

Self directed activities 9

Josh Copeland (SID: 1444772)

2025-10-01

#Packages used:

```
# library(tidyverse)
# library(lubridate)
# library(janitor)
# library(forecast)
```

Q1

Return to the retail sales data, using an estimation sample of 2000q1 to 2018q4 and forecast period of 2019q1 to 2019q4. Use the usual trend model with time trend, GFC trend break at 2008.5 and quarterly dummies with AR(3) specification for deviations from this trend.

```
dt <- read.csv("RetailSales.csv")
Retail_m <- ts(dt$Total, start=c(1982,4), end=c(2024,9), frequency=12)
logRetailSales <- window(log(aggregate(Retail_m, nfrequency=4)),
  start=c(2000,1), end=c(2019,4))

# Setting up the X matrix of deterministic regressors
QD <- seasonaldummy(logRetailSales)
Time <- time(logRetailSales)
gfc <- 2008.5
Time_postgfc <- 1*(Time>gfc)*(Time-gfc)
X <- cbind(Time, Time_postgfc, QD)

# Estimation sample
Y <- window(logRetailSales, end=c(2018,4))
t_est <- which(Time<2019)
n <- length(Y)

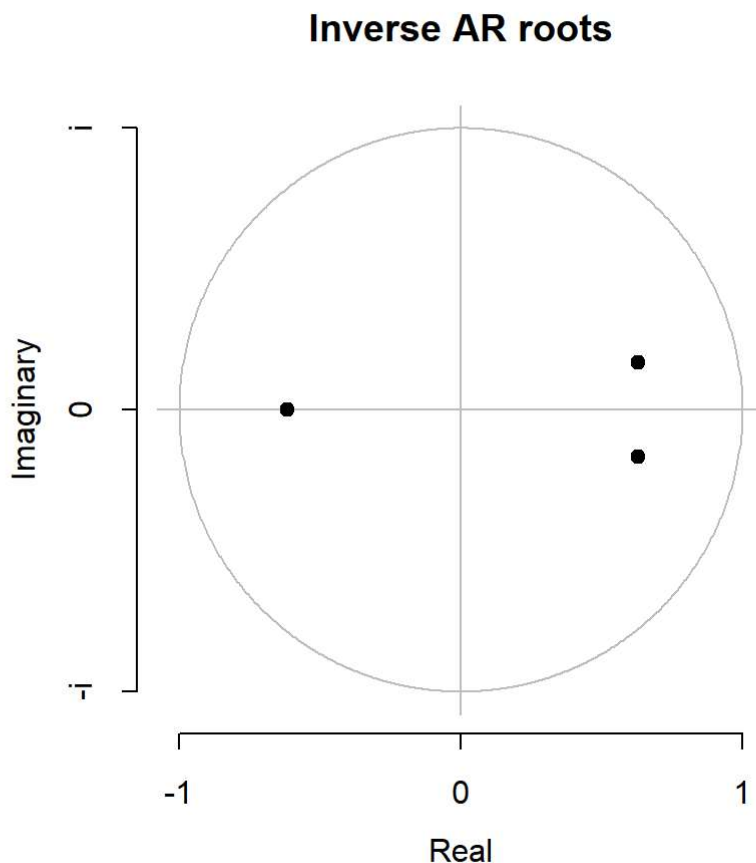
# Forecast sample
Yf <- window(logRetailSales, start=c(2019,1))
t_for <- which(Time>=2019)

eq <- Arima(Y, order=c(3,0,0), xreg=X[t_est,])
```

a) Plot the inverse roots of the estimated model. Comment on these.

The plot below shows the AR model we've estimated is stable and stationary as all its inverse roots are firmly within the unit circle. There is evidence of a strong negative root real root and a complex conjugate pair that introduces cyclical behaviour into the time series.

```
plot(eq)
```

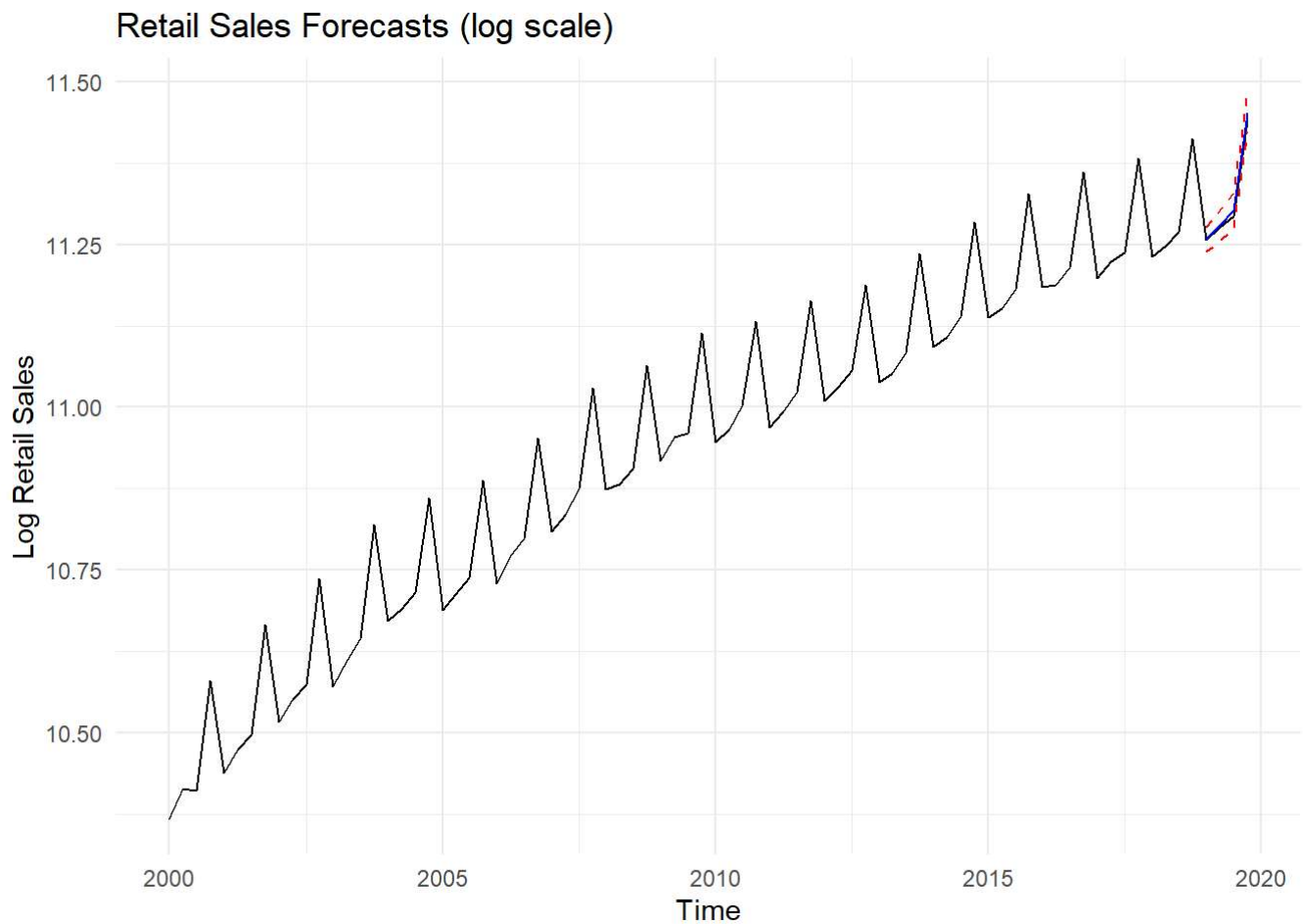


b) Compute the 1,2,3,4-step ahead forecasts from this model, together with 95% prediction intervals based on a normality assumption.

```
eqf <- forecast(eq, xreg=X[t_for,])
Point <- eqf$mean
Lower <- eqf$lower[, "95%"]
Upper <- eqf$upper[, "95%"]

# Turn forecasts into ts objects so they align with logRetailSales
Point_ts <- ts(Point, start=c(2019,1), frequency=4)
Lower_ts <- ts(Lower, start=c(2019,1), frequency=4)
Upper_ts <- ts(Upper, start=c(2019,1), frequency=4)

# Plot
autoplot(logRetailSales) +
  autolayer(Point_ts, series="Forecast", color="blue") +
  autolayer(Lower_ts, series="Lower 95%", linetype="dashed", color="red") +
  autolayer(Upper_ts, series="Upper 95%", linetype="dashed", color="red") +
  ggtitle("Retail Sales Forecasts (log scale)") +
  xlab("Time") + ylab("Log Retail Sales") +
  theme_minimal()
```

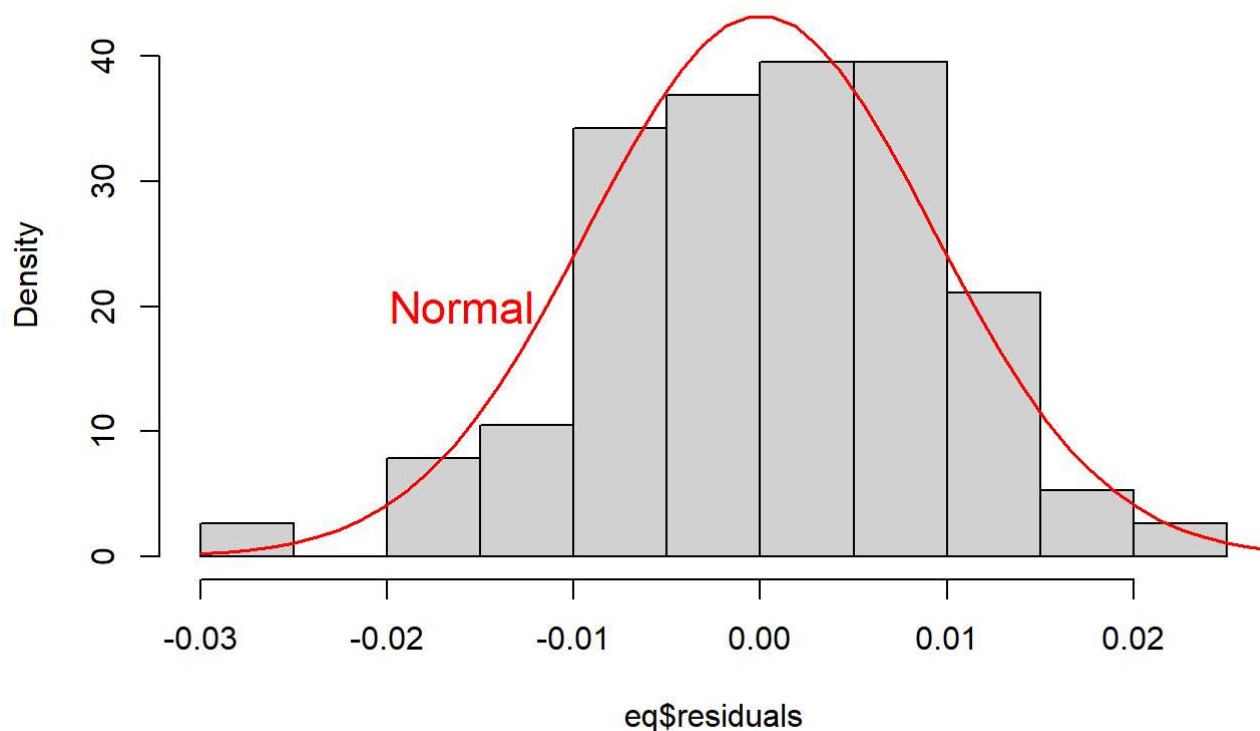


c) Produce a histogram of the residuals. Does normality appear to be a reasonable approximation?

Normality seems to be a reasonable approximation for the distribution of the residuals: the histogram is approximately symmetrical and bell-shaped centered around zero with no significant kurtosis or skewness.

```
hist(eq$residuals, breaks=16, freq=FALSE, ylim=c(0,45))
x <- seq(from=-0.03, to=0.03, length.out=50)
lines(x, dnorm(x, sd=sd(eq$residuals)), col="red", lwd=1.5)
text(x=-0.016, y=20, "Normal", col="red", cex=1.4)
```

Histogram of eq\$residuals

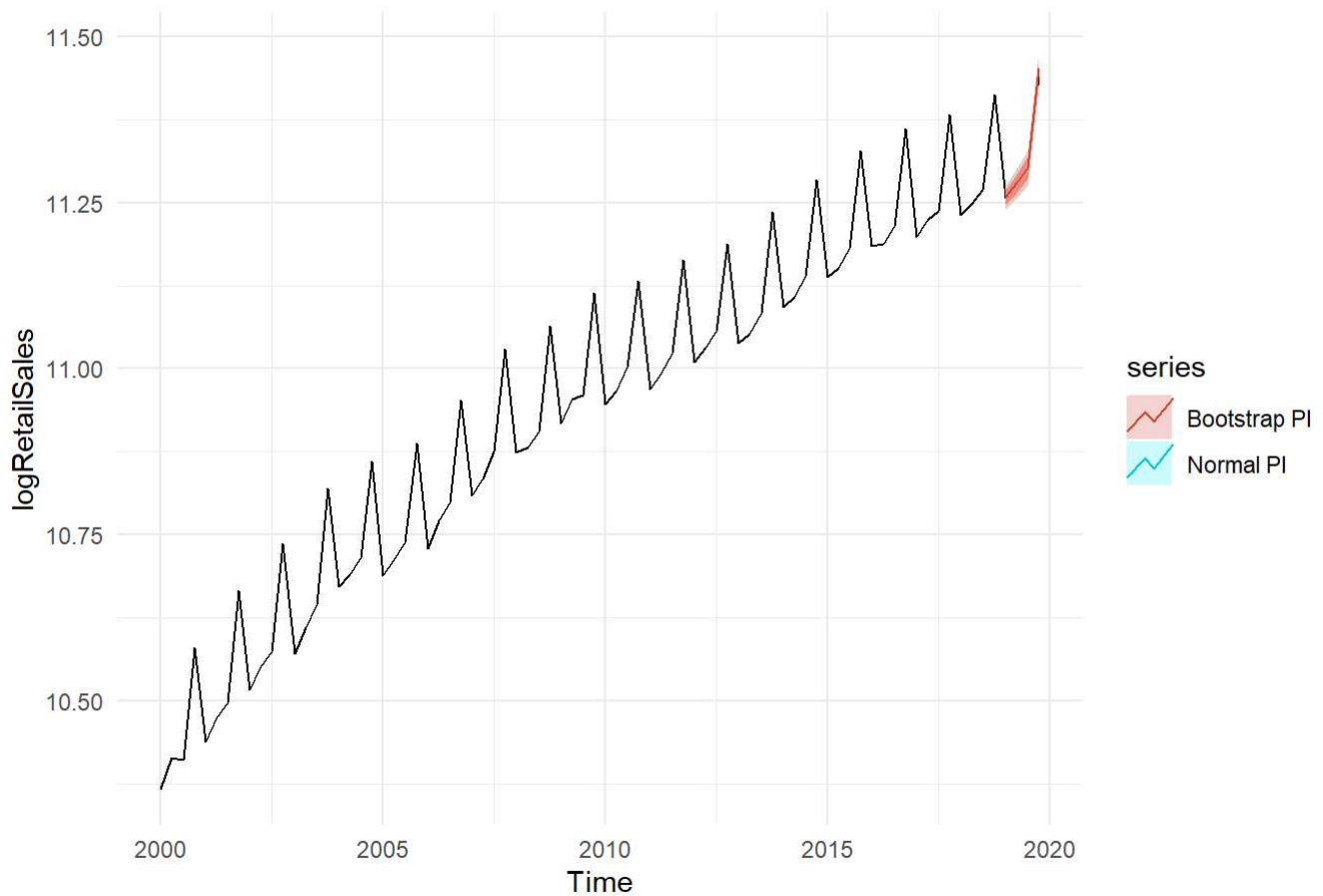


d) Compute bootstrap prediction intervals. Compare with those from part (a), and comment on whether this changes your conclusions

There is very little difference between them, suggesting the normal approximation is indeed reasonable for this dataset.

```
eqfBS <- forecast(eq, xreg=X[t_for,], bootstrap=TRUE)
LowerBS <- eqfBS$lower[, "95%"]
UpperBS <- eqfBS$upper[, "95%"]

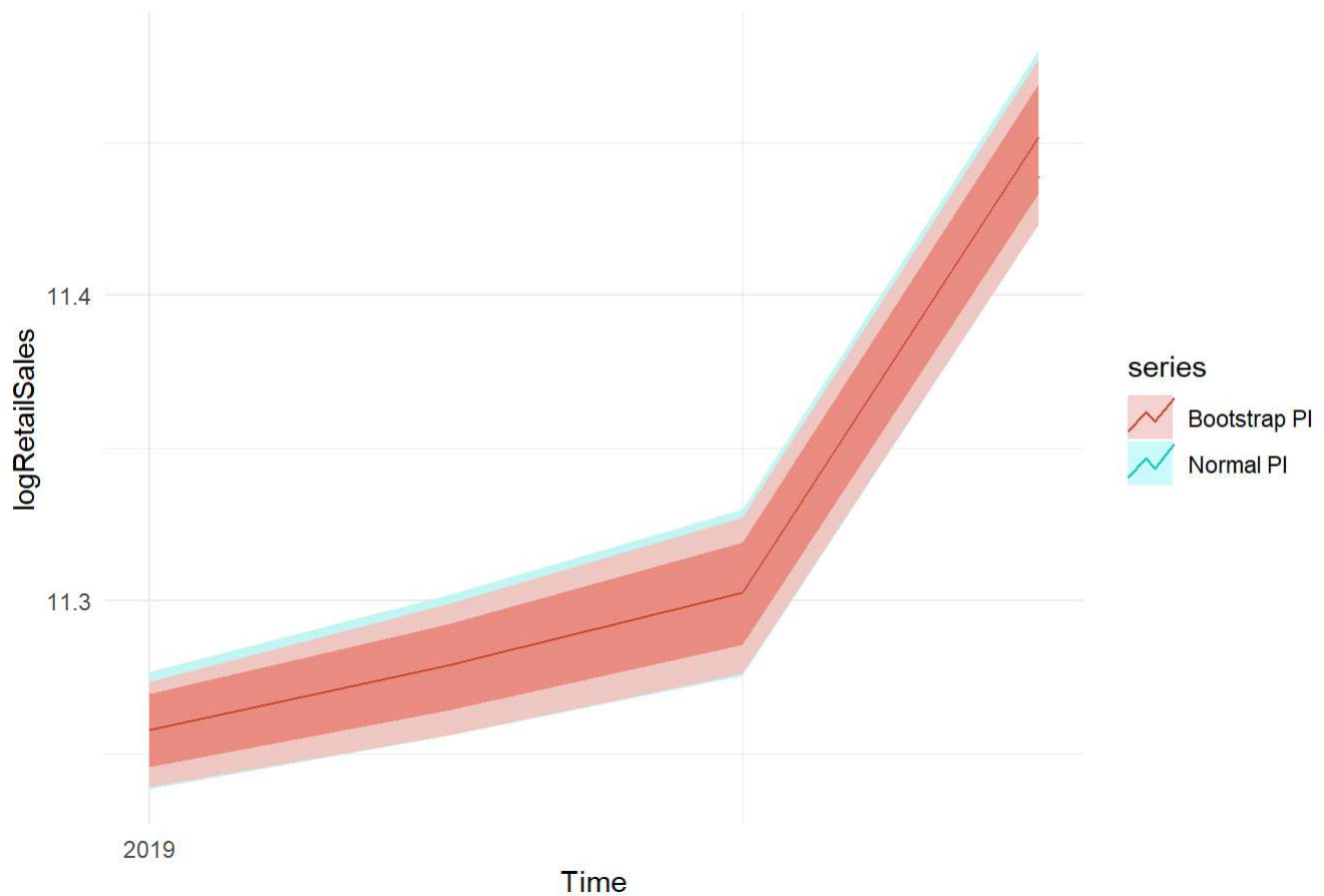
autoplot(logRetailSales) +
  autolayer(eqf, series="Normal PI", PI=TRUE) +
  autolayer(eqfBS, series="Bootstrap PI", PI=TRUE)
```



```
# Last 4 quarters of history
n <- length(logRetailSales)
hist_recent <- window(logRetailSales, start = time(logRetailSales)[n-3])

# Plot restricted history with both sets of intervals
autoplot(hist_recent) +
  autolayer(eqf, series="Normal PI", PI=TRUE) +
  autolayer(eqfBS, series="Bootstrap PI", PI=TRUE) +
  labs(title="Normal vs Bootstrap 95% PIs (last year of history)",
        x="Time", y="logRetailSales") +
  theme_minimal()
```

Normal vs Bootstrap 95% PIs (last year of history)



Q2

The lecture resucrive forecasts/predictions intervals show that regardless of whether you use bootstrap to develop prediction intervals or use a normality assumption, there not that much of a difference between them.

In contrast, comparing direct and recursive forecasts shows the prior tend to have wider intervals. This likely reflects estimating a separate model at each horizon introduces more parameter uncertainty.

```

hmax <- 12

# Reading in data

dt <- read.csv("BAB3mth.csv")
Y <- ts(dt$BAB3, start=c(2010,1), end=c(2024,6), frequency=12)
DY <- diff(Y)
TimeDY <- round(time(DY),3)

## Direct forecasting
ph <- c(3,4,4,4,8,8,8,8,9,10,12,12)

# Storage for forecasts and prediction intervals
Point <- ts(start=c(2024,7), end=c(2025,6), frequency=12)
Lower <- ts(start=c(2024,7), end=c(2025,6), frequency=12)
Upper <- ts(start=c(2024,7), end=c(2025,6), frequency=12)

for (h in 1:hmax){
  # Estimate chosen model for each h
  zeros <- c(rep(0,h-1),
             rep(NA,ph[h]-h+2))
  eq <- Arima(DY,
             order=c(ph[h],0,0),
             fixed=zeros)
  # Forecast h steps ahead
  Point[h] <- forecast(eq, h=h)$mean[h]
  Lower[h] <- Point[h]-1.96*sqrt(eq$sigma2)
  Upper[h] <- Point[h]+1.96*sqrt(eq$sigma2)
}

# AR(3) forecast
eqR <- Arima(DY, order=c(3,0,0))
recf <- forecast(eqR, h=12)

# Chart

dy_dates <- seq(ymd("2010-02-01"), by="month", length.out=length(DY))
fc_dates <- seq(ymd("2024-07-01"), by="month", length.out=hmax)

hist <- data.frame(date = dy_dates, value = as.numeric(DY))
fc <- rbind(
  data.frame(date=fc_dates, mean=as.numeric(Point), lower=as.numeric(Lower), upper=as.numeric(Upper), type="Direct"),
  data.frame(date=fc_dates, mean=as.numeric(recf$mean)[1:hmax], lower=as.numeric(recf$lower[,1])[1:hmax], upper=as.numeric(recf$upper[,1])[1:hmax], type="Recursive")
)

ggplot() +
  geom_line(data=subset(hist, date>=ymd("2023-01-01")), aes(date, value), linewidth=.5, color="grey40") +
  geom_ribbon(data=fc, aes(date, ymin=lower, ymax=upper, fill=type), alpha=.15) +
  geom_line(data=fc, aes(date, mean, color=type), linewidth=1) +
  geom_vline(xintercept=tail(hist$date,1), linetype="dashed", linewidth=.3) +
  labs(x=NULL, y="ΔY", title="Direct vs Recursive Forecasts", subtitle="Shaded areas: 95% int

```

```
ervals", color="Method", fill="Method") +  
  theme_minimal(base_size=12) + theme(axis.text.x=element_text(angle=45, hjust=1), panel.grid.minor=element_blank())
```

Direct vs Recursive Forecasts

Shaded areas: 95% intervals

