# Seleção de modelos para previsão do ipca mensal

Daniel Alvarez

#Pacotes

Serão utilizados os seguintes pacotes.

## Introdução

A teoria econômica estabelece uma forte relação entre a inflação, no Brasil medida pelo IPCA, desemprego e a taxa basica de júros. Isto posto, propõe-se aqui selecionar um modelo capaz de predizer a variação mensal do IPCA e assim construir alguns cenários baseados na taxa de juros e no desemprego da economia brasileira nos proximos 12 meses.

## IPCA

O primeiro passo é a coleta e o tratamento dos dados, as três variáveis utilizadas aqui seram coletadas através dos pacotes rbcb, criado pelo Banco Central do Brasil e do Ipeadatar, criado pelo IPEA. Inicia-se então, coletando a série do IPCA.

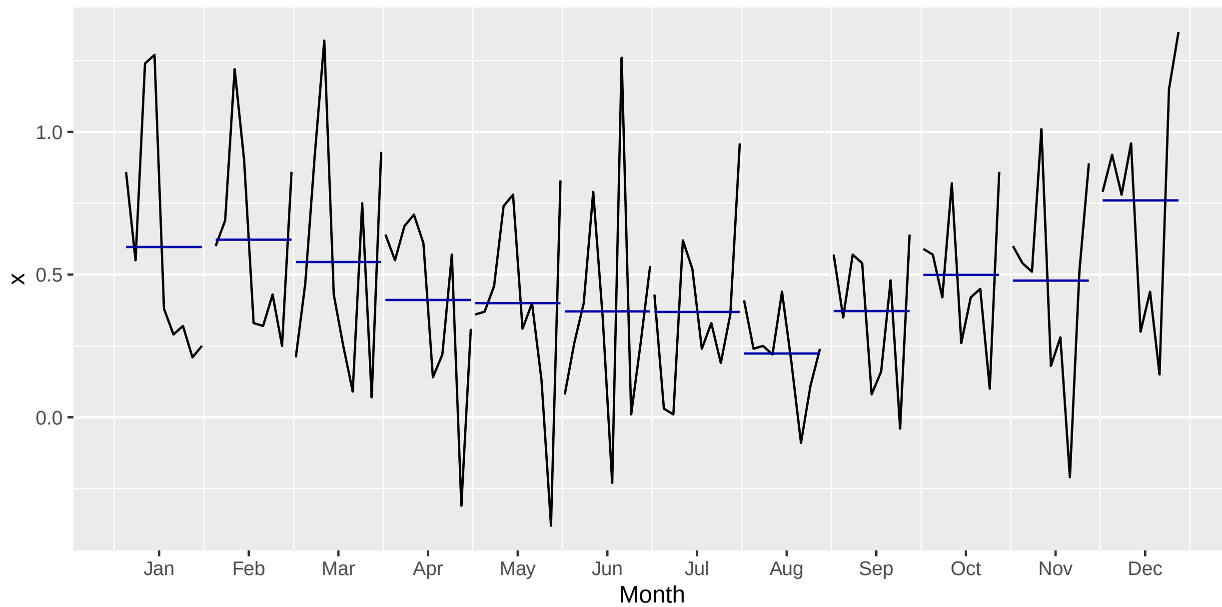
```
ipca <- get_series(433, start_date = as.Date('01/03/2012',format = "%d/%m/%Y"))%>%
mutate(date = yearmonth(date)) %>%
  dplyr::rename(ipca = `433`) %>%
as_tsibble(index=date)
```

O primeiro passo é tratar os dados para que sejam transformados em série temporal.

```
inflacao_mensal = ts(ipca$ipca, start=c(2012,03), freq=12)
```

A primeira análise consiste na detecção de sazonalidade no IPCA. Mediante análise do gráfico nota-se um padrão sazonal bastante pronunciado. E isso será tratado através de dummies sazonais.

```
ggmonthplot(inflacao_mensal)
```




```
dummies <- seasonaldummy(inflacao_mensal)
```

## SELIC

Coleta dos dados da Selic

```
selic <- BETS::BETSget(4189, from ='2012-03-01', frequency = 12)
```

# Desemprego Para o desemprego, será usado a taxa de desocupação aferida pela PNAD contínua.

```
#População desocupada (PNAD-C)
des <- ipeadata("PAN12_TDESOC12")%>%
  select(date,value)%>%
  dplyr::rename(desemprego = value)
```

```
desemprego <- ts(des, start=c(2012,03), freq = 12)
```

```
desemprego_seas <- seas(desemprego[,2])
desemprego  <- desemprego_seas$data[,3]
```

Diferente das outras variáveis apresentadas até aqui, precisamos tratar o desemprego de forma especial. A série é mais curta que as outras duas. Portanto, ao inves de descartar observações no IPCA e na SELIC, optou-se por fazer uma previsão do desemprego para julho/2021.

```
desemprego.forecast <- forecast(auto.arima(desemprego,lambda = 0.5862068 ), h=1, level=95)$mean
x <- data.frame(date = c("2021-07-01"), desemprego_seas= desemprego.forecast)
desemprego.forecast <-as.xts(desemprego.forecast)
desemprego2 <- as.xts(desemprego)
des <- c(desemprego2, desemprego.forecast)
```

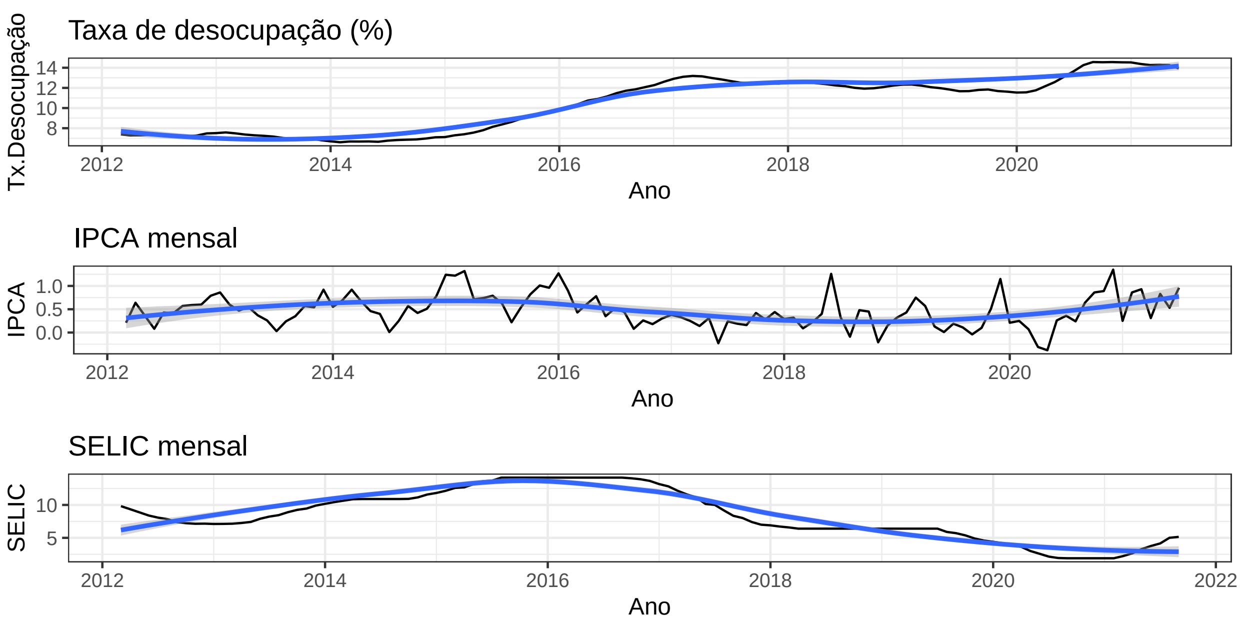
## Breve exploração das séries

```
g1<-autoplot(desemprego)+
  geom_smooth()+
  xlab("Ano")+ylab('Tx.Desocupação')+
  ggtitle("Taxa de desocupação (%)")+
  theme_bw()

g2<-autoplot(inflacao_mensal)+
  geom_smooth()+
  xlab("Ano")+ylab('IPCA')+
  ggtitle("IPCA mensal")+
  theme_bw()

g3<-autoplot(selic)+
  geom_smooth()+
  xlab("Ano")+ylab('SELIC')+
  ggtitle("SELIC mensal")+
  theme_bw()

grid.arrange(g1, g2, g3)
```




Nota-se um comportamento errático na série do ipca, aparentemente sem um componente de tendencia forte. No entanto, para ser possível observar um padrão de tendencia, seria necessário visualizar a séria acumulada nos ultimos 12 meses. Fica pra próxima rs.

Aqui juntamos as três séries em uma uníca base de dados.

```
data <- ts.intersect(inflacao_mensal,des,selic, dummies)
```

## Estimação

Agora iniciamos a parte mais interessante. Não usaremos modelos univariados de séries temporais na presente análise. Vamos deixar as coisas mais interessantes ao incluir nos modelos variáveis exogenas (taxa de juros e desemprego). Mas, devemos começar pelo começo, e para isso devemos começar pela regressão linear.

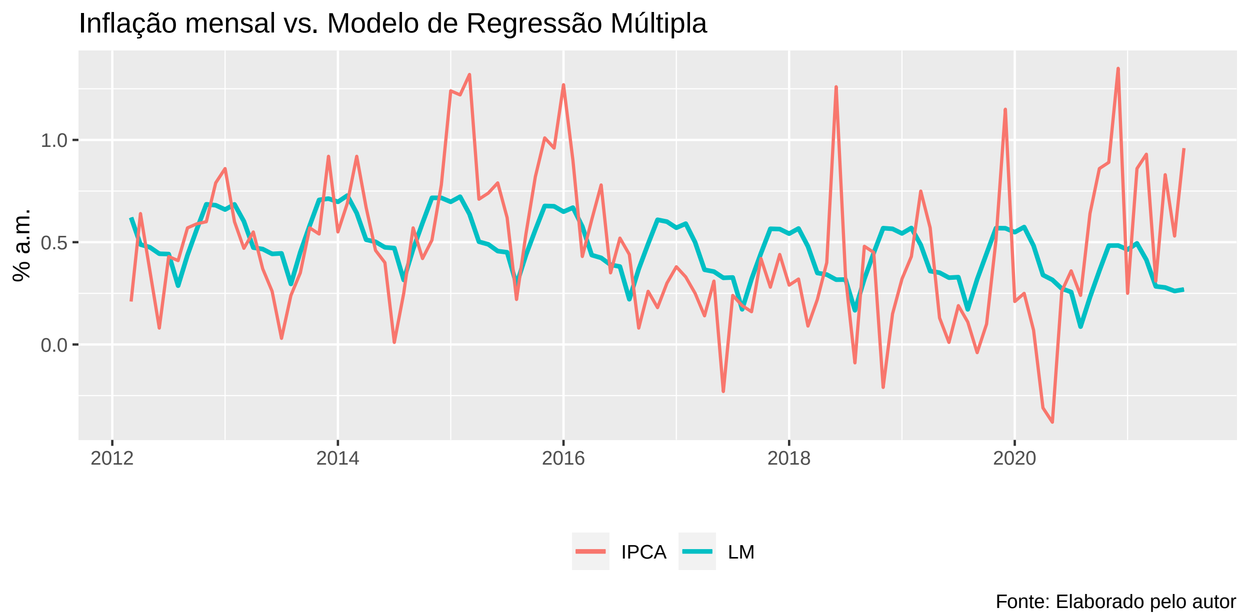
```
lin1 <- tslm(ipca ~ data[,2:13], data = data)
r1<-summary(lin1)
print(xtable(r1), comment=FALSE)
```

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---:|---:|---:|---:|---:|
| (Intercept) | 0.8116 | 0.2091 | 3.88 | 0.0002 |
| data[, 2:13]des | -0.0233 | 0.0135 | -1.73 | 0.0867 |
| data[, 2:13]selic | 0.0061 | 0.0096 | 0.64 | 0.5266 |
| data[, 2:13]dummies.Jan | -0.0203 | 0.1381 | -0.15 | 0.8834 |
| data[, 2:13]dummies.Feb | 0.0068 | 0.1382 | 0.05 | 0.9611 |
| data[, 2:13]dummies.Mar | -0.0786 | 0.1335 | -0.59 | 0.5573 |
| data[, 2:13]dummies.Apr | -0.2103 | 0.1335 | -1.58 | 0.1183 |
| data[, 2:13]dummies.May | -0.2202 | 0.1335 | -1.65 | 0.1020 |
| data[, 2:13]dummies.Jun | -0.2481 | 0.1335 | -1.86 | 0.0660 |
| data[, 2:13]dummies.Jul | -0.2491 | 0.1335 | -1.87 | 0.0649 |
| data[, 2:13]dummies.Aug | -0.4034 | 0.1382 | -2.92 | 0.0043 |
| data[, 2:13]dummies.Sep | -0.2522 | 0.1382 | -1.83 | 0.0709 |
| data[, 2:13]dummies.Oct | -0.1235 | 0.1381 | -0.89 | 0.3734 |

Diferente do que era esperado, a selic não se mostrou estatisticamente significativa para determinar o ipca. Quando o resultado de um refressão segue o lado oposto ao da teoria, é sinal de que algo não está certo. Aqui, no caso, é porque não estamos usando a modelagem correta. Mas vamos dar proseguimento ao processo de forecast por razões didáticas.

### Comparação entre o IPCA e o modelo estimado

```
autoplot(ts(fitted(lin1), start=c(2012,03), freq=12))+
geom_line(size=1, aes(colour='LM'))+
geom_line(aes(y=inflacao_mensal, colour='IPCA'), size=.7)+
xlab('')+ylab('% a.m.')+
labs(title='Inflação mensal vs. Modelo de Regressão Múltipla',
caption='Fonte: Elaborado pelo autor',
colour='')+
theme(legend.position="bottom")
```




Além de não estarmos utilizando o framework metodologico adequado, o com as variáveis que temos em mãos mal mal conseguimos simular as variações do ipca. Quem dirá prevê-las. Para tanto, adicionaremos o úmero índice da produção industrial e o cambio

```
cambio <-BETSget(3697,
                        from = "2012-03-01", frequency = 12)
```

Novamente nos deparamos com o problema de missing value. Desta vez, na série da industria. Para tanto, vamos adotar o mesmo procedimento adotado na construção da série do desemprego.

```
industria <- BETSget(21940,
                     from = "2012-03-01", frequency = 12)
industria.forecast <- forecast(auto.arima(industria,lambda =BoxCox.lambda(industria) ), h=1, level=95)$m
y <- data.frame(date = c("2021-07-01"), industria = industria.forecast)
industria.forecast <-as.xts(industria.forecast)
industria2 <- as.xts(industria)
industria <- c(industria2, industria.forecast)
```

Juntanto no mesmo objeto.

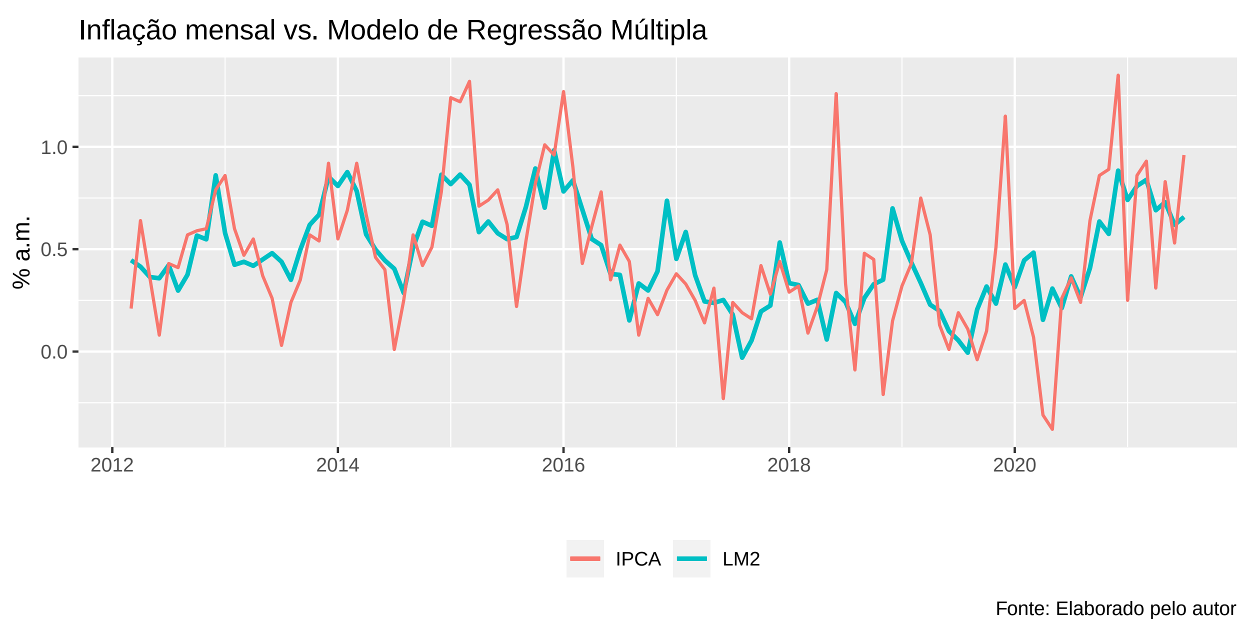
```
data2 <- ts.intersect(inflacao_mensal,des,selic, cambio, industria, dummies)
```

```
lin2 <- tslm(ipca ~ data2[,2:16], data2 = data)
r2<-summary(lin2)
print(xtable(r2), comment=FALSE)
```

Vamos examinar novamente o modelo vs. a série original.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -2.5400 | 0.7331 | -3.46 | 0.0008 |
| data2[, 2:16]des | -0.0432 | 0.0194 | -2.22 | 0.0285 |
| data2[, 2:16]selic | 0.0268 | 0.0088 | 3.06 | 0.0028 |
| data2[, 2:16]cambio | 0.3214 | 0.0542 | 5.93 | 0.0000 |
| data2[, 2:16]industria | 0.0279 | 0.0059 | 4.72 | 0.0000 |
| data2[, 2:16]dummies.Jan | -0.2004 | 0.1348 | -1.49 | 0.1403 |
| data2[, 2:16]dummies.Feb | -0.1607 | 0.1346 | -1.19 | 0.2354 |
| data2[, 2:16]dummies.Mar | -0.2628 | 0.1314 | -2.00 | 0.0482 |
| data2[, 2:16]dummies.Apr | -0.3639 | 0.1311 | -2.78 | 0.0066 |
| data2[, 2:16]dummies.May | -0.3870 | 0.1311 | -2.95 | 0.0040 |
| data2[, 2:16]dummies.Jun | -0.4269 | 0.1313 | -3.25 | 0.0016 |
| data2[, 2:16]dummies.Jul | -0.4423 | 0.1315 | -3.36 | 0.0011 |
| data2[, 2:16]dummies.Aug | -0.5771 | 0.1347 | -4.29 | 0.0000 |
| data2[, 2:16]dummies.Sep | -0.4388 | 0.1347 | -3.26 | 0.0016 |
| data2[, 2:16]dummies.Oct | -0.3047 | 0.1347 | -2.26 | 0.0259 |
| data2[, 2:16]dummies.Nov | -0.3204 | 0.1347 | -2.38 | 0.0193 |

```
autoplot(ts(fitted(lin2), start=c(2012,03), freq=12))+
geom_line(size=1, aes(colour='LM2'))+
geom_line(aes(y=inflacao_mensal, colour='IPCA'), size=.7)+
xlab('')+ylab('% a.m.')+
labs(title='Inflação mensal vs. Modelo de Regressão Múltipla',
caption='Fonte: Elaborado pelo autor',
colour='')+
theme(legend.position="bottom")
```




Embora seja possível notar mais aderencia do modelo ajustado à serie original do ipca, percebe-se que é possível melhora muito. Para tanto, adotaremos o framework correto. Começaremos agora a estimar modelos dentro do escopo de séries temporais. Mas antes vamos interpretar os coeficientes: O desemprego, claramente contribui para a desaceleração do ipca. Tal como, um dolar mais caro contribui para o seu aumento. Chamam atenção o sinal positivo da selic, que deveria impactar negativamente na aceleração da inflação. Mas, devemos considerar o seguinte: estamos analisando a selic e o ipca no mesmo periodo de tempo t1. E, a taxa selic funciona como mecanismo de contenção para a inflação. Ou seja, é aumentada após a detecção de aceleração do ipca. Portanto, o aumento da selic em t1 só causará impacto negativo no ipca futuramente. Quanto ao sinal da produção industrial temos a seguinte hipótese: Estamos produzindo a custos crescentes, o que não é dificil de se eperar no Brasil.

## Modelo Arima.

O primeiro modelo, estimado, será o mais simples possível. Contaremos com a ajuda da função autorima para estimar um modelo univariado. Assim, não precisaremos análisar as funções de autocorrelação e autocorrelação parcial para definir a ordem do modelo. Vamos comparar o modelo estimado com a série original novamente.

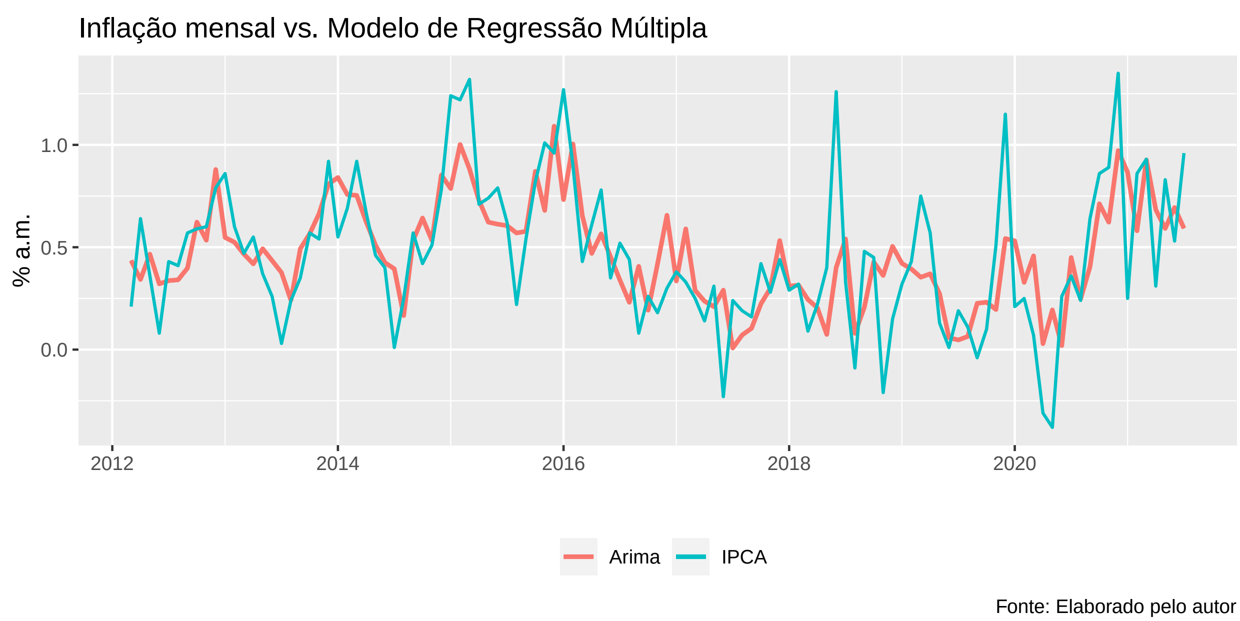
```
arima1<-auto.arima(inflacao_mensal)


comparacao_ts <- cbind(inflacao_mensal, fitted(arima1))
autoplot(comparacao_ts)
```

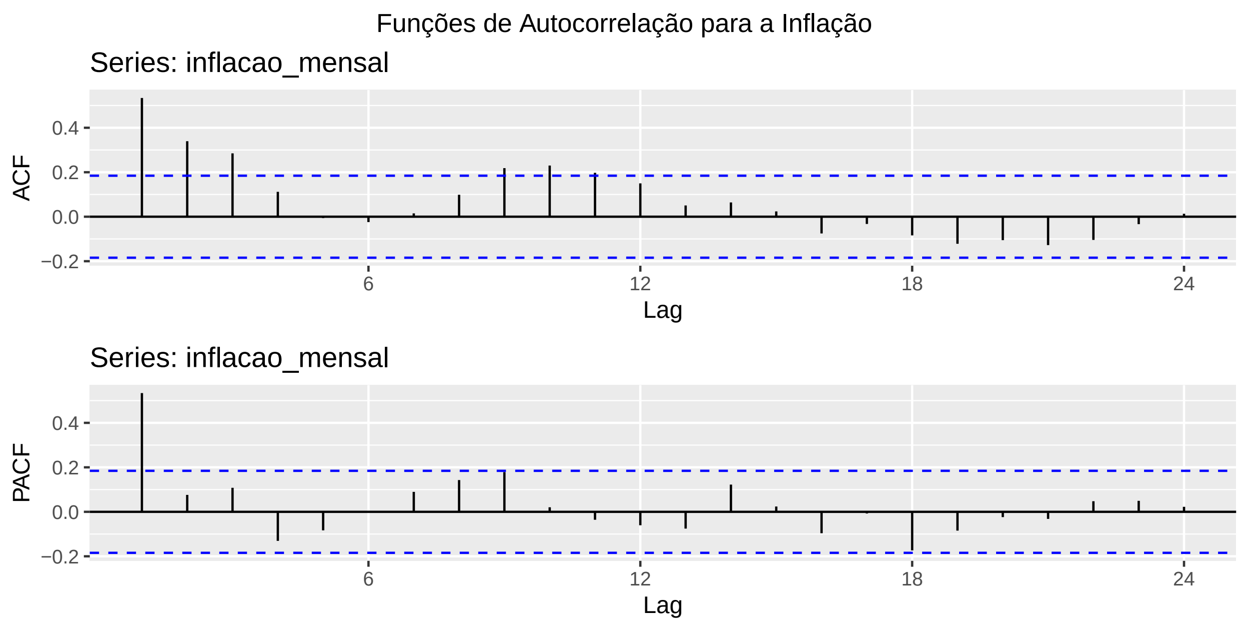

Nota-se uma aderencia muito maior, mas vamos selecionar o modelo sob o qual criaremos cenários usando métricas objetivas a seguir. Antes, vamos brincar mais um pouco e estimar mais alguns modelos. Agora que estamos chegando ao caminho certo, vamos introduzir variáveis exogenas. Aqui vamos, basicamente, combinar o modelo arima com o modelo de regressão linear multípla.

```
arima2 <- auto.arima(data2[,1], xreg = data2[,2:16])
```

```
autoplot(ts(fitted(arima2), start=c(2012,03), freq=12))+
geom_line(size=1, aes(colour='Arima'))+
geom_line(aes(y=inflacao_mensal, colour='IPCA'), size=.7)+
xlab('')+ylab('% a.m.')+
labs(title='Inflação mensal vs. Modelo de Regressão Múltipla',
caption='Fonte: Elaborado pelo autor',
colour='')+
theme(legend.position="bottom")
```




```
a1<-ggAcf(inflacao_mensal)
a2<-ggPacf(inflacao_mensal)
grid.arrange(a1, a2,
            top = "Funções de Autocorrelação para a Inflação",
            layout_matrix = matrix(c(1,1,2,2),
                                   ncol=2, byrow=TRUE))
```



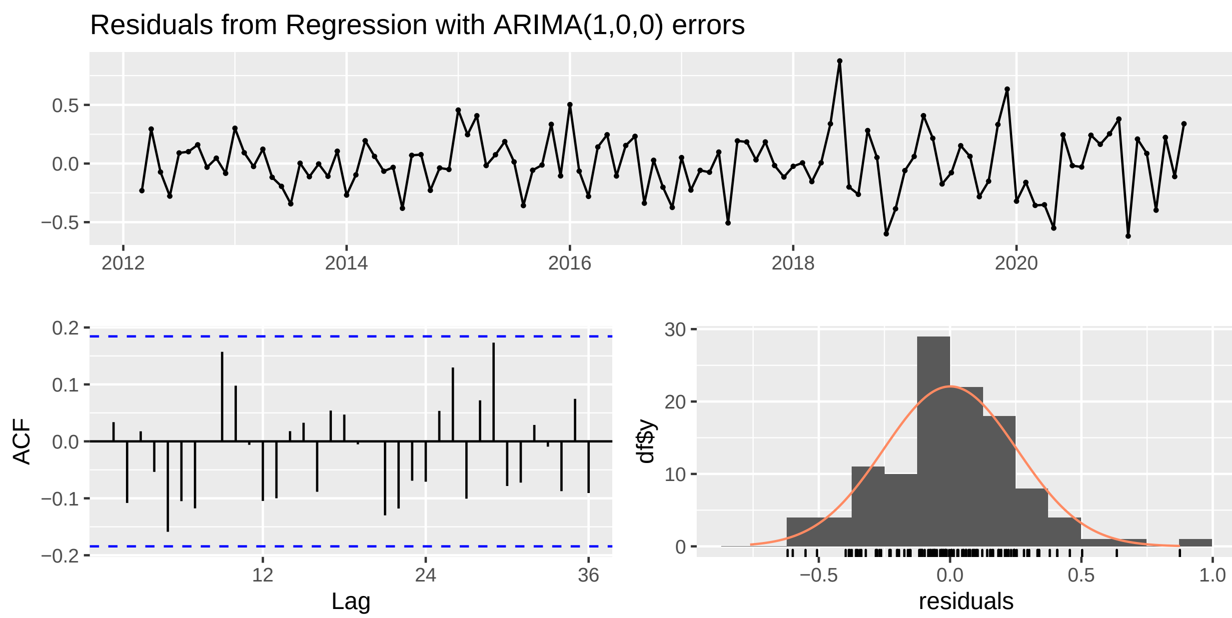

Baseado nas funções de ACF e PACF vamos ambandonar o modelo automatizado e vamos modelar do jeito roots! rs.

```
arima3 <- Arima(data2[,1], xreg = data2[,2:16], order = c(1,0,0))
```

## Diagnóstico dos erros.

Vamos aproveitar e examinar os residuos do nosso modelo.

```
checkresiduals(arima3)
```

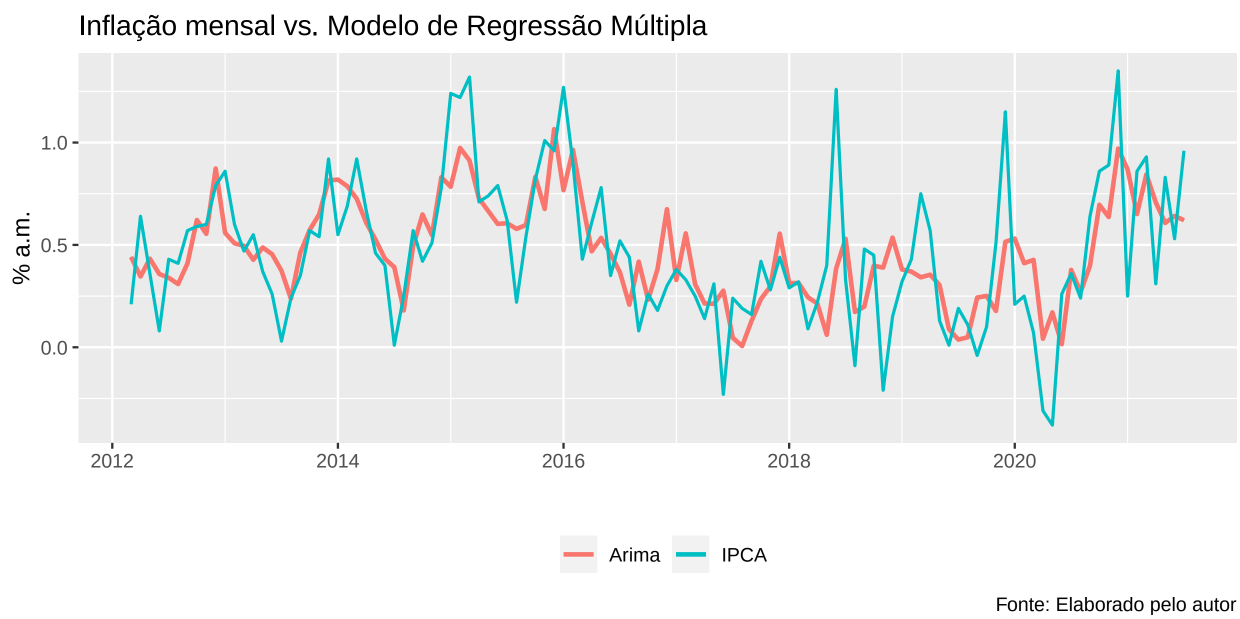


```
Ljung-Box test
```

data: Residuals from Regression with ARIMA(1,0,0) errors Q* = 21.955, df = 6, p-value = 0.001234

Model df: 17. Total lags used: 23

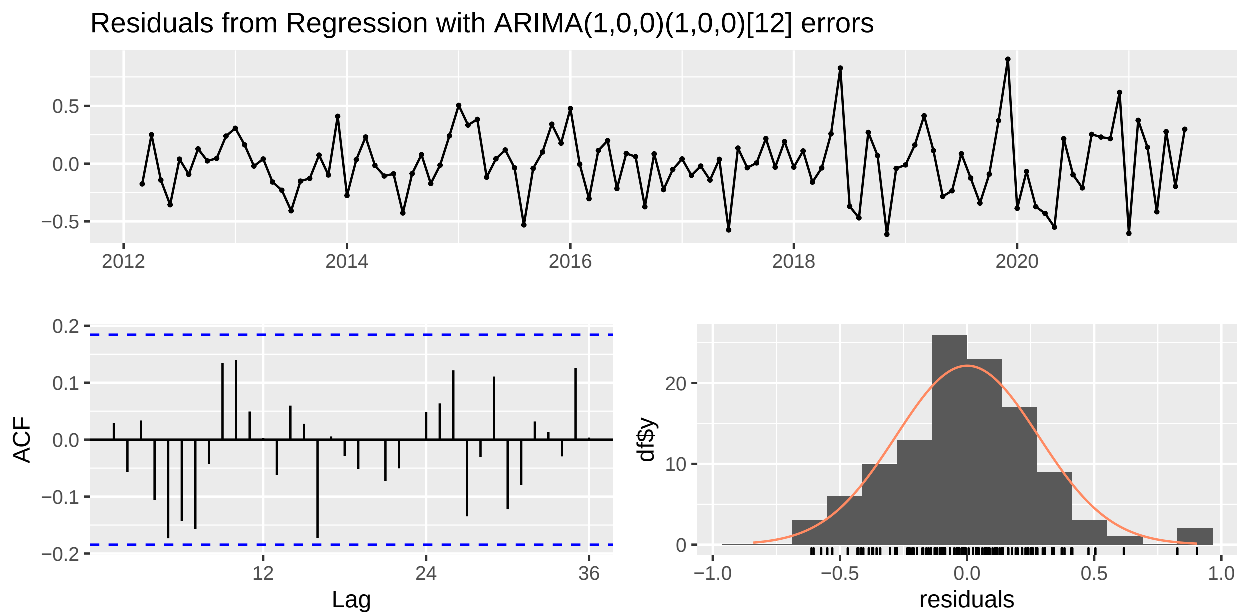
```
autoplot(ts(fitted(arima3), start=c(2012,03), freq=12))+
geom_line(size=1, aes(colour='Arima'))+
geom_line(aes(y=inflacao_mensal, colour='IPCA'), size=.7)+
xlab('')+ylab('% a.m.')+
labs(title='Inflação mensal vs. Modelo de Regressão Múltipla',
caption='Fonte: Elaborado pelo autor',
colour='')+
theme(legend.position="bottom")
```




Podemos ver muito mais aderencia a série original. No entanto, o teste Ljung-Box test sugere que nosso erros são correlacionados. Por isso vamos tentar uma abordagem diferente e modelar a sazonalidade.

```
data2 <- ts.intersect(inflacao_mensal,des,selic, cambio,industria)
```

```
arima4 <- Arima(data2[,1], xreg = data2[,2:5], order = c(1,0,0), seasonal = c(1,0,0) )
```



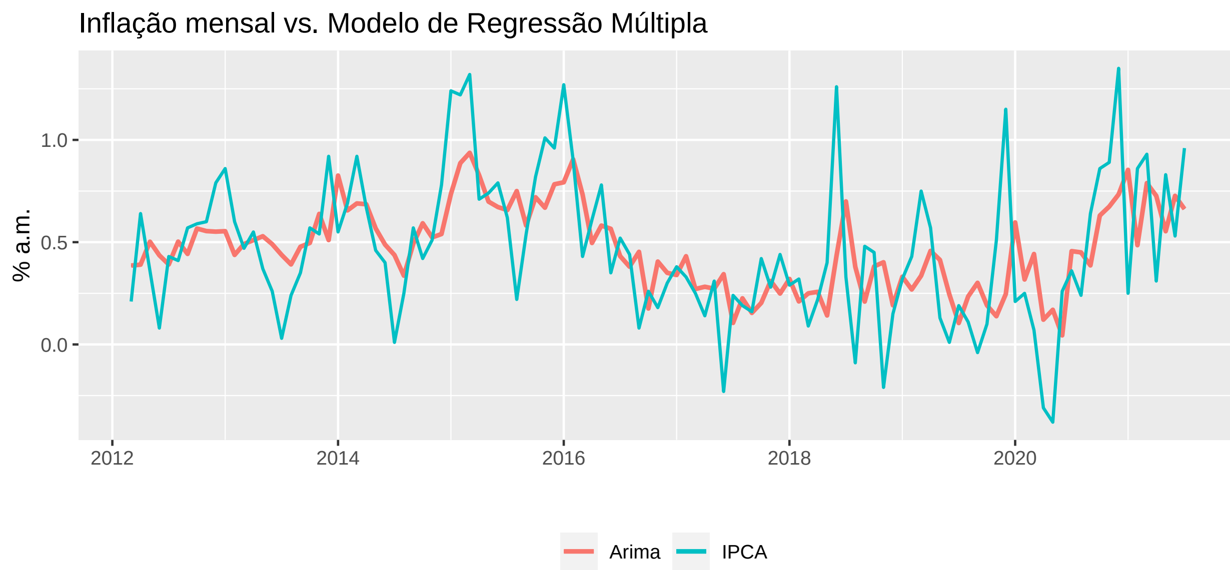

```
Ljung-Box test
```

data: Residuals from Regression with ARIMA(1,0,0)(1,0,0)[12] errors Q* = 23.031, df = 16, p-value = 0.1129

Model df: 7. Total lags used: 23

Bom, conseguimos enxergar que os erros seguem uma distribuição normal e se assemelham a um ruído branco e aceitamos a hiótese nula de que os erros não são correlacionados. Acredito que tenhamos um forte candidato em mãos.

```
autoplot(ts(fitted(arima4), start=c(2012,03), freq=12))+
geom_line(size=1, aes(colour='Arima'))+
geom_line(aes(y=inflacao_mensal, colour='IPCA'), size=.7)+
xlab('')+ylab('% a.m.')+
labs(title='Inflação mensal vs. Modelo de Regressão Múltipla',
caption='Fonte: Elaborado pelo autor',
colour='')+
theme(legend.position="bottom")
```

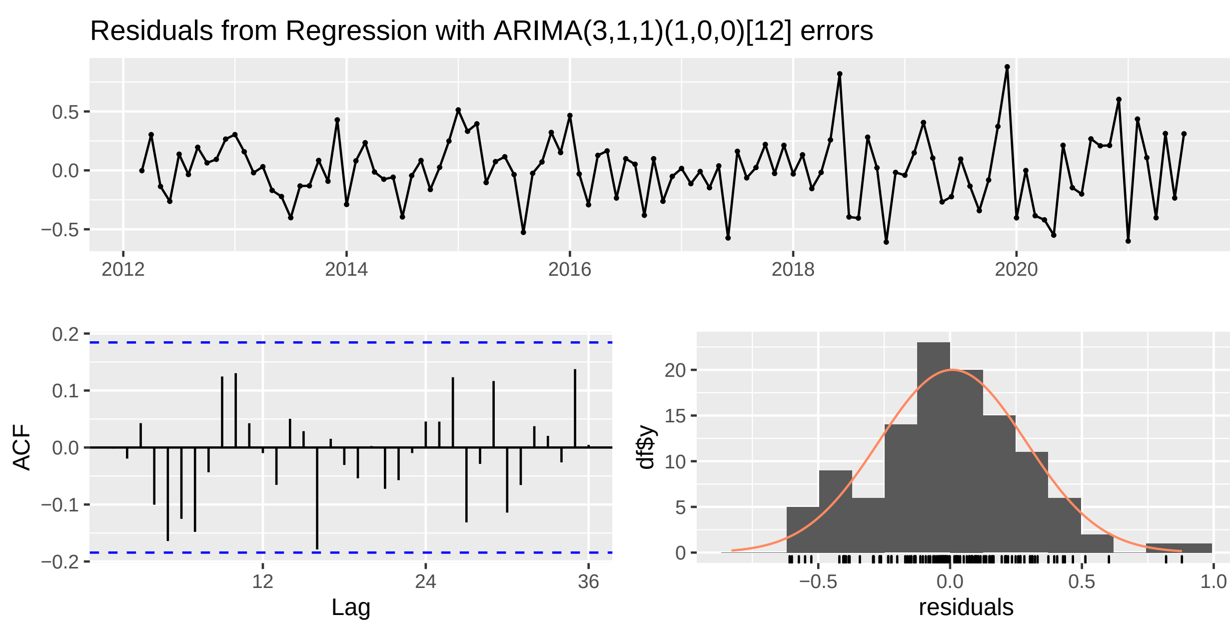



Fonte: Elaborado pelo autor

Vamos tentar um modelo menos parcimonioso agora

```
arima5 <- Arima(data2[,1], xreg = data2[,2:5], order = c(3,1,1), seasonal = c(1,0,0) )
```

```
checkresiduals(arima5)
```

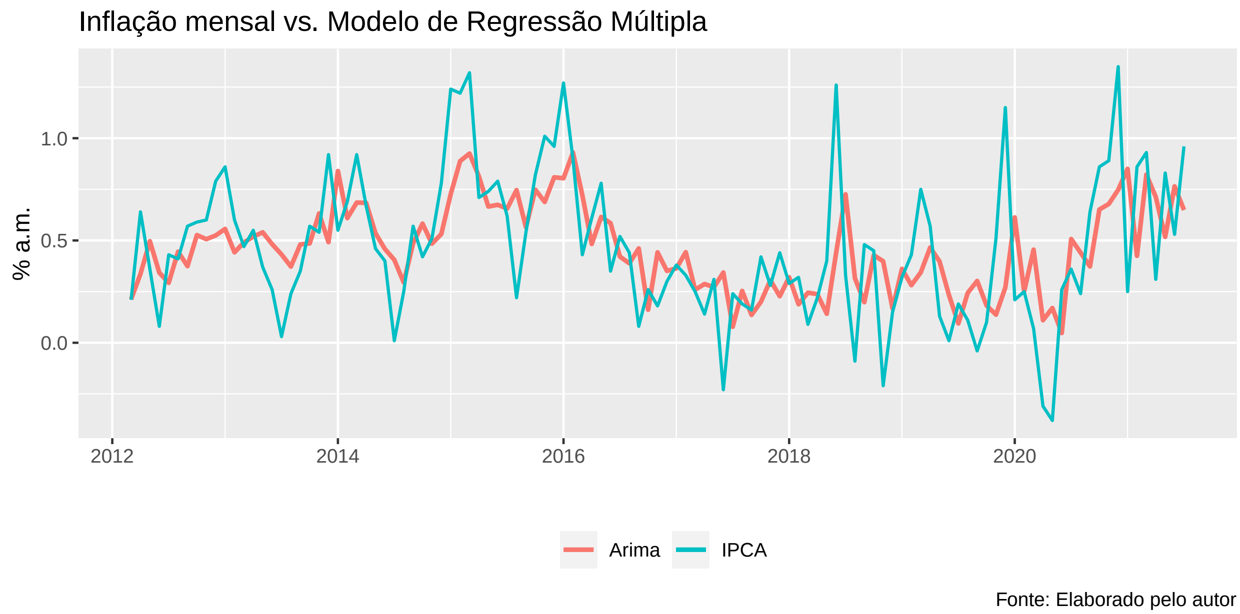



```
Ljung-Box test
```

data: Residuals from Regression with ARIMA(3,1,1)(1,0,0)[12] errors Q* = 20.924, df = 14, p-value = 0.1036

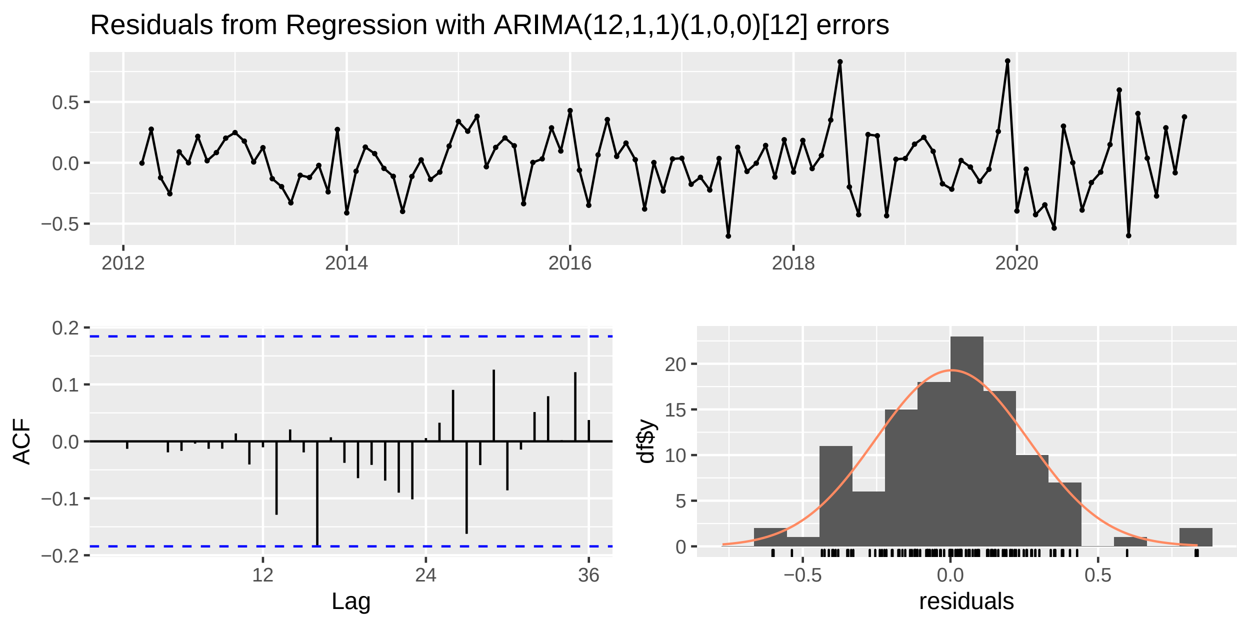
Model df: 9. Total lags used: 23

```
autoplot(ts(fitted(arima5), start=c(2012,03), freq=12))+
geom_line(size=1, aes(colour='Arima'))+
geom_line(aes(y=inflacao_mensal, colour='IPCA'), size=.7)+
xlab('')+ylab('% a.m.')+
labs(title='Inflação mensal vs. Modelo de Regressão Múltipla',
caption='Fonte: Elaborado pelo autor',
colour='')+
theme(legend.position="bottom")
```




```
arima6 <- Arima(data2[,1], xreg = data2[,2:5], order = c(12,1,1), seasonal = c(1,0,0) )
```

```
checkresiduals(arima6)
```

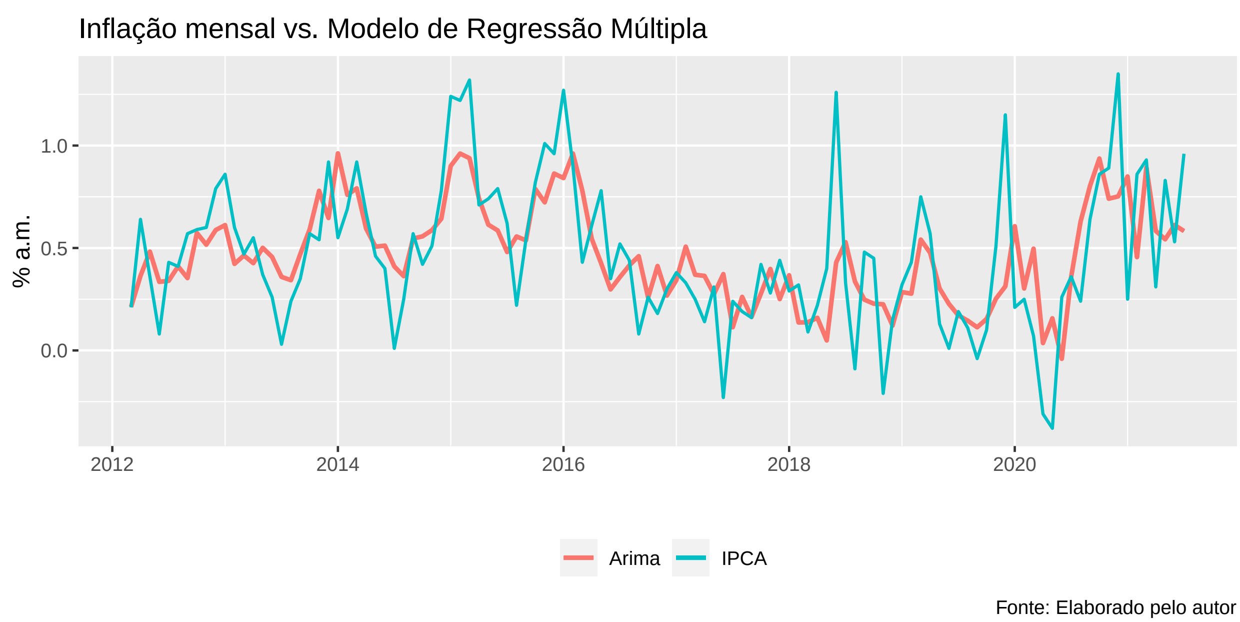


```
Ljung-Box test
```

data: Residuals from Regression with ARIMA(12,1,1)(1,0,0)[12] errors Q* = 11.472, df = 5, p-value = 0.04279

Model df: 18. Total lags used: 23

```
autoplot(ts(fitted(arima6), start=c(2012,03), freq=12))+
geom_line(size=1, aes(colour='Arima'))+
geom_line(aes(y=inflacao_mensal, colour='IPCA'), size=.7)+
xlab('')+ylab('% a.m.')+
labs(title='Inflação mensal vs. Modelo de Regressão Múltipla',
caption='Fonte: Elaborado pelo autor',
colour='')+
theme(legend.position="bottom")
```




## Selecionar o melhor modelo

```
acc1 = accuracy(arima1,  test = inflacao_mensal)
acc2 = accuracy(arima2,  test = inflacao_mensal)
acc3 = accuracy(arima3,  test = inflacao_mensal)
acc4 = accuracy(arima4,  test = inflacao_mensal)
acc5 = accuracy(arima5,  test = inflacao_mensal)
acc6 = accuracy(arima6,  test = inflacao_mensal)
```

```
print(xtable(acc1), comment=FALSE, type = "latex")
```

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | -0.22 | 0.22 | 0.22 | -104.40 | 104.40 | | |

```
print(xtable(acc2), comment=FALSE, type = "latex")
```

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | -0.23 | 0.23 | 0.23 | -107.29 | 107.29 | | |

```
print(xtable(acc4), comment=FALSE, type = "latex")
```

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | -0.18 | 0.18 | 0.18 | -83.69 | 83.69 | | |

```
print(xtable(acc5), comment=FALSE, type = "latex")
```

```
print(xtable(acc6), comment=FALSE, type = "latex")
```

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | -0.00 | 0.00 | 0.00 | -1.22 | 1.22 | | |

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|---|
| Training set | -0.00 | 0.00 | 0.00 | -1.38 | 1.38 | | |

## Fazendo a projeção do IPCA

Vamos selecionar o modelo base no RMSE. Para ser mais exato vamos escolher o modelo com o menor RMSE. A grande questão aqui é que os modelos arima5 e arima6 possuem um RMSE muito pequeno. Dessa forma, vamos selecionar o modelo mais parcimonioso. Ou seja, vamos construir cenários com base no modelo arima5.

Como se trata da criação de um cenário, precisamos, obviamente, criar um cenário. Aqui é onde colocamos a imaginação para funcionar.

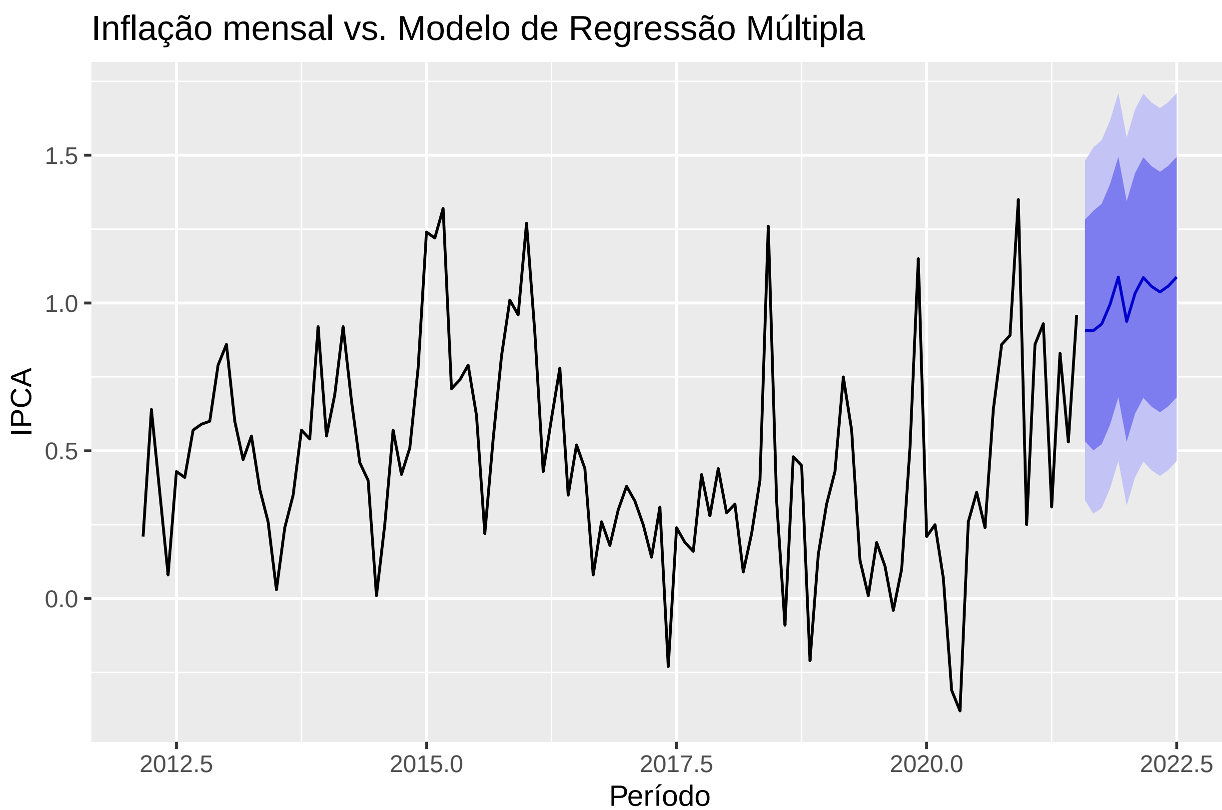
Por se tratar de um exercício informal, podemos adotar as premissas que quisermos. No entanto, no ambiente profissional, esse processo é de grande importância. Para tanto, um bom método de definição de premissas é o método delfi. Para mais informações dá um google aí.

Assim, a seguir criamos a matriz de cenários com base nas seguites premissas: 1 - Queda na taxa de desemprego 2 - Selic chegando a 7,5 pp 3 - Cambio a 5,20 4 - Indice de produção industrial mantendo-se em cerca de 90 pontos;

```
c1 <- data.frame(des=c(13,12.5,11,10,9,9.75,9.5,9.25,9.125,9,8.5,7),
                 selic = c(5.01,5.15,5.25,5.5,6,7,7,7,7,7,7,7),
                 cambio = c(5.2511,5.22,5.10,5.15,5.20,5.20,5.20,5.20,5.20,5.20,5.20,5.20),
                 industria = c(89,90,90,90,89,88,89,91,92,89,90,87))
c1<- as.matrix(c1)
```

```
f_arima <- forecast(arima5, xreg = c1, h = 12)
```

```
autoplot(f_arima)+
  xlab("Período") + ylab("IPCA")+
  labs(title='Inflação mensal vs. Modelo de Regressão Múltipla',
  caption='Fonte: Elaborado pelo autor',
colour='')
```




Fonte: Elaborado pelo autor

```
theme(legend.position="bottom")
```

```
## List of 1
##  $ legend.position: chr "bottom"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

```
f_arima$mean
```

```
##            Jan       Feb       Mar       Apr       May       Jun       Jul
## 2021
## 2022 0.9375720 1.0323410 1.0861160 1.0558177 1.0372872 1.0577435 1.0879000
##            Aug       Sep       Oct       Nov       Dec
## 2021 0.9072292 0.9068081 0.9289647 0.9951340 1.0880282
## 2022
```