Previsão de arrecadação de Imposto sobre circulação de mercadoria e serviços do Estado de Minas Gerais – Brasil

Rodrigo de Oliveira Salles

Faculdade de Ciências da Universidade do Porto

Séries Temporais e Previsão

Resumo

O objetivo deste trabalho foi realizar previsões de arrecadação de impostos sobre circulação de mercadorias e serviços (ICMS) no Estado de Minas Gerais – Brasil. O trabalho utilizou a base de dados disponibilizada pelo Instituto de Pesquisa Económica Aplicada (IPEA), e compreende um período de janeiro de 1993 a dezembro de 2016. Como metodologia utilizou-se os modelos ARIMA, redes neurais artificiais, e por fim a combinação de alguns modelos de previsão. Os dados utilizados foram separados em conjuntos de treino e testes. O primeiro conjunto compreende os dados entre os anos de 1993 e 2015, e o conjunto de testes reúne informações relativas à 2016. Os modelos ARIMA (1,1,1) e ARIMA (3,1,3) apresentaram os melhores desempenhos, com menor erro absoluto médio.

Introdução

Figura 1: *Estado de Minas Gerais – Região Sudeste do Brasil*

Fonte: https://pt.wikipedia.org/wiki/Minas_Gerais

Um dos interesses das ciências económicas é realizar previsões eficientes de variáveis que desempenham papel central em tomadas de decisão de agentes económicos. Para planear aporte de recursos, os governos precisam ter uma previsão de arrecadação de impostos, e no Brasil o imposto sobre circulação de mercadorias e serviços (ICMS) representa uma importante fonte de recursos.

O ICMS representa uma importante fonte de recursos para o Estado de Minas Gerais. Localizado na região Sudeste do Brasil, é o segundo estado mais populoso do pais, com aproximadamente 20,7 milhões de habitantes, e é composto por 853 municípios.

Em um estado com tais dimensões é de extrema importância a gestão eficiente de recursos, e para auxiliar nessa tarefa deve-se utilizar de todas as ferramentas possíveis. E nesse contexto, as séries temporais podem ser de grande valia em previsões de receitas.

O presente trabalho tem por objetivo tentar prever a arrecadação de ICMS para o Estado de Minas Gerais.

Materiais e Métodos

Os dados utilizados no trabalho são de uso público, disponibilizado pelo Instituto de Pesquisa Económica Aplicada (IPEA), e as análises foram feitas em linguagem R, com o auxílio do software RStudio.

A série é composta por 276 observações mensais, compreendidas entre os períodos de janeiro de 1993 e dezembro de 2015. Cada mês possui o valor de arrecadação de ICMS para o Estado de Minas Gerais. Uma amostra dos dados pode ser vista na tabela seguinte:

Data	Valor de ICMS (x mil)	Data	Valor de ICMS (x mil)
1993-01	R$ 3065,845	2015-08	R$ 3129785,458
1993-02	R$ 3496,385	2015-09	R$ 3155190,547
1993-03	R$ 4217,929	2015-10	R$ 3173985,276
1993-04	R$ 6031,502	2015-11	R$ 3316977,968
1993-05	R$ 7366,594	2015-12	R$ 3182727,446

Tabela 1: *Amostras iniciais e finais da série*

A metodologia adotada consiste em primeiramente realizar uma análise exploratória dos dados, com verificação de tendências, ciclos e efeitos de sazonalidade. Em seguida serão executados testes para verificar a estacionaridade da série. Se necessário, serão efetuadas transformações para tornar a série estacionária. Na sequência, serão analisados vários modelos, e escolhidos os modelos que melhor se ajustem aos dados. Por fim os modelos escolhidos serão utilizados para fazer previsões sobre a série adotada.

Os resultados serão comparados em relação aos erros absolutos verificados nas previsões.

Análise Exploratória de Dados

Como foi explicado anteriormente, a série é composta por observações do valor arrecadado pelo ICMS no Estado de Minas Gerais. Os dados começam com o valor referente ao mês de janeiro de 1993, e terminam com a arrecadação para o mês de dezembro de 2015.

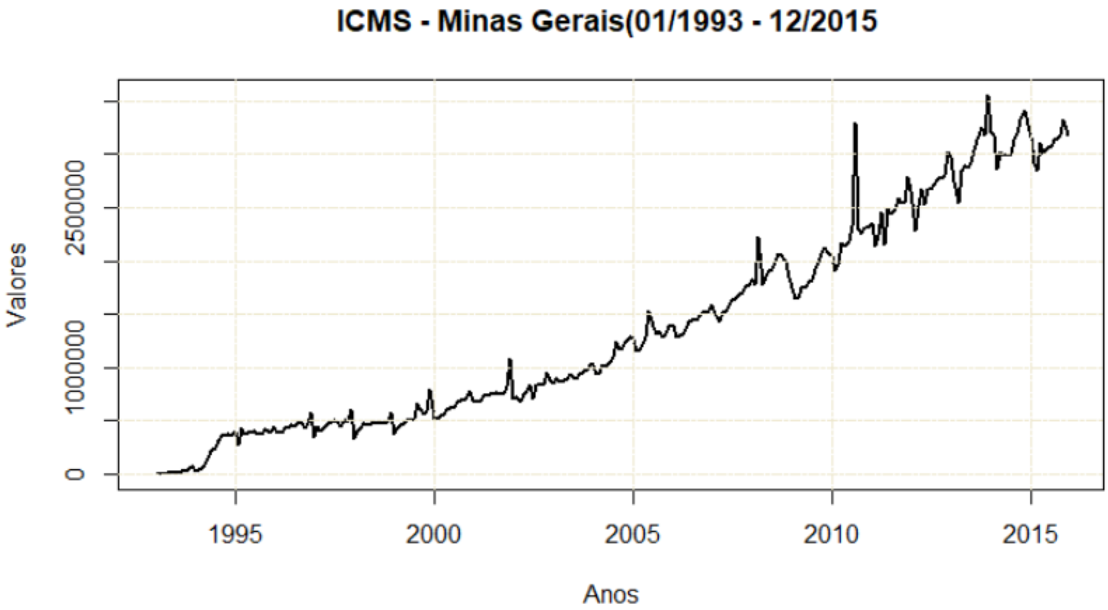

Figura 2: *ICMS do Estado de Minas Gerais. Jan.1993 – dez.2015*

Alguns dados estatísticos básicos da série podem ser observados na tabela seguinte:

Estatística	Dados(R$) x 1000
Mínimo	3096
Mediana	1089820
Média	1368857
1º quartil	486636
3º quartil	2135451
Máximo	3553055
Desvio Padrão	984187.5

Tabela 2: *Estatísticas básicas da série*

Pelo gráfico da figura 2 percebe-se que a série possui uma tendência crescente, e dessa forma pode-se concluir que ela não é estacionária. Nota-se também que a variância não é constante, e a série apresenta um comportamento especialmente turbulento nos últimos anos. Esse comportamento pode ser parcialmente explicado pela crise económica atravessada pelo país nesse período.

O gráfico da função de autocorrelação, que pode ser observada na figura 3, apresenta a forma típica de uma série não estacionária.

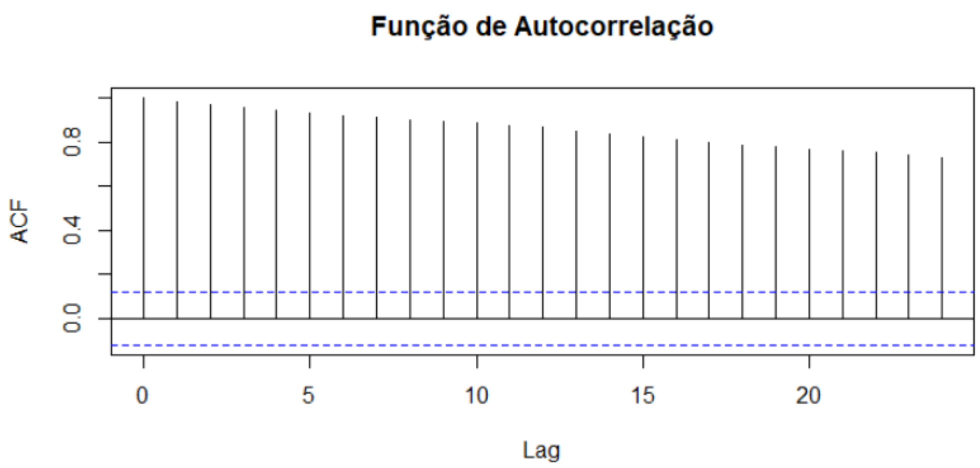

Figura 3: *Função de autocorrelação – ICMS / Minas Gerais*

A função de autocorrelação da série decresce muito lentamente para o zero, indicando os efeitos da tendência e possivelmente de sazonalidade.

A não-estacionaridade da série pode ser confirmado pelo teste da raiz unitária:

```
        Augmented Dickey-Fuller Test

data:  ipeadata$valor
Dickey-Fuller = -2.0473, Lag order = 6, p-value = 0.5561
alternative hypothesis: stationary
```

Figura 4: *Augmented Dickey-Fuller test – teste para raiz unitária*

O teste *Augmented Dickey-Feller* trabalha com a hipótese nula da existência de raiz unitária, sendo a série não-estacionária. Pode-se perceber que o *p-value* apresenta valor bem acima de 0.05(nível de significância), dessa forma o teste indica que a série não é estacionária.

Pode-se aplicar mais um teste que confirme a não estacionaridade. O teste de Kwiatkowski-Phillips-Schmidt (KPSS) adota como hipótese nula a série ser estacionária. O resultado aplicado à série de ICMS é o seguinte:

```
        KPSS Test for Level Stationarity

data:  ipeadata$valor
KPSS Level = 4.5106, Truncation lag parameter = 5, p-value =
0.01
```

Figura 5: *Teste de estacionaridade KPSS*

O p-valor de 0.01 indica a rejeição da hipótese nula, e assim, de acordo com o teste, a série não é estacionária.

Estacionaridade

A estacionaridade é uma importante característica pois indica que propriedades probabilísticas e estatísticas da série temporal permanecem inalteradas ao longo do tempo.

Os testes aplicados anteriormente indicam que a série não é estacionária, e dessa forma, deve-se recorrer a algumas transformações para alcançar o objetivo de tornar a série estacionária.

Assumindo um modelo aditivo, pode-se decompor a série em seus componentes sazonal, aleatório e tendência. Essa decomposição pode ser vista na figura a seguir:

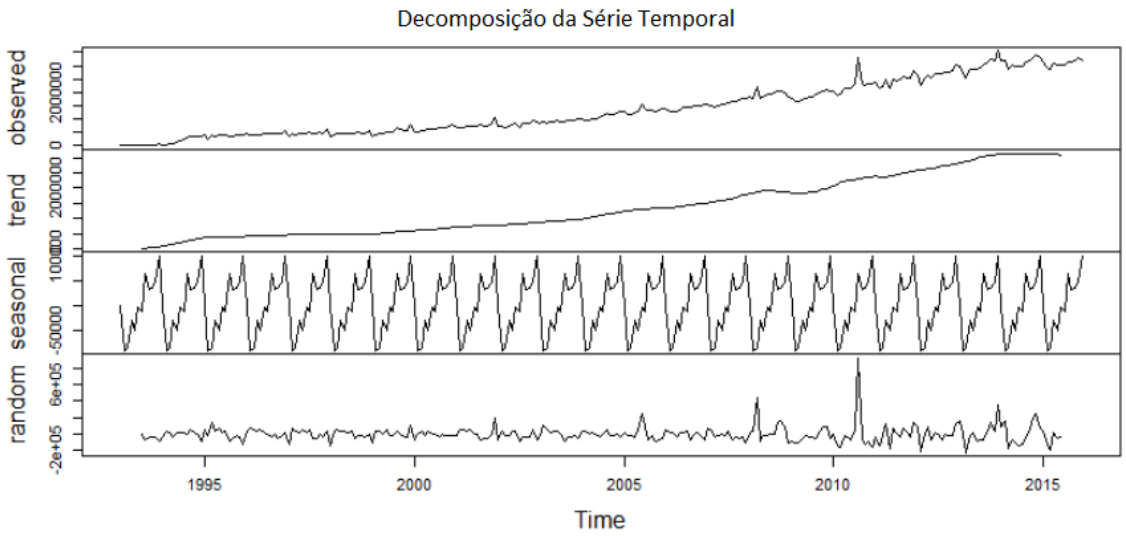

Figura 6: *Decomposição da Série - ICMS*

Ao analisar a decomposição da série pode-se notar uma forte componente de tendência, e um efeito relativo à sazonalidade praticamente desprezível.

Levando-se em consideração os resultados anteriores, vai-se recorrer à diferenciação da série para tentar alcançar a estacionaridade.

A função *ndiffs* nos indica o número de diferenciações necessárias. Ao aplicar essa função à série temos a indicação de que apenas **uma** diferenciação é o suficiente. Existe também a função *nsdiffs*, que indica o número de diferenciações sazonais necessárias. Ao aplicar a função *nsdiffs* temos **zero** como retorno, o que confirma as análises iniciais de efeito sazonal desprezível.

Modelos

Para iniciar a modelação efetuou-se, conforme sugerido anteriormente, uma primeira diferenciação. O resultado da função de autocorrelação e autocorrelação parcial podem ser vistos nas figuras 7, onde os valores de ACF decaem rapidamente para a região próxima à zero.

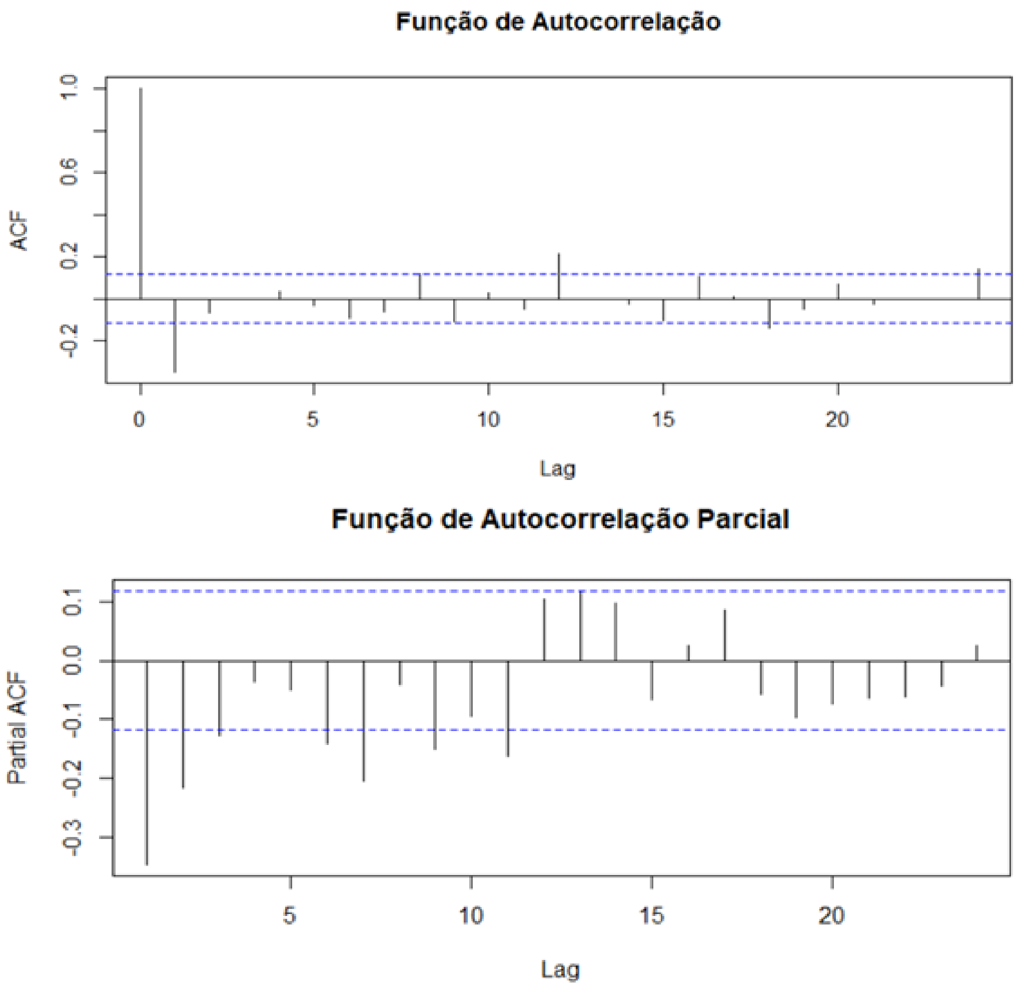

Figura 7: *Funções de autocorrelação e autocorrelação parcial*

O teste da raiz unitária apresenta p-valor baixo, confirmando que uma diferenciação foi, de fato, suficiente para tornar a série estacionária:

```
        Augmented Dickey-Fuller Test

data:  ipea_dif_1
Dickey-Fuller = -9.9316, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Figura 8: *Augmented Dickey-Fuller test para série diferenciada*

O teste KPSS, com p-valor acima de 0.05 indica que não se pode rejeitar a hipótese nula, de a série ser estacionária:

```
        KPSS Test for Level Stationarity

data:  ipea_dif_1
KPSS Level = 0.063815, Truncation lag parameter = 5, p-value = 0.1
```

Figura 9: *KPSS test para série diferenciada*

Para iniciar os testes com os modelos vai-se inicialmente recorrer à função *auto.arima*, para determinar o modelo arima que melhor se ajusta à série. O resultado obtido foi um modelo ARIMA (1,1,1) com drift. O modelo confirma a necessidade de uma diferenciação.

```
ARIMA(1,1,1) with drift

Coefficients:
         ar1      ma1      drift
      0.3518  -0.8315  11746.832
s.e.  0.0832   0.0498   1935.745

sigma^2 estimated as 1.492e+10:  log likelihood=-3610.12
AIC=7228.25   AICc=7228.39   BIC=7242.71
```

Figura 10: *Modelo ARIMA (1,1,1) com drift*

Na figura 11 pode-se verificar os parâmetros AIC e BIC de avaliação do modelo. Na mesma figura pode-se ver que os parâmetros encontrados ar1 e ma1 são estatisticamente significantes:

```
$ttable
           Estimate        SE  t.value p.value
ar1          0.3518    0.0832   4.2282       0
ma1         -0.8315    0.0498 -16.7026       0
constant 11746.8315 1935.7450   6.0684       0

$AIC
[1] 26.28453

$BIC
[1] 26.33714
```

Figura 10: *Modelo ARIMA (1,1,1) com drift*

A partir da análise da figura 11, pode-se concluir, que os resíduos não são correlacionados, são independentes e identicamente distribuídos. Apesar da forma apresentada pelo gráfico Q-Q, pode-se afirmar que esse modelo se ajusta bem aos dados.

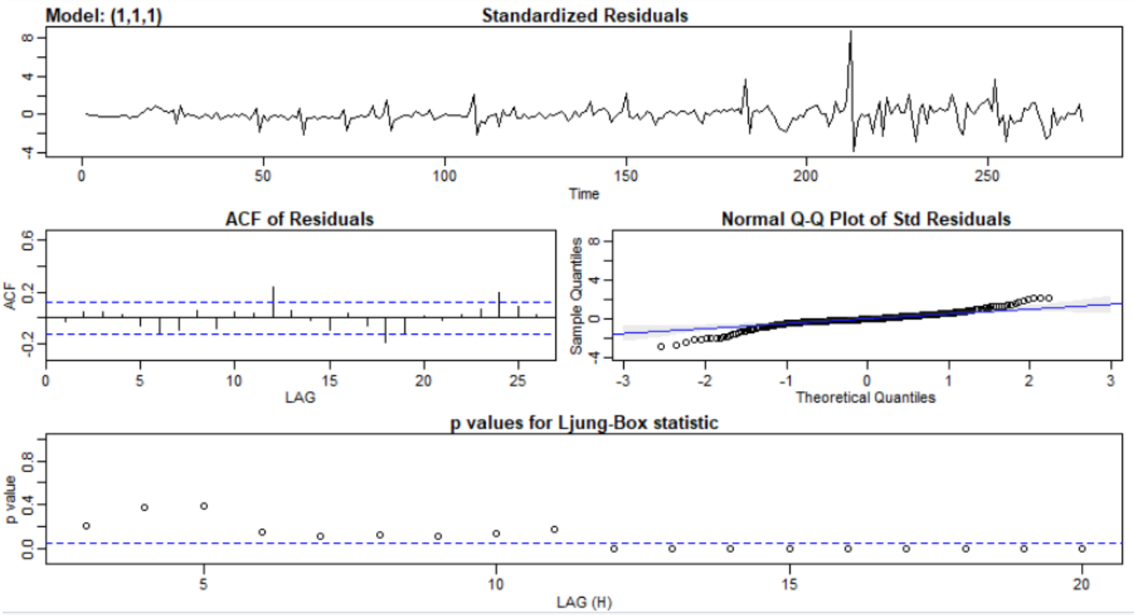

Figura 11: *análise dos resíduos.*

Com a finalidade de efetuar comparações e possivelmente obter melhores modelos, optou-se por fazer vários testes, com diferentes parâmetros AR e MA, mas sempre mantendo uma diferenciação(d=1). Os resultados obtidos foram agrupados na seguinte tabela:

Modelo	AIC	BIC
ARIMA (1,1,1)	26.28453	26.33714
ARIMA (2,1,1)	26.28695	26.35271
ARIMA (1,1,2)	26.28619	26.35194
ARIMA (3,1,1)	26.29265	26.37156
ARIMA (3,1,2)	26.29992	26.39198
ARIMA (3,1,3)	26.23382	26.33904
ARIMA (1,1,2)	26.28618	26.35194
ARIMA (1,1,3)	26.29345	26.37236

Tabela 3: *Comparação de modelos*

Na primeira linha estão os valores de AIC e BIC para o modelo obtido pela função *auto.arima*. Nas linhas seguintes alguns modelos foram testados, e o modelo ARIMA (3,1,3) obteve os menores valores de AIC e BIC.

Uma análise dos parâmetros e resíduos do modelo ARIMA (3,1,3) pode ser vista na figura 12.

```
$ttable
           Estimate        SE  t.value p.value
ar1          0.6352    0.0771   8.2426       0
ar2         -0.9912    0.0316 -31.4064       0
ar3          0.4232    0.0753   5.6182       0
ma1         -1.1382    0.0461 -24.6916       0
ma2          1.2399    0.0318  38.9419       0
ma3         -0.8588    0.0460 -18.6820       0
constant 11734.1681 1851.0512   6.3392       0
```

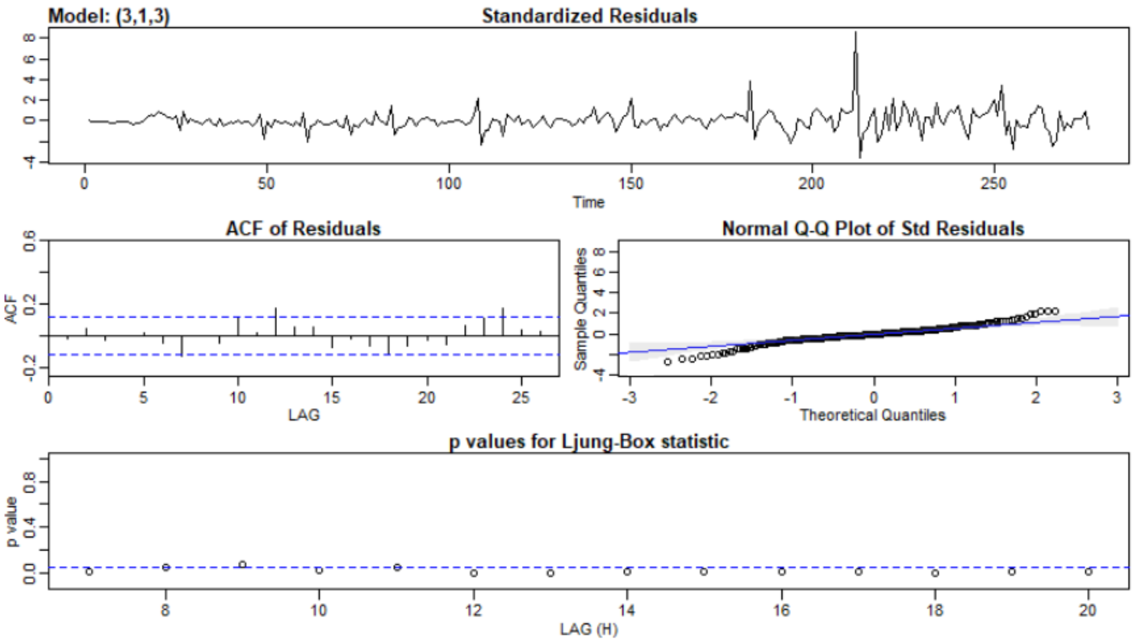

Figura 12: *análise dos resíduos.*

O gráfico da figura 12 nos sugere que os parâmeros são estatisticamente significativos, os resíduos não são correlacionados, são independentes e identicamente distribuídos. Apesar da forma apresentada pelo gráfico Q-Q, em que os extremos se afastam da normalidade.

Modelo com efeitos de sazonalidade

Apesar da indicação da insignificância dos efeitos de sazonalidade, foi feito um teste com um modelo incluindo esse efeito.

O modelo encontrado foi ARIMA (1,1,1) x $(2,0,1)_{12}$. Outras possibilidades foram testadas, mas ocorreram problemas relacionados à estacionaridade da parte AR sazonal.

	Estimate	SE	t.value	p.value
ar1	0.1822	0.1137	1.6027	0.1102
ma1	-0.7328	0.0857	-8.5506	0.0000
sar1	1.0501	0.1065	9.8588	0.0000
sar2	-0.0863	0.0796	-1.0845	0.2791
sma1	-0.8680	0.0928	-9.3585	0.0000
constant	11142.3405	5198.6066	2.1433	0.0330

Figura 13: *ARIMA (1,1,1) x (2,0,1)$_{12}$*

Pode-se perceber que o termo sar2, com p-valor acima de 0.05 não é estatisticamente significante.

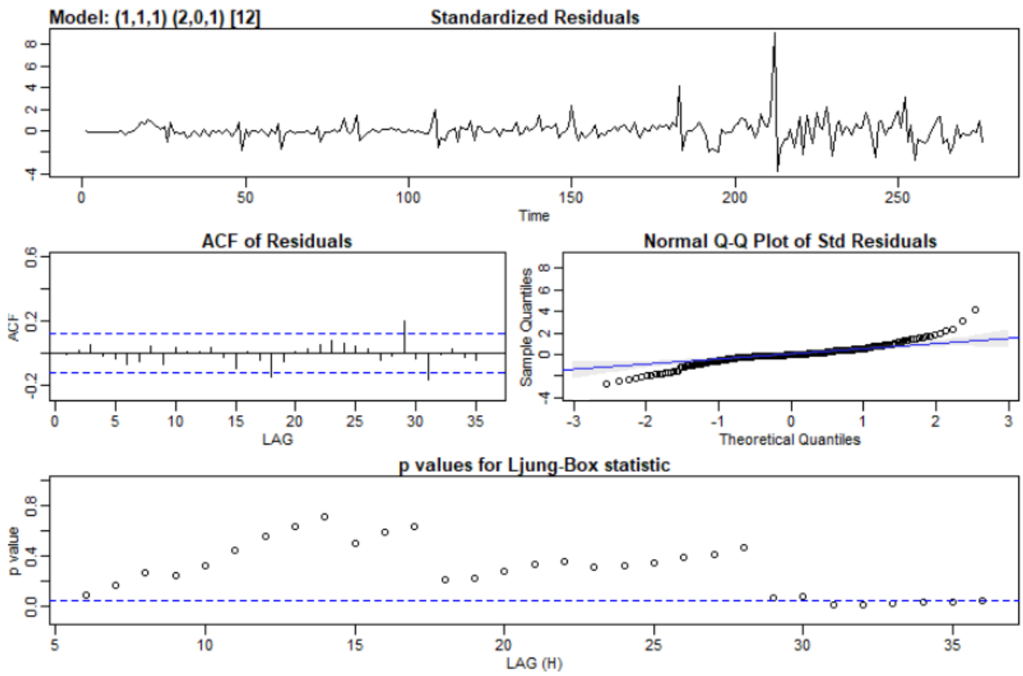

Figura 14: *análise dos resíduos.*

A análise dos resíduos demonstra alguns problemas. O gráfico de autocorrelação demonstra alguns valores fora dos limites, e o p-valor para o teste Ljung-Box apresenta oscilação e valores elevados em relação à proximidade do zero.

Modelo com transformação Box Cox

Para terminar a análise de possibilidades dos modelos Arima, considerou-se um modelo em que os dados sofreram uma transformação Box Cox.

Ao observar o gráfico da série original, na figura 2, pode-se perceber um possível problema de variância. Para tentar estabilizar variância recorreu-se a uma transformação Box Cox. O valor de λ foi obtido com o auxílio da função *BoxCox.lambda(),* e o valor encontrado foi de 0,5993. Esse valor foi utilizado, juntamente com a função *arima()*, para produzir as previsões.

Na figura seguinte, a análise sugere que os resíduos não estão correlacionados:

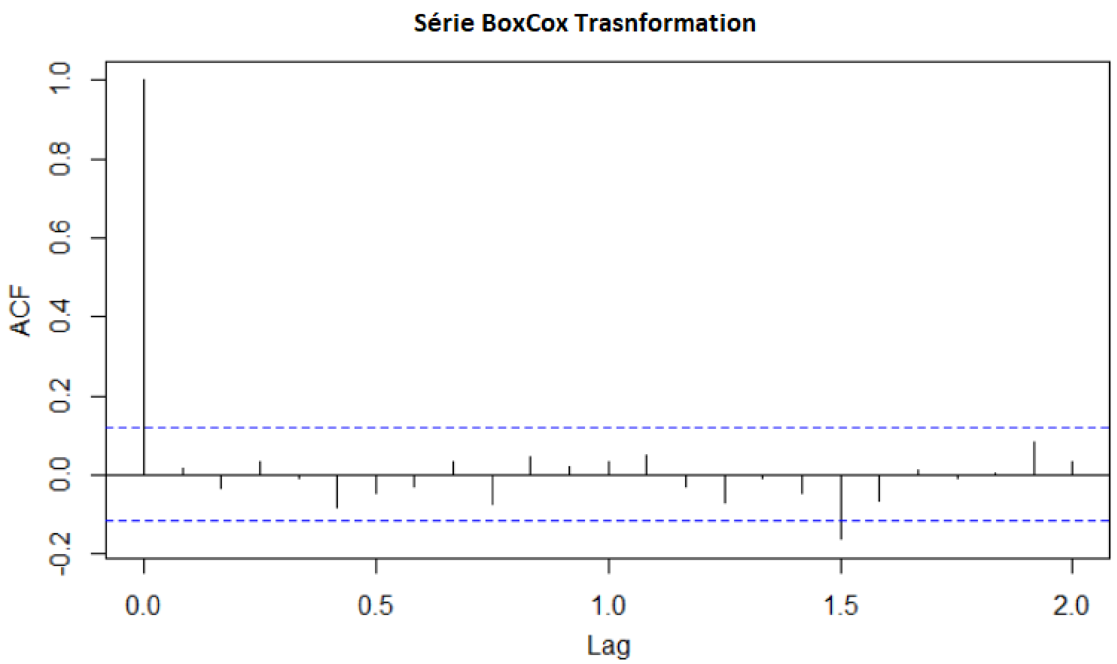

Figura 15: *análise dos resíduos.*

Previsões

Os modelos ajustados anteriormente serão utilizados para realizar previsões sobre a arrecadação de ICMS para o ano de 2016. Os valores reais de arrecadação para esse ano são conhecidos e serão utilizados para avaliar o desempenho dos previsores. Dois outros modelos também serão avaliados: redes neurais artificiais, e a combinação de previsores. O erro absoluto será calculado em cada previsão, com a finalidade de comparar os desempenhos.

Modelo ARIMA (1,1,1) com drift

O primeiro modelo utilizado será o ARIMA (1,1,1), com drift, ajustado automaticamente pela função *auto.arima*. Na figura 16 pode-se observar o gráfico relativo ao modelo.

O modelo parece captar a forma de variação da série. Os erros associados à previsão podem ser vistos na tabela 4.

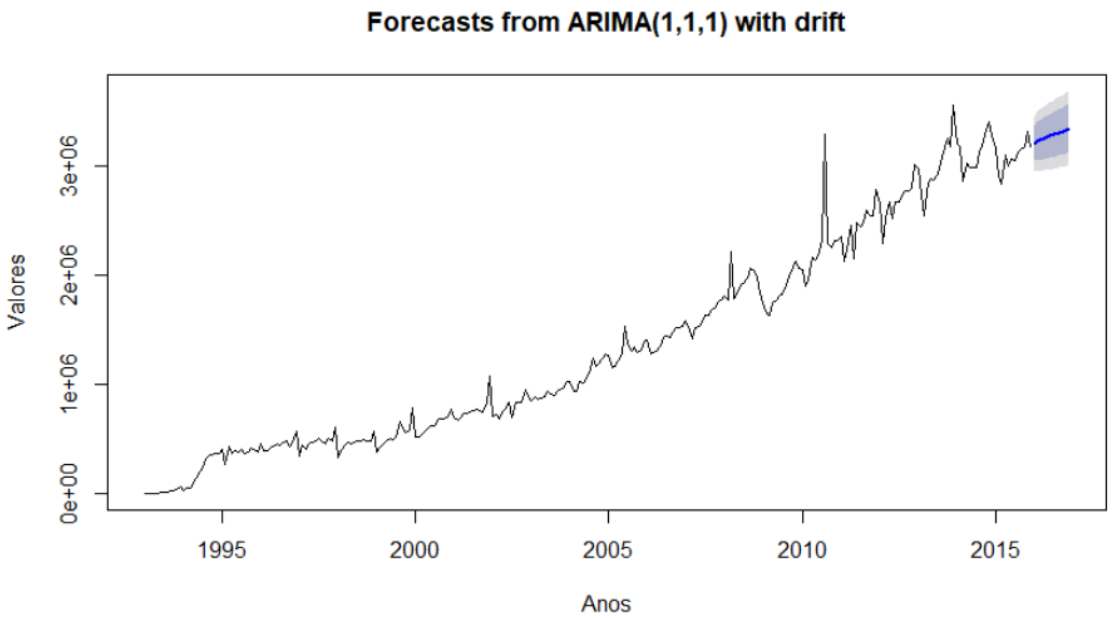

Figura 16: *previsão ARIMA (1,1,1).*

Mês/Ano	Observação	Previsão	Erro absoluto (%)
Janeiro - 2016	3083452,911	3208240,12	4,04
Fevereiro - 2016	3105138,440	3224829,24	3,71
Março - 2016	3220456,532	3238280,22	0,55
Abril - 2016	3388817,086	3250626,32	4,07
Maio - 2016	3389336,355	3262584,83	3,73
Junho - 2016	3401088,078	3274405,88	3,72
Julho - 2016	3474280,801	3286178,29	5,41
Agosto - 2016	3371817,494	3297934,32	2,19
Setembro - 2016	3665961,270	3309684,90	9,71
Outubro - 2016	3503212,918	3321432,06	5.18
Novembro - 2016	3498829,893	3333179,55	4,73
Dezembro - 2016	3925302,600	3344926,21	14,78

Tabela 4: *previsão ARIMA (1,1,1).*

Pode-se verificar que o menor erro foi relativo ao mês de março, e o maior erro ocorreu no último mês, com valor próximo a 15%. O erro absoluto médio para esse ano foi de 5,15%.

Modelo ARIMA (3,1,3) com drift

O próximo modelo testado foi o ARIMA (3,1,3) ajustado manualmente. No gráfico da figura 17 pode-se observar que o modelo parece captar o padrão de mudança da série.

Na tabela 5 pode-se observar os valores observados, reais, e os valores previstos para cada mês, bem como o erro associado a cada previsão.

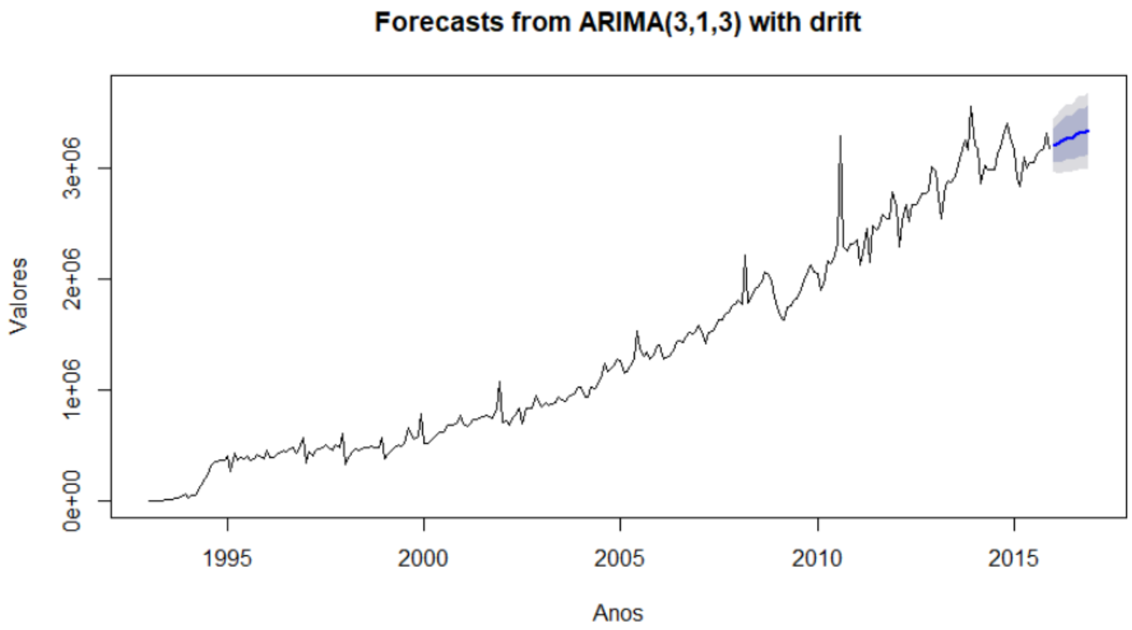

Figura 17: *Previsão ARIMA (3,1,3).*

Mês/Ano	Observação	Previsão	Erro absoluto (%)
Janeiro - 2016	3083452,911	3208479,09	4,05
Fevereiro - 2016	3105138,440	3214922,81	3,53
Março - 2016	3220456,532	3230913,22	0,32
Abril - 2016	3388817,086	3256528,31	3,90
Maio - 2016	3389336,355	3270620,53	3,50
Junho - 2016	3401088,078	3271894,21	3,79
Julho - 2016	3474280,801	3280522,08	5,57
Agosto - 2016	3371817,494	3301649,77	2,08
Setembro - 2016	3665961,270	3318001,33	9,49
Outubro - 2016	3503212,918	3322044,72	5,17
Novembro - 2016	3498829,893	3328291,56	4,87
Dezembro - 2016	3925302,600	3346117,13	14,75

Tabela 5: *previsão ARIMA (3,1,3).*

À semelhança do modelo ARIMA (1,1,1), o menor erro verificado foi para o mês de março, e o maior, para o mês de dezembro. O erro absoluto médio foi de 5,08%.

Modelo com sazonalidade ARIMA (1,1,1) x (2,0,1)$_{12}$

As previsões para modelo que considera a sazonalidade podem ser vistos no gráfico a seguir:

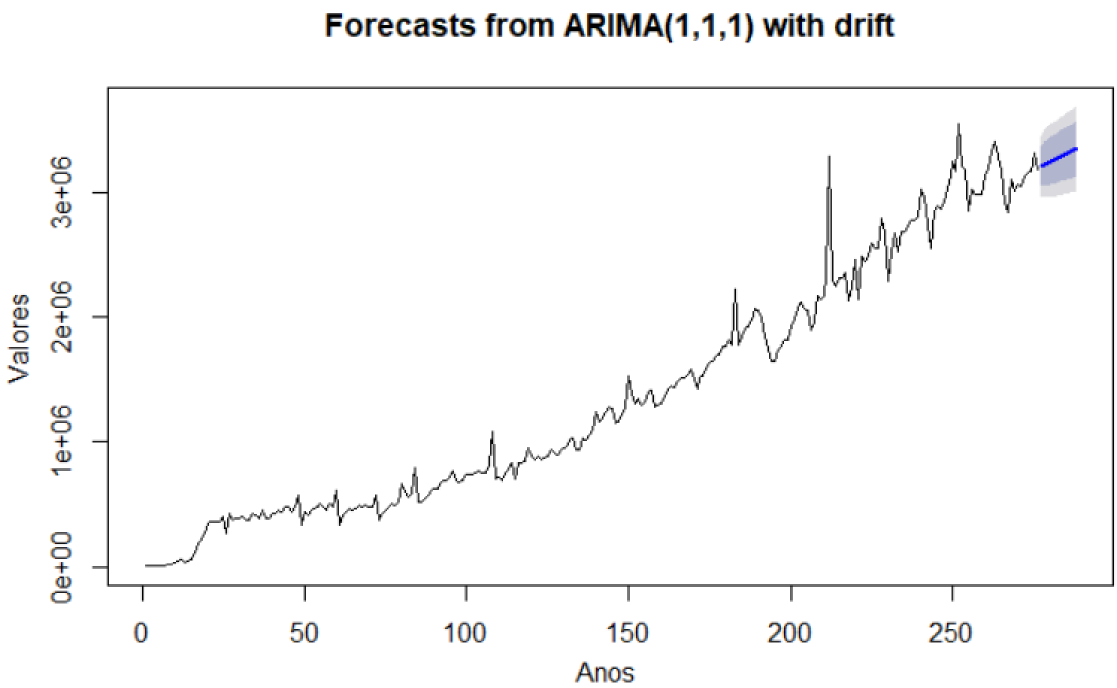

Figura 18: *Previsão ARIMA (1,1,1) com sazonalidade.*

As previsões efetuadas, e os erros do modelo podem ser vistos na tabela seguinte:

Mês/Ano	Observação	Previsão	Erro absoluto (%)
Janeiro - 2016	3083452,911	3208240	4,04
Fevereiro - 2016	3105138,440	3224829	3,85
Março - 2016	3220456,532	3224829	3.59
Abril - 2016	3388817,086	3250626	4,07
Maio - 2016	3389336,355	3262584	3,73
Junho - 2016	3401088,078	3274405	3,72
Julho - 2016	3474280,801	3286178	5,41
Agosto - 2016	3371817,494	3297934	2,19
Setembro - 2016	3665961,270	3309684	9,71
Outubro - 2016	3503212,918	3321432	5.18
Novembro - 2016	3498829,893	3333179	4,73
Dezembro - 2016	3925302,600	3344926	14,78

Tabela 6: *previsão ARIMA (1,1,1) x (2,0,1)$_{12}$*

Pode-se verificar que o menor erro foi relativo ao mês de agosto, e o maior erro ocorreu no último mês, com valor próximo a 15%. O erro absoluto médio para esse modelo foi de 5,41%.

Modelo com transformação Box Cox

O gráfico de previsão e as tabelas de desempenho para o modelo que utilizou a transformação Box Cox, pode ser visto a seguir:

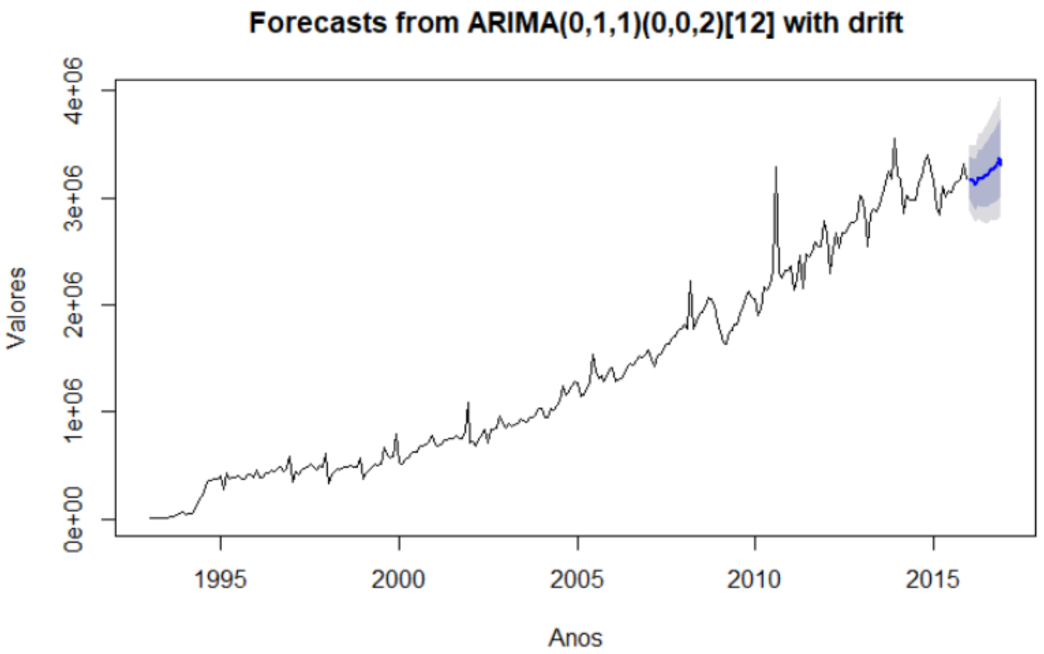

Figura 18: *Previsão ARIMA com transformação Box Cox.*

Mês/Ano	Observação	Previsão	Erro absoluto (%)
Janeiro - 2016	3083452,911	3181254	3,17
Fevereiro - 2016	3105138,440	3156217	1,64
Março - 2016	3220456,532	3120677	3,09
Abril - 2016	3388817,086	3195005	5,71
Maio - 2016	3389336,355	3173920	6,35
Junho - 2016	3401088,078	3204738	5,77
Julho - 2016	3474280,801	3213360	7,51
Agosto - 2016	3371817,494	3255069	3,46
Setembro - 2016	3665961,270	3278613	10,56
Outubro - 2016	3503212,918	3305695	5.18
Novembro - 2016	3498829,893	3366690	5,63
Dezembro - 2016	3925302,600	3308107	15,72

Tabela 7: *previsão ARIMA, com transformação Box Cox.*

O erro absoluto médio para o previsor com a transformação Box Cox foi de 6,14%, sendo o menor erro verificado para o mês de fevereiro, e o maior, a exemplo dos modelos anteriores, foi para o mês de dezembro.

Os próximos modelos testados serão as redes neurais artificias, e a combinação de previsores. Como esses modelos não foram explicados em sala de aula, será feita uma breve descrição dos mesmos.

Redes Neurais Artificiais

As redes neurais artificiais (RNAs) são sistemas paralelos distribuídos compostos por unidades (nodos) que calculam determinadas funções matemáticas. É uma tecnologia enraizada em várias disciplinas: neurociências, matemática, estatística, ciências da computação, e engenharia. Encontram aplicações em diversas áreas, como modelagens, análise de séries, reconhecimento de padrões, processamento de sinais, e séries temporais e previsão. Essa variedade de aplicações se deve a uma importante propriedade: a habilidade de aprender, a partir de dados de entrada.

Uma rede neural artificial pode ser vista como um conjunto de neurónios organizados em camadas: uma camada de entrada(preditores), as camadas intermediárias (camadas ocultas) e a camada de saída, que para o presente trabalho representam as previsões.

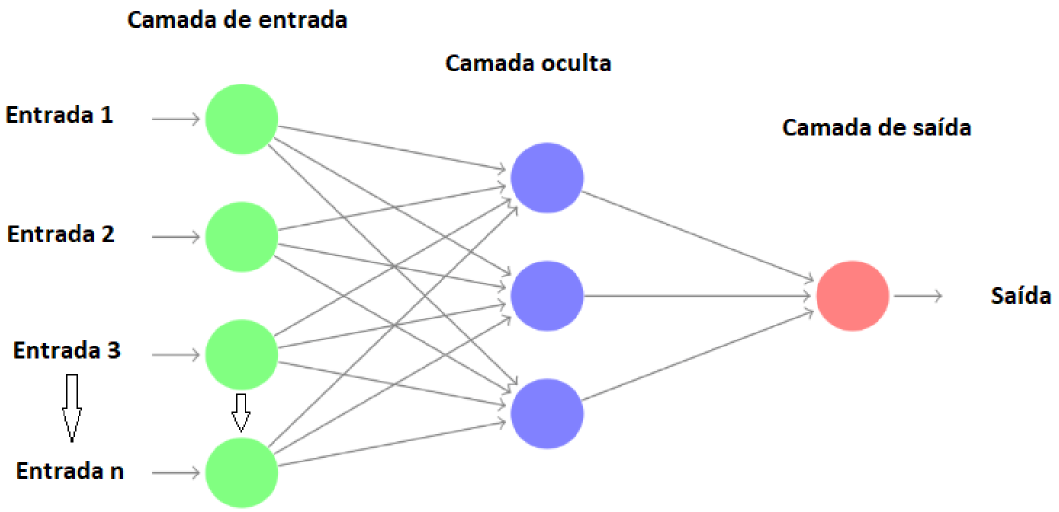

Figura 19: *RNA com uma camada oculta.*

Uma rede com apenas uma camada de entrada e uma camada de saída, sem camada oculta, representa uma regressão linear. Ao acrescentar uma camada oculta a rede neural torna-se não-linear, e assim aumenta muito seu potencial de aplicação.

Na figura 19 é possível ver a configuração de uma rede neural artificial com uma camada oculta.

Para o presente trabalho utilizou-se uma uma estrutura semelhante à apresentada na figura 19, sendo que o número de neurônios nas camadas de entrada e oculta foi determinada automaticamente pela função *nnetar.* O resultado foi um modelo NNAR(1,1,2), com entradas y_{t-1} e y_{t-12} , com dois neurônios na camada oculta. A rede foi treinada com os mesmos valores utilizados anteriormente pelos modelos ARIMA, e a previsão foi feita sobre os valores arrecadados para o ano de 2016. Na figura 20 pode-se observar a previsão para os 12 meses de 2016.

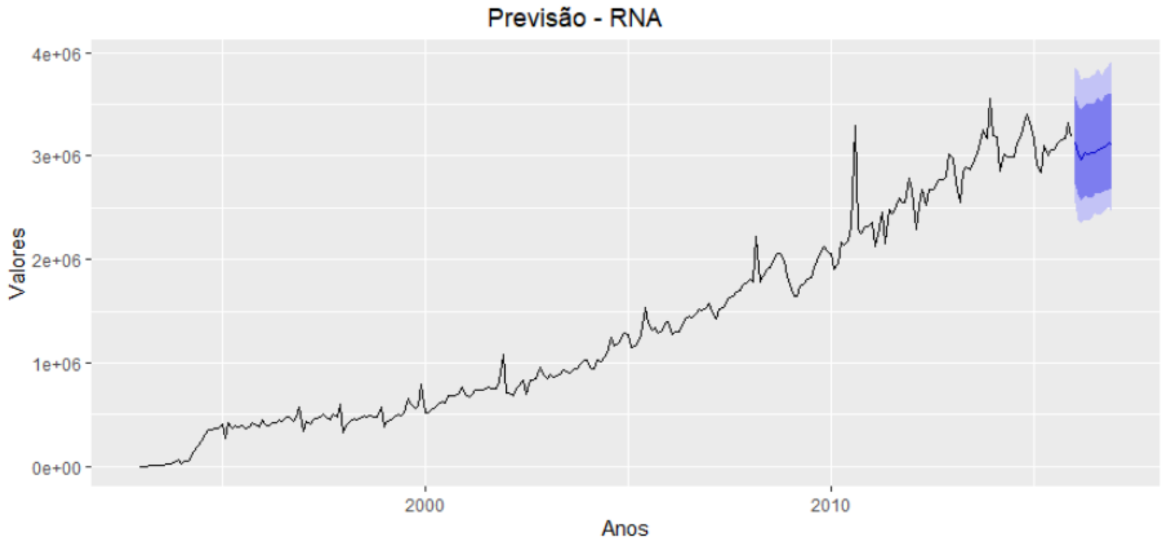

Figura 20: *Previsão feita com auxílio de RNA.*

Pelo gráfico da figura 20 pode-se perceber que o modelo parece captar o padrão de mudanças da série.

A função de autocorrelação dos resíduos das redes neurais artificiais pode ser observada na figura 21. Percebe-se que os resíduos não são correlacionados, ficando restritos à região próxima de zero.

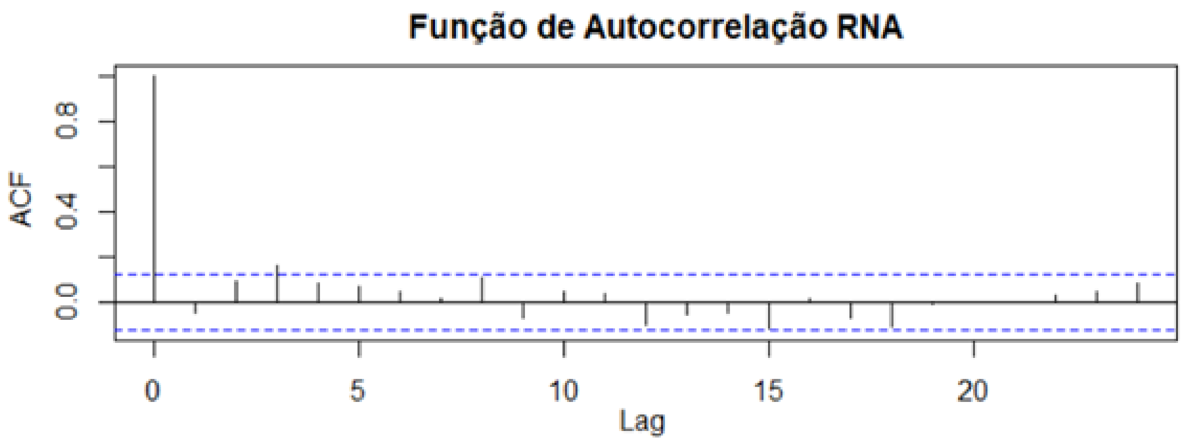

Figura 21: *Função de autocorrelação da RNA.*

A tabela 8 reúne as previsões e os erros para os meses do ano de 2016. Pode-se ver que o menor erro encontrado foi para o mês de janeiro, e o maior foi para o mês de dezembro.

O erro absoluto médio foi de 10,41%. Valor maior do que o registrado pelos modelos ARIMA.

Mês/Ano	Observação	Previsão	Erro absoluto (%)
Janeiro - 2016	3083452,911	3124299,54	1,32
Fevereiro - 2016	3105138,440	3028088,53	2,48
Março - 2016	3220456,532	2968597,88	7,82
Abril - 2016	3388817,086	3026798,21	10,68
Maio - 2016	3389336,355	3019229,20	10,91
Junho - 2016	3401088,078	3032314,29	10,84
Julho - 2016	3474280,801	3036832,48	12,59
Agosto - 2016	3371817,494	3060156,51	9,24
Setembro - 2016	3665961,270	3076458,44	16,08
Outubro - 2016	3503212,918	3088057,65	11,85
Novembro - 2016	3498829,893	3132048,12	10,48
Dezembro - 2016	3925302,600	3111517,52	20,72

Tabela 8: *previsões RNA NNAR (1,1,2).*

Combinação de previsores

Uma interessante possibilidade para melhorar o desempenho das previsões é a combinação de previsores. Cada previsor produz um resultado, para a mesma série, e o resultado final é a combinação dos resultados individuais.

Para tentar aproximar o valor de ICMS para o ano de 2016 foram escolhidos os seguintes modelos: ETS, ARIMA, NNAR e TBATS. Como os modelos ARIMA e NNAR já foram explicados anteriormente, vai-se comentar brevemente os modelos ETS e TBATS.

ETS (Error, Trend, seasonal)

A suavização exponencial é um método de previsão de séries temporais para dados univariados.

Métodos de séries temporais como a família de métodos Box-Jenkins desenvolvem um modelo em que a previsão é uma soma linear ponderada de observações ou atrasos recentes.

Os métodos de previsão de suavização exponencial são semelhantes, pois uma previsão é uma soma ponderada de observações passadas, mas o modelo usa explicitamente um peso decrescente para observações passadas.

Coletivamente, os métodos às vezes são chamados de **modelos ETS**, referindo-se à modelagem baseada em Erro, Tendência e Sazonalidade.

TBATS

O modelo BATS é um método de suavização exponencial combinado com transformação Box-Cox e modelo ARMA. A transformação Cox-Box tem a finalidade de lidar com dados não lineares. Alysha M. (2010) demonstrou que o modelo BATS pode melhorar o desempenho das previsões em comparação com modelo simples de espaço de estados.

Os modelos BATS não se saem bem quando a série apresenta altas frequências, assim Alysha M. (2011) propôs o modelo TABTS, que é um modelo BATS combinado com trigonometria sazonal. De forma reduzida, o método de suavização exponencial levou a um modelo de espaço de estados, que por sua vez levou a um modelo BATS, que foi, por fim, complementado com a trigonometria sazonal, formando o modelo TBATS.

Combinação de previsores

Os previsores anteriormente descritos foram combinados, e o resultado de previsão é uma média aritmética simples dos resultados individuais.

Os modelos foram treinados com a mesma série utilizada para os modelos anteriores. Pelos gráficos da figura 22 pode-se perceber que os resíduos não estão correlacionados:

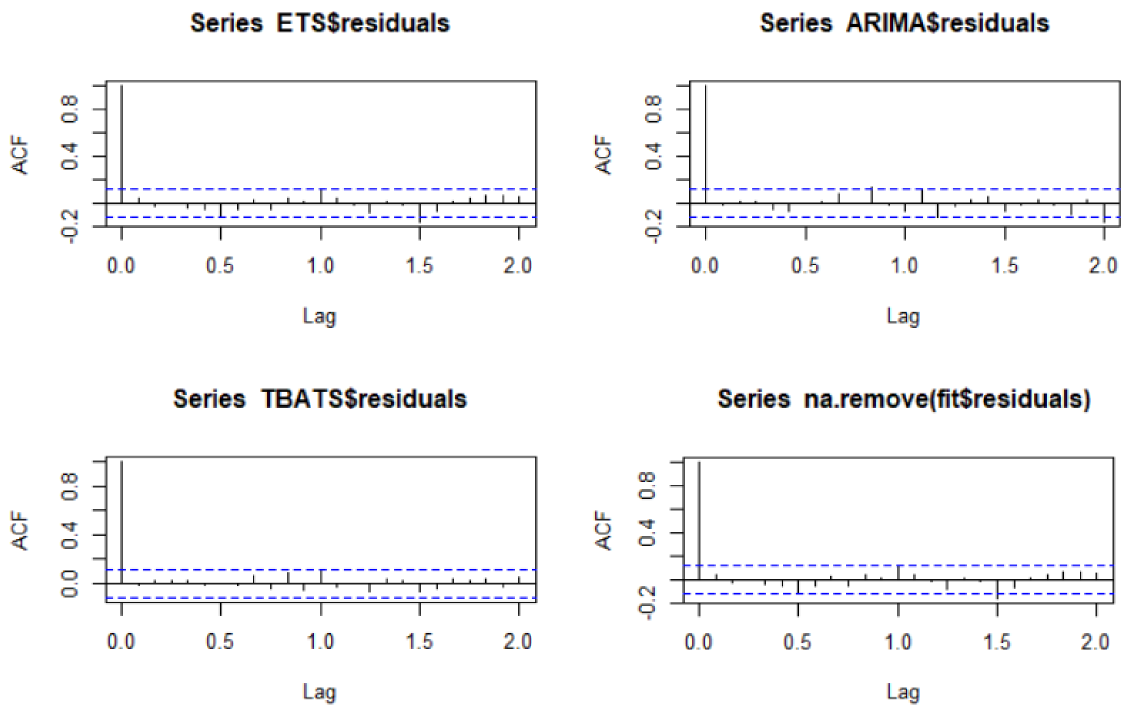

Figura 22: *Funções de autocorrelação.*

Os resultados de predição, bem com os erros podem ser observados na tabela seguinte:

Mês/Ano	Observação	Previsão	Erro absoluto (%)
Janeiro - 2016	3083452,911	3171219.02	2,46
Fevereiro - 2016	3105138,440	3082142,21	0,74
Março - 2016	3220456,532	3064450,32	4,84
Abril - 2016	3388817,086	3138345,45	7,39
Maio - 2016	3389336,355	3125730,65	7,77
Junho - 2016	3401088,078	3146907,21	7,47
Julho - 2016	3474280,801	3165513,33	8,88
Agosto - 2016	3371817,494	3219027,90	4,53
Setembro - 2016	3665961,270	3215372,54	12,29
Outubro - 2016	3503212,918	3222647,53	8,00
Novembro - 2016	3498829,893	3266345,11	6,64
Dezembro - 2016	3925302,600	3276248,31	16,53

Tabela 9: *previsões – Combinação de previsores.*

Pode-se perceber que o menor erro de previsão foi para o mês de fevereiro, e o maior para o mês de dezembro. Ao contrário da espectativa, a combinação de modelos apresentou desempenho inferior aos modelos ARIMA, com erro absoluto médio de 7.29%.

Conclusões

Dentre os modelos analisados o melhor desempenho foi alcançado pelos modelos ARIMA, com erro absoluto médio de aproximadamente 5,1%. Conforme sugerido pela função *auto.arima*, o modelo que considera a sazonalidade não apresentou melhoria de desempenho. O modelo que considerou a transformação Box Cox, para estabilizar a variância também não apresentou melhoras significativas em relação aos modelos ARIMA (1,1,1) e ARIMA (3,1,3).

As redes neurais artificiais e a combinação de previsores apresentaram maiores erros, e em especial para os meses finais da série de testes.

As crises financeiras e as mudanças de legislação tornaram espacialmente difíceis as previsões dos últimos anos da série.

Apesar de todas as dificuldades e instabilidades os modelos ARIMA se mostraram eficientes, e seriam as opções para realizar previsões fora da série considerada. Em especial o modelo ARIMA (3,1,3), que apresentou o menor erro absoluto médio nas previsões.

Referências

ALMEIDA, T. R. C. de. Previsão de Arrecadação Tributária na Crise: Alisamento Exponencial de Holt-Winters e SARIMA. **Revista de Estatística UFOP**, v. 6, 2017.

BOX, G.; JENKINS, G.; REINSEL, G. **Time Series Analysis**: Forecasting and Control. 4. ed. Nova Iorque: John Wiley & Sons, 2008.

Costa, K. H. P. (2013) *Expectativas acerca da arrecadação de icms no RN: modelagem em séries temporais* (Relatório de pesquisa/2013) Natal, RN, Disciplina de TCC1, Curso de administração, Faculdade de Natal (FAL).

R. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice. Second Edition.

R.H. Shumway, D.S. Stoffer, Time Series Analysis and its applications. Fourth Edition.

Castanho, B. J. S. (2011) *Modelos para previsão de receitas tributárias: o ICMS do estado do Espírito Santo*. Dissertação de mestrado, Universidade Federal do Espírito Santo, Vitória, ES, Brasil.