ST 534 Fall 2025 Project: Quarterly Aircraft Incidents

Joy Zhou, Felicia Chen, Chris Aguilar

2025-11-30

Contents

1 Introduction

This data ('aviation_incident_data_clean.csv') represents the quarterly aircraft incident count for all 50 US states going back to Q1 1983. This data was selected because claims were made that aviation incidents had become more common after January 2025, so we wanted to see if 2025 quarterly data followed a trend consistent with trends seen in previous years.

1.1 Data source

Source: Case Analysis and Reporting Online (CAROL), National Transportation Safety Board, retrieved 2025-10-28, (https://carol.ntsb.gov/).

1.2 Data import and project source

The next section includes R code for importing the data. All project code is also available at (https://github.com/chris-aguilar/st__534__project).

2 Exploratory Data Analyis

We start by doing some exploratory data (EDA) analysis on our data.

2.1 Data import and preview

We'll need to import the aviation incident data first. Then we'll preview the first 10 quarters.

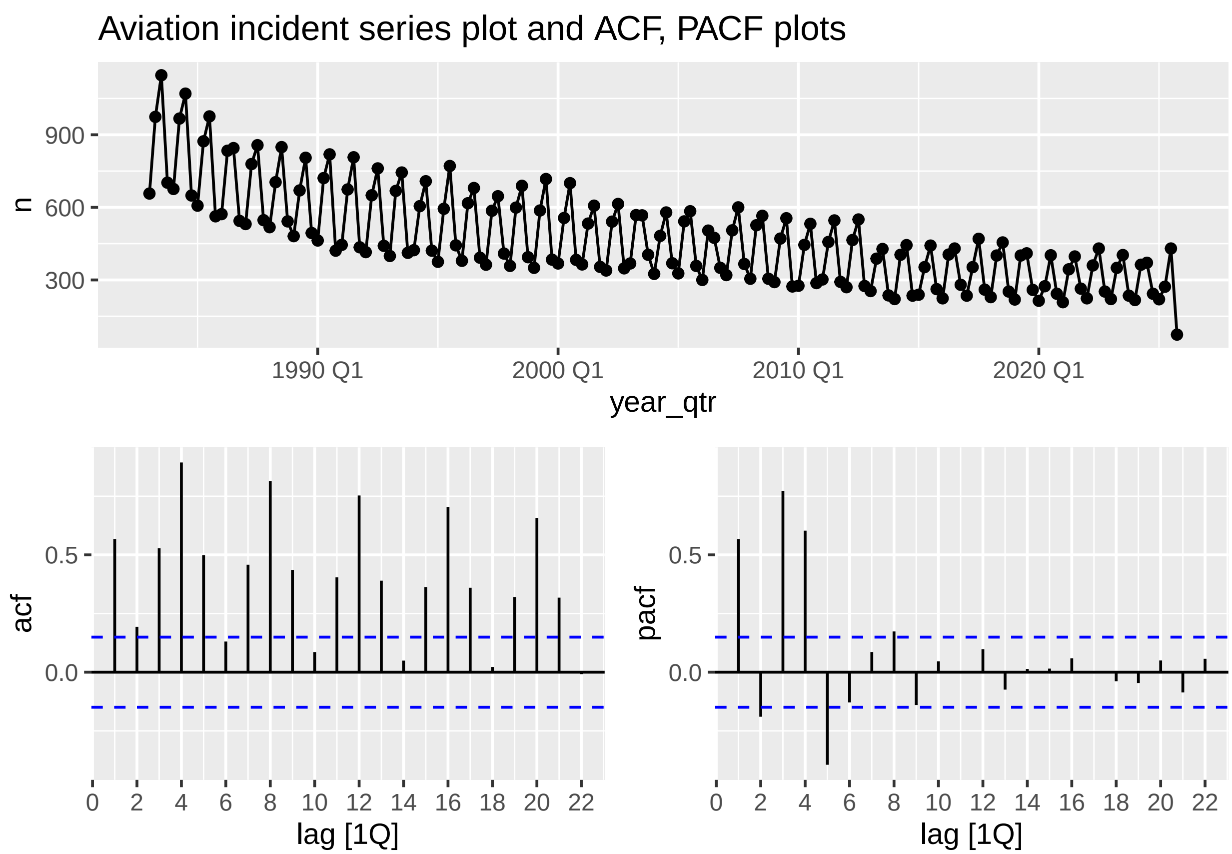
```
library(dplyr)
library(readr)
library(fpp3)
library(knitr)

aviation_data <- read_csv("aviation_incident_data_clean.csv") |>
  mutate(year_qtr = yearquarter(year_qtr)) |>
  as_tsibble(index=year_qtr)

kable(aviation_data |> head(10), col.names = c("Quarter", "Events"))
```

Quarter	Events
1983 Q1	657
1983 Q2	974
1983 Q3	1146
1983 Q4	702
1984 Q1	676
1984 Q2	967
1984 Q3	1070
1984 Q4	649
1985 Q1	607
1985 Q2	873

2.2 Data visualization

We begin by plotting the aviation incident time series itself along with the relevant ACF and PACF plots.

```
aviation_data |>
  gg_tsdisplay(n, plot_type = "partial") +
  labs(title = "Aviation incident series plot and ACF, PACF plots")
```



The last point in the series is an unusual observation. This point corresponds to Q4 2025. This data for analysis was collected 2025-10-28, meaning the final observation only has four weeks of data, not an entire quarter. So we'll drop that point from analysis.

It also seems like the incident variance is not constant over time. This is shown by the smaller amplitudes toward the end of the series. So we'll make use of a log-transformation to stabilize variance.

Next, we notice the ACF plot has large spikes at lag multiples of 4 and these die out very slowly. We're very likely going to need a seasonal differencing at lag of 4 to account for this strong seasonal pattern.

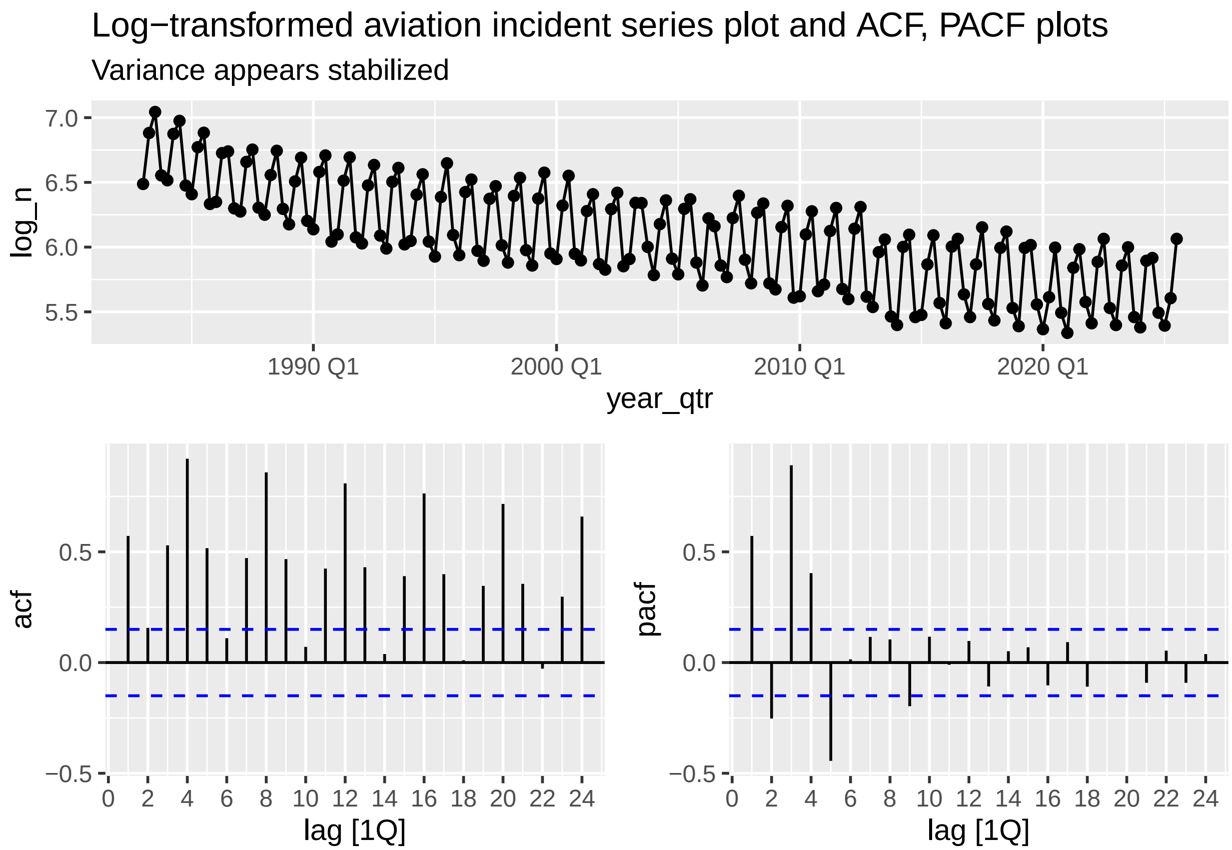
We also see there's a downward trend. We might need lag-1 differencing to account for this, if our seasonally-differenced series does not become stationary.

We'll filter away the final point, create a log-transformed incident count, a seasonal difference of the log-transform, a lag-1 difference of the log transform, and a combination of the differences on the log-transform.

We'll start by visualizing the series with a log-transform first.

```
# creating differenced series at lags 4, 1, and 4+1
aviation_data <- aviation_data |>
  filter_index(. ~ "2025 Q3") |> # everything up to Q3 2025 to omit partial Q4
  mutate(
    log_n = log(n), # for variance stabilization
    sdiff_logn = difference(log_n, 4), # seasonal difference
    diff_logn = difference(log_n), # lag-1 difference
    double_diff_logn = difference(diff_logn, 4) # seasonal diff on simple diff
    )

aviation_data |>
```

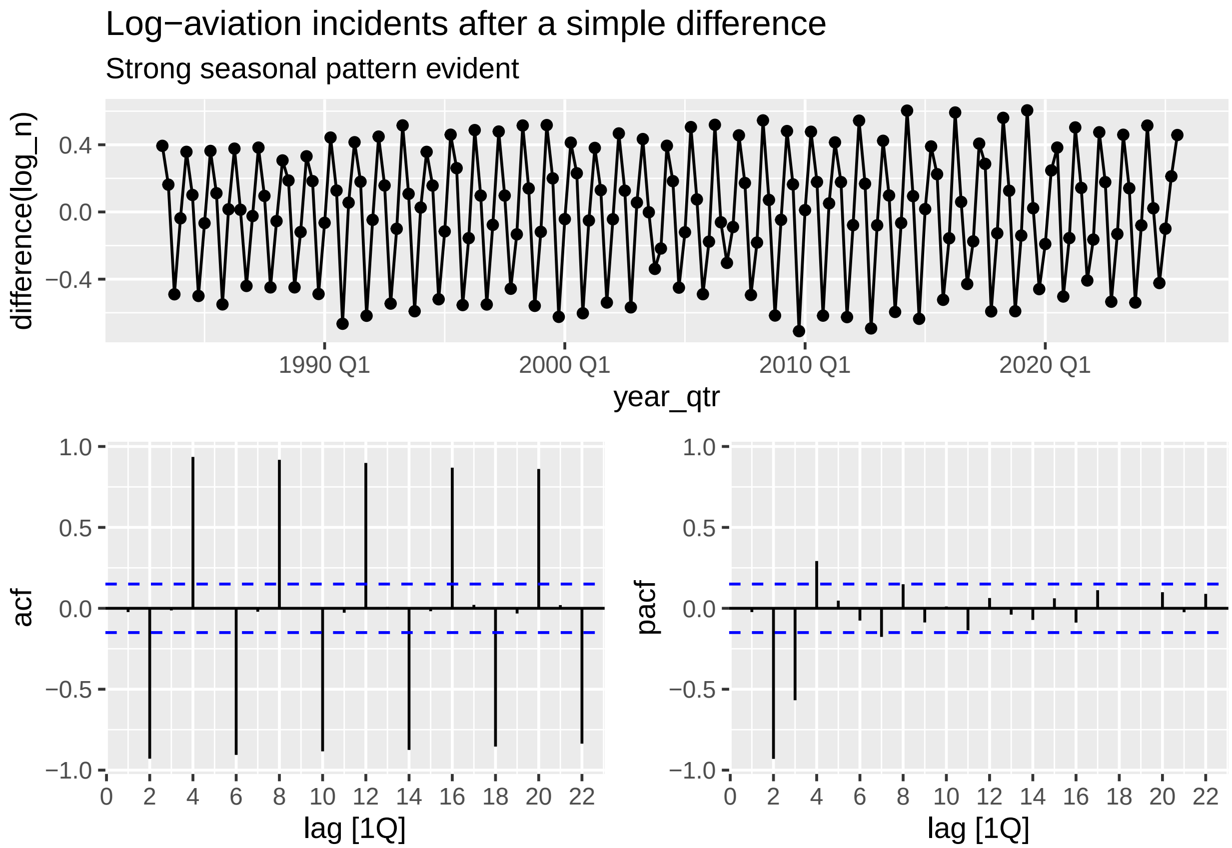
```r
gg_tsdisplay(log_n, plot_type = "partial", lag_max = 24) +
  labs(title = "Log-transformed aviation incident series plot and ACF, PACF plots",
       subtitle = "Variance appears stabilized")
```



After transforming our response variable, it's clear to see variance now appears stable.

Next, we check a lag-1 difference to see if any other seasonal pattern is masked in our ACF/PACF plots.

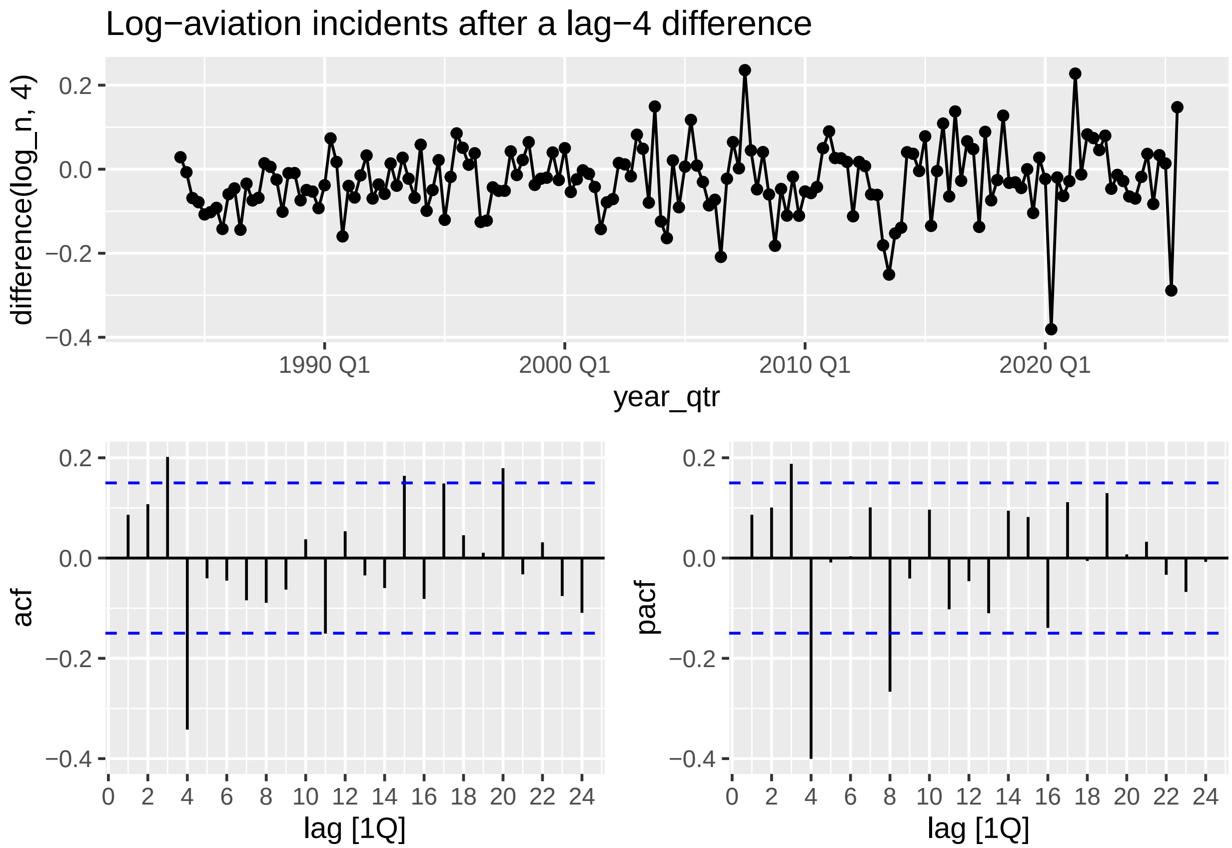
```r
aviation_data |>
  gg_tsdisplay(difference(log_n), plot_type = "partial") +
  labs(title = "Log-aviation incidents after a simple difference",
       subtitle = "Strong seasonal pattern evident")
```



It's clear that there's a strong seasonal pattern based on the differenced series and ACF plots.

Next, we check visually after incorporating a seasonal lag of 4 but no lag-1 difference. We're curious to see if the series becomes stationary after a single seasonal difference.

```
aviation_data |>
  gg_tsdisplay(difference(log_n, 4), plot_type = "partial", lag_max = 24) +
  labs(title = "Log-aviation incidents after a lag-4 difference")
```



The series looks pretty close to stationary seeing as how there appears to be no trend or pattern in the differenced series plot, and the most of the seasonal ACF spikes disappear. We'll check with an ADF test.

```
library(tseries)

# lag-4 adf
aviation_data |>
  filter(!is.na(sdiff_logn)) |> # no NA's allowed for ADF test
  pull(sdiff_logn) |>
  adf.test(k = 4) # test at lag 4 due to quarterly seasonality
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  pull(filter(aviation_data, !is.na(sdiff_logn)), sdiff_logn)
## Dickey-Fuller = -6.6688, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

Based on the ADF test, a single seasonal difference at lag 4 is enough to make our series stationary.

We also check to see if we can reject the Ljung-Box null hypothesis of white noise.

```
aviation_data |>
  filter(!is.na(sdiff_logn)) |> # base-R Ljung-Box test expects no NAs
  pull(sdiff_logn) |>
  Box.test(lag = 8, type = "Ljung-Box") # 2 * seasonal period = 8
```

```
##
##  Box-Ljung test
##
## data:  pull(filter(aviation_data, !is.na(sdiff_logn)), sdiff_logn)
## X-squared = 33.825, df = 8, p-value = 4.369e-05
```

The Ljung-Box test suggests the lag-4 differenced series is not white noise, so we should do some modeling to make use of more information in the series.

Looking at seasonal lag components of the **Lag-4 difference series, ACF and PACF plots**, we notice that the ACF plot has significant spikes into the teens and 20. We also notice the PACF plot has no more significant spikes after lag 8. The fact that the ACF plot continues to have spikes at later lags but the PACF plot stops after lag 8 suggests a non-seasonal AR portion. But also, the PACF still looks a little bit like it dies off slowly. So this suggests both a seasonal AR and MA component will be needed as well.

So, since we took one seasonal difference, we fix $D = 1$. For simplicity's sake, we're going to guess $p = 3$ based on the significant PACF spike at lag 3. So we'll just play around with $P = [0, 1, 2]$, and leave $q = 0$.

We leave $Q = 1$ to account for the seasonal ARMA component, and since P and Q are rarely greater than 1. However, we allow $P = [0, 1, 2]$ to vary because of the PACF.

We'll also play with including a constant or not in the model. This is because the original series showed some slight downward trend, even though the lag-4 differenced series became stationary. With 171 observations, we have plenty of data to play with.

We'll also make use of a variation of the Hyndman-Khandakar algorithm to automatically select a model by choosing parameters that minimize $AICc$, a corrected version of the AIC.

Thus, the candidate models using the log-transformed incident series are:

- $ARIMA(3, 0, 0) \times (0, 1, 1)_4$ sans constant,
- $ARIMA(3, 0, 0) \times (0, 1, 1)_4$ with a constant,
- $ARIMA(3, 0, 0) \times (1, 1, 1)_4$ sans constant,
- $ARIMA(3, 0, 0) \times (1, 1, 1)_4$ with a constant,
- $ARIMA(3, 0, 0) \times (2, 1, 1)_4$ sans constant,
- $ARIMA(3, 0, 0) \times (2, 1, 1)_4$ with a constant,
- Some auto-selected seasonal ARIMA model, fixing $D = 1$.

We'll go through the model fitting exercise next using these models.

3 Modeling

We will first split the data into training and test sets, using Q1 1983 - Q4 2024 for training and Q1 2025 - Q3 2025 for testing.

```
aviation_train <- aviation_data |> filter(year(year_qtr) < 2025)
aviation_test <- aviation_data |> filter(year(year_qtr) >= 2025)
```

We will also use the following utility functions for the Ljung-Box white noise test.

```
get_dof_from_arima <- function(fit_arima, model_name) {
  return (
    fit_arima |>
      select(ends_with(model_name)) |>
```

```r
      tidy(.) |>
      filter(term != "constant") |>
      NROW()
  )
}

# Ljung-Box test function
# Lag selected based on https://robjhyndman.com/hyndsight/ljung-box-test/
# As the data is quarterly, we can use h = min(2m, T/5) = min(2*4, 171/5) =5
test_ljung_box <- function(fit_arima, model_name) {
  return (
    augment(fit_arima) |>
      filter(.model == model_name) |>
      features(.innov, ljung_box, lag=8,
               dof=get_dof_from_arima(fit_arima, model_name))
  )
}
```

3.1 Model fitting

We will now fit the candidate models proposed during exploratory data analysis on the log-transformed data.

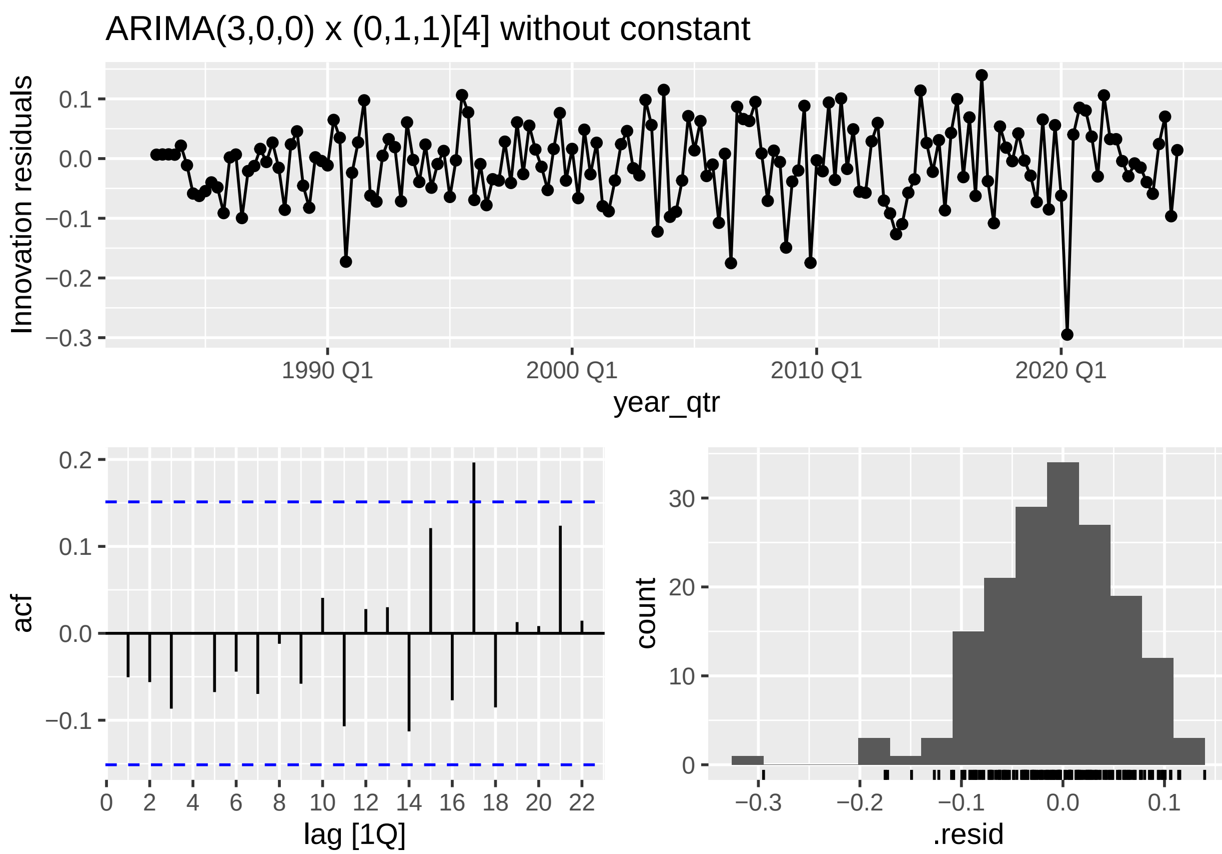
- $ARIMA(3,0,0) \times (0,1,1)_4$ sans constant,
- $ARIMA(3,0,0) \times (0,1,1)_4$ with a constant,
- $ARIMA(3,0,0) \times (1,1,1)_4$ sans constant,
- $ARIMA(3,0,0) \times (1,1,1)_4$ with a constant,
- $ARIMA(3,0,0) \times (2,1,1)_4$ sans constant,
- $ARIMA(3,0,0) \times (2,1,1)_4$ with a constant,
- Some auto-selected seasonal ARIMA model, fixing $D = 1$.

```r
aviation_fit <-
  aviation_train |>
  model(arima300_011 = ARIMA(log_n ~ 0 + pdq(3, 0, 0) + PDQ(0, 1, 1, period=4)),
        arima300_011_c = ARIMA(log_n ~ 1 + pdq(3, 0, 0) + PDQ(0, 1, 1, period=4)),
        arima300_111 = ARIMA(log_n ~ 0 + pdq(3, 0, 0) + PDQ(1, 1, 1, period=4)),
        arima300_111_c =  ARIMA(log_n ~ 1 + pdq(3, 0, 0) + PDQ(1, 1, 1, period=4)),
        arima300_211 = ARIMA(log_n ~ 0 + pdq(3, 0, 0) + PDQ(2, 1, 1, period=4)),
        arima300_211_c = ARIMA(log_n ~ 1 + pdq(3, 0, 0) + PDQ(2, 1, 1, period=4)),
        auto = ARIMA(log_n ~ PDQ(D=1, period=4), stepwise = FALSE, approx = FALSE)
        )
```

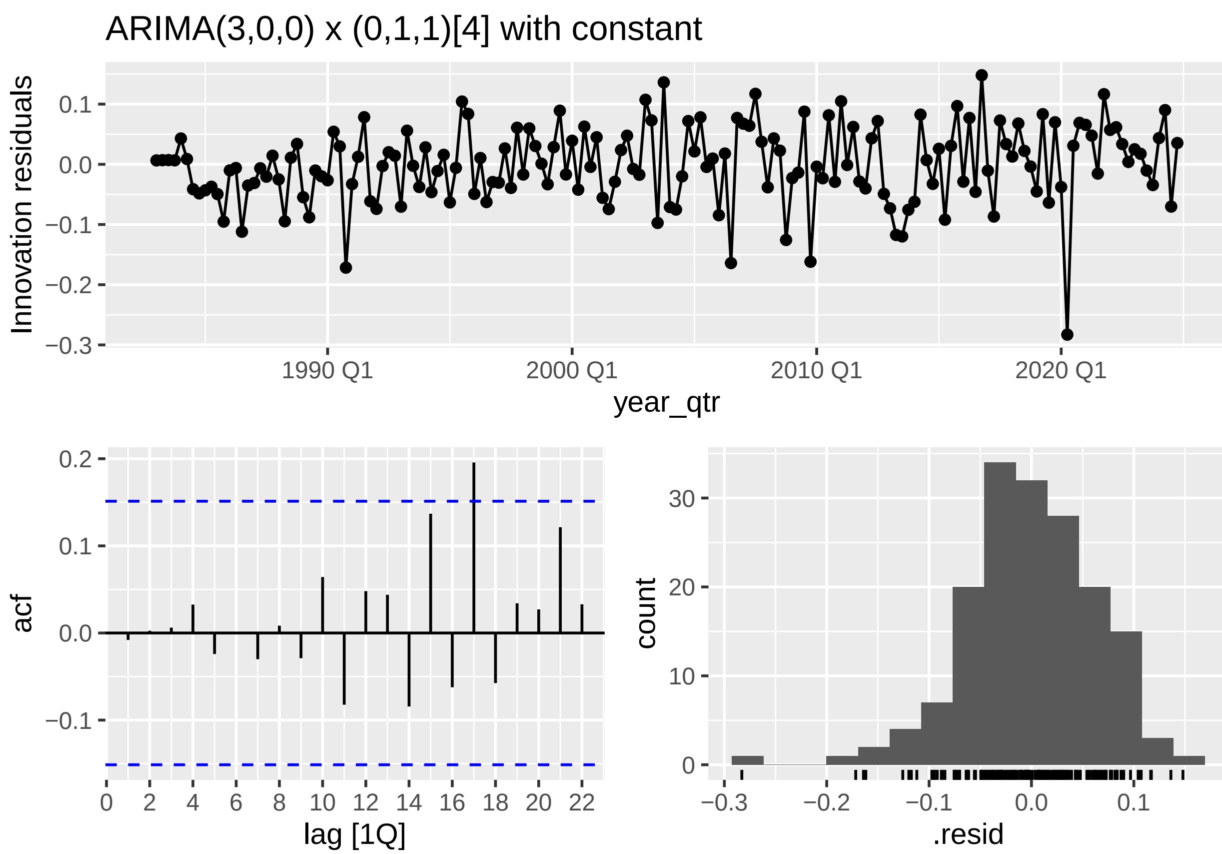
We will first examine the $ARIMA(3,0,0) \times (0,1,1)_4$ models.

3.2 $ARIMA(3,0,0) \times (0,1,1)_4$

```r
aviation_fit |>
  select(arima300_011) |>
  ggtime::gg_tsresiduals() +
  labs(title = "ARIMA(3,0,0) x (0,1,1)[4] without constant")
```




```r
aviation_fit |>
  select(arima300_011_c) |>
  ggtime::gg_tsresiduals() +
  labs(title = "ARIMA(3,0,0) x (0,1,1)[4] with constant")
```



The residual pattern of the constant-inclusive model is closer to normal than the no-constant model (which skews left). More autocorrelation remains in the no-constant ACF than the constant-inclusive model, though no quarterly autocorrelation remains in either. Both contain an autocorrelation spike at lag 17, but this is potentially natural variation given the higher lag order and singular occurrence. We also note a sharp residual difference at Q1 2020, coinciding with the COVID-19 pandemic onset.

```
kable(rbind(test_ljung_box(aviation_fit, "arima300_011"),
      test_ljung_box(aviation_fit, "arima300_011_c")))
```

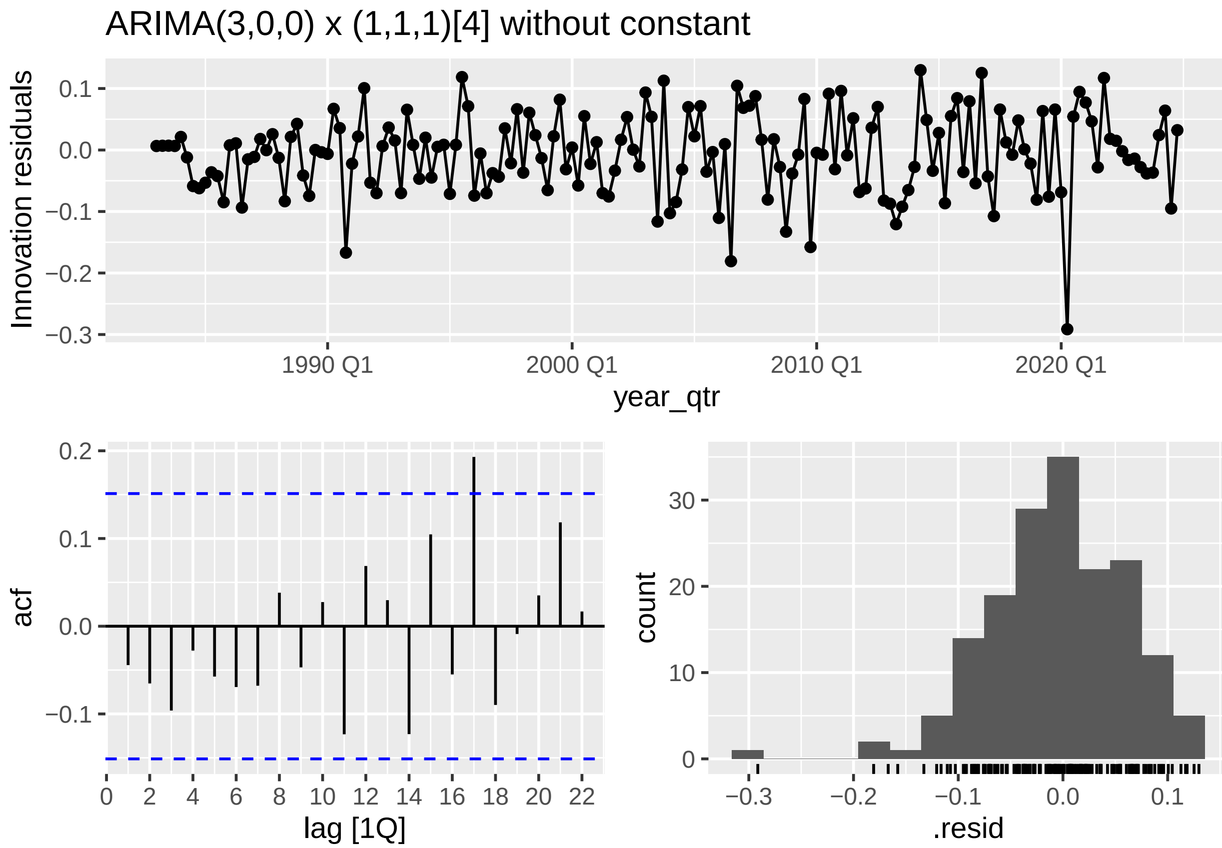
.model	lb_stat	lb_pvalue
arima300_011	4.3114853	0.3654891
arima300_011_c	0.4798399	0.9754337

The Ljung-Box white noise test results in p-values well above the $\alpha = .05$ threshold for both models, though the constant-inclusive model has a much higher p-value (.97). The residuals of both models are thus indistinguishable from white noise. Overall, the constant-inclusive model seems to have the better fit.

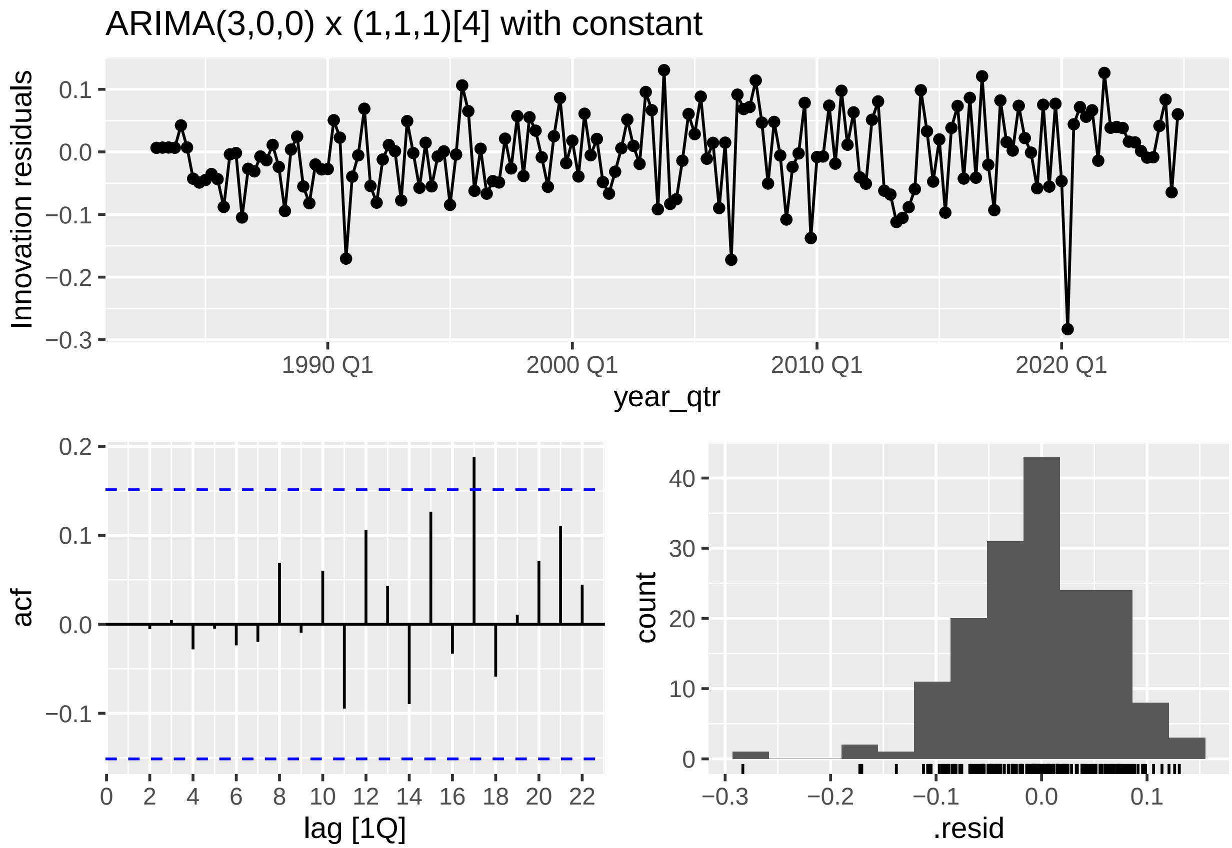
3.3 $ARIMA(3,0,0) \times (1,1,1)_4$

We will now examine an $ARIMA(3,0,0) \times (1,1,1)_4$ sans constant, and an $ARIMA(3,0,0) \times (1,1,1)_4$ with a constant.

```
aviation_fit |>
  select(arima300_111) |>
  ggtime::gg_tsresiduals() +
  labs(title = "ARIMA(3,0,0) x (1,1,1)[4] without constant")
```




```r
aviation_fit |>
  select(arima300_111_c) |>
  ggtime::gg_tsresiduals() +
  labs(title = "ARIMA(3,0,0) x (1,1,1)[4] with constant")
```




```
kable(rbind(test_ljung_box(aviation_fit, "arima300_111"),
      test_ljung_box(aviation_fit, "arima300_111_c")))
```

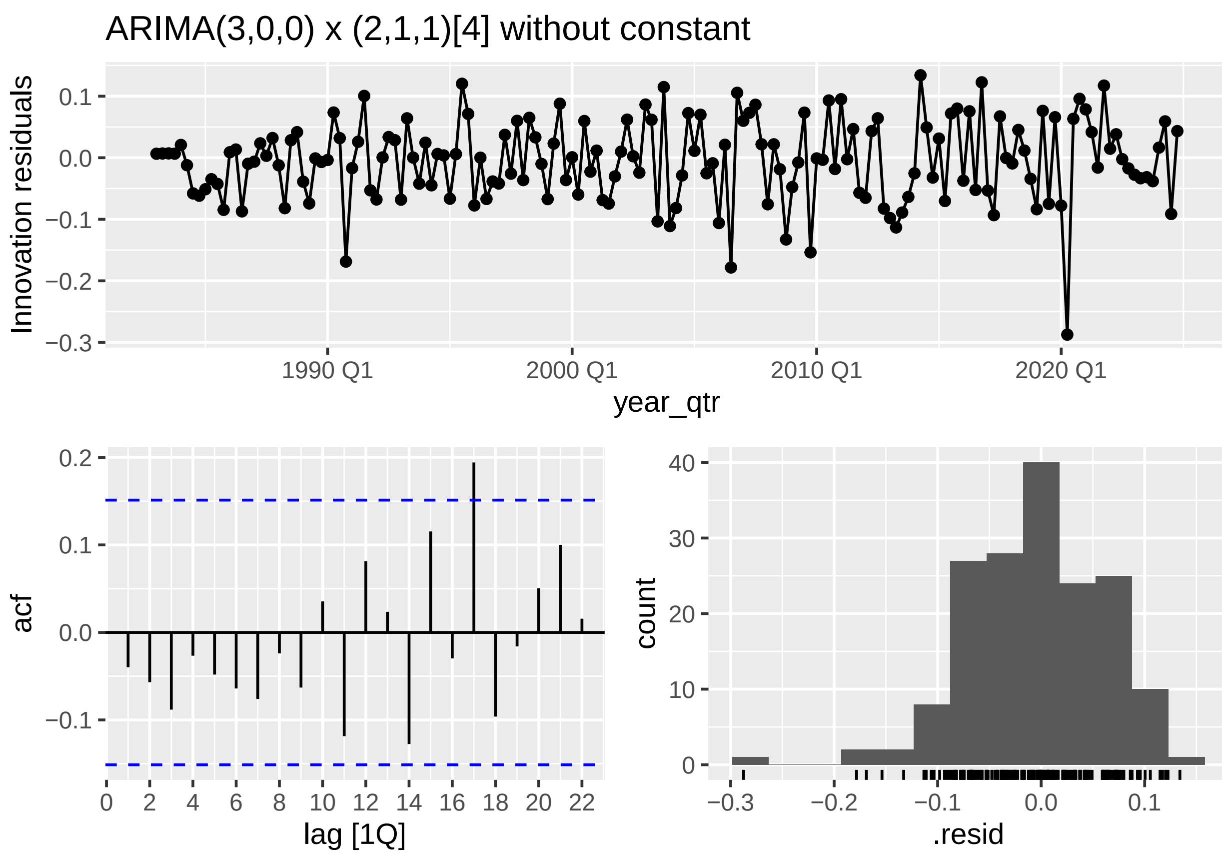
.model	lb_stat	lb_pvalue
arima300_111	5.304052	0.1508397
arima300_111_c	1.171298	0.7598962

Like the previous $ARIMA(3,0,0) \times (0,1,1)_4$ models, we see isolated residual ACF spikes at lag 17 and a large residual in Q1 2020. We also observe a closer-to-normal distribution of residuals with the constant-inclusive model. From the Ljung-Box test, we can reasonably conclude the residuals of both models behave like white noise, with the constant-inclusive model providing a high probability of white noise. Compared to the previous models, we observe slightly more residual autocorrelation, but the models appear to fit adequately, favoring the constant-inclusive model.

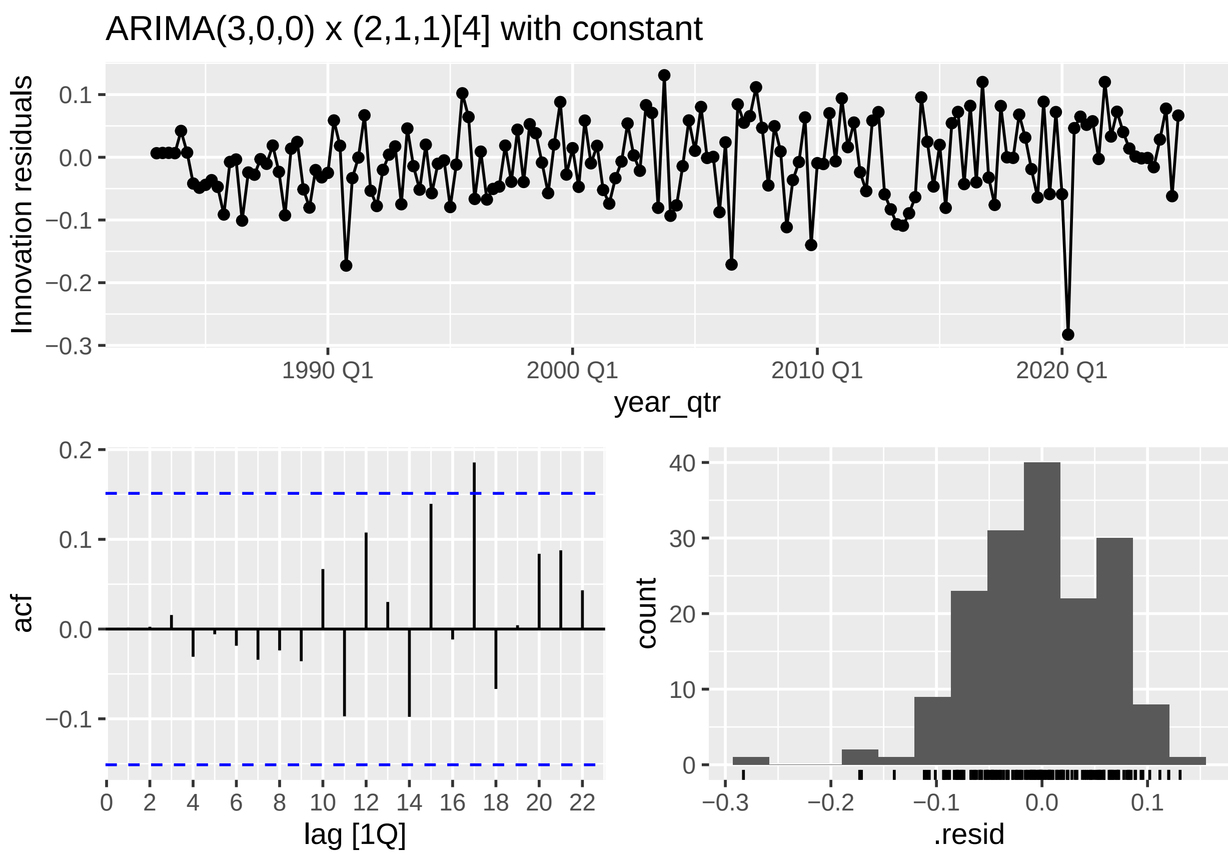
3.4 $ARIMA(3,0,0) \times (2,1,1)_4$

Next, we will inspect the $ARIMA(3,0,0) \times (2,1,1)_4$ models (with and without a constant). This model has one more seasonal AR term than the previous set.

```
aviation_fit |>
  select(arima300_211) |>
  gg_tsresiduals() +
  labs(title = "ARIMA(3,0,0) x (2,1,1)[4] without constant")
```




```r
aviation_fit |>
  select(arima300_211_c) |>
  gg_tsresiduals() +
  labs(title = "ARIMA(3,0,0) x (2,1,1)[4] with constant")
```




```r
kable(rbind(test_ljung_box(aviation_fit, "arima300_211"),
      test_ljung_box(aviation_fit, "arima300_211_c")))
```

.model	lb_stat	lb_pvalue
arima300_211	4.5563794	0.1024695
arima300_211_c	0.5852419	0.7463050

The residuals for the $ARIMA(3,0,0) \times (2,1,1)_4$ models are very similar to those previously examined, with a large residual in Q1 2020 and the constant-inclusive model possessing overall lower autocorrelation and a nearer-to-normal residual arrangement. The Ljung-Box check has a fairly low p-value of ~.10 for the non-constant-inclusive model, but a quite large one (~.75) for the constant-inclusive model, suggesting that the constant-inclusive model residuals are white noise.

3.5 Model comparison

Overall, the models with constants fit the data more closely. From visual inspection, the first model with a constant - $ARIMA(3,0,0) \times (0,1,1)_4$ - appeared to fit best. This model was also automatically selected by the Hyndman-Khandakar algorithm.

```r
aviation_fit |> select(auto)
```

```
## # A mable: 1 x 1
##                                auto
##                             <model>
## 1 <ARIMA(3,0,0)(0,1,1)[4] w/ drift>
```

```r
kable(glance(aviation_fit) |> arrange(AICc) |> select(.model:BIC))
```

.model	sigma2	log_lik	AIC	AICc	BIC
arima300_011_c	0.0043334	214.5120	-417.0240	-416.4890	-398.4248
auto	0.0043334	214.5120	-417.0240	-416.4890	-398.4248
arima300_111_c	0.0042786	215.2349	-416.4698	-415.7518	-394.7707
arima300_211_c	0.0042143	216.2414	-416.4828	-415.5538	-391.6839
arima300_011	0.0046127	208.7728	-407.5457	-407.1659	-392.0463
arima300_111	0.0045943	209.0625	-406.1250	-405.5900	-387.5258
arima300_211	0.0045684	209.6366	-405.2732	-404.5552	-383.5741

The constant-inclusive $ARIMA(3,0,0) \times (0,1,1)_4$ model has the lowest AIC, AICc, and BIC values, and is also the most parsimonious of the models with constants.

We can also compare the accuracy of the 2025 forecasts.

```r
kable(
  aviation_fit |> forecast(h=3) |>
    accuracy(aviation_test) |>
    select(!RMSSE & !MASE) |> # remove NaN cols
    arrange(RMSE)
  )
```

.model	.type	ME	RMSE	MAE	MPE	MAPE	ACF1
arima300_011	Test	-0.0296566	0.1339688	0.1107522	-0.5641826	1.940522	-0.6395015
arima300_011_c	Test	-0.0037086	0.1342079	0.1219105	-0.1108435	2.130034	-0.6319503
auto	Test	-0.0037086	0.1342079	0.1219105	-0.1108435	2.130034	-0.6319503
arima300_111	Test	-0.0334170	0.1361920	0.1109762	-0.6300485	1.945734	-0.6416006
arima300_211	Test	-0.0343844	0.1364869	0.1100254	-0.6498594	1.926219	-0.6344749
arima300_111_c	Test	-0.0048673	0.1381927	0.1249095	-0.1334598	2.181585	-0.6295503
arima300_211_c	Test	-0.0072219	0.1383949	0.1227787	-0.1795580	2.139480	-0.6153429

The $ARIMA(3,0,0) \times (0,1,1)_4$ models also minimize test RMSE. As AIC and BIC favor the model with a constant, we will select it as our preferred model for forecasting.

```r
aviation_fit |> select(arima300_011_c) |> report()
```

```
## Series: log_n
## Model: ARIMA(3,0,0)(0,1,1)[4] w/ drift
##
## Coefficients:
##          ar1     ar2     ar3     sma1  constant
##       0.2425  0.1610  0.2485  -0.7446   -0.0088
## s.e.  0.0789  0.0783  0.0779   0.0877    0.0014
##
## sigma^2 estimated as 0.004333:  log likelihood=214.51
## AIC=-417.02   AICc=-416.49   BIC=-398.42
```

3.6 Final model

Our preferred model for the log number of airline incidents is therefore:

$$\phi(B)\,(1-B^4)Z_t = -.0088 + \Theta(B^4)a_t$$

where $\phi(B) = 1 - .2425B - .1610B^2 - .2485B^3$, and $\Theta(B^4) = 1 - .7446B^4$.

Table 1: 3-year forecasts from ARIMA(3,0,0) x (0,1,1)[4]

year_qtr	n	.mean
2025 Q4	t(N(5.4, 0.0047))	231.558
2026 Q1	t(N(5.3, 0.0048))	196.137
2026 Q2	t(N(5.8, 0.005))	317.349
2026 Q3	t(N(5.9, 0.0056))	378.996
2026 Q4	t(N(5.4, 0.0063))	222.320
2027 Q1	t(N(5.3, 0.0065))	197.463
2027 Q2	t(N(5.7, 0.0067))	301.329
2027 Q3	t(N(5.9, 0.0069))	368.071
2027 Q4	t(N(5.4, 0.0075))	217.679
2028 Q1	t(N(5.2, 0.0076))	191.077
2028 Q2	t(N(5.7, 0.0077))	293.287
2028 Q3	t(N(5.9, 0.0079))	358.779

4 Forecasting

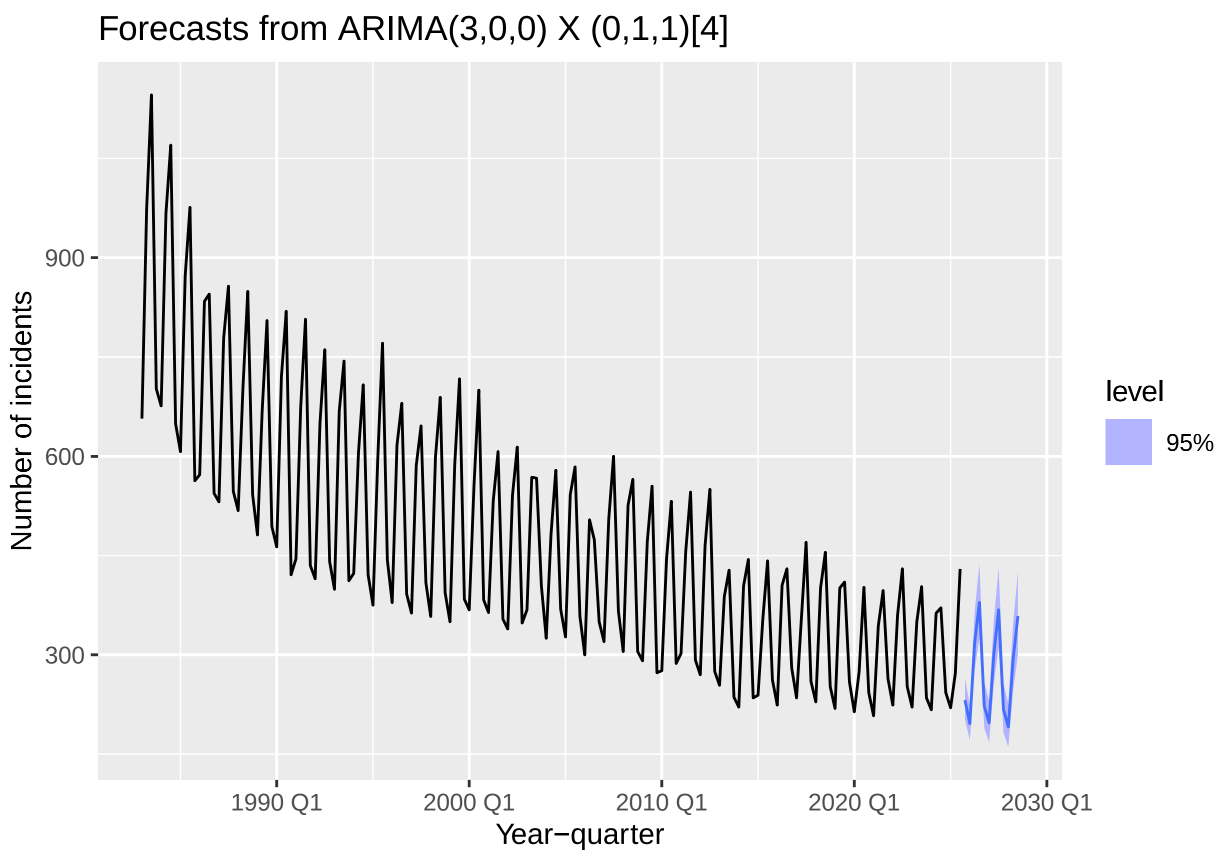
We will use the best-fitting model, $ARIMA(3,0,0) \times (0,1,1)_4$ with a constant, to generate forecasts for the next 3 years forecast.

```
library(fpp3)

fit <- aviation_data |>
  model(arima300_011_c = ARIMA(log(n) ~ 1 + pdq(3, 0, 0) + PDQ(0, 1, 1, period=4)
    )
  )

# Forecast the next 12 quarters (3 years)
fc <- fit |>
  forecast(h = 12)
```

```
library(knitr)

fc |>
  as_tibble() |>
  select(year_qtr, n,.mean)|>
  kable(digits = 3, caption = "3-year forecasts from ARIMA(3,0,0) x (0,1,1)[4]")
```

Using `autoplot()` function to generate a plot of 3-year ahead forecasts

```
fc |>
  autoplot(level=95) +
  autolayer(aviation_data, n) +
  labs(
    title = "Forecasts from ARIMA(3,0,0) X (0,1,1)[4]",
    y = "Number of incidents",
    x = "Year-quarter"
  )
```



The model predicts that aviation incident counts will remain stable over the next three years, with seasonal variation following the historical quarterly pattern. The forecast doesn't indication increase in aviation incidence frequency in the near future. This may reflect steady aviation safety practice.

5 Conclusion

5.1 Summary

In this project, we analyzed quarterly aviation incident data from 1983 to 2025 across all 50 states using exploratory data analysis (EDA), time series modeling, and forecasting. EDA revealed long-term trends, seasonal patterns, and non-stationarity in the data, guiding the selection of candidate ARIMA models. After evaluating multiple models using AIC, AICc, BIC, and residual diagnostics, the seasonal ARIMA model $ARIMA(3,0,0) \times (0,1,1)_4$ with a constant was chosen for forecasting. Using this model, we we generated forecasts of aviation incident counts for the next three years.

5.2 Core findings

5.2.1 Exploratory Data Analysis (EDA)

The aviation incident series displays non-constant variance and a pronounced long-term downward trend across the study period. Quarterly seasonality is evident, with Q3 consistently recording the highest number of incidents, likely due to increased flight activity during this period. To stabilize the variance, we applied a log transformation.

The ACF of the original series exhibits strong seasonal spikes at lags 4, 8, 12, etc., along with a slow decay, confirming non-stationarity. Based on the ADF test, applying a seasonal difference at lag 4 to the log-transformed series is sufficient to achieve stationarity. However, the Ljung–Box test indicates that this differenced series is not white noise, suggesting the further modeling is required.

In addition, the ACF and PACF plots of the seasonally differenced series show notable spikes around lags 3 and 4, indicating the presence of both non-seasonal and seasonal AR and MA components. Based on these patterns, we proposed six candidate ARIMA models, along with an automatically selected seasonal ARIMA model using the `Hyndman-Khandakar algorithm`, for further evaluation.

5.2.2 Modeling

We split the dataset into a training set and a test set to evaluate the performance of our candidate models. The Ljung-Box white noise test indicated that residuals of all candidate models are white noise, indicating that each model provided an adequate fit to the training data.

Among the seasonal ARIMA models considered, the model $ARIMA(3, 0, 0) \times (0, 1, 1)_4$ with a constant demonstrated the best overall performance. It achieved the lowest AIC (–417.0240), AICc (–416.4890), and BIC (–398.4248), and was also the model automatically selected by the `Hyndman-Khandakar algorithm`. Additionally, its residuals were closer to normally distributed, further supporting its suitability. Although the constant-free version of the model achieved the lowest RMSE (0.1339688), the constant-inclusive model had substantially better AIC and BIC values. Given the emphasis on model parsimony and overall fit, we selected the $ARIMA(3, 0, 0) \times (0, 1, 1)_4$ with a constant as the final model for forecasting.

Including a constant in a seasonally differenced ARIMA model introduces a drift term, allowing the model to capture any remaining long-term trend in the data. The constant-inclusive model provided a better fit based on AIC, AICc, and BIC, suggesting that a small underlying trend remained after seasonal differencing.

5.2.3 Forecasting

Using selected $ARIMA(3, 0, 0) \times (0, 1, 1)_4$ with a constant, we generated quarterly forecast for the next three years. The forecast retains the strong seasonal pattern observed historically, with seasonal peaks persisting in Q3. Forecasts indicate continued decline in incident counts,

The 95% prediction intervals widen gradually, reflecting increasing uncertainty over longer forecast horizons, which is typical for ARIMA models.

5.3 Overall interpretation and recommendations

This study provides a view of aviation incident trends in the United States over four decades. The significant long-term decline in incident counts reflects advancements in aviation safety standards, regulatory oversight, pilots training, and aircraft technology. However, the consistent peak in Q3, suggests that operational factors such as increased flight activity during summer months continue to influence incident frequency.

The selected ARIMA model effectively captures these historical dynamics and offers reliable short-term forecasts. While the projected decline in incidents is encouraging, the widening prediction intervals highlight inherent uncertainty in long-range forecasting, emphasizing the need for continuous monitoring and model updates.

Furthermore, these findings can provide useful information for resource allocation, safety audits, and policy planning by aviation authorities and airlines. Anticipating seasonal peaks may help optimize staffing and maintenance schedules during high-traffic periods.

5.4 Limitations of the project

This analysis relied solely on aggregated incident counts, excluding explanatory variables such as weather, aircraft age, maintenance practices, and operational workload. Incorporating these factors in future models could improve predictive accuracy and provide deeper insights into causal mechanisms.

Additionally, exploring advanced modeling approaches, such as SARIMAX for external regressors or machine learning techniques like LSTM, could provide deeper insights and improve long-term forecasting performance.