

## 4 Aplicações

A fim de se comparar o desempenho e a aderência dos modelos apresentados no capítulo 3, três bases de dados serão utilizadas. A primeira base, apresentada na tabela 4.1, foi amplamente estudada na literatura de Reservas IBNR (cf. Mack, 1993; England & Verrall, 2002; de Jong, 2006; entre outros), e, aqui, será denotada pela forma abreviada AFG. A segunda, mostrada na tabela 4.2, também foi estudada em Taylor & Ashe (1983), Verrall (1991) e Mack (1993), e que será chamada de MC1. Por fim, a terceira base, apresentada na tabela 4.3, foi utilizada no trabalho de de Jong & Zehwirth (1983). Esta base, aqui chamada por DJZ, refere-se aos montantes pagos, em milhar de libra, por um seguradora inglesa não identificada.

Tabela 4.1: Dívidas gerais facultativas (excluindo ambiental) do estudo da evolução histórica de perdas (1991 - unidade milhar de dólar)–AFG.

Ano de Origem <i>w</i>	Desenvolvimento <i>d</i>									
	0	1	2	3	4	5	6	7	8	9
1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	106	4179	1111	5270	3116	1817	-103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

As séries univariadas criadas a partir do “empilhamento” dos anos de acidente das bases AFG, MC1 e DJZ estão representadas nas figuras 4.1(a), 4.1(b) e 4.1(c), respectivamente, em suas escalas originais, e nas figuras 4.2(a), 4.2(b) e 4.2(c) em escala logarítmica. Em uma primeira análise gráfica da série DJZ, notam-se claros indícios de periodicidade: o padrão de decaimento dos valores da série parece repetir-se a cada ano de acidente. Já nas séries AFG e

Tabela 4.2: Triângulo de *runoff* extraído de Mack (1993).

Ano de Origem <i>w</i>	Desenvolvimento <i>d</i>									
	0	1	2	3	4	5	6	7	8	9
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	
3	290507	1001799	926219	1016654	750816	146922	495992	280405		
4	310608	1108250	776189	1562400	272482	352053	206286			
5	443160	693190	991983	769488	504851	470639				
6	396132	937085	847498	805037	705960					
7	440832	847631	1131398	1063269						
8	359480	1061648	1443370							
9	376686	986608								
10	344014									

Tabela 4.3: Triângulo de *runoff* para uma seguradora inglesa (unidade milhar de libra)–DJZ.

Ano de Origem <i>w</i>	Desenvolvimento <i>d</i>				
	0	1	2	3	4
1	753.5	648.9	311.7	173.5	71.3
2	642.3	648.4	249.7	206.5	
3	715.8	661.1	309.4		
4	841.6	862.6			
5	968.8				

MC1 esse comportamento não fica tão evidente: cada ano de acidente aparenta ter padrões distintos um dos outros.

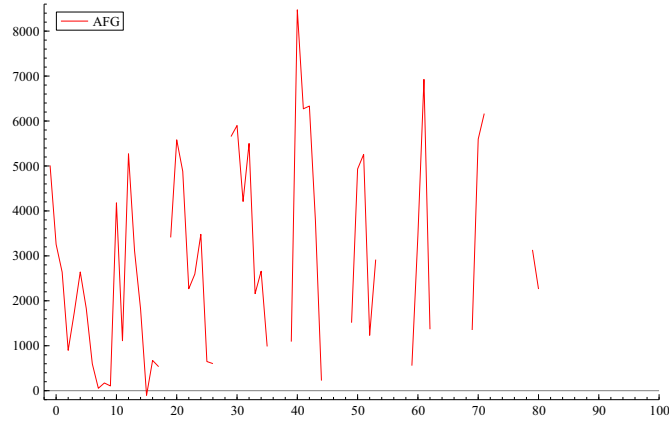
O número significativo de valores faltantes do triângulo pode dificultar a estimação da função de verossimilhança, pois a informação de Fisher correspondente, que é diretamente ligada à curvatura (cf. Migon & Gamerman, 2001, cap.2), perde contribuições. Com o intuito de contornar parcialmente este problema, decidiu-se pela combinação entre o algoritmo *Expectation-Maximization* (EM) adaptado ao modelo em EE (cf. Durbin & Koopman, 2001; e Shumway & Stoffer, 2006) com o otimizador do tipo *quasi-Newton* BFGS para a maximização da verossimilhança.

O algoritmo EM é um procedimento iterativo composto por dois passos: o primeiro passo envolve a avaliação da esperança da densidade  $p(\mathbf{Y}, \boldsymbol{\alpha} | \boldsymbol{\Psi})$  condicionada a  $p(\boldsymbol{\alpha} | \mathbf{Y}, \tilde{\boldsymbol{\Psi}})$ , com  $\boldsymbol{\Psi} = (\sigma_\varepsilon^2, \sigma_\mu^2, \sigma_\gamma^2)'$  e  $\tilde{\boldsymbol{\Psi}}$  sendo os hiperparâmetros da iteração corrente; e a segunda etapa consiste na maximização desta esperança com relação aos elementos de  $\boldsymbol{\Psi}$ . As expressões dos valores de  $\boldsymbol{\Psi}$  para cada iteração, no caso particular de um modelo estrutural de nível local com peri-

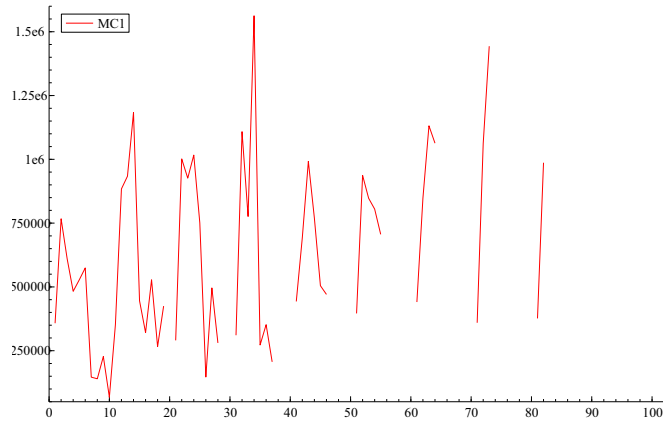
odacidade (sazonalidade), são dadas por:

$$\begin{aligned}\bar{\sigma}_\varepsilon^2 &= \frac{1}{n} \sum_{t=1}^n \{\hat{\varepsilon}_t^2 - \text{Var}(\varepsilon_t|\mathbf{Y})\} \\ \bar{\sigma}_\mu^2 &= \frac{1}{n-1} \sum_{t=2}^n \{\hat{\mu}_{t-1}^2 - \text{Var}(\mu_{t-1}|\mathbf{Y})\} \\ \bar{\sigma}_\gamma^2 &= \frac{1}{n-1} \sum_{t=2}^n \{\hat{\gamma}_{t-1}^2 - \text{Var}(\gamma_{t-1}|\mathbf{Y})\}\end{aligned}$$

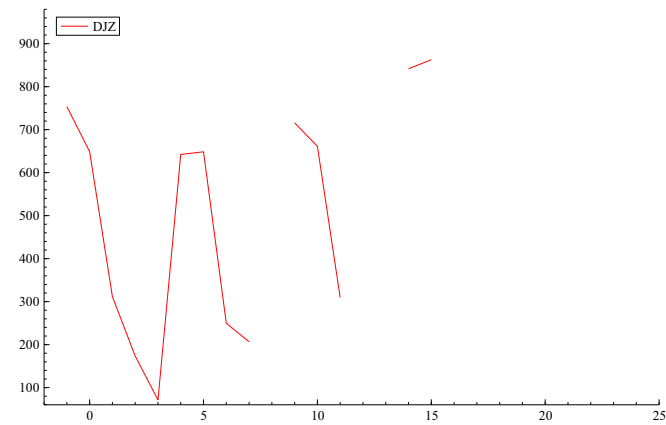
Na sequência, serão apresentados os resultados das implementações dos métodos aplicados nas três bases de dados. A estimação dos modelos foi realizada utilizando-se a linguagem *Ox* com o pacote de Espaço de Estado *SsfPack* (cf. Doornik, 2001 e Koopman *et al.*, 1999).



4.1(a): Dados AFG

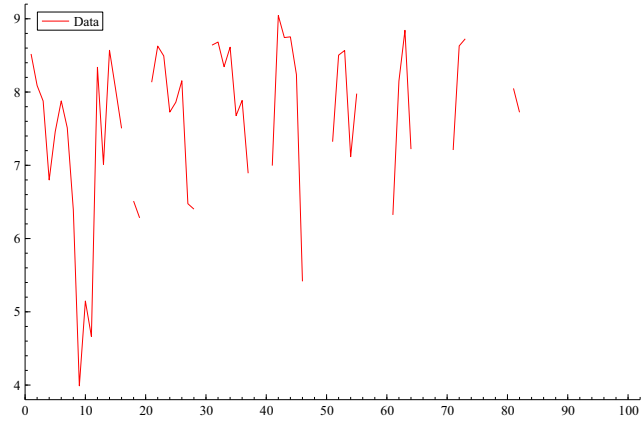


4.1(b): Dados MC1

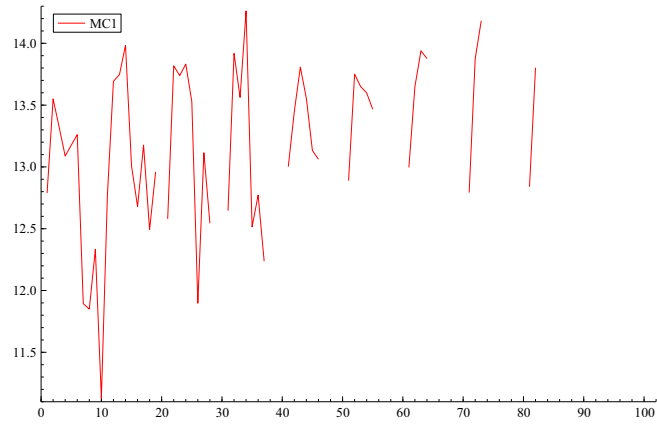


4.1(c): Dados DJZ

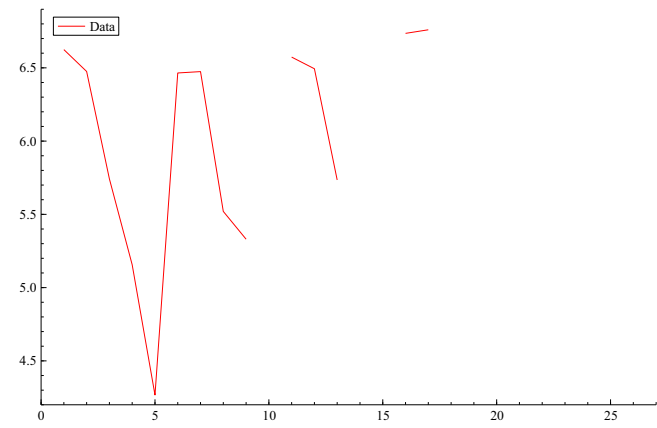
Figura 4.1: Séries empilhadas na escala real.



4.2(a): Dados AFG



4.2(b): Dados MC1



4.2(c): Dados DJZ

Figura 4.2: Séries empilhadas na escala logarítmica.

## 4.1

**Série AFG: resultados**

A estimação dos parâmetros do modelo para os dados AFG, na escala original, revelou nível e periodicidade estocásticos, enquanto que para a escala logarítmica estas mesmas componentes mostraram-se determinísticas, de acordo com as desprezíveis relações sinal/ruído mostradas na Tabela 4.4. Tal mudança de comportamento, no que diz respeito às componentes de nível das escalas analisadas, pode ser justificado por causa de a razão sinal/ruído do dado AFG em escala original ter se mostrado não muito “alta”, facilitando assim a eliminação de variabilidade pela transformação logarítmica, compressora natural de escala. Vale ainda dizer que este fenômeno não compromete e tampouco torna conflitantes as análises particulares para cada escala, pois o comportamento estocástico do nível da série em escala original, no máximo, indicou um fraquíssimo movimento não-constante entre os 1º e 5º anos de acidente (cf. Figura 4.3). Quanto à periodicidade, apenas confirma-se o mesmo padrão de decaimento já esperado, assim, como antes, obtido em Atherino & Fernandes (2007), para cada ano de origem.

Tabela 4.4: Parâmetros estimados e relações sinal/ruído para a base de dados AFG.

Parâmetro	Escala	
	(original)	(log)
Log-ver.	-407.41	-62.96
$\sigma_\varepsilon^2$	$2.15 \times 10^6$	$6.59 \times 10^{-1}$
$\sigma_\xi^2$	$1.64 \times 10^4$	$1.82 \times 10^{-13}$
$\sigma_\omega^2$	$2.05 \times 10^5$	$2.39 \times 10^{-10}$
$\sigma_\xi^2/\sigma_\varepsilon^2$	$7.62 \times 10^{-3}$	$2.77 \times 10^{-13}$
$\sigma_\omega^2/\sigma_\varepsilon^2$	$9.55 \times 10^{-2}$	$3.63 \times 10^{-10}$

As reservas IBNR estimadas estão expostas na Tabela 4.7.

Tabela 4.5: Comparação entre os modelos (dados AFG).

	Escala		
	(original)	(log)	Chain Ladder
MAPE (%)	87.13	127.47	127.82
EQM	$3.51 \times 10^6$	$4.87 \times 10^6$	$9.65 \times 10^6$
Pseudo R2 (%)	31.92	23.61	12.05
AIC	15.29	2.76	-
BIC	15.76	3.24	-

Quanto aos diagnósticos, constata-se bom comportamento das inovações para o modelo na escala original, sem indícios de correlação serial, como está

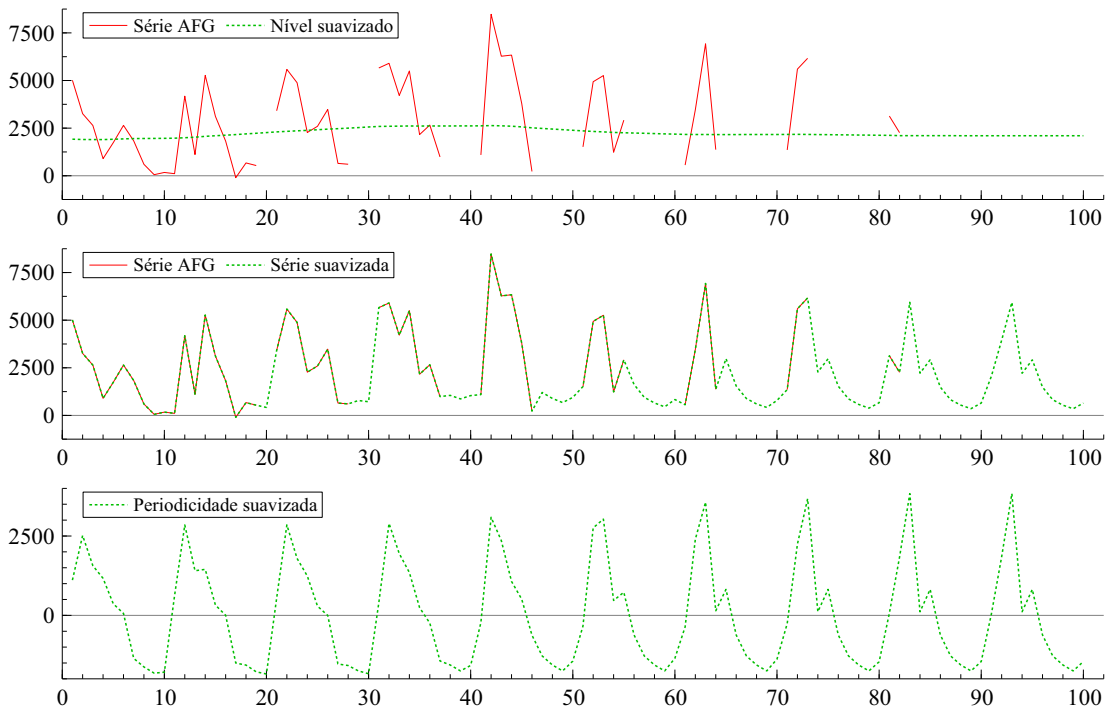


Figura 4.3: Resultados do modelo estrutural nos dados sem transformação (dados AFG).

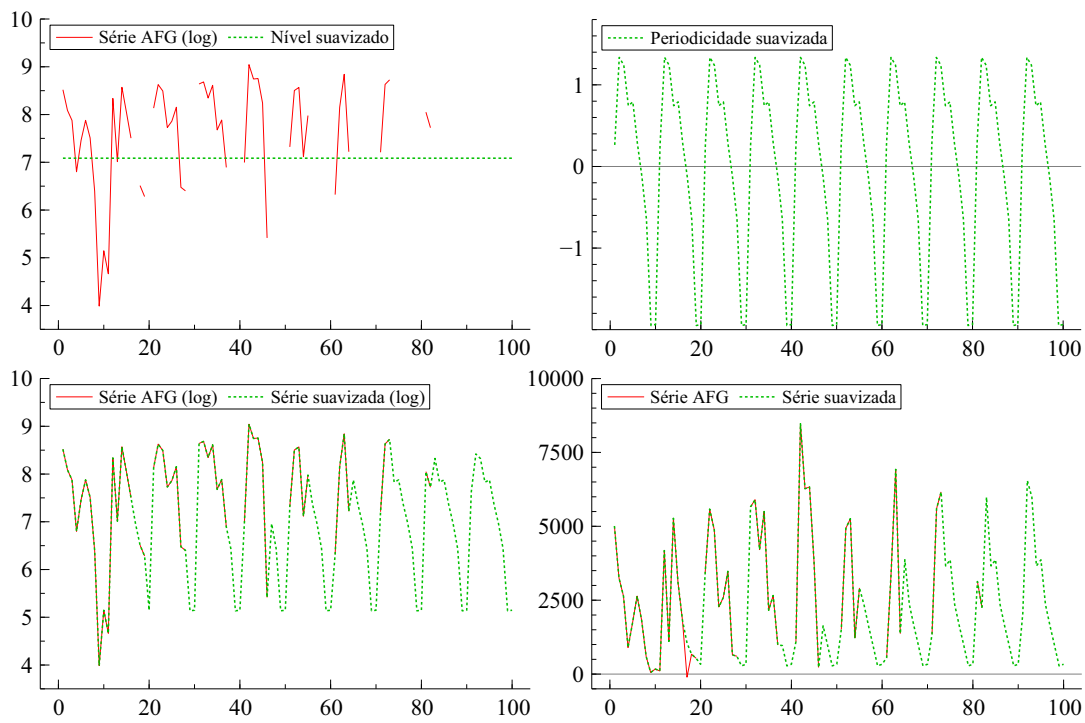


Figura 4.4: Resultados do modelo estrutural nos dados com transformação (dados AFG).

Tabela 4.6: Comparação entre os modelos (dados AFG) – Out of Sample.

	Escala		
	(original)	(log)	Chain Ladder
MAPE (%)	206.26	177.94	246.85
EQM	$2.44 \times 10^6$	$2.53 \times 10^6$	$8.45 \times 10^6$
Pseudo R2 (%)	57.62	36.23	49.01

Tabela 4.7: Reservas AFG calculadas e CV em %.

Ano de Origem	Chain Ladder	Mod. Est. (original)	Mod. Est. (log)
2	154 (134.0%)	733 (290.6%)	1966 (102.4%)
3	617 (101.0%)	2011 (171.6%)	611 (101.6%)
4	1636 (45.7%)	3584 (125.3%)	1579 (80.3%)
5	2747 (53.5%)	4378 (125.0%)	3213 (68.1%)
6	3649 (54.9%)	5201 (124.2%)	5565 (58.5%)
7	5435 (40.6%)	7867 (95.3%)	9433 (54.3%)
8	10907 (49.1%)	9961 (86.7%)	13093 (47.7%)
9	10650 (59.5%)	15472 (64.2%)	19076 (45.4%)
10	16339 (150.4%)	19447 (58.5%)	25624 (42.1%)
Total	52135 (51.6%)	68654 (48.9%)	80159 (24.2%)

retratado em suas FAC e FACP (cf. Figura 4.5). O modelo em log também apresentou o mesmo comportamento (cf. Figura 4.6). Segundo a própria Tabela 4.5, ambos os modelos superaram o método *Chain-Ladder* tanto em erro médio quadrático quanto em pseudo-R2. Entretanto, o comportamento dos resíduos auxiliares<sup>1</sup> – um instrumento para detecção de observações aberrantes – apontou a presença de observações *outliers*, tornando-se necessária uma nova análise introduzindo intervenções ao modelo. Essa análise será discutida em detalhes na próxima subseção.

#### 4.1.1

##### Análise de intervenções nos dados AFG

Para cada uma das escalas, original e em log, foram estimados dois modelos com intervenções (totalizando assim, com os modelos previamente estimados, seis modelos): um com “menos intervenções”; e o outro com “mais intervenções”. Com o intuito de facilitar a análise, deste ponto em diante, cada modelo será referenciado seguindo a nomenclatura abaixo:

- Modelo I-a — modelo em escala original sem intervenções;
- Modelo I-b — modelo em escala original com 5 intervenções;

<sup>1</sup>Foram considerados *outliers* todas as observações cujos resíduos auxiliares superaram três unidades em valor absoluto.



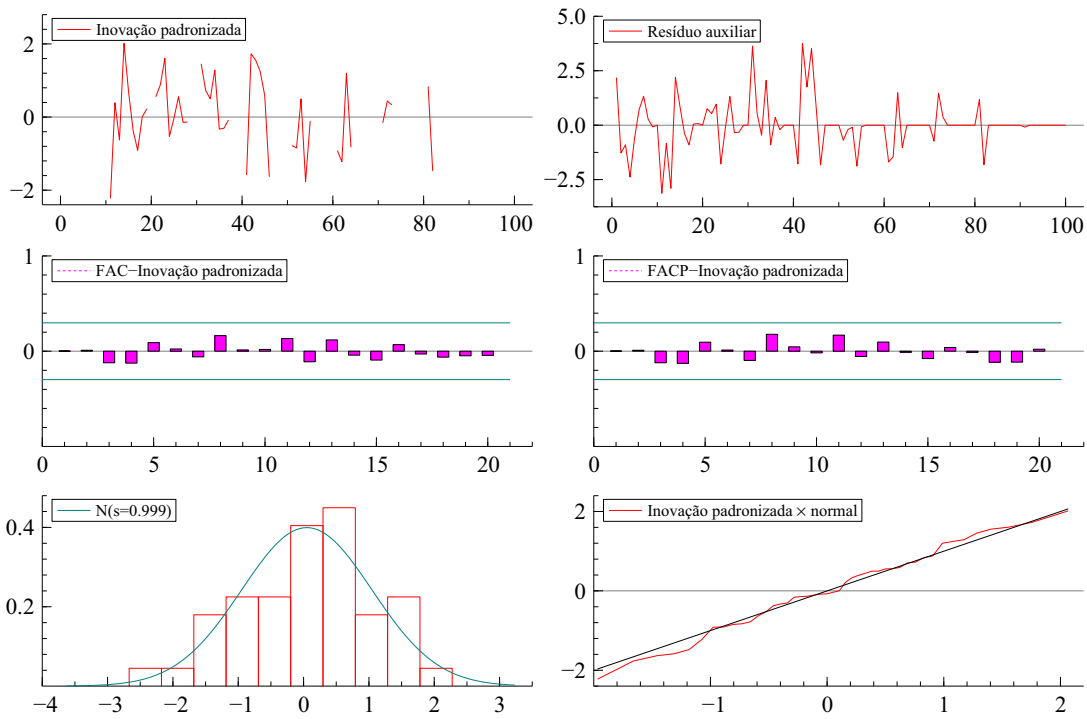


Figura 4.5: Diagnósticos – modelo estrutural nos dados AFG sem transformação.

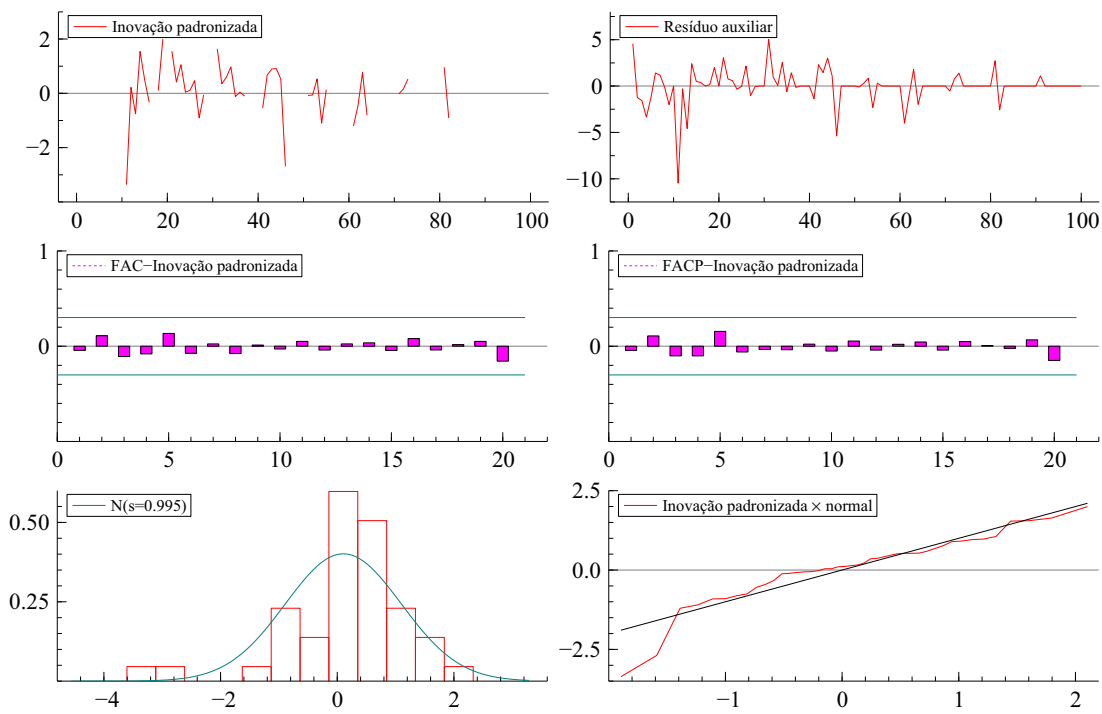


Figura 4.6: Diagnósticos – modelo estrutural nos dados AFG com transformação.

- Modelo I-c — modelo em escala original com 8 intervenções;
- Modelo II-a — modelo em escala log sem intervenções;
- Modelo II-b — modelo em escala log com 7 intervenções;
- Modelo II-c — modelo em escala log com 10 intervenções.

A Tabela 4.8 mostra os instantes que sofreram intervenções em cada um dos modelos.

Tabela 4.8: Intervenções em cada modelo para os dados AFG.

Modelo	# Interv.	Instantes
I-a	0	NA
I-b	5	11,13,31,42 e 44
I-c	8	4,11,13,14,31,34,42 e 44
II-a	0	NA
II-b	7	11,13,21,31,44,46 e 60
II-c	10	4,9,11,13,21,31,34,44,46 e 61

Primeiramente, serão analisados os parâmetros estimados de cada modelo. Segundo a Tabela 4.9, há indícios de que a componente de nível é determinística para todos os modelos, já que foram estimados valores próximos de zero para a variância da componente de nível ( $\sigma_\xi^2$ ) (excetuando-se o modelo I-a). Entretanto, apesar de o nível ter mostrado comportamento determinístico, a componente periódica manteve-se com um padrão estocástico, de acordo com as relações sinal/ruído expressivamente não-nulas (0.24 para o modelo I-b e 1.23 para o I-c).

Tabela 4.9: Parâmetros estimados e relações sinal/ruído para a base de dados AFG - modelos com intervenções.

Parâmetro	Escala			
	(normal)		(log)	
	I-b	I-c	II-b	II-c
Log-ver.	-392.66	-380.27	-39.89	-17.39
$\sigma_\varepsilon^2$	$9.89 \times 10^5$	$3.00 \times 10^5$	$1.99 \times 10^{-1}$	$9.06 \times 10^{-187}$
$\sigma_\xi^2$	$1.03 \times 10^{-4}$	0	$9.89 \times 10^{-13}$	$1.64 \times 10^{-4}$
$\sigma_\omega^2$	$2.37 \times 10^5$	$3.68 \times 10^5$	$1.34 \times 10^{-2}$	$7.48 \times 10^{-2}$
$\sigma_\xi^2/\sigma_\varepsilon^2$	$1.04 \times 10^{-10}$	0	$4.95 \times 10^{-12}$	—
$\sigma_\omega^2/\sigma_\varepsilon^2$	$2.39 \times 10^{-1}$	1.23	$6.71 \times 10^{-2}$	—

As estatísticas de comparação *dentro da amostra* entre os modelos encontram-se na Tabela 4.10, na qual se explicitam MAPE, EQM, Pseudo-R2, AIC, BIC e resultados dos testes de razão de verossimilhança para a relevância

das intervenções. Para o cálculo do MAPE, do EQM e do Pseudo-R2, foram excluídas as observações da primeira linha do triângulo – referente ao período difuso do filtro de Kalman – e também as da primeira coluna – à qual o método *Chain-Ladder* não se aplica por construção. Comparando-se, com essas medidas, os modelos em EE e o método *Chain-Ladder*, os primeiros mostraram-se mais aderentes aos dados do que o último. Até mesmo os modelos I-a e II-a, que não possuem intervenções, foram considerados mais adequados em relação ao *Chain-Ladder*. Finalmente, cita-se que as intervenções se mostraram estatisticamente significantes através de testes de razão de verossimilhança (RV) aplicados aos modelos I-a & I-b, I-a & I-c e I-b & I-c, com todos os testes tendo suas respectivas hipóteses nulas<sup>2</sup> sendo fortemente rejeitadas (todos os p-valores inferiores a  $10^{-4}$ ).

Tabela 4.10: Comparação entre os modelos com intervenções (dados AFG).

	I-b	I-c	II-b	II-c	CL
MAPE (%)	63.68	55.23	42.26	31.03	127.82
EQM	$2.16 \times 10^6$	$1.06 \times 10^6$	$2.66 \times 10^6$	$2.04 \times 10^6$	$9.65 \times 10^6$
Pseudo R2 (%)	50.45	55.96	43.19	54.46	12.05
AIC	14.93	14.59	2.18	1.47	–
BIC	15.59	15.36	2.91	2.31	–
Teste RV (1)	29.5 (0.000)	54.28 (0.000)	46.14 (0.000)	91.14 (0.000)	– –
Teste RV (2)	– –	24.78 (0.000)	– –	45.00 (0.000)	– –

Também foi realizada uma validação *fora da amostra*, em que cada modelo é estimado novamente sem levar em conta as observações pertencentes à diagonal; tais observações serão utilizadas para a validação do modelo. Assim como no procedimento dentro da amostra, para efeito de comparação entre os modelos e o método *Chain-Ladder*, não serão levadas em conta as observações da primeira linha e a da primeira coluna. As estatísticas MAPE, EQM e Pseudo-R2 para o procedimento fora da amostra encontram-se na Tabela 4.11. Os resultados confirmam a inferioridade do método *Chain-Ladder* a todos os modelos em EE analisados, quanto ao poder preditivo. Os modelos em EE com melhor poder preditivo foram os que possuíam o maior número de intervenções: o modelo I-c obteve o menor EQM e o maior Pseudo-R2, enquanto que o modelo II-c obteve o menor MAPE; ressalte-se, porém, que essas diferenças não são expressivas ao ponto de se determinar qual modelo é superior.

<sup>2</sup>Para o teste RV(1), a hipótese nula  $H_0$  é “Os coeficientes associados às intervenções são todos zero”, e os modelos reduzido e completo são, respectivamente, dados por I-a e I-b (ou I-c) para o caso da escala original, e por II-a e II-b (ou II-c) para a escala em log. Para o RV(2), os modelos comparados são os com “menos intervenções” com os de “mais

Tabela 4.11: Comparação entre os modelos com intervenções (dados AFG) – Fora da Amostra.

	I-b	I-c	II-b	II-c	CL
MAPE (%)	161.64	149.12	161.56	142.76	246.85
EQM	$1.44 \times 10^6$	$1.27 \times 10^6$	$1.47 \times 10^6$	$1.33 \times 10^6$	$2.75 \times 10^6$
Pseudo R2 (%)	70.49	73.55	69.00	72.98	27.23

Os diagnósticos dos modelos em EE estão representados na Tabela 4.12. Todos eles foram implementados sobre as *inovações padronizadas*, que são definidas, para cada  $t$ , por  $v_t^S = \frac{v_t}{\sqrt{F_t}}$  (vide 3.1.1) e, sob as hipóteses básicas do modelo linear Gaussiano univariado, devem se comportar como variáveis aleatórias i.i.d.  $N(0, 1)$ . Todos os modelos apresentaram inovações padronizadas não-correlacionadas segundo o teste de Ljung-Box. Os testes de heterocedasticidade também não evidenciaram violações nesse quesito, com exceção do modelo II-b que apresentou certos problemas de inconstância na variância. As estatísticas de Durbin-Watson também foram próximas de 2 (com exceção do modelo II-a), bem como o teste do sinal de Cox-Stuart, que também não acusou presença de quaisquer tendências nas inovações. Já o teste Jarque-Bera rejeitou a hipótese de normalidade apenas no modelo II-a (o resultado desse critério relativo ao modelo II-b, por causa de seu comportamento possivelmente heterocedástico, deve ser visto com cautela).

Tabela 4.12: Testes & Diagnósticos para os dados AFG.

	I-a	I-b	I-c	II-a	II-b	II-c
Heterocedasticidade (20)	1.225 (0.655)	0.952 (0.913)	1.535 (0.346)	0.589 (0.245)	0.343 (0.021)	1.079 (0.867)
Ljung-Box (15 lags) (inov. padr.)	11.898 (0.999)	8.962 (1.000)	11.660 (0.999)	7.287 (1.000)	12.526 (0.998)	24.530 (0.748)
Ljung-Box (15 lags) (quadrado inov. padr.)	8.042 (1.000)	7.442 (1.000)	10.411 (1.000)	4.199 (1.000)	9.587 (1.000)	16.690 (0.976)
Jarque-Bera	0.733 (0.693)	1.700 (0.427)	0.486 (0.784)	23.070 (0.000)	13.981 (0.001)	1.569 (0.456)
Durbin-Watson	1.778	1.935	2.058	1.639	2.121	2.247
Cox-Stuart	7 (0.134)	6 (0.052)	8 (0.286)	8 (0.286)	7 (0.134)	14 (0.134)

Na escala original, o melhor modelo, segundo qualquer um dos critérios da Tabela 4.10 (log-verossimilhança, AIC, BIC etc.), foi o I-c; enquanto que para a escala logarítmica foi o modelo II-c. Graficamente, as componentes intervenções”, seja qual for a escala.

estimadas e os diagnósticos dos referidos modelos podem ser observados nas Figuras 4.7 e 4.9 (modelo I-c), e 4.8 e 4.10 (modelo II-c). A fim de se escolher qual o modelo mais adequado dentre estes dois, alguns cuidados precisam ser tomados: suas log-verossimilhanças e seus critérios de informação AIC e BIC não são comparáveis devido às observações estarem em escalas diferentes. A medida MAPE também é fortemente influenciada pela escala dos dados: mesmo com as previsões transformadas de volta para a escala original dos modelos em escala logarítmica, o MAPE, por ser uma medida percentual, cria distorções nos erros absolutos.

Na literatura atuarial, existe a prática de se utilizar o coeficiente de variação (CV) teórico para a reserva total como uma medida de comparação entre modelos, cf. Mack (1993), England & Verrall (2002), Taylor (2000), de Jong (2006), entre outros autores. O CV é uma medida percentual que é função do erro quadrático médio teórico (cujo cálculo torna-se viável pelo método dos blocos ou, para modelos em escalas originais, pelo método de acumulador) e também do valor esperado da reserva. Logo, se dois modelos possuírem o mesmo erro médio quadrático previsto, porém com um deles tendo estimado uma reserva mais alta, este será considerado “melhor”, pois terá um CV menor. Além disso, cumpre observar que, se pressupostos básicos de um dado modelo são violados – o que se atestaria mediante prática de diagnósticos – e/ou o mesmo demonstrar pouca habilidade em reproduzir os dados – o que se refletiria em baixa performance sob critérios de poder preditivo –, poucos argumentos restariam a favor do uso de medidas teóricas, como o CV, para comparação de modelos. A fim de se evitar tais distorções e possíveis ambiguidades, sugere-se o uso de medidas como o CV, ou o erro médio quadrático teórico (oriundo dos métodos discutidos nas seções metodológicas dessa Tese – seções 3.2 e 3.3) apenas como medidas de precisão *nominal* da estimação da reserva dos modelos, considerados como os “mais adequados” de acordo com análises dentro da amostra, como as que se praticaram até então. Para tais modelos, com os dados AFG, essas informações, junto com as reservas estimadas (parciais e total) encontram-se na Tabela 4.13.

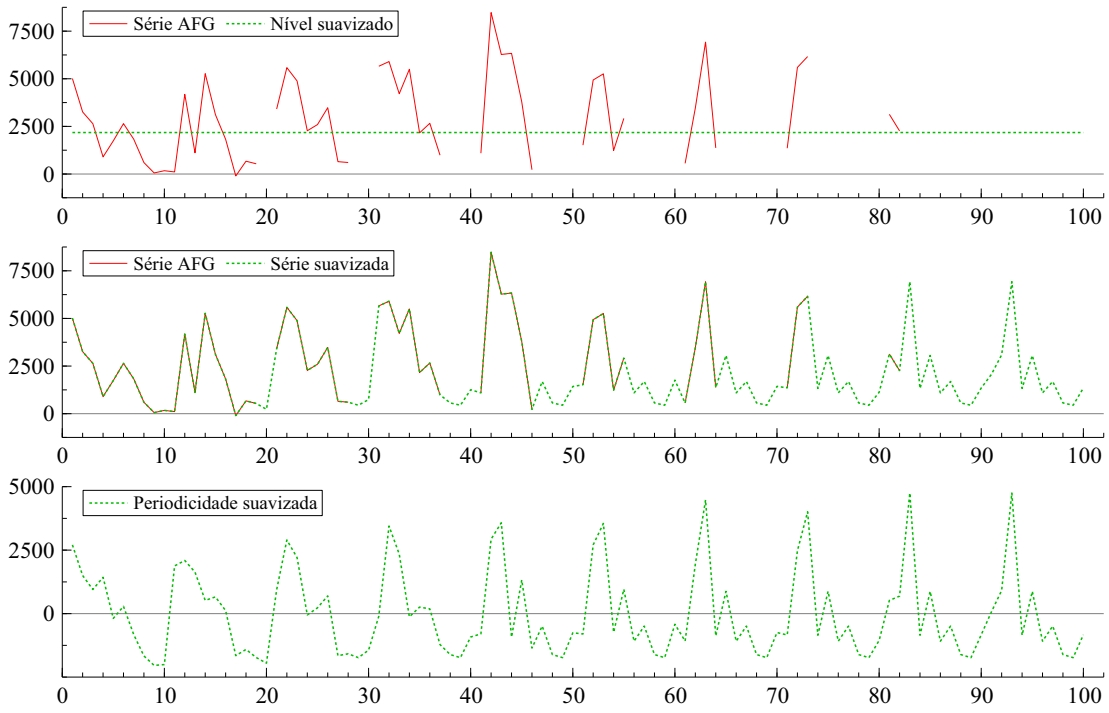


Figura 4.7: Resultados dados AFG – Modelo I-c (8 intervenções).

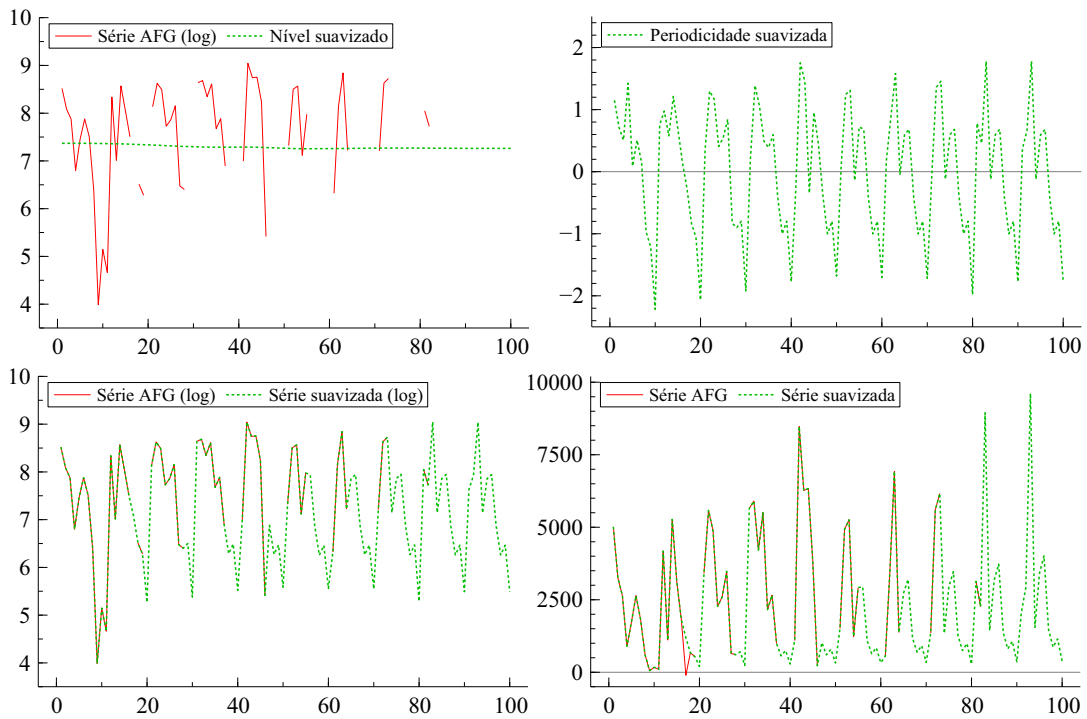


Figura 4.8: Resultados dados AFG – Modelo II-c (10 intervenções).

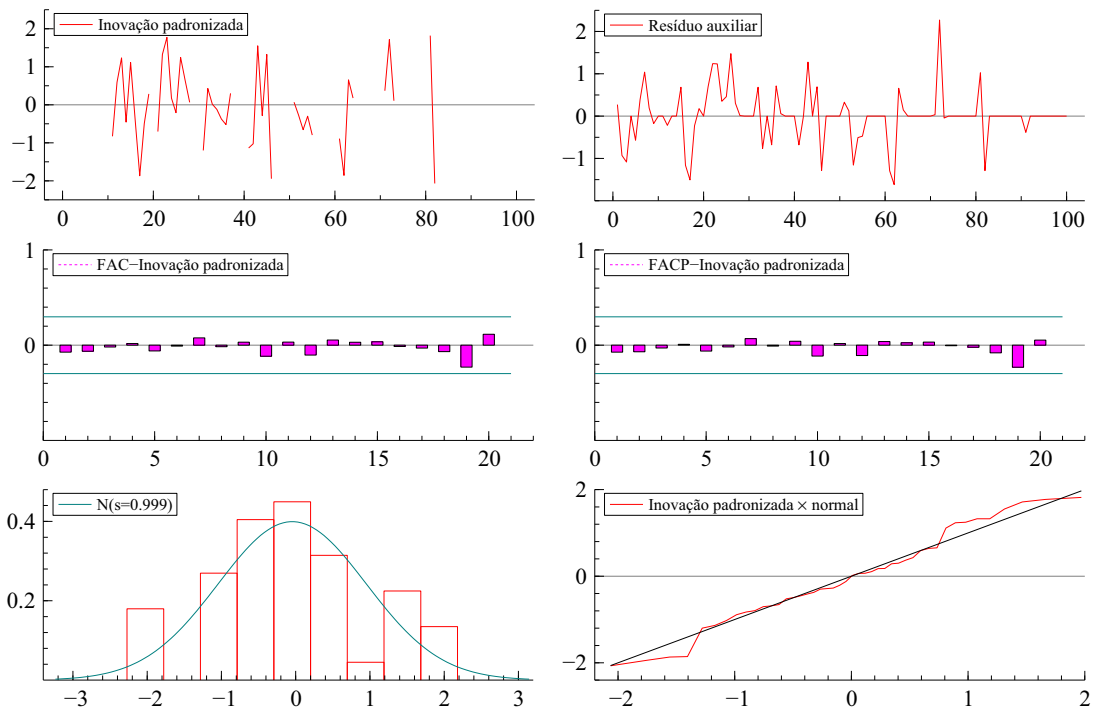


Figura 4.9: Diagnósticos dados AFG – Modelo I-c (8 intervenções).

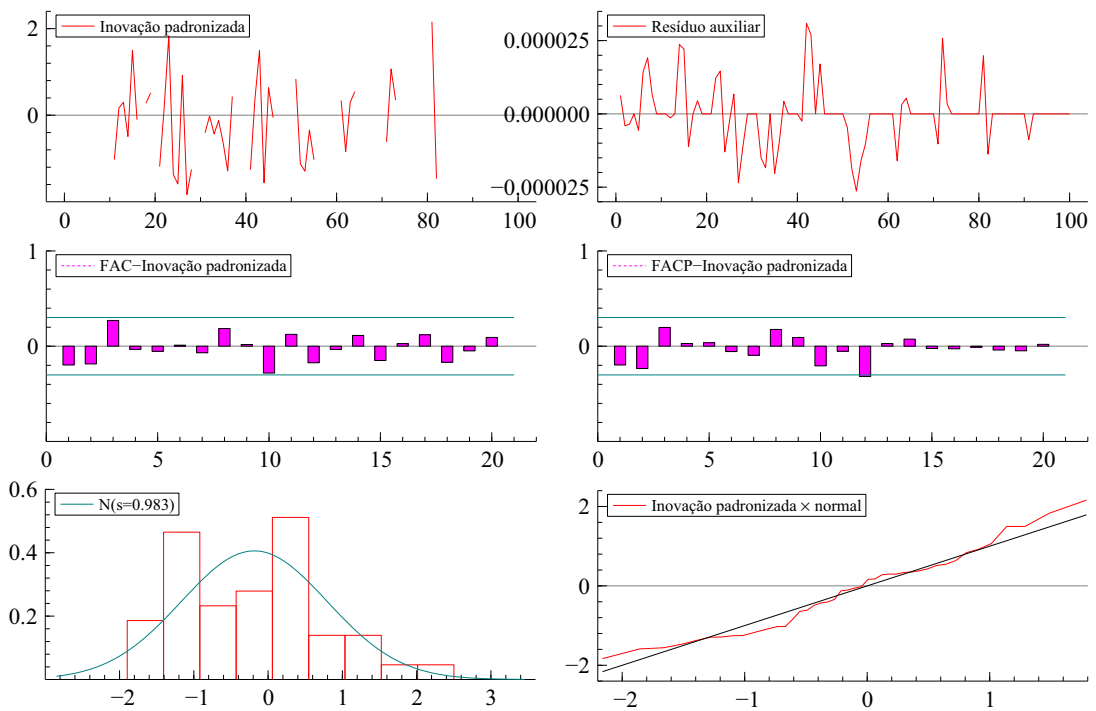


Figura 4.10: Diagnósticos dados AFG – Modelo II-c (10 intervenções).

Tabela 4.13: Reservas Estimadas para os modelos com intervenções e seus respectivos CV (%) – AFG.

Linha	I-c	II-c	Chain Ladder
2	226 (461.5%)	200 (24.0%)	154 (134.0%)
3	1185 (112.4%)	937 (22.1%)	617 (101.0%)
4	2264 (67.3%)	1597 (20.9%)	1636 (45.7%)
5	4119 (40.5%)	2733 (19.4%)	2747 (53.5%)
6	5544 (32.2%)	5837 (18.4%)	3649 (54.9%)
7	8270 (22.7%)	9046 (18.7%)	5435 (40.7%)
8	9286 (21.1%)	11051 (22.2%)	10907 (49.1%)
9	16436 (12.4%)	20882 (20.8%)	10650 (59.5%)
10	19526 (10.9%)	25394 (25.2%)	16339 (150.4%)
Total	66856 (14.9%)	77677 (17.1%)	52135 (51.6%)

## 4.2

### Série MC1: resultados

Seguindo com os exercícios empíricos desta Tese, serão apresentados os resultados com o MC1 para os dados em escala logarítmica<sup>3</sup>.

Tabela 4.14: Intervenções em cada modelo para os dados MC1.

Modelo	# Interv.	Instantes
II-a	0	–
II-b	7	4,6,17,26,34,35 e 73
II-c	10	4,6,7,14,17,26,27,34,35 e 73

Perseguindo o mesmo roteiro adotado na subseção anterior, inicialmente, apresentam-se as estimativas das variâncias na Tabela 4.15. Pelas mesmas e correspondentes “razões sinal-ruído”, a periodicidade mostra-se determinística para qualquer um dos modelos sob questão. Já as componentes de nível, noutra direção, mostraram-se com comportamento razoavelmente estocástico, ao contrário do ocorrido nos dados AFG.

No tocante à comparação dos modelos sob a ótica de poder preditivo, o modelo II-c é o que se mostra sistematicamente superior na comparação que envolve também, mais uma vez, o *Chain-Ladder*, como é comprovado na Tabela 4.16. A validação fora da amostra, cujos resultados encontram-se na Tabela 4.17, também sugere que o modelo II-c seja o mais adequado, embora o método *Chain-Ladder* tenha o erro médio quadrático inferior aos dos modelos em EE. Quanto aos critérios de informação e aos testes de RV, todas

<sup>3</sup>Dada a enorme quantidade de *outliers* sugeridos pelos resíduos auxiliares para a escala original, além de alguns problemas encontrados no processo de estimação de parâmetros, a análise só se efetivou para os dados em escala logarítmica.



Tabela 4.15: Parâmetros estimados e relações sinal/ruído para a base de dados MC1.

	II-a	II-b	II-c
Log-Ver	-24.67	-4.69	5.90
$\sigma_\varepsilon^2$	$1.10 \times 10^{-1}$	$3.72 \times 10^{-2}$	$2.48 \times 10^{-2}$
$\sigma_\xi^2$	$2.47 \times 10^{-4}$	$1.67 \times 10^{-3}$	$6.44 \times 10^{-4}$
$\sigma_\omega^2$	$3.43 \times 10^{-12}$	$3.23 \times 10^{-12}$	$1.18 \times 10^{-11}$
$\sigma_\xi^2/\sigma_\varepsilon^2$	$2.24 \times 10^{-3}$	$4.50 \times 10^{-2}$	$2.60 \times 10^{-2}$
$\sigma_\omega^2/\sigma_\varepsilon^2$	$3.11 \times 10^{-11}$	$8.68 \times 10^{-11}$	$4.78 \times 10^{-10}$

as intervenções consideradas mostraram-se, mais uma vez, estatisticamente significantes (ao nível de 5% ou menos). Em relação aos diagnósticos, a Tabela 4.18 e, graficamente, a Figura 4.12, revelam que o modelo II-c é o que mais se adequa aos pressupostos básicos da forma em EE linear Gaussiana, apesar de indício de que ainda persista estrutura de correlação nas inovações padronizadas segundo o teste de Ljung-Box – apesar de o teste de independência de Cox-Stuart apontar adequado comportamento.

Tabela 4.16: Comparação entre os modelos (dados MC1).

	II-a	II-b	II-c	CL.
MAPE (%)	35.51%	19.43%	14.92%	50.47%
EQM	$6.17 \times 10^{10}$	$2.75 \times 10^{10}$	$1.72 \times 10^{10}$	$1.17 \times 10^{11}$
Pseudo R2	52.11%	66.33%	64.51%	29.04%
AIC	1.37	0.82	0.44	–
BIC	1.84	1.48	1.10	–
Teste RV (1)	–	39.96	61.14	–
	–	(0.000)	(0.000)	–
Teste RV (2)	–	–	21.18	–
	–	–	(0.000)	–

Tabela 4.17: Comparação entre os modelos (dados MC1) – Fora da Amostra.

	II-a	II-b	II-c	CL.
MAPE (%)	33.73%	24.10%	21.81%	26.97%
EQM	$6.21 \times 10^{10}$	$4.25 \times 10^{10}$	$4.64 \times 10^{10}$	$3.44 \times 10^{10}$
Pseudo R2	82.96%	86.67%	89.40%	80.93%

As evidências reunidas, no parágrafo anterior, levam à conclusão de que os dados são melhores descritos pelo modelo com “mais intervenções”. Na Tabela 4.19, encontram-se as reservas estimadas do mesmo, juntamente com o CV teórico oriundo do método dos blocos.

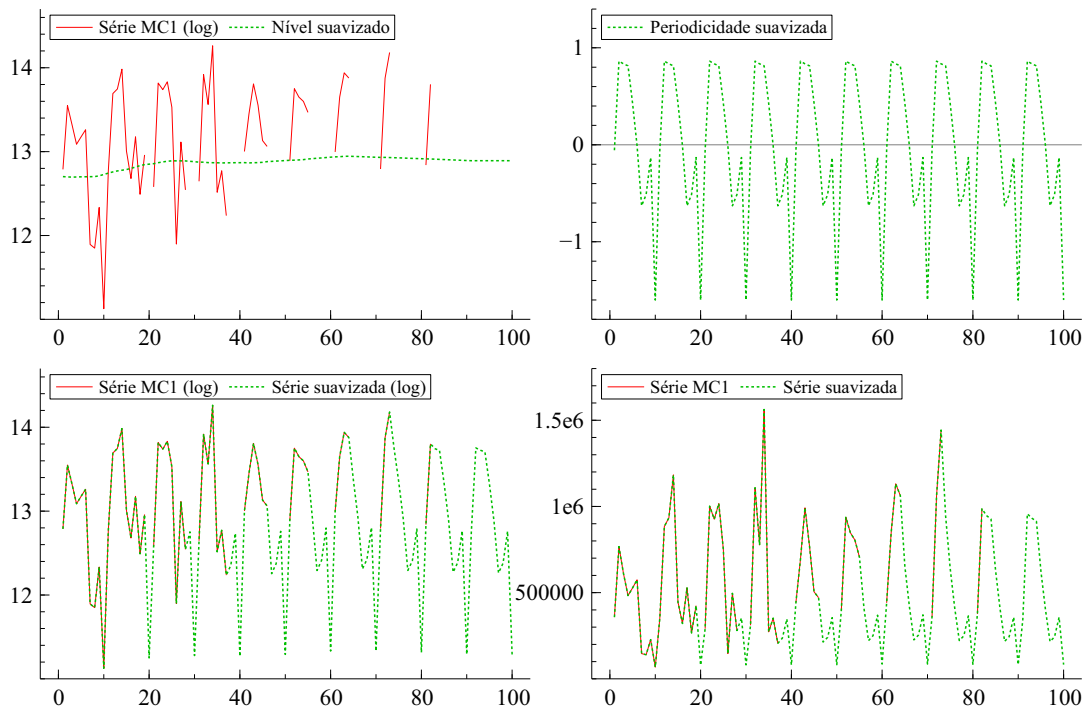


Figura 4.11: Resultados do modelo estrutural nos dados MC1 com transformação (10 intervenções).

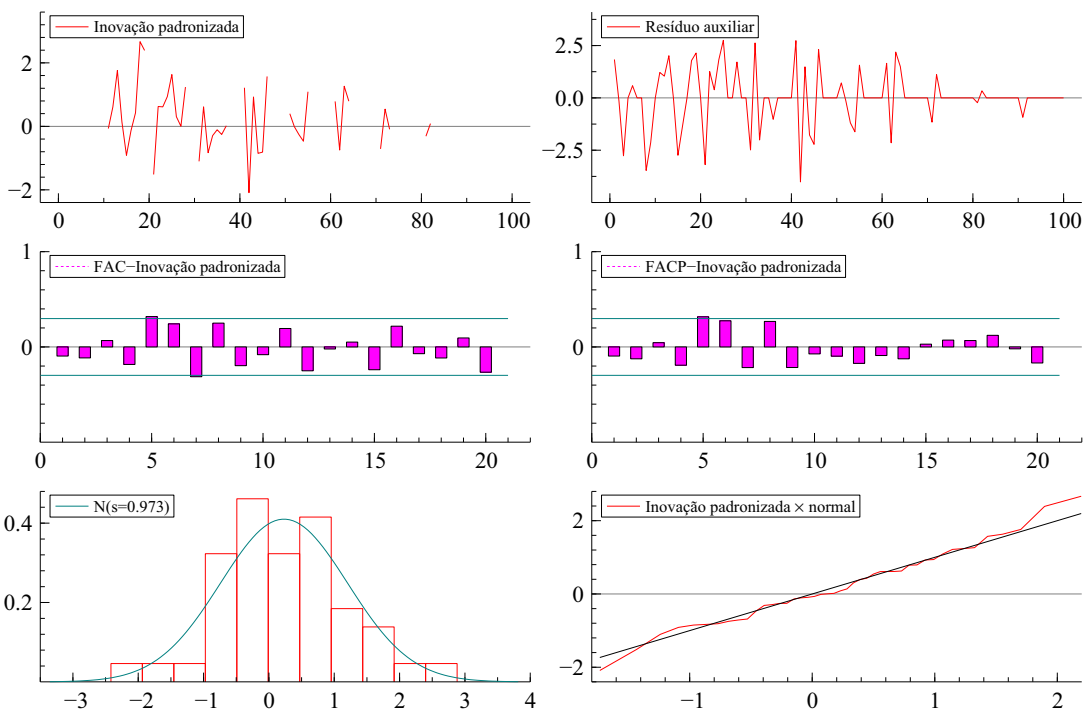


Figura 4.12: Diagnósticos – modelo estrutural nos dados MC1 com transformação (10 intervenções).

Tabela 4.18: Testes &amp; Diagnósticos para os dados MC1.

	II-a	II-b	II-c
Heterocedasticidade (20)	0.235 (0.002)	0.369 (0.031)	0.556 (0.198)
Ljung-Box (15 lags) (inov. padr.)	22.679 (0.091)	14.020 (0.524)	32.980 (0.005)
Ljung-Box (15 lags) (quadrado inov. padr.)	16.028 (0.380)	12.258 (0.659)	14.303 (0.503)
Jarque-Bera	4.893 (0.087)	1.268 (0.530)	0.395 (0.821)
Durbin-Watson	2.355	1.916	2.053
Cox-Stuart	9.000 (0.523)	8.000 (0.286)	12.000 (0.523)

Tabela 4.19: Reservas Estimadas para os modelos com intervenções e seus respectivos CV (%) – MC1.

Linha	II-c	Chain Ladder
2	78,904 (23.3%)	94,634 (79.8%)
3	433,790 (17.3%)	469,510 (25.9%)
4	663,310 (13.7%)	709,640 (18.8%)
5	891,770 (12.0%)	984,890 (26.5%)
6	1,336,400 (10.8%)	1,419,500 (29.0%)
7	2,009,900 (10.3%)	2,177,600 (25.6%)
8	2,919,600 (10.4%)	3,920,300 (22.3%)
9	3,810,800 (10.8%)	4,279,000 (22.7%)
10	4,726,900 (12.1%)	4,625,800 (29.5%)
Total	16,871,000 (7.1%)	18,681,000 (13.1%)

### 4.3

#### Série DJZ: resultados

Os parâmetros estimados do modelo estrutural, para a base de dados DJZ, encontram-se na tabela 4.20. As estimações, em ambas as escalas (original e em log), confirmaram a intuição gráfica citada anteriormente, apontando uma periodicidade determinística do padrão de pagamentos de sinistros IBNR (vide as desprezíveis razões sinal/ruído na Tabela 4.20). Mas, ressalte-se que o padrão exibido de decaimento, ao longo de um dado ano de origem, é similar aos provenientes de outras modelagens, com este mesmo triângulo, na literatura (cf. de Jong & Zehnwirth, 1983; e Atherino & Fernandes, 2007).

Os primeiros painéis das Figuras 4.13 e 4.14 mostram um crescimento do nível ao longo do tempo, o que pode ser interpretado, dada a ordenação usada, como um aumento do número de sinistros IBNR com o avanço dos anos de acidente.

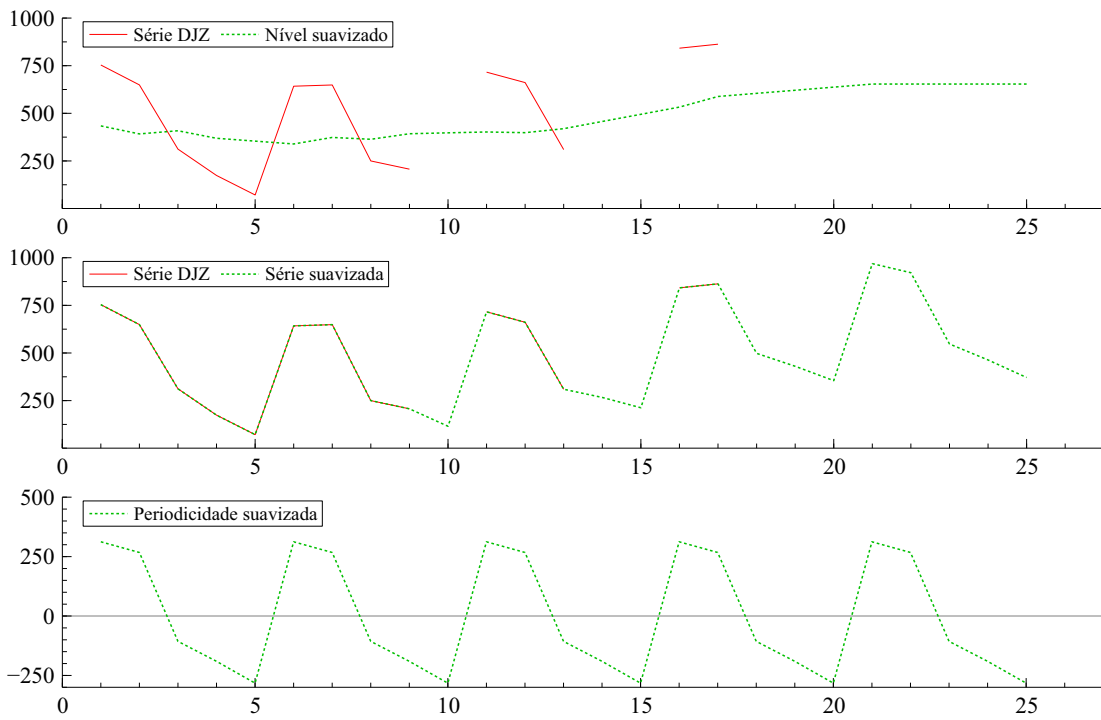


Figura 4.13: Resultados do modelo estrutural nos dados sem transformação (dados DJZ).

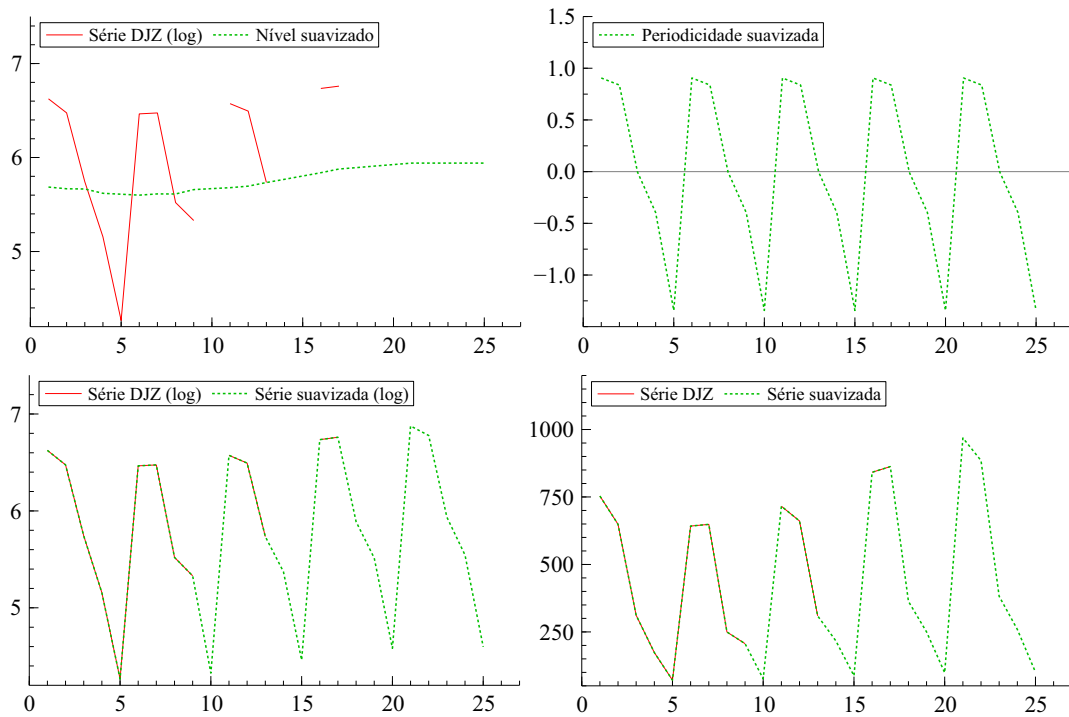


Figura 4.14: Resultados modelo estrutural nos dados com transformação (dados DJZ).

Tabela 4.20: Parâmetros estimados e relações sinal/ruído para a base de dados DJZ.

Parâmetro	Escala	
	(normal)	(log)
Log-ver.	-58.44	3.84
$\sigma_\varepsilon^2$	$3.31 \times 10^2$	$5.65 \times 10^{-3}$
$\sigma_\xi^2$	$1.84 \times 10^3$	$3.05 \times 10^{-3}$
$\sigma_\omega^2$	$1.81 \times 10^{-5}$	$1.52 \times 10^{-8}$
$\sigma_\xi^2/\sigma_\varepsilon^2$	5.55	0.54
$\sigma_\omega^2/\sigma_\varepsilon^2$	$5.47 \times 10^{-8}$	$2.69 \times 10^{-6}$

No processo de comparação dos modelos entre si e com o *Chain-Ladder*, apenas foram consideradas, face à escassez de dados, medidas de poder preditivo e critérios de informação. Pela Tabela 4.21, não há indícios claros de um método “vencedor”. A validação, fora da amostra, também não sugere um método com melhores propriedades preditivas (vide Tabela 4.22), porém as medidas foram calculadas levando-se em conta apenas três observações da diagonal principal. As reservas estimadas, junto com os correspondentes CVs, estão na Tabela 4.23.

Tabela 4.21: Comparação entre os modelos (dados DJZ).

	Mod. Est. (normal)	Mod. Est. (log)	Chain Ladder
MAPE (%)	14.08%	12.12%	13.00%
EQM	$3.76 \times 10^3$	$3.57 \times 10^3$	$2.19 \times 10^3$
Pseudo R2	95.64%	96.13%	96.87%

Tabela 4.22: Comparação entre os modelos (dados DJZ) – Fora da Amostra.

	Mod. Est. (normal)	Mod. Est. (log)	Chain Ladder
MAPE (%)	12.95%	16.54%	14.87%
EQM	$4.53 \times 10^3$	$1.20 \times 10^4$	$2.55 \times 10^3$
Pseudo R2	99.20%	99.94%	99.97%

Tabela 4.23: Reservas estimadas e seus respectivos CV (%) – DJZ.

Ano de Origem	Mod. Est. (normal)	Mod. Est. (log)	Chain Ladder
2	115 (46.0%)	76 (13.1%)	66 (55.5%)
3	478 (19.4%)	304 (7.4%)	268 (23.3%)
4	1283 (10.4%)	712 (7.7%)	693 (11.6%)
5	2302 (11.7%)	1628 (10.4%)	1696 (8.1%)
Total	4179 (8.9%)	2720 (7.7%)	2723 (8.2%)