

Uma pequena comparação de prevsiões: ARIMA Vs SSA

Daniel L. P. Esteves dos Reis

Contents

O objetivo deste pequeno relatório é demonstrar o potencial da técnica Singular Spectrum Analysis (SSA). Para tanto, vamos comparar a metodologia SSA com a metodologia ARIMA na capacidade de prever o PIB Agropecuário. Os dados foram obtidos junto ao IPEADATA e estão em frequência trimestral compreendendo o período entre 1996T1 até 2020T1, consistindo em 97 observações. Para um ajuste inicial dos modelos, foram separadas 88 observações, deixando 9 trimestres para fins de previsão fora da amostra.

Os pacotes necessários e os dados utilizados são:

```
library(Rssa)
library(ipeadatar)
library(dplyr)
library(forecast)
library(ggplot2)
library(Metrics)
library(reshape2)
```

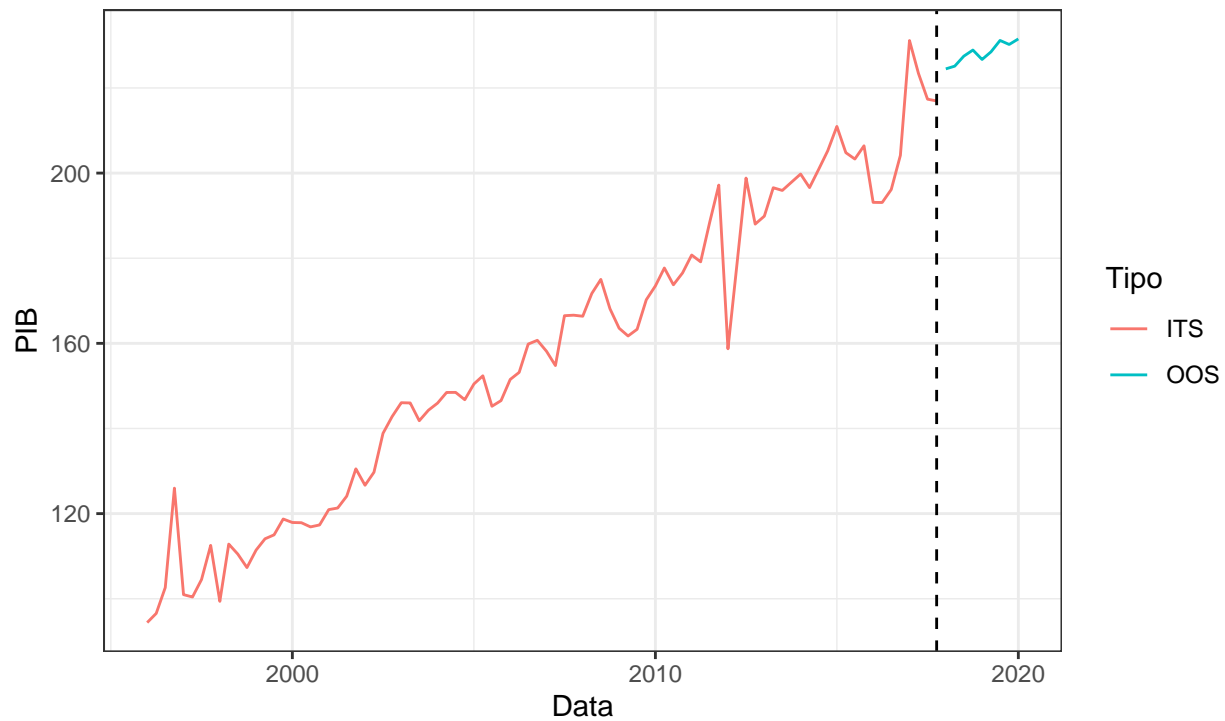
Para a obtenção da série temporal de interesse:

```
.SERIES <- ipeadatar::search_series(terms = "PIB - agropecuária")
.CODE <- .SERIES$code[6]
```

```
MDF <- ipeadatar::ipeadata(code = .CODE)
MDF <- data.frame("Data" = MDF$date, "dPIB" = MDF$value,
                  "Tipo" = c(rep("ITS", 88), rep("OOS", 9)))
```

```
ggplot(data = MDF, mapping = aes(x = Data, y = dPIB, colour = Tipo)) +
  geom_line() +
  geom_vline(xintercept = MDF$Data[88], lty = 2) +
  ylab("PIB") +
  xlab("Data") +
  ggtitle("Produto Interno Bruto (PIB) - Agropecuária (R$ (milhões)) \n1996T1 - 2020T1 (97 Observações)") +
  labs(caption = "Fonte: IPEADATA (SCN104_PIBAGPAS104)") +
  theme_bw()
```

Produto Interno Bruto (PIB) – Agropecuária (R\$ (milhões))
1996T1 – 2020T1 (97 Observações: 88/9)



Fonte: IPEADATA (SCN104_PIBAGPAS104)

Os dados serão rearranjados da seguinte forma:

```
y <- MDF$dPIB
l <- sum(MDF$Tipo == "ITS")
h <- sum(MDF$Tipo == "OOS")

y.arima <- y.ssa <- MDF$dPIB
```

As previsões serão feitas em *rolling window* para um passo à frente, da seguinte forma:

```
for(i in 1:h){
  mod <- forecast::auto.arima(y = y[1:(l+i-1)], stepwise = FALSE)
  y.arima[l+i] <- forecast(mod, h = 1)$mean

  s.pib <- Rssa::ssa(x = y[1:(l+i-1)], kind = "1d-ssa", L = 60)
  f.pib <- predict(s.pib, groups = c(1:6), len = 1)

  y.ssa[l+i] <- sum(unlist(f.pib[1:6]))
}
```

De posse dos valores previstos pelos dois modelos e da série original, pode-se calcular o *Root Mean Squared Errors* (RMSE):

```
e.arima <- rmse(actual = y[1:(l+h)], predicted = y.arima[1:(l+h)])
e.ssa <- rmse(actual = y[1:(l+h)], predicted = y.ssa[1:(l+h)])
```

```
data.frame("ARIMA" = e.arima, "SSA" = e.ssa) %>% round(2) %>% print()
```

```
##   ARIMA  SSA
## 1   3.05 1.63
```

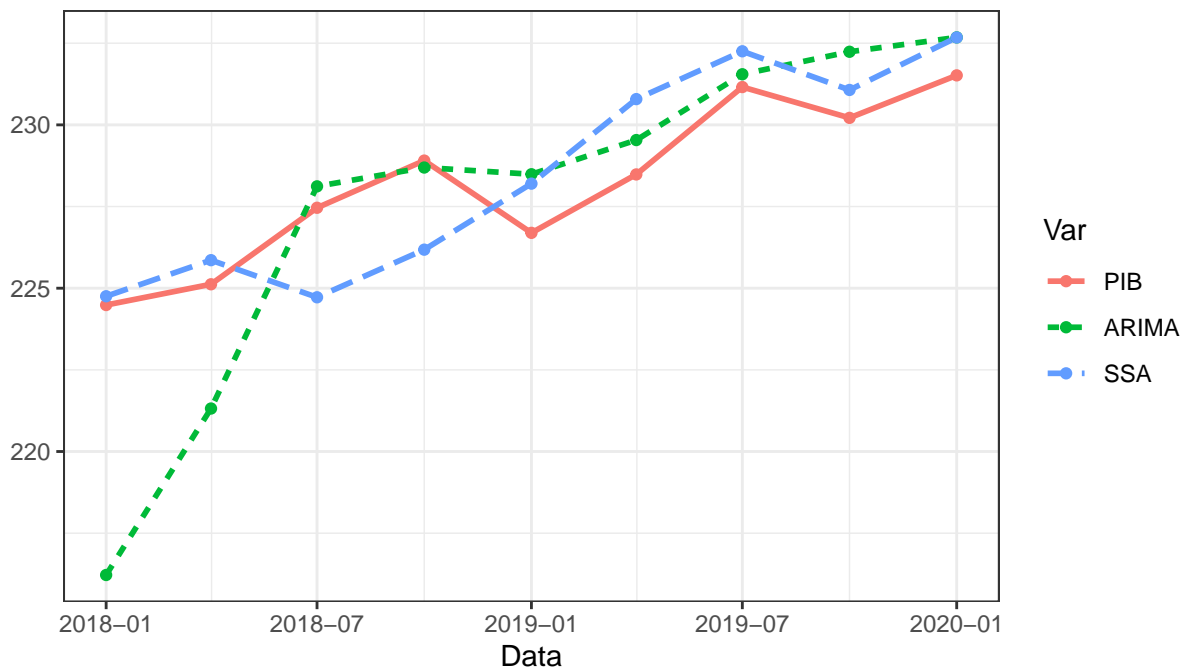
Por fim, vejamos as previsões:

```
sMDF <- data.frame("Data" = tail(MDF$Data,h),
                  "PIB" = tail(y, h),
                  "ARIMA" = tail(y.arima,h),
                  "SSA" = tail(y.ssa,h))

df <- reshape2::melt(data = sMDF, id.vars = "Data")
colnames(df) <- c("Data", "Var", "Val")

ggplot(data = df, mapping = aes(x = Data, y = Val, group = Var)) +
  geom_line(aes(linetype = Var, color = Var), lwd = 1) +
  geom_point(aes(color = Var)) +
  ylab(" ") +
  xlab("Data") +
  ggtitle("PIB Agro. e Previsões ARIMA e SSA \n2018T1 - 2020T1 (9 Observações)") +
  labs(caption = "RMSE ARIMA = 3,05 \nRMSE SSA = 1,62 \nFonte: Elaboração própria.") +
  theme_bw()
```

PIB Agro. e Previsões ARIMA e SSA 2018T1 – 2020T1 (9 Observações)



RMSE ARIMA = 3,05
RMSE SSA = 1,62
Fonte: Elaboração própria.

A análise sugere que as previsões geradas pelo modelo SSA foram mais precisas que aquelas geradas pelo modelo ARIMA, capturando melhor a tendência crescente do PIB Agropecuário brasileiro.