Задание 1. Прогнозирование месячных сезонных рядов по SAR(I)MA

Для выполнения задания были взяты данные с предыдущего задание с именем “GBR\_Dairy.tsv”.

Данные были просмотрены на корректность и разделены на два массива для обучения и тестирования наших данных.

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

df <- read.delim("GBR\_Dairy.tsv")

View(df)

h <- 24

train\_len <- length(df$dairy) - h

data <- c(df$dairy[1:train\_len])

test <- c(df$dairy[(train\_len + 1): length(df$dairy)])

plot.ts(data, type = "o", pch = 18, col = "#0000ff")

grid()

Graphical user interface

Description automatically generated

Было построено две модели. Одна модель согласно подобранным параметрам, вторая модель построена с помощью auto.arima(). Обе модели были просмотрены на статистики и для каждой из них был построен график автогрегрессии остатков расчётных значений

Первая модель:

model1 <- auto.arima(train\_ts, seasonal = TRUE)

summary(model1)

print(model1)

acf(resid(model1))

Chart, histogram

Description automatically generated

Text

Description automatically generated

Вторая модель:

model2 <- arima(train\_ts, c(2, 1, 1), list(order = c(2, 1, 1)), period = 12)

summary(model2)

print(model2)

acf(resid(model2))

Text

Description automatically generated

Text

Description automatically generated

Chart, histogram

Description automatically generated

Посмотрев на лаги ещё раз можно сделать предположение, что автоарима лучше подобрала параметры, хотя и параметры второй модели подходят. Смущает лишь, что у второй модели на графике лага приближение некоторых значений ближе к границе по сравнению с первой моделью.

Исходя из вышесказанного для дальнейшей работы была взята первая модель и построен прогноз

pr <- predict(model1, n.ahead = h)

yp <- pr$pred

str(yp)

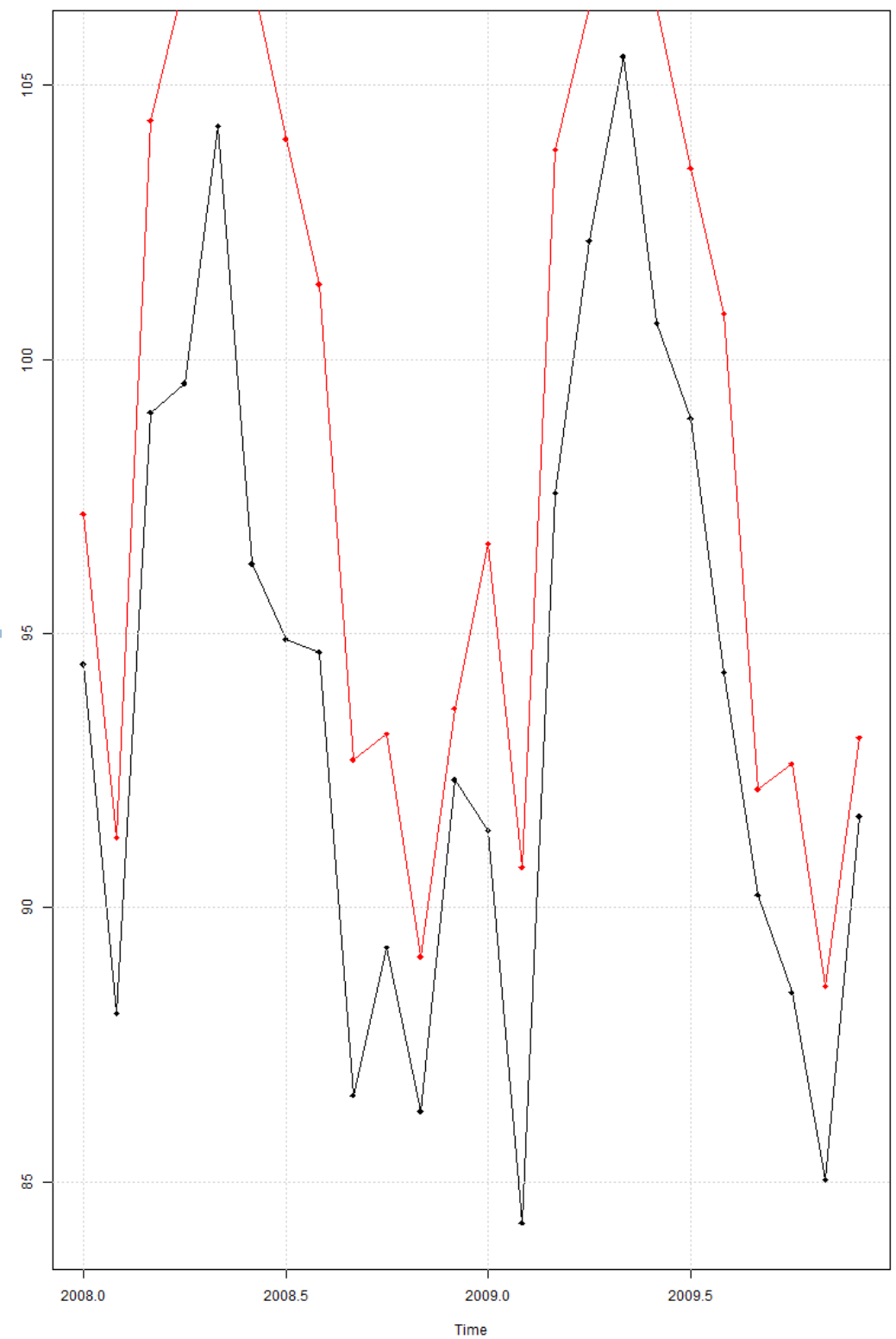
Далее был построен график фактического ряда и точечного прогноза

test\_ts <- ts(data = test, start = c(2008, 1), frequency = 12)

plot(test\_ts, type = "o", pch = 18)

grid()

lines(yp, col = "#ff0000", type = "o", pch = 18)



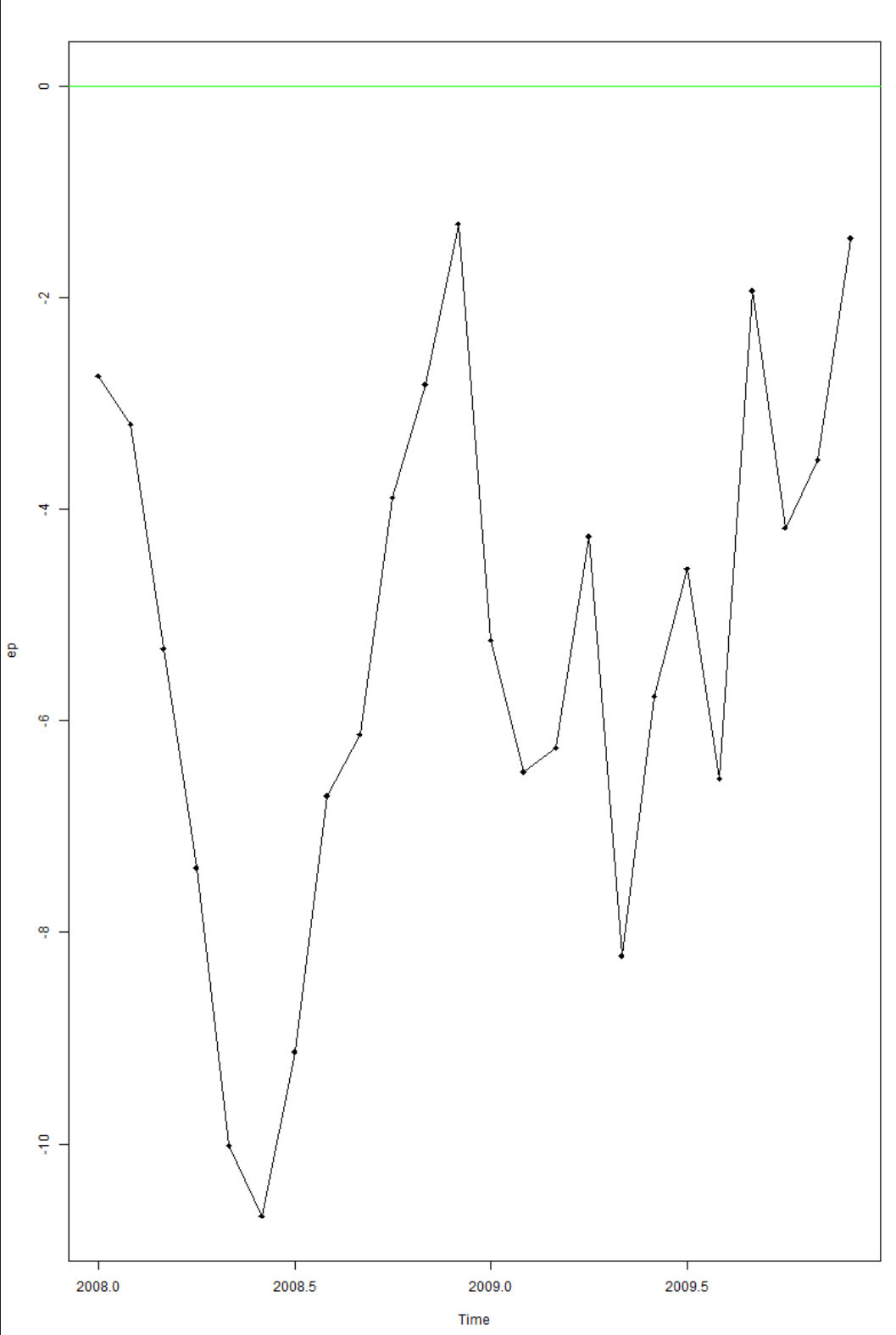
Далее был построен график ошибок прогноза

ep <- test\_ts - yp

plot(ep, type = "o", pch = 18, ylim = range(ep, 0))

abline(h = 0, col = "#00FF00")

где abline – линия нуля(нулевой ошибки)



Теперь к графику фактического ряда и точечного ряда добавляем границы 90%го интервального прогноза.

Построим верхнюю и нижнюю границу интервального прогноза

bottom\_boundary <- pr$pred - pr$se \* qnorm(1 - 0.1 / 2)

top\_boundary <- pr$pred + pr$se \* qnorm(1 - 0.1 / 2)

Построим необходимый график

full\_ts <- ts(data= df$dairy, start = c(1990, 1), frequency = 12)

plot(full\_ts, type = "o", pch = 18, xlim = c(2005, 2010))

grid()

lines(yp, col = "blue", type = "o", pch = 18, lty = 2)

lines(bottom\_boundary, col = "#FF0000", lty = 2, lwd = 2)

lines(top\_boundary, col = "#00FF00", lty = 2, lwd = 2)

