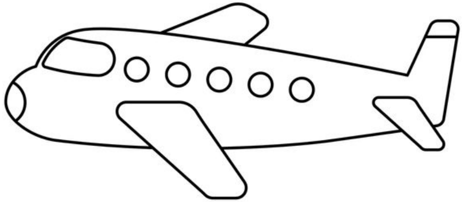
Market Forecasting Project

By George Barahona

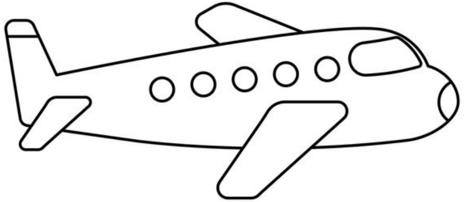
TABLE OF CONTENTS

Ansett Dataset

- Background: Ansett Airlines (Australian Airline, no longer exists)
  - Previously a strong player among domestics carriers
  - Covering diverse range of routes, contributions to regional connectivity
  - Data sheds light on dynamics of domestic air travel in Australia



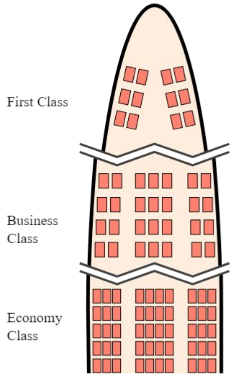
Ansett Dataset in R



- Features a major pilots' industrial dispute
  - Resulted in some weeks having zero passengers
  - Two changes in definition of passenger classes
- Weekly tsibble with one value (Passengers)
- Passengers: Total air passengers travelling with Ansett
- Each series is uniquely identified using two keys:
  - Airports: The airports that passengers are travelling between (both directions)
  - Class: The class of the ticket

Nature of Forecasting Problem

- Objective
  - Predict the number of passengers per class (e.g., Economy, Business) for future periods
- Nature of Problem
  - Understand the demand patterns across different classes of service
  - Analyzing how passenger preferences shift between classes over time
  - See a trend around holidays and seasonality



METHODS

DRM

Dynamic Regressions Model

ADF

Augmented Dickey-Fuller

ACF

Autocorrelation Function

1

2

3

Cleaning Data

```
###Clean and organize data
data("ansett")
# Check for missing values
colSums(is.na(ansett))
#if you want to fill missing values with the mean for numeric columns:
ansett <- na.omit(ansett)
#Convert into proper date format & tsibble
ansett_ts <- ansett %>%
  mutate(Week = as.Date(Week, format = "%Y W%V")) %>%
  as_tsibble(index = Week, key = c("Class", "Airports")) %>%
  fill_gaps(Passengers=0)
```

```
> colSums(is.na(ansett))
      Week   Airports        Class Passengers
         0          0            0          0
```

```
ansett_ts <- ansett_ts %>%
  mutate(Passengers = if_else(Passengers == 0, NA_real_, Passengers))

ansett_ts <- VIM::kNN(ansett_ts, variable = "Passengers")
```

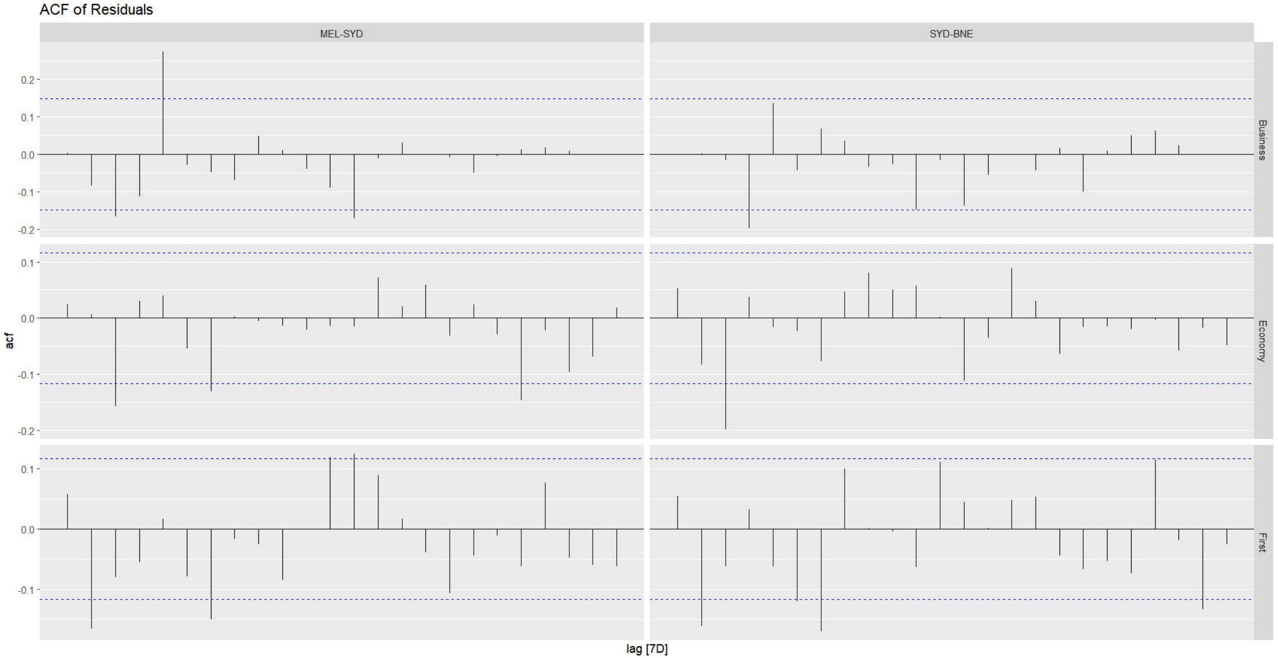
ADF

```
        Augmented Dickey-Fuller Test

data:  passengers_ts
Dickey-Fuller = -1.8871, Lag order = 4, p-value = 0.6234
alternative hypothesis: stationary
```

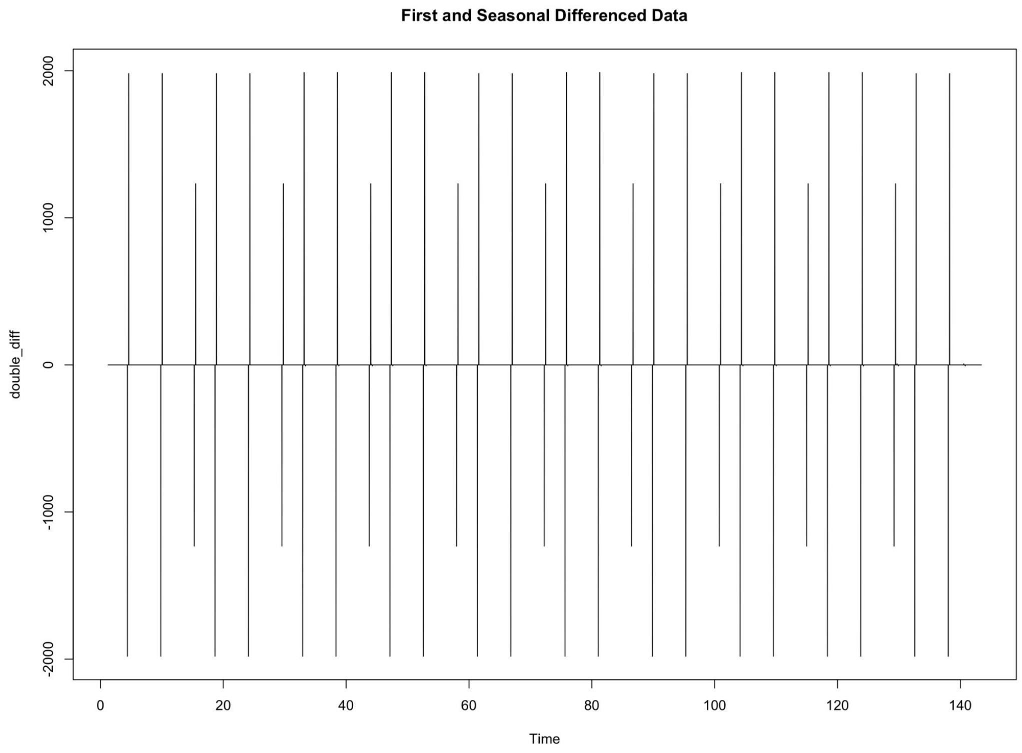
- P-value is greater than 0.5
- Reject the null hypothesis
- Non stationary

```
stationarity_results <- dynamic_data %>%

  features(Passengers, feature_set(tsfeatures::augmented_dickey_fuller))

print(stationarity_results)
```

ACF



- Confirming it is nonstationary
- More than 5% lies outside the threshold
- Not white noise



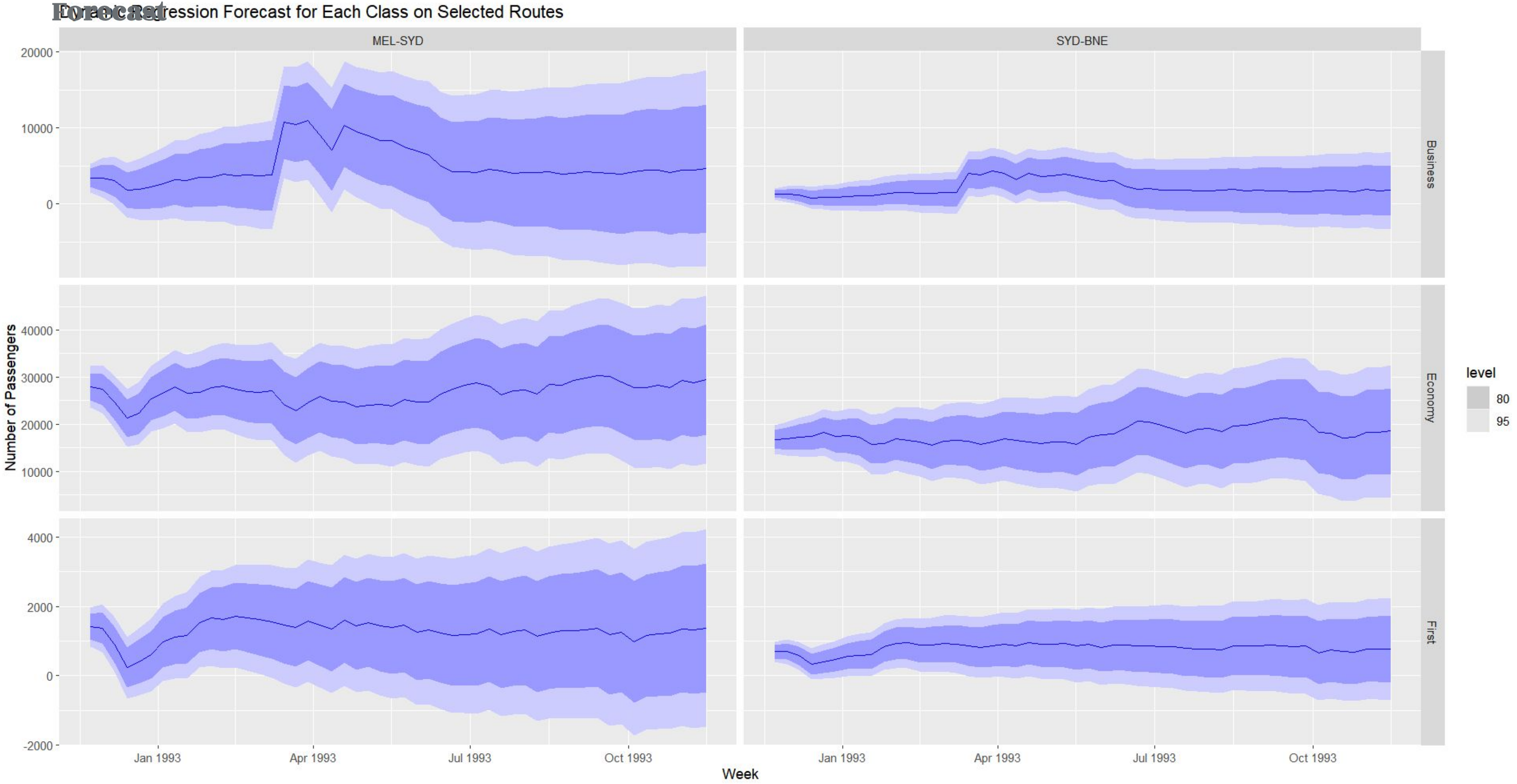

```
double_diff <- diff(diff(ts_data, differences =
1), lag = 12, differences = 1)
plot(double_diff, main="First and Seasonal
Differenced Data")
```

Forecast Code

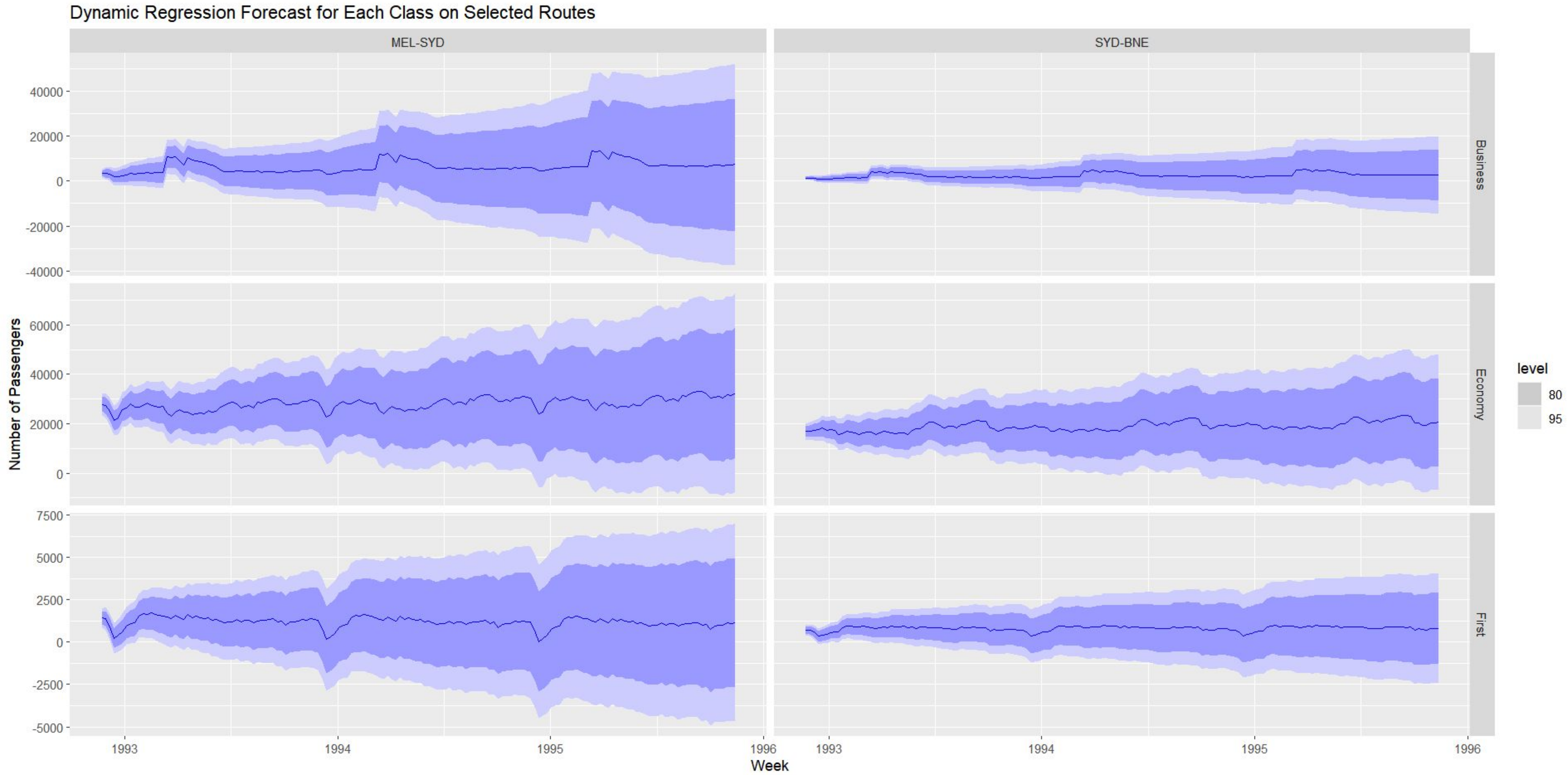
```r
# Adding a hypothetical regressor: Marketing Spend
ansett_ts <- ansett_ts %>%
  mutate(MarketingSpend = rnorm(n = n(), mean = 1000, sd = 300))

# Filter data for specific airports and include all classes
dynamic_data <- ansett_ts %>%
  filter(Airports %in% c("MEL-SYD", "SYD-BNE")) %>%
  group_by(Airports, Class, Week) %>%
  summarise(Passengers = sum(Passengers, na.rm = TRUE),
            MarketingSpend = mean(MarketingSpend), .groups = 'drop') %>%
  as_tsibble(index = Week, key = c("Airports", "Class"))
```

```r
# Fit a Dynamic Regression model using ARIMA with Marketing Spend as a regressor
dynamic_models <- dynamic_data %>%
  model(DynReg = ARIMA(Passengers ~ MarketingSpend + pdq(0, 1, 1) + PDQ(0, 1, 1, period = 52))

# Forecasting new dataset
future_data <- new_data(dynamic_data, n = 52)

# Add new values which is Marketing Spend
future_data <- future_data %>%
  mutate(MarketingSpend = rnorm(n = n(), mean = 1000, sd = 5))

# forecast using the model
dynamic_forecasts <- dynamic_models %>%
  forecast(new_data = future_data)

# Plot forecast
autoplot(dynamic_forecasts) +
  facet_grid(Class ~ Airports, scales = "free_y") +
  labs(title = "Dynamic Regression Forecast for Each Class on Selected Routes",
       y = "Number of Passengers", x = "Week")
```

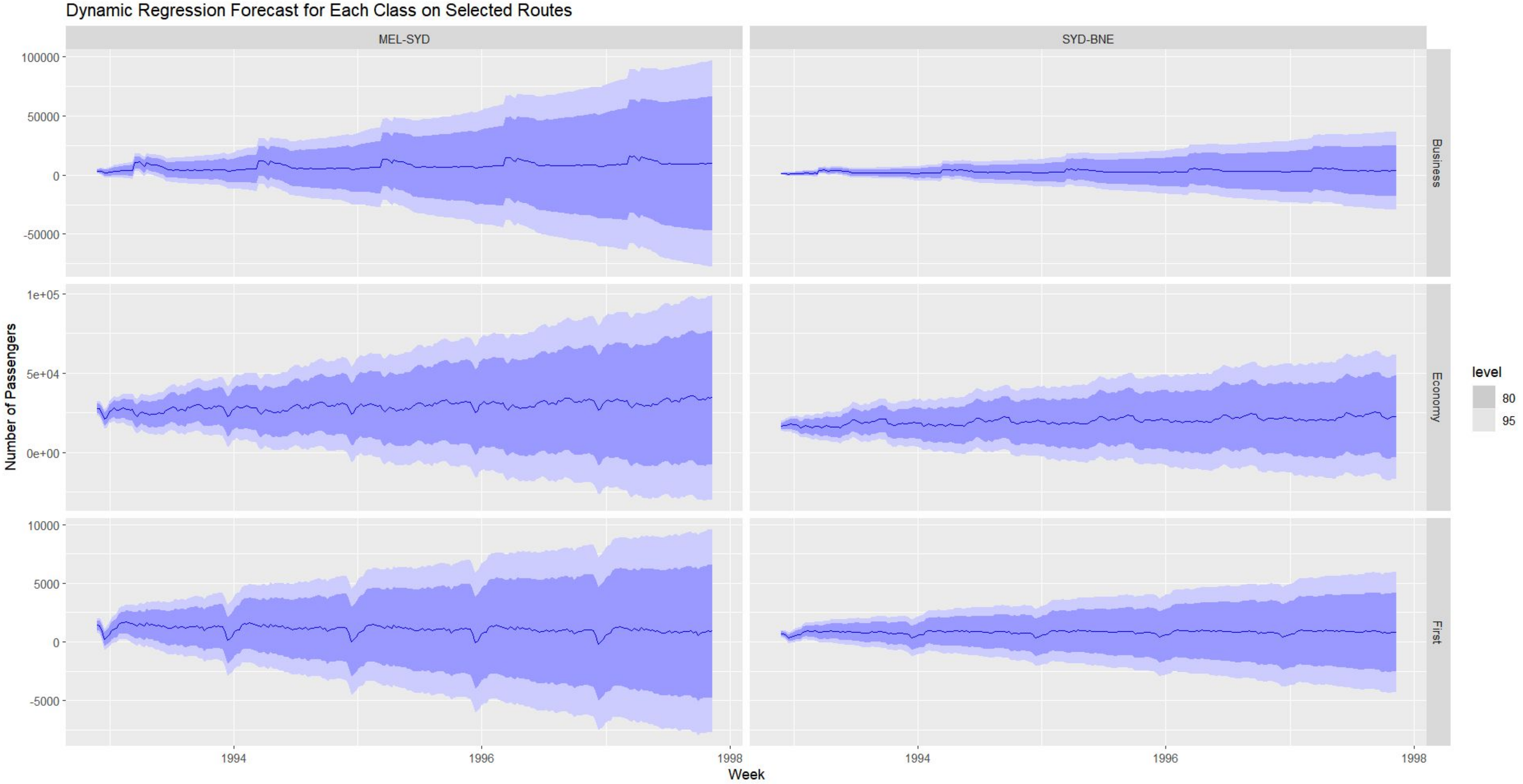
First Year Forecast



Third Year Forecast



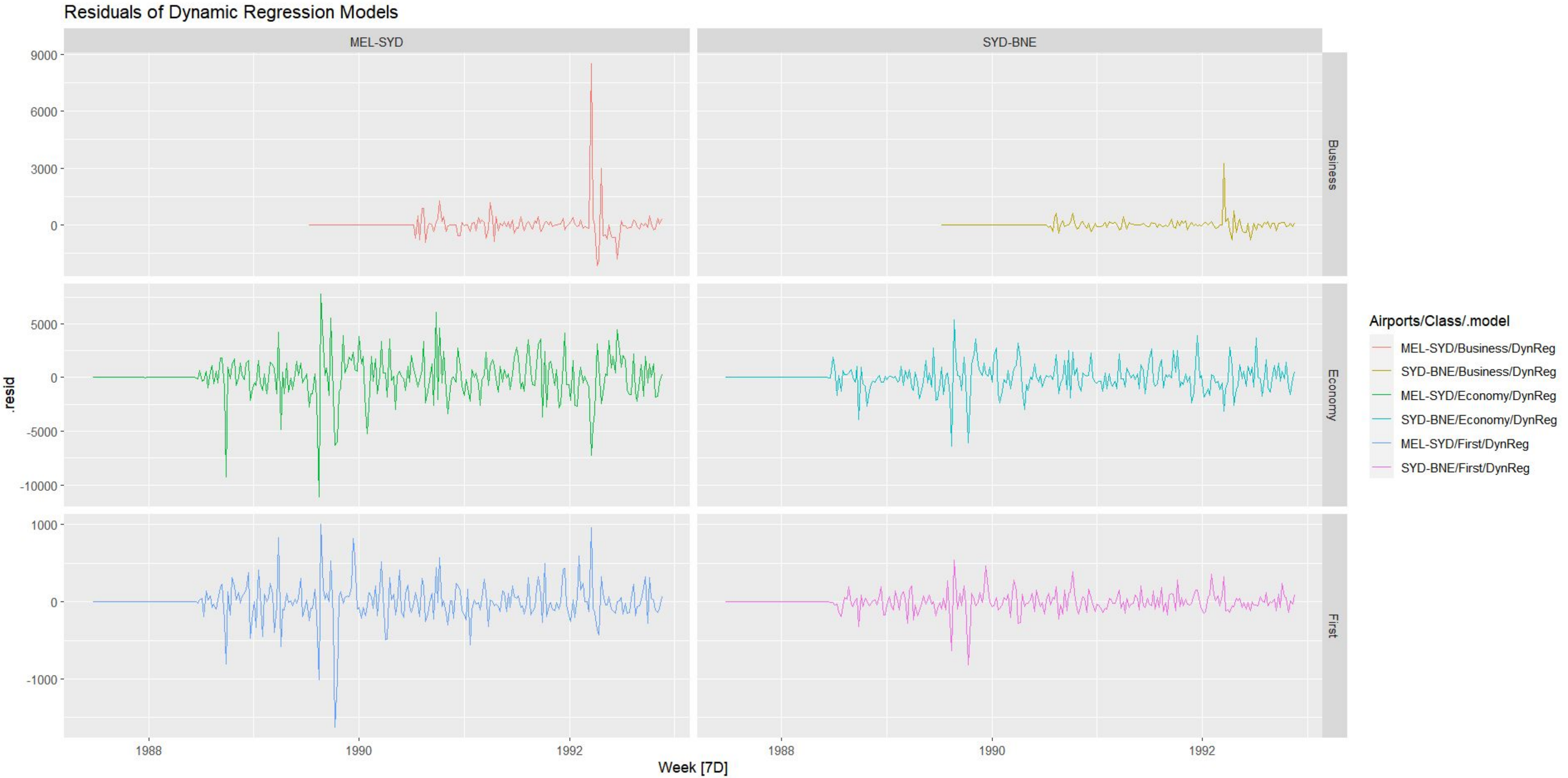
Fifth Year Forecast



Residuals Code

```
# Plot residuals
residuals_data <- residuals(dynamic_models)

autoplot(residuals_data) +
  facet_grid(Class ~ Airports, scales = "free_y") +
  labs(title = "Residuals of Dynamic Regression Models")
```

Residuals



Residuals of Dynamic Regression Models
MEL-SYD
SYD-BNE
Business
Economy
First
.resid
9000
6000
3000
0
5000
0
-5000
-10000
1000
0
-1000
1988
1990
1992
Week [7D]
Airports/Class/.model
MEL-SYD/Business/DynReg
SYD-BNE/Business/DynReg
MEL-SYD/Economy/DynReg
SYD-BNE/Economy/DynReg
MEL-SYD/First/DynReg
SYD-BNE/First/DynReg

```
> augment(dynamic_models) %>%
+   features(.innov, features = guerrero)
# A tibble: 6 × 4
  Airports Class    .model lambda_guerrero
  <chr>    <chr>    <chr>            <dbl>
1 MEL-SYD  Business DynReg           0.446
2 MEL-SYD  Economy  DynReg           0.794
3 MEL-SYD  First    DynReg           0.802
4 SYD-BNE  Business DynReg           0.573
5 SYD-BNE  Economy  DynReg           0.936
6 SYD-BNE  First    DynReg           0.850
```

Guerrero

Actionable Recommendations

- New Marketing Strategies and Operations
  - Adjust number of seats and flights in business class to meet demand for the airports chosen
  - Pricing strategies (during holiday season)
  - Enhance business class services
  - Collecting additional data to depict new trends and changes in consumer behavior

THANK YOU