

Prova Prática I  
Universidade Federal do Ceará  
Centro de Ciências  
Departamento de Estatística e Matemática Aplicada  
Bacharelado em Estatística  
Análise de Séries Temporais

Antônio Arthur Silva de Lima

30 de agosto de 2024

## Sumário

Série . . . . .	2
Modelagem . . . . .	3
Primeiro Modelo . . . . .	3
Segundo modelo . . . . .	7
Terceiro Modelo . . . . .	10
Modelo Final . . . . .	13

## Série

Primeiro, vamos entender o comportamento da série de **Recursos Federais - SUS** através do seu gráfico, e tirar nossas primeiras conclusões.

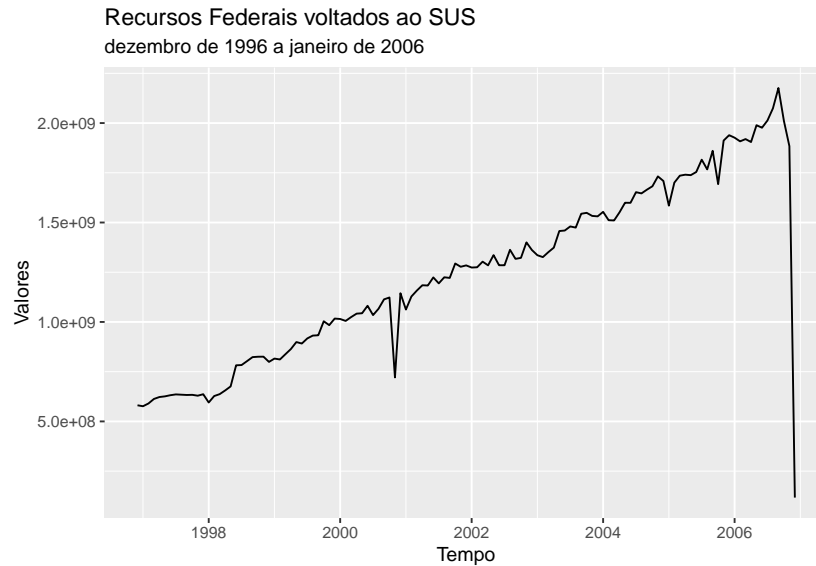


Figura 1: Gráfico da série original

A partir do gráfico, podemos ver que a série é claramente não estacionária, pois possui uma forte tendência geral de crescimento. Também podemos concluir que não há sazonalidade na série, e também destacar que em alguns meses, como os meses finais de 2001 e 2006, o SUS sofreu com quedas drásticas no recebimento de recursos por parte do governo federal.

Podemos confirmar a hipótese de não estacionariedade utilizando os testes de **Dickey-Fuller** e **Kwiatkowski-Phillips-Schmidt-Shin (KPSS)**, como mostram as saídas logo abaixo.

```
##
## Augmented Dickey-Fuller Test
##
## data: recursos_ts
## Dickey-Fuller = -3.1366, Lag order = 4, p-value = 0.1045
## alternative hypothesis: stationary

##
## KPSS Test for Level Stationarity
##
## data: recursos_ts
## KPSS Level = 2.331, Truncation lag parameter = 4, p-value = 0.01
```

Também devemos checar o comportamento das funções de autocorrelação e autocorrelação parcial da série, pois são a partir dessas funções que iremos estimar a ordem dos polinômios  $\phi(B)$  e  $\theta(B)$ .

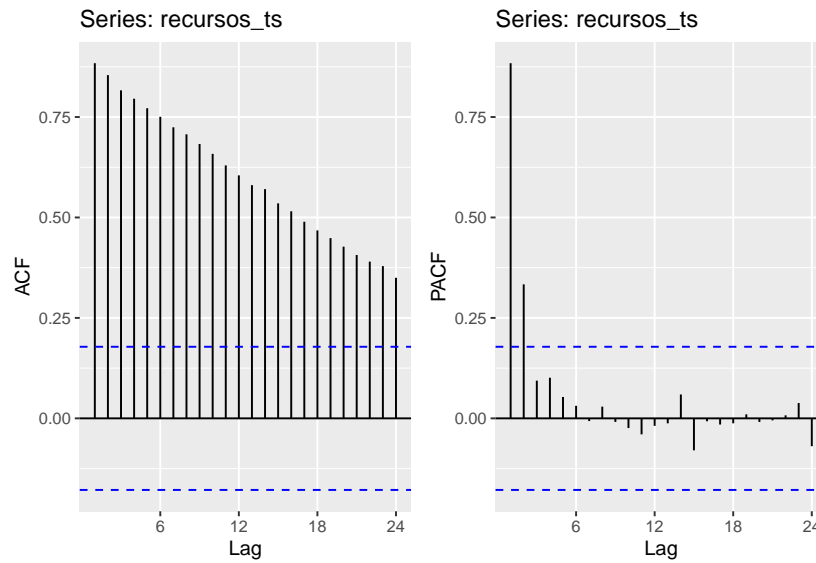


Figura 2: Funções de Autocorrelação

E por fim, dividimos a série em conjuntos de treino e teste, onde ajustamos modelos ao conjunto de treino para, posteriormente, avaliar previsões sobre o conjunto de teste.

```
treino = recursos_ts |> head(-12)
teste = recursos_ts |> tail(12)
```

## Modelagem

### Primeiro Modelo

Como já vimos anteriormente, temos uma série não estacionária, cujas autocorrelações decaem muito lentamente. Isso é um indicativo de que possivelmente devemos realizar diferenças na série. Tomando a primeira diferença no conjunto de treino, obtemos aparentemente uma série estacionária, na qual a variância parece ser estável (apesar de alguns pontos bruscos de mudança) e a média, em torno de 0, como mostram os gráficos da Figura 3.

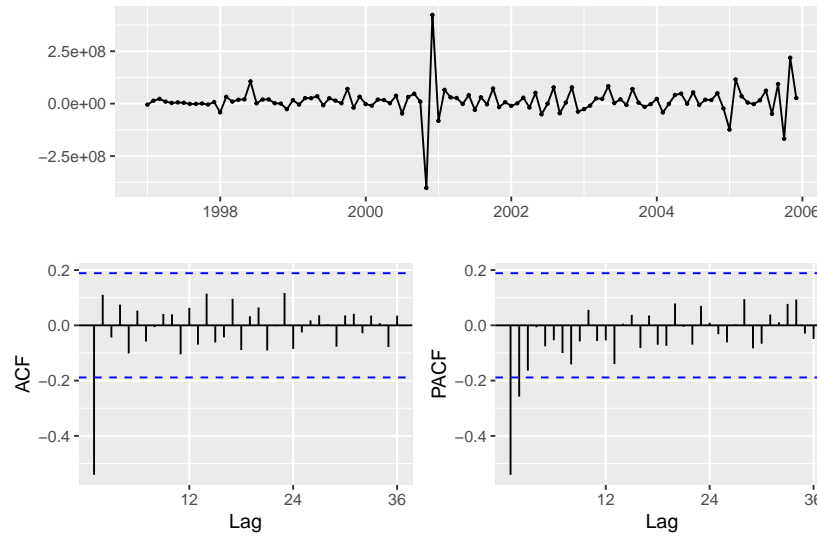


Figura 3: Primeira Diferença

Para validar a estacionariedade encontrada, usamos novamente os testes estatísticos anteriormente apresentados, e vemos que, de fato, a série é estacionária.

```
##
## Augmented Dickey-Fuller Test
##
## data: diff(treino)
## Dickey-Fuller = -5.8957, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary

##
## KPSS Test for Level Stationarity
##
## data: diff(treino)
## KPSS Level = 0.069699, Truncation lag parameter = 4, p-value = 0.1
```

Além disso, vemos que na função de autocorrelação somente o primeiro lag é significativo, enquanto na autocorrelação parcial temos que os dois primeiros lags são significativos. Todos os lags após os citados, em cada função, são não significativos. Isso sugere que a ordem dos polinômios seja de  $p = 2$  e  $q = 1$ .

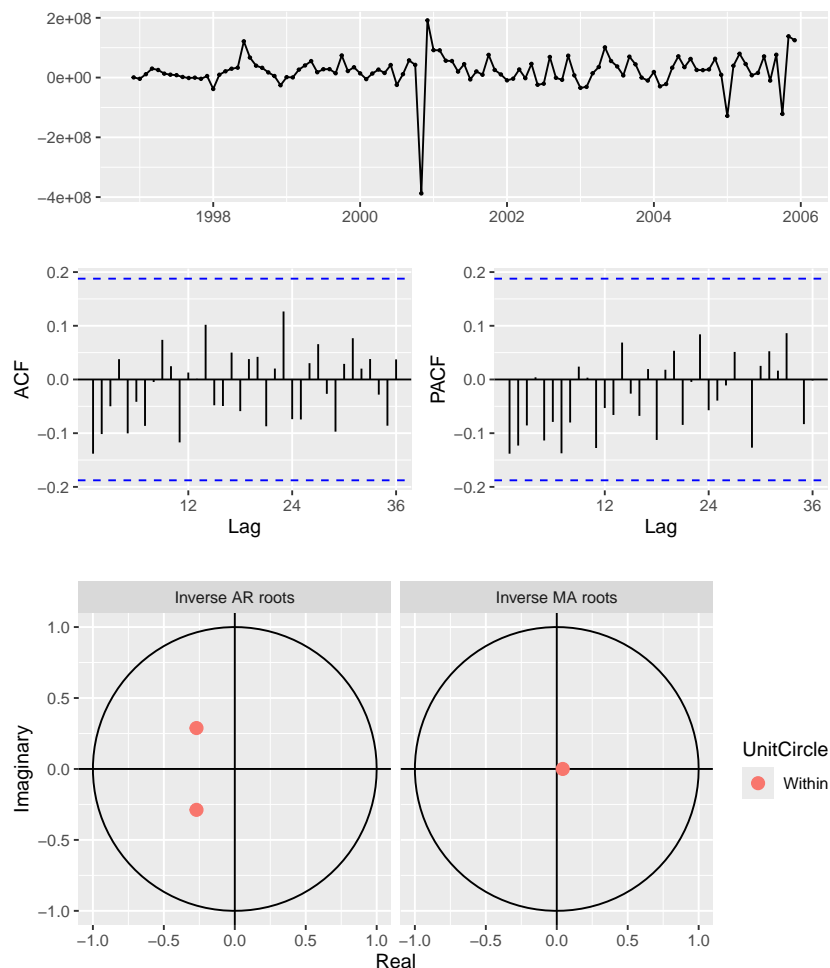
A partir disso, poderíamos então sugerir como primeiro modelo no método de Box-Jenkins, um  $ARIMA(2,1,1)$ , já que realizamos uma diferença. Ajustamos esse modelo no conjunto treino com o seguinte código:

```
fit1 = treino |> Arima(order = c(2, 1, 1)); fit1
```

```
## Series: treino
## ARIMA(2,1,1)
##
## Coefficients:
##          ar1      ar2      ma1
##       -0.5398 -0.1559 -0.0420
## s.e.   0.3482  0.1955  0.3415
```

```
##
## sigma^2 = 3.976e+15: log likelihood = -2091.52
## AIC=4191.04 AICc=4191.43 BIC=4201.77
```

Verificamos então o comportamento dos resíduos do modelo através de gráficos e testes de estacionariedade a seguir.



```
##
## Augmented Dickey-Fuller Test
##
## data: fit1$residuals
## Dickey-Fuller = -5.5308, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary

##
## KPSS Test for Level Stationarity
##
## data: fit1$residuals
## KPSS Level = 0.10095, Truncation lag parameter = 4, p-value = 0.1

##
## Ljung-Box test
```

```
##
## data: Residuals from ARIMA(2,1,1)
## Q* = 12.727, df = 19, p-value = 0.8522
##
## Model df: 3. Total lags used: 22

##
## Shapiro-Wilk normality test
##
## data: fit1$residuals
## W = 0.73712, p-value = 1.109e-12
```

Temos que todos os lags do nosso modelo ficaram dentro do intervalo de confiança construído. Além disso, os resíduos aparentemente são estacionários, e confirmamos a hipótese através dos testes. Ainda, vemos que as raízes inversas dos polinômios estão todas dentro do círculo unitário. Apesar disso, o pressuposto de normalidade não foi atendido.

Podemos ver a qualidade do ajuste através de gráficos, como na Figura 4, e também com medidas de erro no conjunto de teste. Vemos boa performance no conjunto treino, mas previsões ruins comparado ao conjunto teste.

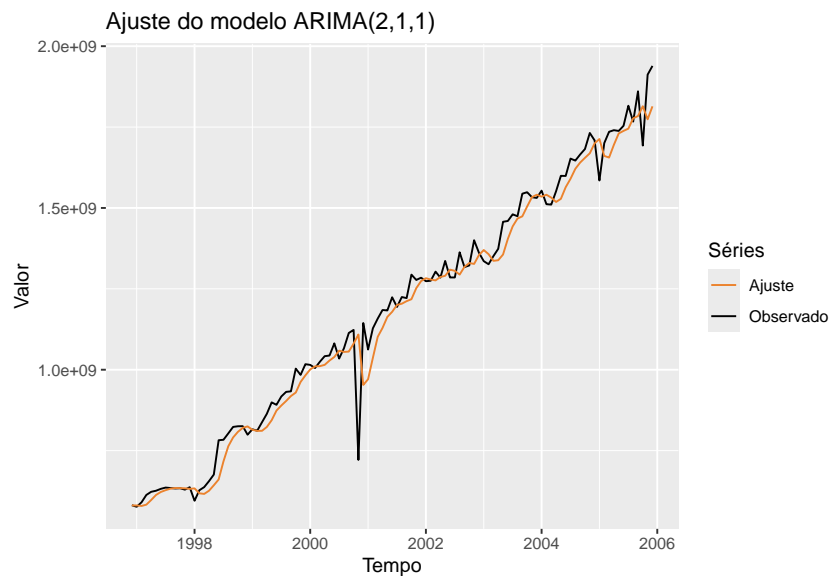
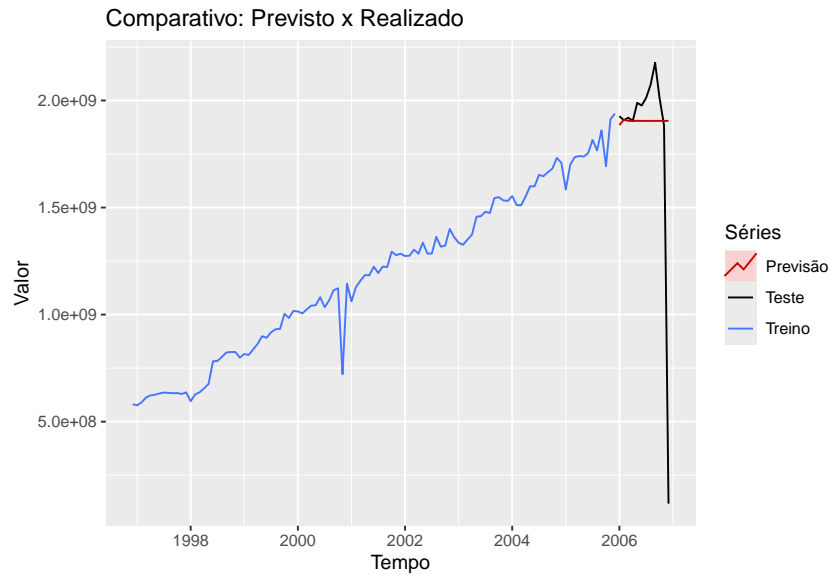


Figura 4: Gráfico do Ajuste Modelo 1

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Treino	21507990	61885254	38974439	1.6662	3.4740	0.2633	-0.1382	—
Teste	-78550440	527263219	223223773	-123.9411	130.9599	1.5079	0.0610	1.0212



Por fim, podemos então testar a significância dos parâmetros, conforme a tabela abaixo, e vemos que mesmo com a qualidade do ajuste atestada, os mesmos são não significativos, nos levando a considerar uma segunda iteração na metodologia Box-Jenkins.

Coef	Val	S.E.	Tcalc	Ttab
ar1	-0.5398	0.3482	1.5503	1.9826
ar2	-0.1559	0.1955	0.7976	1.9826
ma1	-0.0420	0.3415	0.1229	1.9826

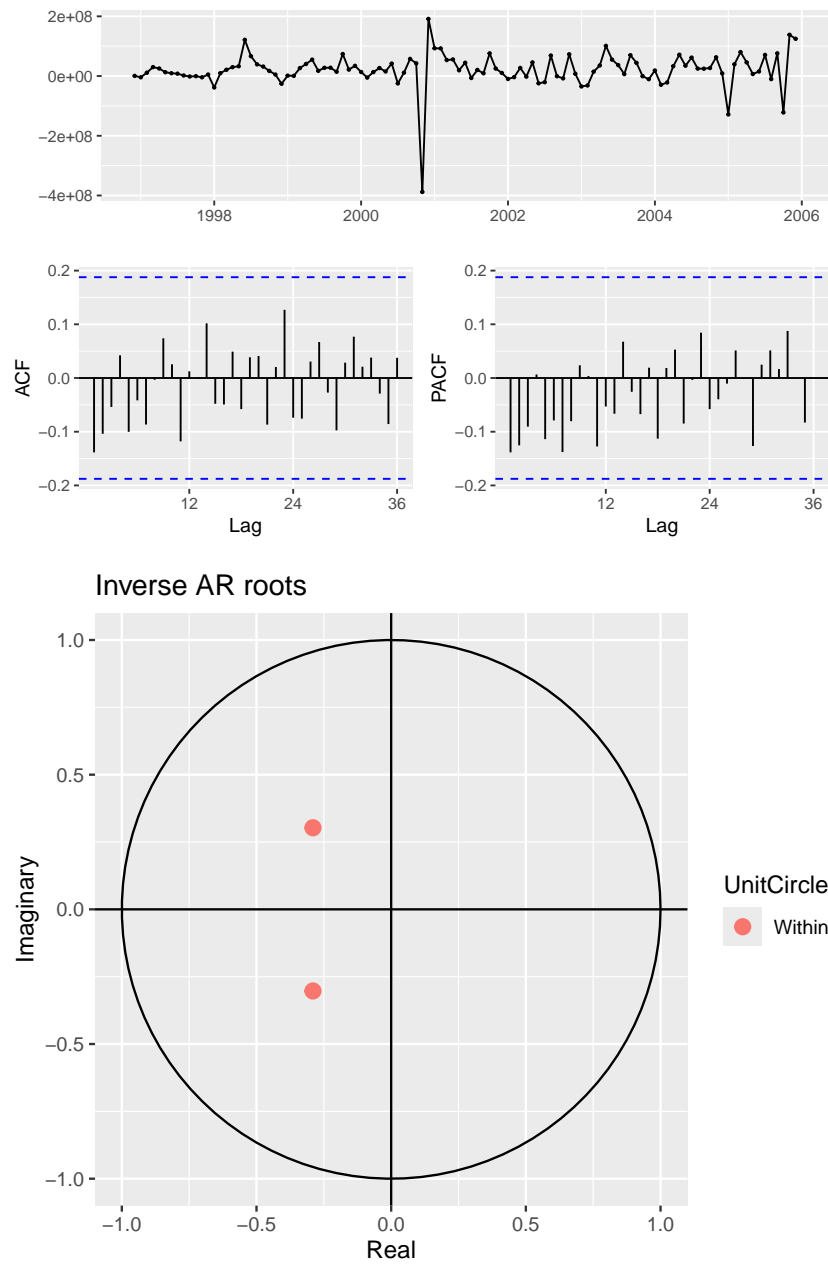
## Segundo modelo

Pela Figura 3, na função de autocorrelação, vemos que apenas o primeiro lag é significativo. Poderíamos então testar o primeiro modelo desconsiderando a influência da parte de médias móveis, isto é, fazer um  $ARIMA(2,1,0)$ :

```
fit2 = treino |> Arima(order = c(2, 1, 0)); fit2
```

```
## Series: treino
## ARIMA(2,1,0)
##
## Coefficients:
##          ar1      ar2
##      -0.5808  -0.1761
## s.e.   0.0967   0.1005
##
## sigma^2 = 3.939e+15: log likelihood = -2091.52
## AIC=4189.05  AICc=4189.28  BIC=4197.1
```

Verificamos então o comportamento dos resíduos do modelo através de gráficos e testes de estacionariedade a seguir.



```
##
## Augmented Dickey-Fuller Test
##
## data: fit2$residuals
## Dickey-Fuller = -5.5395, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary

##
## KPSS Test for Level Stationarity
##
## data: fit2$residuals
## KPSS Level = 0.10056, Truncation lag parameter = 4, p-value = 0.1

##
```



```
## Ljung-Box test
##
## data: Residuals from ARIMA(2,1,0)
## Q* = 12.856, df = 20, p-value = 0.8835
##
## Model df: 2. Total lags used: 22

##
## Shapiro-Wilk normality test
##
## data: fit2$residuals
## W = 0.73717, p-value = 1.112e-12
```

Novamente, vemos que a normalidade não foi satisfeita. Vejamos então o gráfico do ajuste, na Figura 5, que foi muito semelhante ao do primeiro modelo.

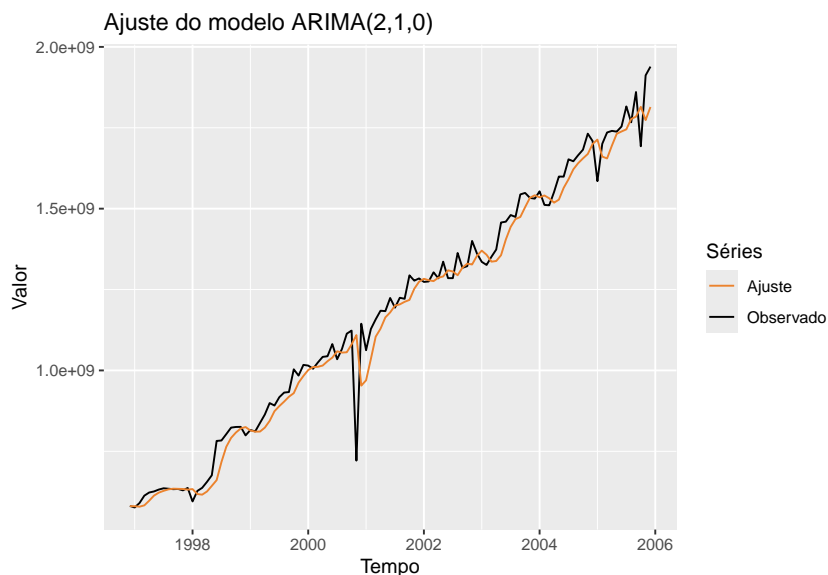
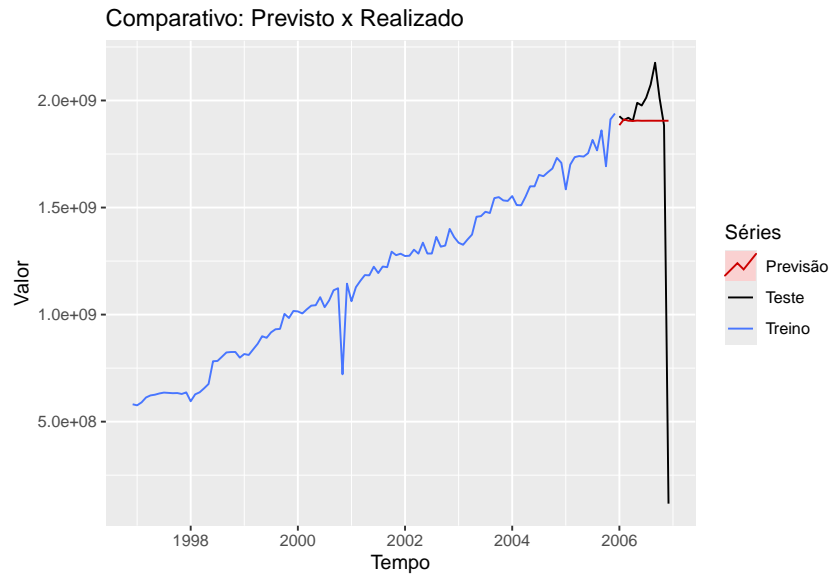


Figura 5: Gráfico do Ajuste Modelo 2

Realizamos então a previsão para 12 meses à frente, e vemos a qualidade das mesmas comparando-as ao conjunto de teste.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Treino	21360521	61888128	38898398	1.6532	3.4670	0.2628	-0.1384	—
Teste	-79240900	527376859	223032183	-124.0244	130.9994	1.5066	0.0610	1.0214



Novamente, obtivemos bom ajuste no conjunto de treino, mas previsões ruins no conjunto de teste. Validamos então a significância dos parâmetros estimados com a tabela a seguir:

Coef	Val	S.E.	Tcalc	Ttab
ar1	-0.5808	0.0967	6.0068	1.9824
ar2	-0.1761	0.1005	1.7532	1.9824

Agora, temos significância para o primeiro parâmetro autorregressivo, mas não para o segundo, o que nos leva a sugerir um terceiro modelo.

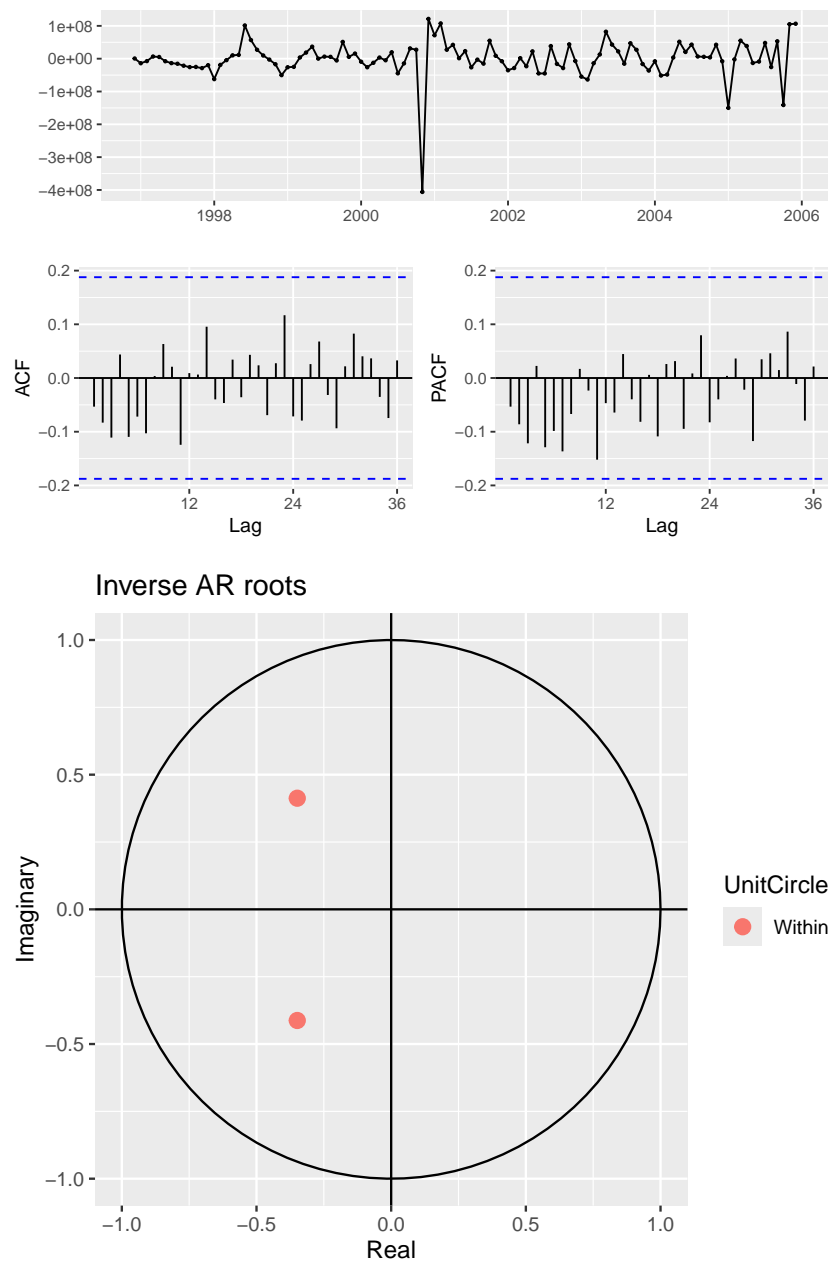
### Terceiro Modelo

Anteriormente, tivemos que o segundo modelo proposto foi um pouco melhor que o primeiro. Então, poderíamos testar o segundo modelo acrescido de uma constante, a fim de retirar os efeitos que possam estar afetando a média da série, isto é, deixando-a significativamente diferente de 0. No R, isso é feito da seguinte forma:

```
fit3 = treino |> Arima(order = c(2, 1, 0), include.drift = T); fit3
```

```
## Series: treino
## ARIMA(2,1,0) with drift
##
## Coefficients:
##          ar1          ar2          drift
##        -0.6973    -0.2918    12302230
## s.e.      0.0941     0.0973     2815012
##
## sigma^2 = 3.428e+15: log likelihood = -2083.61
## AIC=4175.21  AICc=4175.6  BIC=4185.94
```

Verificamos então o comportamento dos resíduos do modelo através de gráficos e testes de estacionariedade a seguir.



```
##
## Augmented Dickey-Fuller Test
##
## data: fit3$residuals
## Dickey-Fuller = -5.4984, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary

##
## KPSS Test for Level Stationarity
##
## data: fit3$residuals
## KPSS Level = 0.077966, Truncation lag parameter = 4, p-value = 0.1

##
```

```
## Ljung-Box test
##
## data: Residuals from ARIMA(2,1,0) with drift
## Q* = 11.514, df = 20, p-value = 0.9318
##
## Model df: 2. Total lags used: 22

##
## Shapiro-Wilk normality test
##
## data: fit3$residuals
## W = 0.74435, p-value = 1.728e-12
```

Novamente, com exceção do pressuposto de normalidade dos resíduos, todos os outros foram atendidos. Vejamos então o gráfico do ajuste para o modelo com constante, na Figura 6.

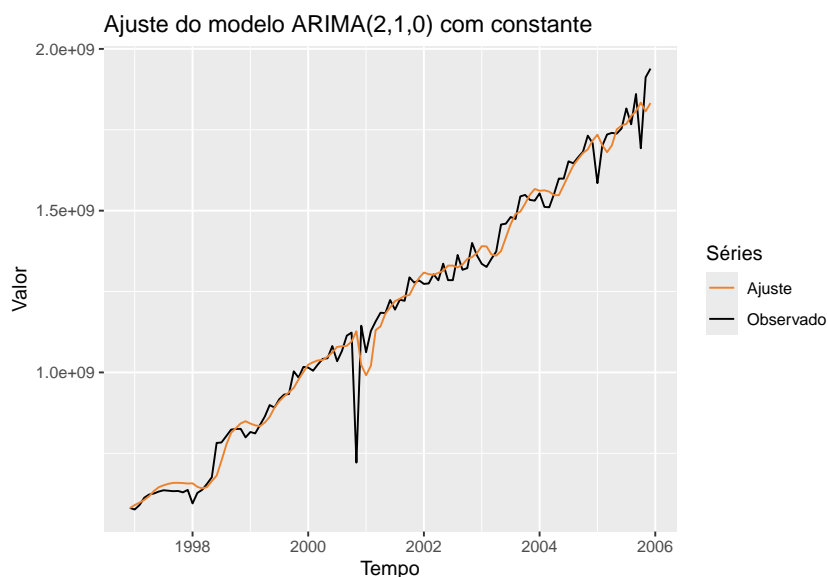
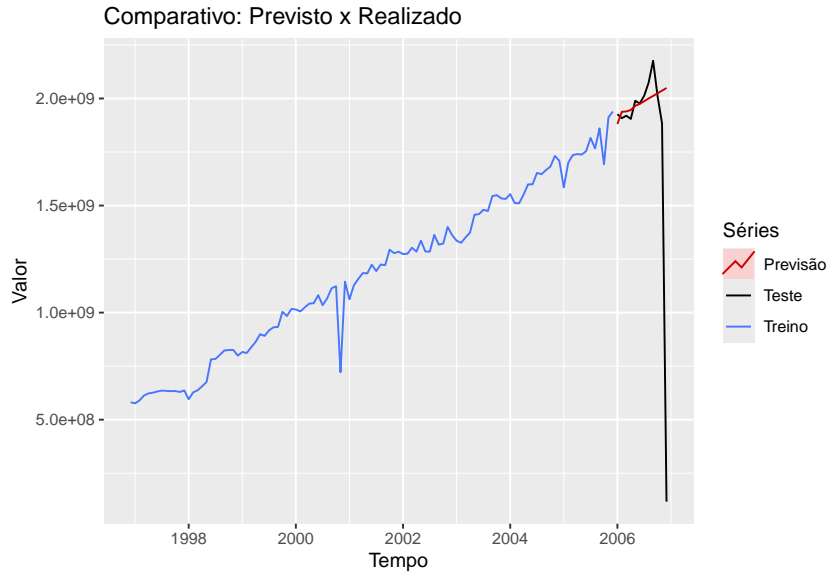


Figura 6: Gráfico do Ajuste Modelo 3

Novamente, há semelhança com os outros modelos, porém, olhando atentamente, notamos uma leve melhora na qualidade do ajuste. Para a previsão no conjunto de teste, temos a seguinte tabela e gráfico:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Treino	-88624.34	57464622	33754287	-0.3804	3.1034	0.2280	-0.0533	—
Teste	-154571078.93	562364307	210419663	-137.4205	140.0944	1.4214	0.0825	1.0914



A previsão, de modo geral, conseguiu captar o padrão de crescimento antes do decaimento dos valores próximo ao final da série, ou seja, conseguiu ser um pouco melhor que as fornecidas pelos modelos anteriores.

Por fim, precisamos atestar que todos os parâmetros estimados para o modelo são significativos, ou seja, diferentes de 0.

Coef	Val	S.E.	Tcalc	Ttab
ar1	-0.6973	0.0941	7.4115	1.9826
ar2	-0.2918	0.0973	2.9987	1.9826
drift	12302230.3215	2815011.6418	4.3702	1.9826

Assim, ficam validados os pressupostos do modelo, bem como a melhora na qualidade do ajuste em relação aos anteriores.

### Modelo Final

O último passo da iteração do método de Box-Jenkins nos forneceu o modelo mais apropriado para modelar a série proposta. A especificação do mesmo é dada:

$$(1 - 0,6973B - 0,2918B^2)(1 - B)Z_t = 12302230 + a_t$$

com média  $\hat{\mu} = 11286449540$  e  $\hat{\sigma}^2 = 3.428e + 15$ .

Por fim, o próximo passo seria ajustar o modelo estimado para o treino à série completa, a fim de que pudéssemos realizar previsões futuras.