

CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ
Rio de Janeiro - RJ - Brasil
ISSN 1413-9022 / v. 30, p. 01 - 17, 2011

UTILIZAÇÃO DO SOFTWARE ESTATÍSTICO R NA PREVISÃO DE SÉRIES TEMPORAIS: SÉRIE PRÊMIO DE RESSEGURO

Rafael Ferraz de Moraes
CODIN – Companhia de Desenvolvimento Industrial do Estado do Rio de Janeiro
rferraz@codin.rj.gov.br

Paulo Roberto de Holanda Sales
IME/UERJ
pholanda@br.inter.net

José Fabiano da Serra Costa
IME/UERJ
fabiano@ime.uerj.br

Resumo

O objetivo dos autores no presente artigo é a utilização de modelos estocásticos da família SARIMA multiplicativo, de Box & Jenkins, para análise e previsão de séries temporais. Para tanto, é utilizado o software R, um pacote estatístico desenvolvido em ambiente livre e muito utilizado nas universidades. Para o estudo de caso de aplicação e funcionalidade do software, nas fases de identificação e estimação de modelos, foi utilizada a série temporal Prêmio de Resseguro de jan/2001 a dez/2008. Dessa forma, são descritos os modelos utilizados, os passos básicos com a linguagem de programação R, os resultados obtidos e as considerações finais.

Palavras-chave: *Série Temporal, Identificação e Estimação de Modelos, Previsão, R.*

1. Introdução

Em um mundo cada vez mais globalizado e competitivo as análises estatísticas ficaram cada vez mais importantes, impondo parâmetros e bases para tomar decisões visando o planejamento de futuras ações. Dentro deste contexto, a Análise de Séries Temporais é uma importante ferramenta, onde se sobressai a metodologia desenvolvida por Box & Jenkins (1976), a partir da família de modelos estocásticos SARIMA Multiplicativo $(p, d, q) \times (P, D, Q)_s$.

Para se obter estimativas de valores futuros (previsões) com maior praticidade e rapidez, pode-se utilizar os métodos de previsão chamados automáticos, pela adoção de um pacote estatístico. Dentre os softwares estatísticos utilizáveis destaca-se o R (R PROJECT, 2010), um software livre, muito utilizado pelas universidades e com crescente aceitação entre as empresas. O R foi criado por Ross Ihaka e Robert Gentleman e, desenvolvido por esforço colaborativo. Sua utilização é bastante difundida na literatura recente (RIBEIRO & MINGOTI, 2006; FOX, 2005; MINEO & PONTILO, 2006; PETERNELLI *et al.*, 2009; VALENÇA & SANTOS, 2010). A intenção desse artigo é mostrar uma aplicação da previsão de uma série temporal, Prêmio de Resseguro, utilizando o software R. Os dados são referentes ao período janeiro de 2001 a dezembro de 2008, num total de 96 observações (MEIRELES *et al.*, 2009).

2. Modelos SARIMA – Box & Jenkins

Uma metodologia bastante utilizada na análise e previsão de modelos paramétricos é conhecida como abordagem de BOX & JENKINS (1976) e BOX, JENKINS & REISEL (1994), que consiste em ajustar modelos auto-regressivos integrados de médias móveis sazonais multiplicativos, SARIMA $(p,q,d) \times (P, D, Q)_s$, a um conjunto de dados.

A metodologia Box & Jenkins é aplicada aos processos estocásticos que sejam estacionários. Se o processo estocástico não for estacionário, este pode se tornar estacionário por meio de sucessivas diferenciações da série original. Quando a série recebe a influência de fatores sazonais, outro tipo de correlação passa a ter importância: a correlação entre os instantes de tempo distantes entre si por s ou múltiplos de s , onde s representa o período da sazonalidade.

A estratégia para a construção do modelo é calcada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são: (i) uma classe geral de modelos é considerada para a análise (especificação); (ii) identificação de um ou mais modelos, a partir da análise das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP), bem como outros critérios, envolvendo diferenças simples e/ou sazonais; (iii) estimação dos parâmetros dos modelos identificados; (iv) testes de verificação dos modelos ajustados, por meio de uma análise de seus resíduos, para saber se são adequados para os fins em vista

Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação. Um procedimento muitas vezes utilizado é identificar não só um único modelo, mas alguns modelos que serão então estimados e verificados. Se o propósito é previsão, escolher-se-á entre os modelos ajustados o melhor, por exemplo, no sentido de ser parcimonioso e fornecer o menor erro quadrático médio de previsão (EQM) e o menor erro quadrático médio do limite superior da previsão, como critério probabilístico.

Em geral, os modelos postulados devem atender ao critério da parcimônia, i. é, devem possuir um número mínimo de parâmetros.

No caso dos Modelos SARIMA Multiplicativo $(p,d,q) \times (P,D,Q)_s$, o efeito sazonal implica que a observação de um determinado mês, num determinado ano, está relacionada com a observação deste mesmo mês em um outro ano. Supõe-se que a t -ésima observação Z_t é referente a um mês logo, a observação do mesmo mês referente ao ano seguinte pode ser modelada por:

$$(1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps})(1 - B^s)^D Z_t = (1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}) \alpha_t, \quad (1)$$

e sua forma reduzida:

$$\Phi(B^s) \Delta_s^D Z_t = \Theta(B^s) \alpha_t. \quad (2)$$

Onde:

$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}$	operador auto-regressivo sazonal de ordem P, estacionário;
$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}$	operador de médias móveis sazonal de ordem Q, invertível;
Δ_s^D	o operador diferença sazonal, sendo D o número indicador de “diferenças sazonais” e $s = 12$;
α_t	pode ser, eventualmente o ruído branco; neste caso, a FAC do processo Z_t é zero para todos os “lags” não sazonais e o modelo é denominado sazonal puro.

Suponha, agora, que o processo α_t satisfaça um modelo ARIMA (p,d,q),

$$\varphi(B)\alpha_t = \theta(B)a_t, \quad (3)$$

onde $\varphi = (1-B)^d \phi(B)$ e α_t é um processo de ruído branco. Demonstra-se que Z_t satisfaz o modelo:

$$\phi(B)\Phi(B^s)(1-B^s)^D(1-B)^d Z_t = \theta(B)\Theta(B^s)a_t, \quad (4)$$

onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$; e $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$;

Este modelo é denominado ARIMA sazonal multiplicativo (SARIMA) ordem (p,d,q) x (P,Q,D)_s.

3. Análise e Previsão da Série de Resseguros

O software R é altamente expansível com o uso dos pacotes, que são bibliotecas para funções específicas ou áreas de estudo específicas (R PROJECT, 2010). Para introduzir os dados no ambiente do programa, utiliza-se o comando: `x0=read.table('projeto-final0.txt',head=T)`. O programa lerá os dados salvos no formato.txt e com o argumento head=T, informará que o nome da variável está na primeira linha. Após este procedimento segue-se com os comandos:

```
attach(x0)
names(x0)
[1] "RESSEGURO"
RESSEGURO
[1] 94.400 122.719 110.706 113.372 114.044 229.749 95.772 135.221 101.839
[10] 132.209 122.853 298.744 151.357 148.462 220.101 192.725 131.031 208.893
[19] 182.622 328.293 212.464 229.653 202.419 237.201 311.806 306.934 224.254
[28] 155.987 263.550 200.763 258.990 206.041 185.079 221.988 243.856 285.778
[37] 292.186 297.411 186.652 193.167 205.084 203.405 200.298 308.577 230.973
[46] 219.486 249.566 255.434 259.378 346.757 292.361 179.206 210.351 178.104
[55] 201.548 207.876 284.457 203.825 203.352 323.484 285.264 341.893 210.487
[64] 240.621 172.327 155.456 228.120 278.024 325.535 179.514 201.553 275.696
[73] 351.535 326.501 270.676 253.098 200.744 230.187 291.207 290.359 318.500
[82] 208.319 205.051 309.143
```

É desejável que a série temporal seja estacionária de 1ª ordem (MORETTIN & TOLLOI, 1979). Assim, antes de se identificar o(s) modelo(s), deve-se verificar a existência de não estacionariedades: Tendência (T_t) e Sazonalidade (S_t), estimando-se as médias anuais que correspondem a Tendência (T_t) e as médias mensais que correspondem a Sazonalidade (S_t), Tabela 1 e nos Gráficos 1 e 2.

Tabela 1: Médias anuais e Médias mensais da série prêmio de resseguro

Ano	Mês												Média Tt
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	
2001	94.400	122.719	110.706	113.372	114.044	229.749	95.772	135.221	101.839	132.209	122.853	298.744	139.302
2002	151.357	148.462	220.101	192.725	131.031	208.893	182.622	328.293	212.464	229.653	202.419	237.201	203.769
2003	311.806	306.934	224.254	155.987	263.550	200.763	258.990	206.041	185.079	221.988	243.856	285.778	238.752
2004	292.186	297.411	186.652	193.167	205.084	203.405	200.298	308.577	230.973	219.486	249.566	255.434	236.853
2005	259.378	346.757	292.361	179.206	210.351	178.104	201.548	207.876	284.457	203.825	203.352	323.484	240.892
2006	285.264	341.893	210.487	240.621	172.327	155.456	228.120	278.024	325.535	179.514	201.553	275.696	241.207
2007	351.535	326.501	270.676	253.098	200.744	230.187	291.207	290.359	318.500	208.319	205.051	309.143	271.277
MedSt	249.418	270.097	216.463	189.739	185.305	200.937	208.365	250.627	236.978	199.285	204.093	283.640	224.579

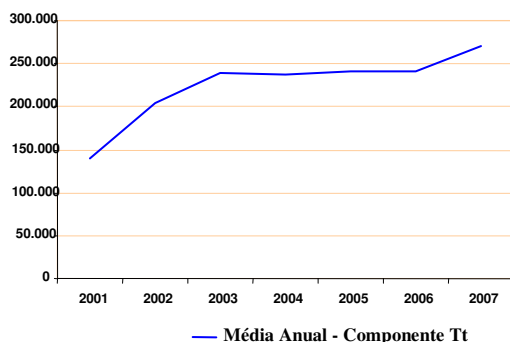
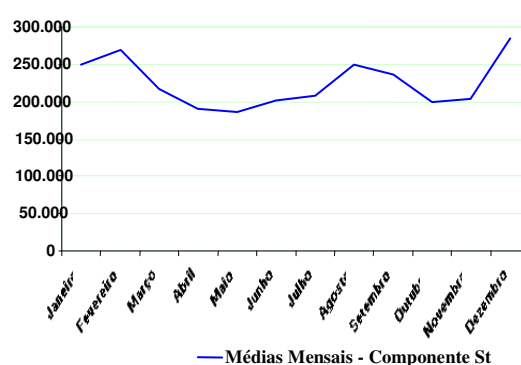
Gráfico 1: Média Anual Comp. T_t da série

Gráfico 2 - Média Mensal Comp. St da série



Comumente para se obter uma série estacionária tomam-se diferenças sucessivas da série original. Nos modelos ARIMA (p, d, q), de Box & Jenkins e, em situações onde a série temporal não apresenta comportamento explosivo, será suficiente tomar uma ou duas diferenças simples ($d = 1$ ou $d = 2$).

Analisando o Gráfico 1, constata-se a componente de Tendência na série. As diferenças sucessivas de 1º e 2º ordem são obtidas através dos seguintes comandos:

```
diff(RESSEGURO)
plot(ts(diff(RESSEGURO)),col=2,main='RESSEGURO - 1a diferença',xlab='meses',ylab='Valores em Reais (valores e milhares)') ;
diff(diff(RESSEGURO))
plot(ts(diff(diff(RESSEGURO))),col=2,main='RESSEGURO - 2a diferença',xlab='meses',ylab='Valores em Reais (valores e milhares)').
```

Os valores das diferenças sucessivas de 1ª e 2ª ordem, estão nos Gráficos 3 e 4. Pode-se observar que a tendência na série Prêmio de Resseguro foi eliminada. Os dois gráficos se assemelham, por conseguinte a decisão do número de diferenças será tomada na fase de estimação do modelo, a partir do Critério de Akaike (AIC), da Função de Verossimilhança (*Likelihood*) e da Variância dos Resíduos (AKAIKE, 1974; SHUMWAY & STOFFER, 2000; HARVEY, 1993; SOUZA & CAMARGO, 2004).

Gráfico 3: 1ª Diferença Simples da série

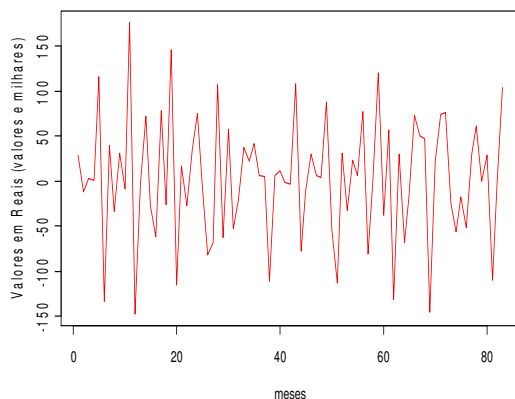
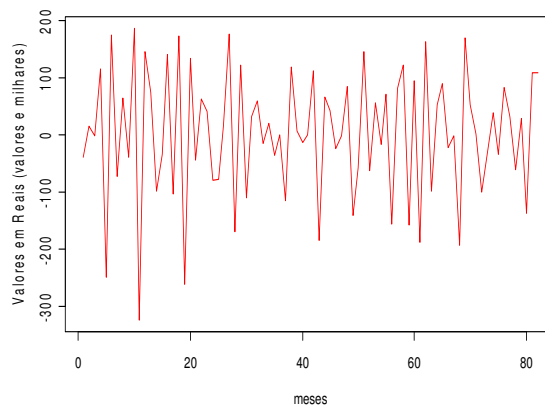


Gráfico 4: 2ª Diferença Simples da série



Analisando a Sazonalidade (S_t), na Tabela 1 e no Gráfico 2, observa-se oscilações consideráveis, concluindo-se pela existência de sazonalidade. Para eliminar a sazonalidade será adotada as diferenças sazonais,. Como na diferença simples, na diferença sazonal será suficiente tomar uma ou duas diferenças sazonais da série ($D = 1$ ou $D = 2$), ou seja, modelo SARIMA Multiplicativo $(p, d, q) \times (P, D, Q)$.

Na série em estudo, apenas uma diferença sazonal é suficiente para eliminar a sazonalidade (transformá-la em estacionária). Os valores da 1ª Diferença Sazonal, da 1ª Diferença Simples e 1ª Sazonal e da 2ª Diferença Simples e 1ª Sazonal (Gráficos 5, 6 e 7) são obtidos com seguintes comandos:

```
diff(RESSEGURO,lag=12)
plot(ts(diff(RESSEGURO),lag=12),col=2,main='RESSEGURO - 1a diferença simples e 1a
sazonal',xlab='meses',ylab='Valores em Reais (valores e milhares)') ;
diff(diff(RESSEGURO),lag=12)
plot(ts(diff(diff(RESSEGURO),lag=12)),col=2,main='RESSEGURO - 2a diferença simples e 1a
sazonal',xlab='meses',ylab='Valores em Reais (valores e milhares)') ;
diff(diff(diff(RESSEGURO)),lag=12)
plot(ts(diff(diff(diff(RESSEGURO)),lag=12)),col=2,main='RESSEGURO - 2a diferença simples e 1a
sazonal',xlab='meses',ylab='Valores em Reais (valores e milhares)') .
```

Gráfico 5: 1ª Diferença Sazonal da série

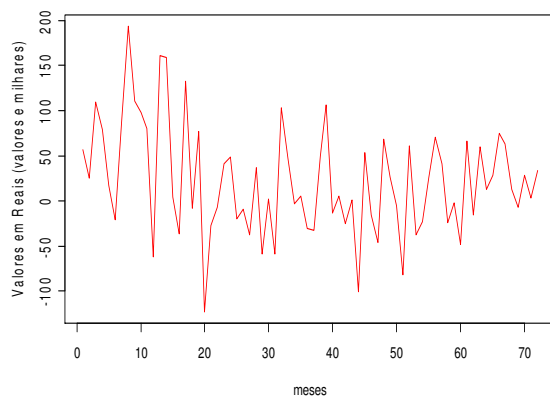


Gráfico 6: 1ª Diferença simples e 1ª Sazonal da série

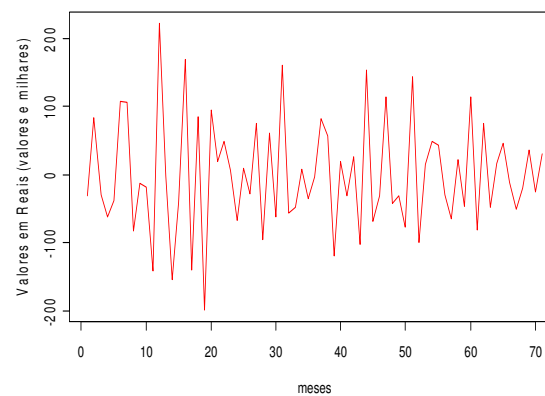
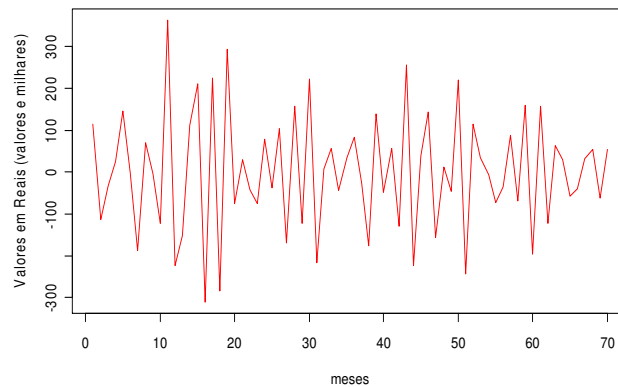


Gráfico 7: 2ª Diferença Simples e 1ª Sazonal da série

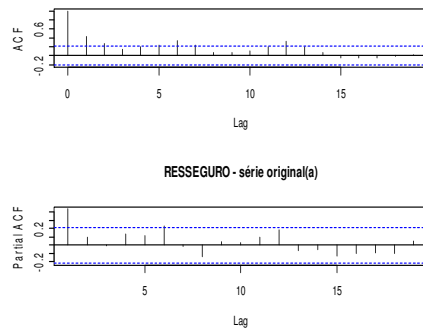
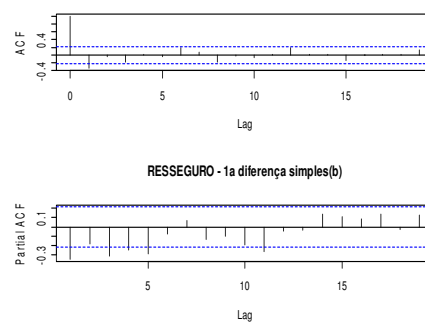
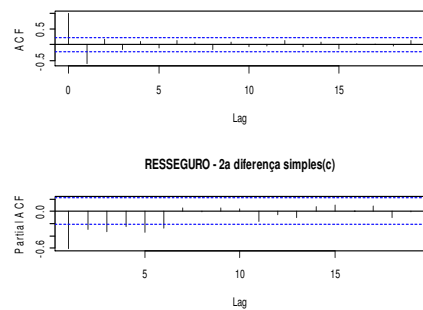
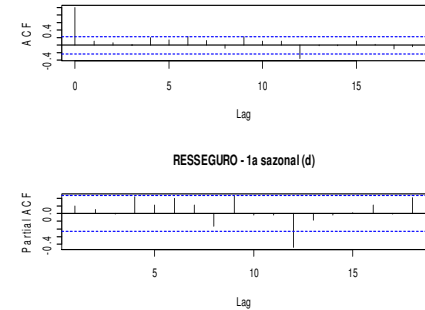
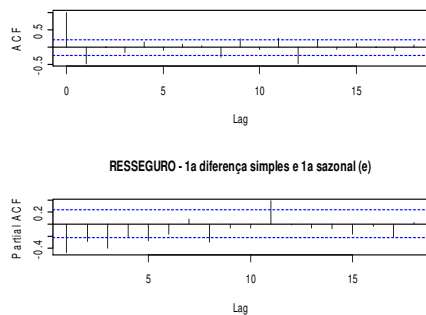
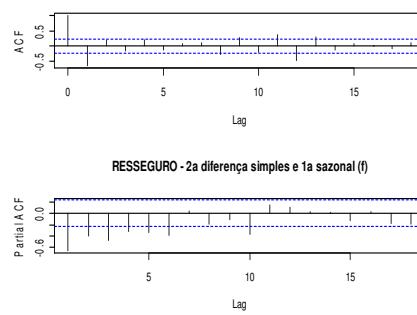


A fase (ii), identificação do(s) modelo(s), será feita por meio da análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série, consoante os valores teóricos da Tabela 1. O modelo selecionado na etapa de identificação poderá ser corrigido na etapa de estimação, quando a significância dos parâmetros será analisada, com vistas à obtenção das previsões (MEIRELES *et al.*, 2009). Deve-se nesta fase considerar o critério da parcimônia (BOX, JENKINS & REINSEL, 1994).

As estimativas das FAC e FACP da série original e da série diferenciada sazonalmente de primeira ordem (Tabela 2 e nos Gráficos de 8 a 13) foram obtidas com os seguintes comandos.

```
par(mfrow=c(2,1))
a=acf(RESSEGURO,main='RESSEGURO - série original(a)')
ap=pacf(RESSEGURO,main='RESSEGURO - série original(a)')
A
ap ;
a1=acf(diff(RESSEGURO),main='RESSEGURO - 1a diferença simples(b)')
ap1=pacf(diff(RESSEGURO),main='RESSEGURO - 1a diferença simples(b)')
a1
ap1 ;
a2=acf(diff(diff(RESSEGURO)),main='RESSEGURO - 2a diferença simples(c)')
ap2=pacf(diff(diff(RESSEGURO)),main='RESSEGURO - 2a diferença simples(c)')
a2
ap2 ;
a12=acf(diff(RESSEGURO,lag=12),main='RESSEGURO - 1a sazonal (d)')
ap12=pacf(diff(RESSEGURO,lag=12),main='RESSEGURO - 1a sazonal (d)')
a12
ap12 ;
a112=acf(diff(diff(RESSEGURO),lag=12),main='RESSEGURO - 1a diferença simples e 1a sazonal (e)')
ap112=pacf(diff(diff(RESSEGURO),lag=12),main='RESSEGURO - 1a diferença simples e 1a sazonal (e)')
a112
ap112 ;
a212=acf(diff(diff(diff(RESSEGURO)),lag=12),main='RESSEGURO - 2a diferença simples e 1a sazonal (f)')
ap212=pacf(diff(diff(diff(RESSEGURO)),lag=12),main='RESSEGURO - 2a diferença simples e 1a sazonal (f)')
a212
ap212
```

Gráfico 8: Autocorrelação (FAC) e Autocorrelação Parcial (FACP) da (a) série original (Z(t))

Gráfico 9: Autocorrelação (FAC) e Autocorrelação Parcial (FACP) da (b) 1ª Diferença Simples ($\Delta Z(t)$)Gráfico 10: Autocorrelação (FAC) e Autocorrelação Parcial (FACP) da (c) 2ª Diferença Simples ($\Delta^2 Z(t)$)Gráfico 11: Autocorrelação (FAC) e Autocorrelação Parcial (FACP) (d) 1ª Diferença Sazonal ($\Delta_s Z(t)$)Gráfico 12: Autocorrelação (FAC) e Autocorrelação Parcial (FACP) (e) 1ª Diferença Simples e 1ª Sazonal $\Delta_1 \Delta_s Z(t)$ Gráfico 13: Autocorrelação (FAC) e Autocorrelação Parcial (FACP) 2ª Diferença Simples e da (f) 1ª Sazonal ($\Delta^2 \Delta_s Z(t)$)

Na fase (iii), estimação dos parâmetros dos modelos selecionados, após a estimação das funções FAC e FACP (Tabela 2) foram selecionados inicialmente os seguintes modelos com suas devidas descrições para a fase de estimação:

AR (1)	$(1 - \phi_1 B^1) Z_t = a_t$
AR (6)	$(1 - \phi_1 B^1 - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6) Z_t = a_t$
ARI (1,1)	$(1 - \phi_1 B^1) \Delta^1 Z_t = a_t$
ARI (6,1)	$(1 - \phi_1 B^1 - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6) \Delta^1 Z_t = a_t$
IMA (1,1)	$\Delta^1 Z_t = (1 - \theta_1 B^1) a_t$
IMA (2,1)	$\Delta^2 Z_t = (1 - \theta_1 B^1) a_t;$
SAR (1,0,0) x (0,1,0)	$(1 - \phi B) \Delta_{12} Z_t = a_t$
SIMA (0,1,1) x (0,1,0)	$\Delta^1 \Delta_{12}^1 Z_t = (1 - \theta_1 B^1) a_t$
SIMA (0,2,1) x (0,1,0)	$\Delta^2 \Delta_{12}^1 Z_t = (1 - \theta_1 B^1) a_t$

Tabela 2: Autocorrelações (FAC) e Autocorrelações Parciais (FACP) da (a) série original ($Z(t)$), da (b) 1ª Diferença Simples ($\Delta Z(t)$), da (c) 2ª Diferença Simples ($\Delta^2 Z(t)$), da (d) 1ª Diferença Sazonal ($\Delta_s Z(t)$), da (e) 1ª Diferença Simples e 1ª Sazonal $\Delta_1 \Delta_s Z(t)$ e 2ª Diferença Simples e da (f) 1ª Sazonal ($\Delta_2 \Delta_s Z(t)$), da série prêmio de resseguro

	Z(t)		$\Delta Z(t)$		$\Delta^2 Z(t)$		$\Delta_s Z(t)$		$\Delta_1 \Delta_s Z(t)$		$\Delta_2 \Delta_s Z(t)$	
	a		b		c		d		e		f	
	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF
0	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000
1	0,436	0,436	-0,346	-0,346	-0,608	-0,608	0,101	0,101	-0,475	-0,475	-0,659	-0,659
2	0,273	0,102	-0,042	-0,184	0,180	-0,301	0,059	0,050	0,002	-0,290	0,208	-0,399
3	0,149	-0,003	-0,188	-0,317	-0,147	-0,328	0,002	-0,009	-0,146	-0,398	-0,154	-0,473
4	0,197	0,139	0,020	-0,249	0,102	-0,249	0,212	0,212	0,162	-0,221	0,191	-0,321
5	0,229	0,119	-0,043	-0,293	-0,111	-0,347	0,138	0,104	-0,088	-0,273	-0,144	-0,339
6	0,349	0,228	0,188	-0,077	0,131	-0,277	0,226	0,197	0,093	-0,171	0,072	-0,390
7	0,234	-0,018	0,066	0,075	0,058	0,051	0,131	0,109	0,059	0,083	0,101	0,050
8	0,070	-0,138	-0,182	-0,138	-0,159	-0,008	-0,082	-0,161	-0,282	-0,297	-0,289	-0,191
9	0,074	0,045	-0,020	-0,105	0,086	0,048	0,223	0,223	0,237	-0,064	0,274	-0,112
10	0,105	0,030	-0,073	-0,193	-0,064	0,045	0,104	-0,014	-0,060	-0,068	-0,208	-0,375
11	0,194	0,094	0,023	-0,267	-0,031	-0,164	0,098	-0,017	0,259	0,378	0,358	0,148
12	0,317	0,187	0,196	-0,047	0,127	-0,054	-0,358	-0,437	-0,464	-0,015	-0,471	0,116
13	0,196	-0,059	0,020	-0,035	-0,062	-0,102	-0,005	-0,082	0,197	-0,071	0,304	0,036
14	0,070	-0,053	0,018	0,141	0,082	0,073	-0,007	-0,021	-0,053	-0,083	-0,134	0,022
15	-0,044	-0,126	-0,149	0,110	-0,133	0,099	0,107	0,006	0,106	-0,162	0,089	-0,129
16	-0,045	-0,105	-0,014	0,089	0,039	0,004	0,034	0,107	0,015	-0,035	-0,004	0,027
17	-0,044	-0,096	0,006	0,139	0,004	0,094	-0,089	-0,005	-0,089	-0,216	-0,085	-0,184
18	-0,010	-0,103	0,017	-0,026	-0,037	-0,105	-0,031	0,208	0,072	0,017	0,103	-0,186
19	0,014	0,047	0,139	0,125	0,138	-0,008	0,000	0,000	0,000	0,000	0,000	0,000

As funções de FAC e FACP dos resíduos dos modelos são importantes na verificação se os mesmos são ruído branco – fase (iv). Das referidas análises, foram selecionados dois modelos: modelo ARI(5,1) e o modelo IMA(1,1), (Tabela 3).

Tabela 3: Estimativas dos parâmetros dos modelos selecionados: ARI (5,1) e IMA (1,1) da série

Modelos	Estimativa dos Coeficientes	Função de Verossimilhança (Likelihood)	Critério de Akaike (AIC)	Variância da Estimativa dos Coeficientes	Erro Padrão das Estimativas dos Coeficientes	Variância dos Resíduos
ARI(5,1)	$\phi_1 = -0,6273$	-449,73	911,46	2.934	0,1058	2.887,37
	$\phi_2 = -0,5231$				0,1176	
	$\phi_3 = -0,5430$				0,1155	
	$\phi_4 = -0,4122$				0,1176	
	$\phi_5 = -0,3015$				0,1071	
IMA (1,1)	$\theta_1 = -0,8142$	-453,36	910,72	3.208	0,0652	3.111,69

Os comandos necessários são dados por:

```
ari51=arima(RESSEGURO,order=c(5,1,0))
ari51
residuals(ari51)
var(residuals(ari51))
resari51=acf(ari51$residuals,main='Autocorrelação dos Resíduos - Modelo ARI(5,1)')
respari51=pacf(ari51$residuals,main='Autocorrelação Parcial dos Resíduos - Modelo ARI(5,1)')
resari51
respari51 ;
ima11=arima(RESSEGURO,order=c(0,1,1))
ima11
residuals(ima11)
var(residuals(ima11))
resima11=acf(ima11$residuals,main='Autocorrelação dos Resíduos - Modelo IMA(1,1)')
```

```
respima11=pacf(ima11$residuals,main='Autocorrelação Parcial dos Resíduos - Modelo IMA(1,1)')
resima11
respima11 .
```

Os valores são apresentados nos Gráficos 14 e 15.

Gráfico 14: Autocorrelações (FAC) e Autocorrelações Parciais (FACP) residuais do modelo: (a) ARI (5,1) da série

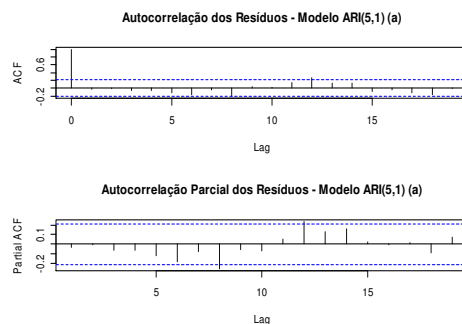
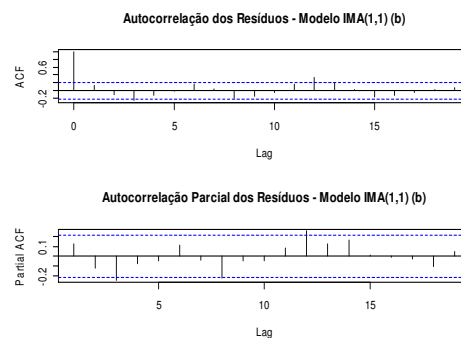


Gráfico 15: Autocorrelações (FAC) e Autocorrelações Parciais (FACP) residuais dos modelos: IMA (1,1) da série



Na fase (iv) – testes de verificação dos modelos – a partir da análise da FAC e da FACP dos resíduos dos modelos, é confirmada a hipótese de que os mesmos representam estimativas de ruído branco.

A próxima etapa é a fase de previsão, onde ocorrerá a decisão pelo melhor modelo. Como critérios, foram utilizados o EMQ da previsão, valor de janeiro até dezembro de 2008, separados no início da análise e não utilizados nas fases de identificação e estimação, de forma comparar, sem tendenciosidade, as previsões e também o EQM do limite superior dos valores da previsão (média do quadrado da diferença entre cada limite superior do valor previsto e o valor real observado) como critério probabilístico de decisão, a partir do menor erro padrão da estimativa.

Quanto menor for o EQM, melhor será o grau de ajustamento do modelo para os dados da série temporal. Os comandos básicos das previsões, dos 2 modelos 12 passos à frente e seus respectivos limites superiores e inferiores, são apresentados a seguir:

```
prevari51=predict(ari51,n.ahead=12,se.fit=T)
prevari51 ;
liari51=prevari51$pred-1.96*prevari51$se
liari51 ;
lsari51=prevari51$pred+1.96*prevari51$se
lsari51 ;
previma11=predict(ima11,n.ahead=12,se.fit=T)
previma11 ;
liima11=previma11$pred-1.96*previma11$se
liima11 ;
lsima11=previma11$pred+1.96*previma11$se
lsima11 .
```

Os comandos das previsões um passo à frente requerem mais aprofundamento, já que os valores referentes ao ano de 2008 foram excluídos das respectivas fases anteriores, de forma a não haver redundância na seleção do modelo na fase de previsão. Com os dados do ano de 2008, estimam-se, para cada novo dado, os novos parâmetros para os modelos e, posteriormente, obtêm-se as previsões e seus limites de confiança.

Os valores que serão levados em consideração, tanto para as previsões quanto para o cálculo dos limites, serão os primeiros valores retornados. Como exemplo, o primeiro dado referente ao ano de 2008 é o mês de janeiro.

O arquivo 'projetofinal1.txt' é o novo arquivo de dados, salvo com o primeiro dado do ano de 2008 -231.328 - sendo assim repete-se este procedimento a cada passo até a última observação, referente ao mês de dezembro de 2008. Seguindo os comandos de entrada de dados, estimação do modelo com o novo dado, previsão, limite superior e limite inferior, lembrando que os valores retornados e são sempre os primeiro valores, sublinhados nesta primeira interação.

```
x1=read.table('projetofinal1.txt',head=T)
attach(x1)
names(x1)
[1] "RESSEGURO"
RESSEGURO
[1] 94.400 122.719 110.706 113.372 114.044 229.749 95.772 135.221 101.839
[10] 132.209 122.853 298.744 151.357 148.462 220.101 192.725 131.031 208.893
[19] 182.622 328.293 212.464 229.653 202.419 237.201 311.806 306.934 224.254
[28] 155.987 263.550 200.763 258.990 206.041 185.079 221.988 243.856 285.778
[37] 292.186 297.411 186.652 193.167 205.084 203.405 200.298 308.577 230.973
[46] 219.486 249.566 255.434 259.378 346.757 292.361 179.206 210.351 178.104
[55] 201.548 207.876 284.457 203.825 203.352 323.484 285.264 341.893 210.487
[64] 240.621 172.327 155.456 228.120 278.024 325.535 179.514 201.553 275.696
[73] 351.535 326.501 270.676 253.098 200.744 230.187 291.207 290.359 318.500
[82] 208.319 205.051 309.143 231.328
ari151=arima(RESSEGURO,order=c(5,1,0))
ari151
Call:
arima(x = RESSEGURO, order = c(5, 1, 0))
Coefficients:
    ar1    ar2    ar3    ar4    ar5
-0.6435 -0.5220 -0.5259 -0.4142 -0.3013
s.e.  0.1049  0.1177  0.1149  0.1180  0.1073
sigma^2 estimated as 2944: log likelihood = -455.27, aic = 922.54
$pred
Time Series:
Start = 86
End = 96
Frequency = 1
[1] 265.9507 264.0926 246.0104 241.2772 263.8429 251.6392 258.2532 255.9075
[9] 252.4623 250.6813 255.7967
$se
Time Series:
Start = 86
End = 96
```

```
Frequency = 1
[1] 54.26164 57.60592 59.16483 59.57120 60.42162 61.73765 66.89636 69.49085
[9] 71.25669 72.33695 73.61075
liari151=prevari151$pred-1.96*prevari151$se
liari151
Time Series:
Start = 86
End = 96
Frequency = 1
[1] 159.5978 151.1850 130.0473 124.5177 145.4165 130.6334 127.1363 119.7055
[9] 112.7992 108.9009 111.5196
lsari151=prevari151$pred+1.96*prevari151$se
lsari151
Time Series:
Start = 86
End = 96
Frequency = 1
[1] 372.3035 377.0002 361.9734 358.0368 382.2692 372.6450 389.3701 392.1096
[9] 392.1254 392.4617 400.0737
```

Nas Tabelas 4 a 8 e nos Gráficos 16 a 19 estão os resultados dos modelos ARI (5,1) e IMA (1,1), com os valores previstos doze passos à frente e um passo à frente, e os valores de EQM da previsão, o EQM do limite superior dos valores da previsão.

Tabela 4: Valores da autocorrelação (FAC) e autocorrelação parcial (FACP) residual, estimados dos modelos: ARI (5,1) e IMA (1,1) da série

ARI (5,1)	A	ACF	1	0,033	0,007	0,061	0,057	0,116	-0,17	-0,04	0,207	0,038	0,016	0,152	0,277	0,125	0,123	0,078	0,055	-0,11	0,169	0,021
		PACF	0	0,033	0,009	0,061	0,062	0,122	0,189	0,076	-0,26	0,055	0,072	0,053	0,238	0,135	0,164	0,025	0,008	0,012	0,091	0,073
IMA (1,1)	B	ACF	1	0,128	0,102	0,263	-0,12	0,013	0,174	0,041	0,203	0,141	0,048	0,161	0,336	0,185	0,022	0,165	0,124	0,055	0,014	0,069
		PACF	0	0,128	0,121	0,241	0,074	80	0,11	0,042	0,223	-0,05	0,048	0,085	0,252	0,125	0,164	0,012	0,012	0,027	0,102	0,049

Tabela 5: Valores da previsão doze passos à frente

[illegible]

Gráfico 16: Série prêmio de resseguro mais a previsão doze passos à frente com seus limites

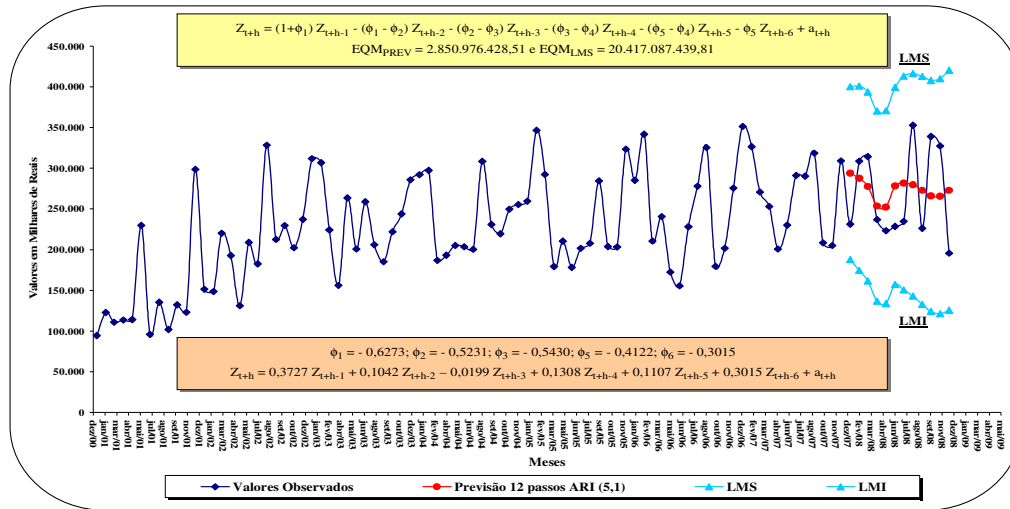


Tabela 6: Valores da previsão um passo à frente

T	Z _{t+h}	Z _{t+h} (-)	Z _t (h)	Z _{t+h} (+)	s do erro de Previsão	et (h)	et (h) LS
85	231.328	187.866,90	294.035,50	400.204,20	54.167,67	-62.707,24	-168.875,94
86	308.783	159.597,80	265.950,70	372.303,50	54.261,64	42.832,17	-63.520,63
87	314.382	171.882,50	277.952,80	384.023,00	54.117,49	36.429,05	-69.641,15
88	236.982	161.577,60	267.279,30	372.981,00	53.929,44	-30.297,54	-135.999,24
89	223.198	139.792,60	245.044,40	350.296,10	53.699,88	-21.846,47	-127.098,17
90	228.493	156.853,60	261.587,70	366.321,90	53.435,78	-33.094,41	-137.828,61
91	234.652	141.528,60	245.899,90	350.271,10	53.250,64	-11.247,45	-115.618,65
92	352.895	163.903,10	267.717,30	371.531,50	52.966,43	85.177,37	-18.636,83
93	226.232	197.036,40	301.659,60	406.282,80	53.379,19	-75.427,14	-180.050,34
94	339.073	141.424,50	246.572,80	351.721,00	53.647,07	92.500,60	-12.647,60
95	327.451	152.138,90	258.317,40	364.495,90	54.172,70	69.133,47	-37.045,03
96	195.644	179.696,80	286.200,20	392.703,60	54.338,48	-90.555,74	-197.059,14
EQM _{PREV} =	3.682.616.669,18						
EQM _{LMS} =	14.795.669.478,37						

Gráfico 17: Série Prêmio de Resseguro mais a previsão um passo à frente com seus limites

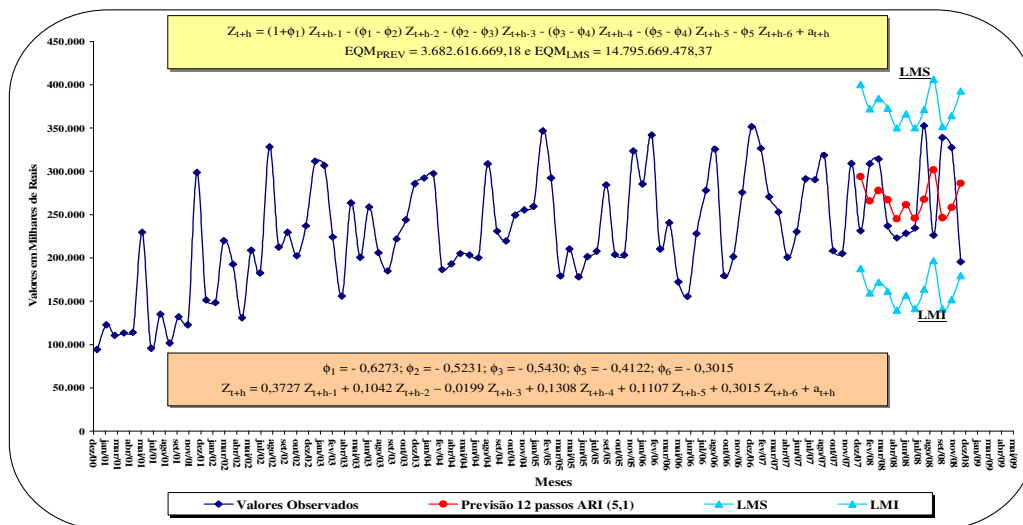


Tabela 7: Valores da previsão doze passos à frente

[illegible]

Gráfico 18: Série Prêmio de Resseguro mais a previsão doze passos à frente com seus limites

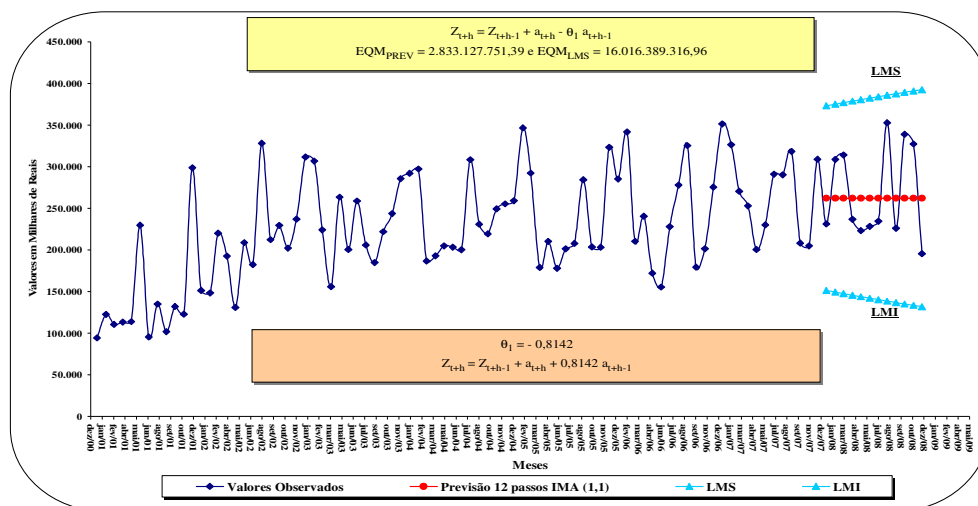
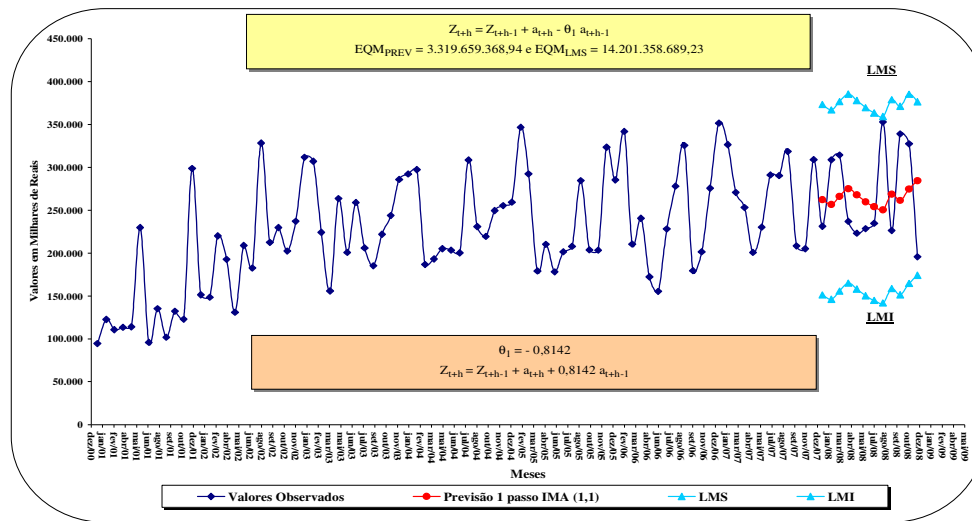


Tabela 8: Valores da previsão um passo à frente

[illegible]

Gráfico 19: Série Prêmio de Resseguro mais a previsão um passo à frente com seus limites.



Os EQMs da previsão e do limite superior da previsão do modelo IMA (1,1) possuem valores menores do que modelo ARI (5,1), tanto 12 passos à frente, quanto um passo à frente. Outro critério em que o modelo IMA (1,1) também se enquadra, diz respeito à parcimônia, logo o modelo IMA (1,1) é o que melhor fornece previsões.

4. Considerações Finais

A proposta dos autores no artigo foi mostrar uma aplicação da metodologia de Box & Jenkins, na obtenção de previsões de uma série temporal, Prêmio de Resseguro, pela utilização do software R, referente ao período janeiro de 2001 a dezembro de 2008. Cumpridos os estágios do ciclo iterativo da metodologia, a série foi considerada estacionária após a eliminação das componentes de Tendência e Sazonalidade, por meio das diferenças simples e sazonal de primeira ordem.

Na análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) foram selecionados inicialmente, na fase (ii) da metodologia, dez modelos; estes foram levados às fases seguintes (iii) e (iv), de onde foram selecionados dois modelos: autorregressivo integrado ARI(5,1) e médias móveis IMA(1,1); os dois foram, então, levados para a fase de previsão dos valores mensais de 2008, separados no início da análise e não utilizados nas fases de identificação e estimação, de forma a melhor comparar, sem tendência e redundância, as previsões com os valores observados.

Na fase de previsão foram utilizados os critérios do Erro Quadrático Médio (EQM) da previsão, de janeiro até dezembro de 2008, e o EQM do limite superior dos valores da previsão de forma a se ter um critério probabilístico de decisão, a partir do

menor erro padrão da estimativa. Na análise final dos resultados, no modelo IMA (1,1) os valores dos EQMs das previsões foram menores que os do modelo ARI (5,1), tanto previsões 12 passos à frente, quanto um passo à frente. Além disso, o modelo IMA (1,1) atende ao critério da parcimônia.

Deve-se destacar que estão apresentados no artigo, ao longo da aplicação, os comandos básicos de entrada, representação gráfica, identificação, estimação e previsão dos dados, pelo uso do software R.

Referências

AKAIKE, H. A New Look at the Statistical Model Identification. **IEEE Transactions on Automatic Control**, Boston, v.19, n.6, p.716-723, 1974.

BOX, G. E. P., JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. New York: Holden Day, 1976.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. **Time Series Analysis: Forecasting and Control**. 3.ed. New Jersey: Prentice Hall, 1994.

FOX J. The R Commander: A Basic-Statistics Graphical User Interface to R. **Journal of Statistical Software**, **14(9)**, 1–42. 2005.

HARVEY, A. C., **Time Series Models**, 2nd Edition, Harvester Wheatsheaf, Londres, 1993.

MEIRELES, E. A. A., RESENDE, M. T., MORAES, R. F.. Resseguro: Monopólio, Cenário Atual e Expectativas para o Mercado Brasileiro. **Monografia de Graduação** do Bacharelado em Estatística, IME/UERJ, 2009.

MINEO, A. M., PONTILO, A. Using R via PHP for Teaching Purposes: R-php. **Journal of Statistical Software**, v. **17**, issue **4**. 2006.

MORETTIN, P. A., TOLLOI, C. M. C. **Análise de Séries Temporais**, Ed. Edgard Blucher, 2ª Ed., 2006.

PETERNELLI, L. A., FERREIRA, F. M., ROCHA, R. B., BARROS, W. S., BARBOSA, M. H. P. Analysis of The Inbreeding and Coancestry Coefficients for Any Ploidy Level using the Statistical Package R. **Bragantia** vol.68 no.4, Campinas, 2009.

R PROJECT. **The R Project for Statistical Computing**. Disponível em: <<http://www.r-project.org/>> Acesso em: dezembro de 2010.

RIBEIRO, R. L. F., MINGOTI, S. L. Apresentando um Programa Computacional em R e Geo-R para Análise Espaço-Temporal via Geoestatística. In. VIII SIMPEP, 2006, Bauru, **Anais...SP**: UNESP, 2006.

SHUMWAY, R. H., STOFFER, D. S. **Time Series Analysis and Its Applications**. New York: Springer, 2nd Edition, 2000.

SOUZA, R. C., CAMARGO, M. E. **Análise e Previsão de Séries Temporais: Os Modelos ARIMA**. Gráfica e Editora Regional, 2ª ed., Rio de Janeiro, 2004.

VALENÇA, D. M., SANTOS, P. B. Estimação em Modelos Paramétricos para Dados de Sobrevida Correlacionados no Software R: Simulação e Aplicação. In: XLII SBPO 2010, Bento Gonçalves. **Anais...RS**: UFSM, 2010.

AN USAGE OF STATISTICAL SOFTWARE R SERIES IN TIME SERIES FORECASTING: REINSURANCE PREMIUM SERIES

Abstract

Our objective with this paper is the usage of Box & Jenkins' SARIMA multiplicative family stochastic models for analysis and forecasting of time series. To that end we employ the R Series, an open source statistic software suite adopted by many universities. The Reinsure Premium time series ranging from January, 2001 through December, 2008 were used for the software application and features use case study in identification and model forecasting. In this way, we describe the models used, principles of R Series programming language, results and final considerations.

Key-words: *Time Series, Model Identification and Estimation, Forecasting, R Series Software.*