf*

05

Model Diagnostic and Evaluation

Google Stock Price – Model Diagnostic and Evaluation



Model Diagnostic and Evaluation

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.2798	0.065	-4.289	0.000	-0.408	-0.152
ar.L2	-0.8129	0.062	-13.121	0.000	-0.934	-0.691
ma.L1	0.2091	0.064	3.273	0.001	0.084	0.334
ma.L2	0.8357	0.058	14.377	0.000	0.722	0.950
sigma2	0.0002	3.86e-06	62.616	0.000	0.000	0.000

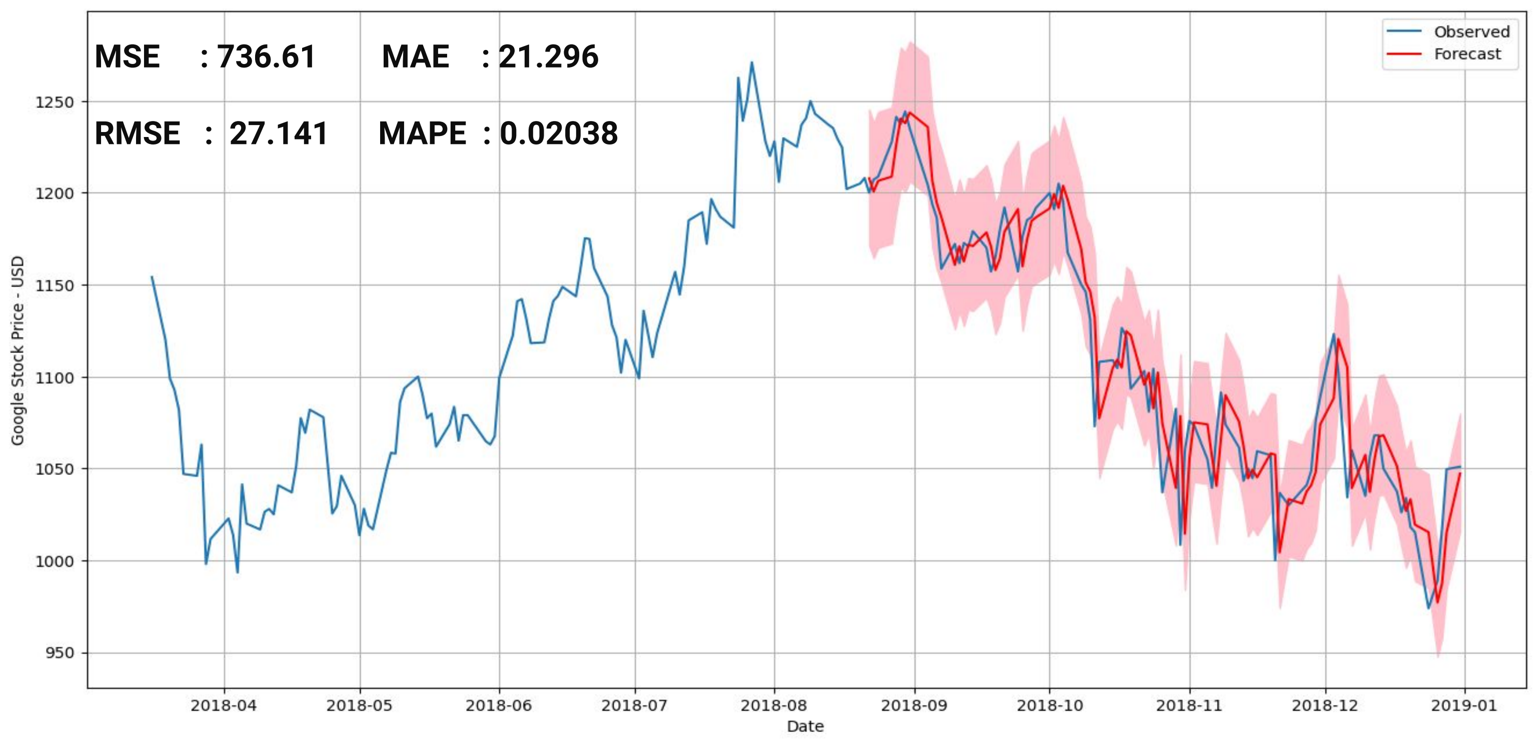
Ljung-Box (L1) (Q):	0.06	Jarque-Bera (JB):	6463.73
Prob(Q):	0.81	Prob(JB):	0.00
Heteroskedasticity (H):	0.94	Skew:	0.27
Prob(H) (two-sided):	0.55	Kurtosis:	14.10

06

Forecasting with ARIMA models

Forecasting with ARIMA Models

Forecasting with ARIMA Models

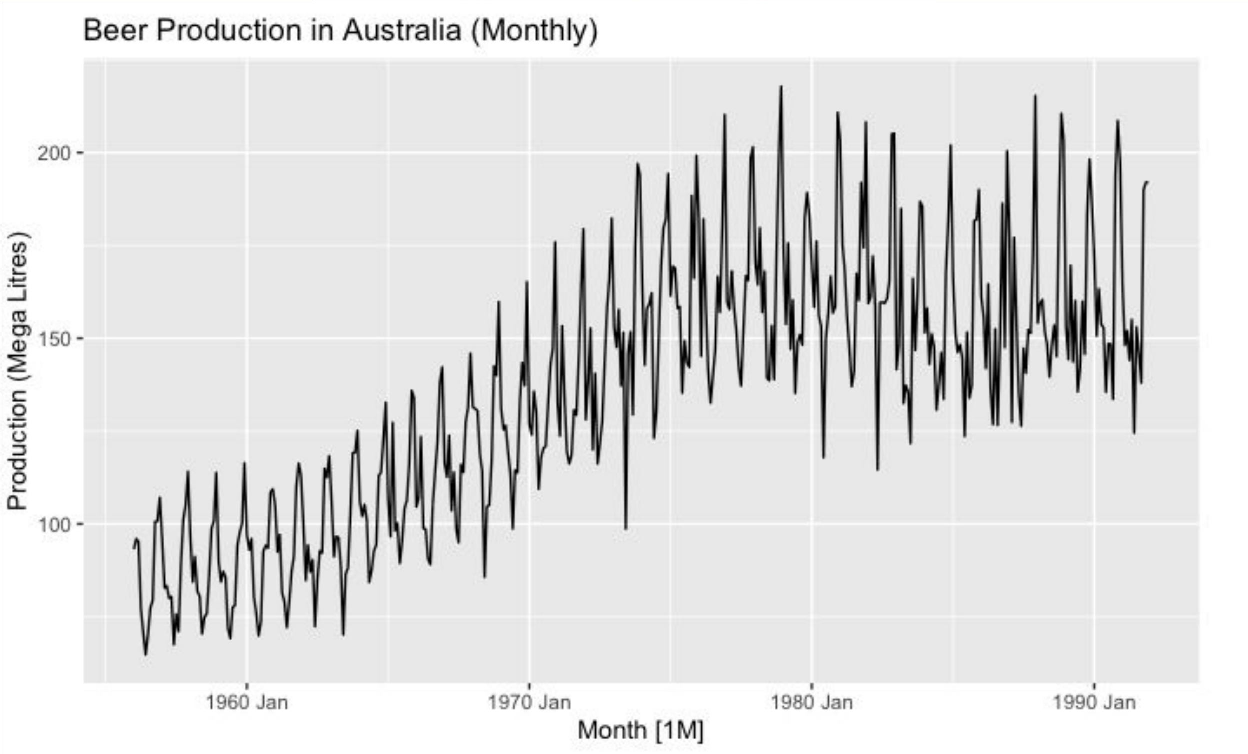


07

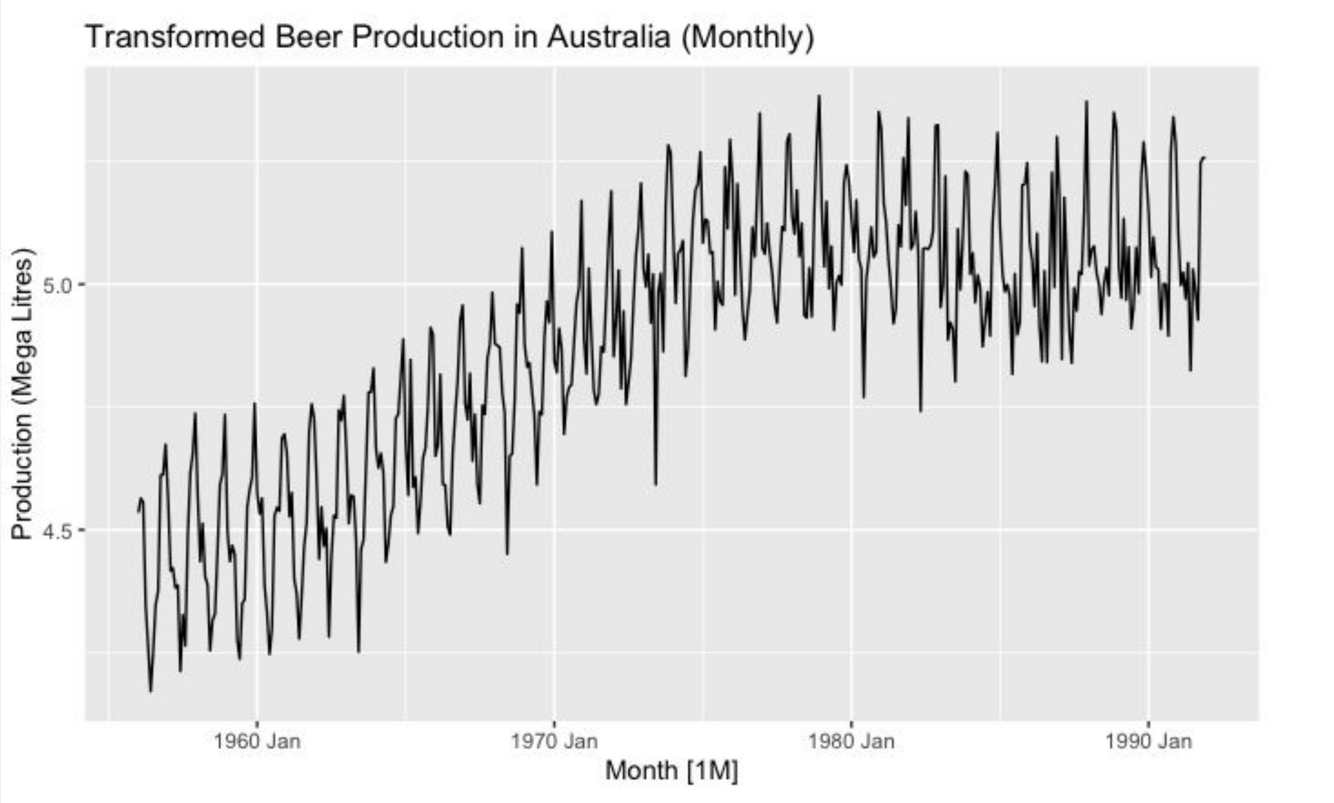
Seasonal ARIMA

Australian Beer Production

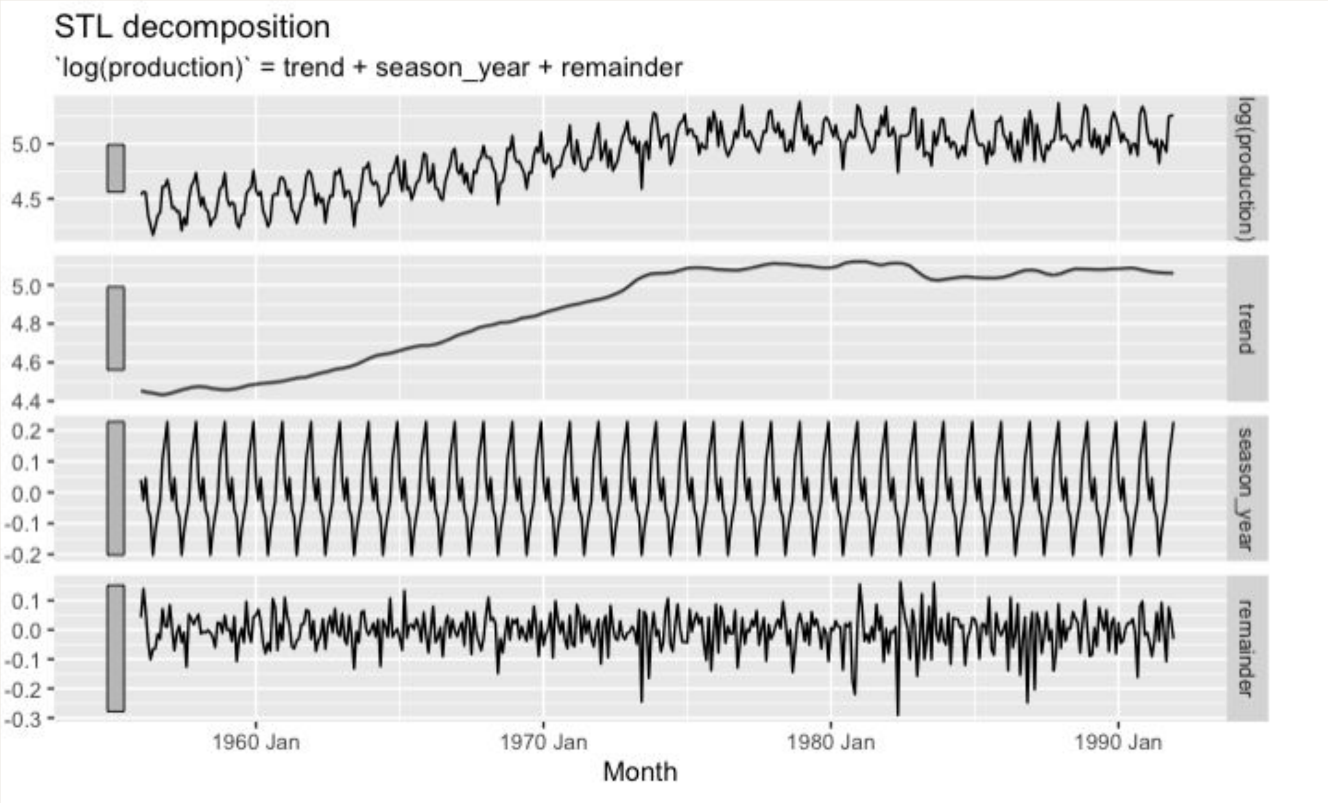
$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P, D, Q)_m}_{\substack{\uparrow \\ \text{Seasonal part} \\ \text{of the model}}}$$



After Log Transformation



STL Decomposition



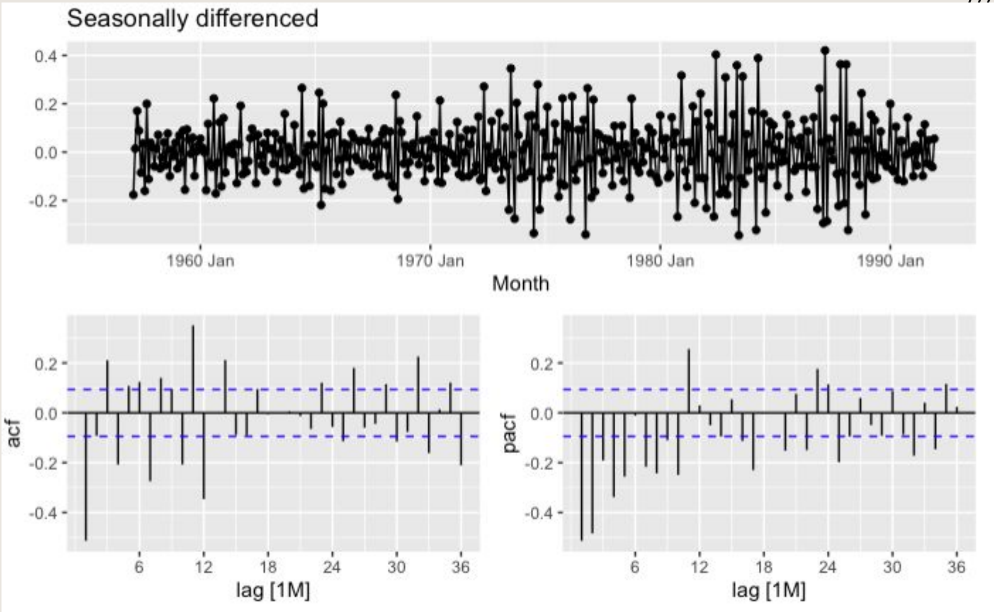
```{r warning=FALSE}
beer_ts |>
  gg_tsdisplay(difference(log(production), 12)|>difference(),
               plot_type='partial', lag=36) +
  labs(title="Seasonally differenced", y="")
```

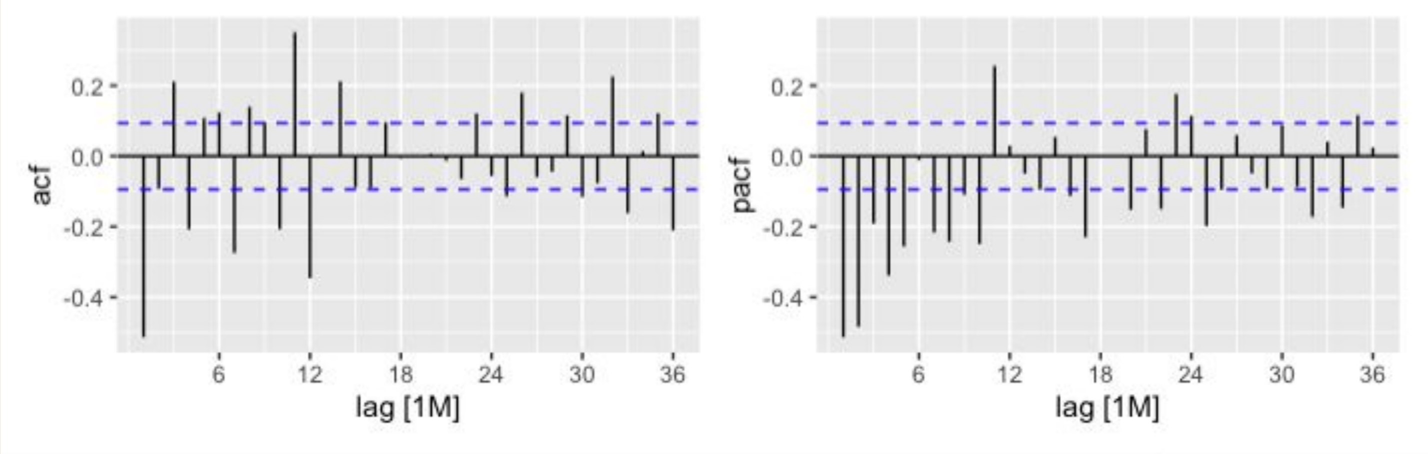
```{r}
beer_ts |>
  mutate(log_prod = difference(log(production), 12)) |>
  features(log_prod, unitroot_ndiffs)
```

A tibble: 1 × 1

ndiffs <int>
1

1 row






```{r}
fit <- beer_ts |>
  model(arima0111013 = ARIMA(log(production) ~ 0 + pdq(0,1,11)+PDQ(0,1,3)),
        arima1110210 = ARIMA(log(production) ~ 0 + pdq(11,1,0)+PDQ(2,1,0)),
        arima314111 = ARIMA(log(production) ~ 0 + pdq(3,1,4)+PDQ(1,1,1)),
        arima310110 = ARIMA(log(production) ~ 0 + pdq(3,1,0)+PDQ(1,1,0)),
        arima313112 = ARIMA(log(production) ~ 0 + pdq(3,1,3)+PDQ(1,1,2)),
        auto = ARIMA(log(production), stepwise = FALSE, approximation = FALSE))
```

```{r}
glance(fit) |> arrange(AICc) |> select(.model:BIC)
```

A tibble: 6 x 6

.model <chr>	sigma2 <dbl>	log_lik <dbl>	AIC <dbl>	AICc <dbl>	BIC <dbl>
arima314111	0.003823228	564.2942	-1108.588	-1108.0492	-1068.2097
arima0111013	0.004370602	542.1744	-1054.349	-1053.1578	-993.7808
auto	0.004494746	532.8018	-1051.604	-1051.3310	-1023.3384
arima313112	0.004549141	531.4799	-1042.960	-1042.4205	-1002.5810
arima1110210	0.004844871	524.4654	-1020.931	-1019.8912	-964.4006
arima310110	0.007452891	432.2530	-854.506	-854.3607	-834.3166

6 rows

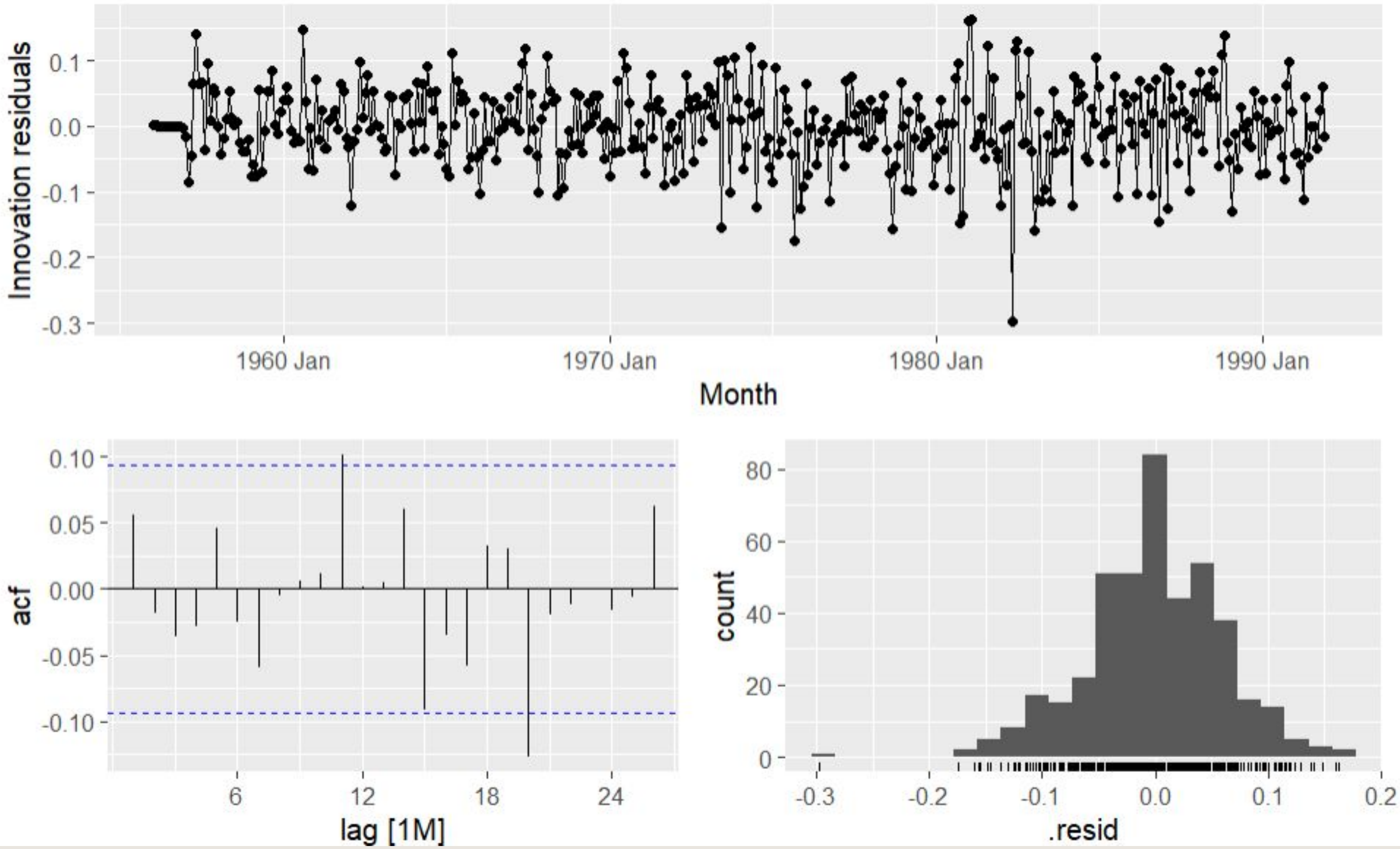
```{r}
fit|>
  accuracy()|>arrange(RMSE)
```

A tibble: 6 x 10

.model <chr>	.type <chr>	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>	MASE <dbl>	RMSSE <dbl>	ACF1 <dbl>
arima314111	Training	-0.13755617	8.652973	6.346916	-0.2215620	4.620051	0.6769872	0.6855461	0.055546810
arima0111013	Training	-0.18539222	9.171343	6.678233	-0.2791862	4.890190	0.7123267	0.7266149	-0.006852976
arima313112	Training	-0.21795962	9.397983	6.811001	-0.2921734	4.987327	0.7264883	0.7445708	0.017259665
auto	Training	-0.22563239	9.399475	6.804963	-0.2888263	4.978114	0.7258443	0.7446890	0.005041232
arima1110210	Training	-0.07435153	9.648806	6.959582	-0.2517469	5.101980	0.7423365	0.7644427	0.024689270
arima310110	Training	-0.02959673	12.319454	9.023652	-0.3454046	6.489179	0.9624984	0.9760292	-0.031233961

6 rows

```{r}
fit |> select(arima314111) |> gg_tsresiduals()
```

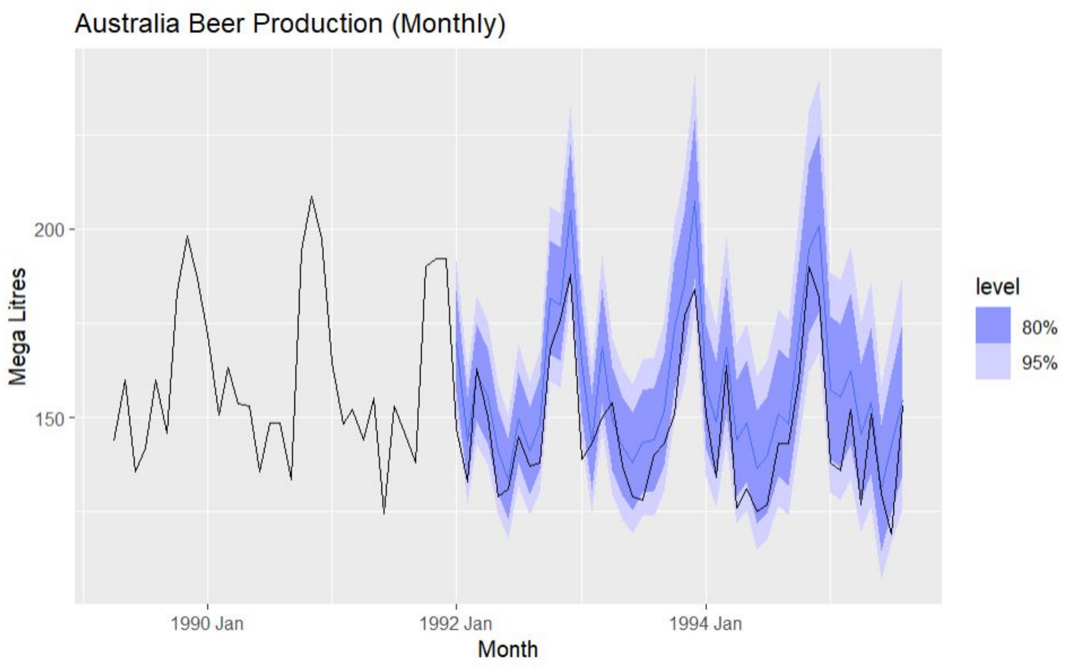


```{r}
augment(fit) |>
  filter(.model == "arima314111") |>
  features(.innov, ljung_box, lag=36)
```

A tibble: 1 × 3

.model <chr>	lb_stat <dbl>	lb_pvalue <dbl>
arima314111	42.29477	0.2176564

1 row



08

Conclusion

Conclusion

Advantage of ARIMA model:

- ARIMA models offer the flexibility to capture various types of patterns and behaviors in the data, such as seasonality, cycles, or trends.
- ARIMA models are very easy to implement, as they have only three parameters and some statistical assumptions.
- ARIMA models are widely used and supported by many software packages and libraries, such as R.
- ARIMA models provide confidence interval and error measure in forecasting, such as, standard errors or root mean squared error.

Disadvantages of ARIMA model:

- ARIMA models cannot handle non linear relationships or complex dynamics, as they are linear models.
- ARIMA models are not suitable for very short or very long time series as they may not have enough information or become unstable over time.
- Outliers and missing values can affect the model estimation and forecasting performance in ARIMA model.

Thanks!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution

Questions?

Resources

- https://online.stat.psu.edu/stat510/lesson/4/4.1
- https://mfe.baruch.cuny.edu/wp-content/uploads/2014/12/TS_Lecture1_2019.pdf
- https://otexts.com/fpp3/
- https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06
- https://medium.com/@ooemma83/how-to-interpret-acf-and-pacf-plots-for-identifying-ar-ma-arma-or-arima-models-498717e815b6
- https://myweb.uiowa.edu/cavaaugh/doc/pub/aicaicc.pdf
-