

## 5 Resultados

### 5.1. Lee-Carter

O modelo demográfico LC foi ajustado aos dados da Inglaterra e País de Gales de 1950 a 2003. A modelagem foi feita segregada por sexo, para preservar a diferença de mortalidade entre homens e mulheres. Primeiramente, foi feito o ajuste do modelo demográfico previsto na eq.(1), com a estimação dos parâmetros  $\alpha_x$ ,  $\beta_x$  e  $\kappa_t$ . O passo seguinte foi a projeção de  $\kappa_t$  para 25 anos subsequentes à última observação. Por fim, foi obtida a previsão das taxas de mortalidade para 25 anos.

Os parâmetros  $\hat{\alpha}_x$  obtidos com base na eq.(3) estão ilustrados na Fig. 11. No período 1950-2003, a curva masculina encontra-se acima da feminina para todas as idades, refletindo o fato esperado da mortalidade ter sido maior, em média, na população masculina. Um ponto interessante a ser observado é o fato da curva possuir um crescimento mais brando na faixa etária dos 20 aos 30 anos (maior destaque na população masculina). Uma justificativa seria o aumento geralmente existente da mortalidade nessas idades devido a acidentes e violência.

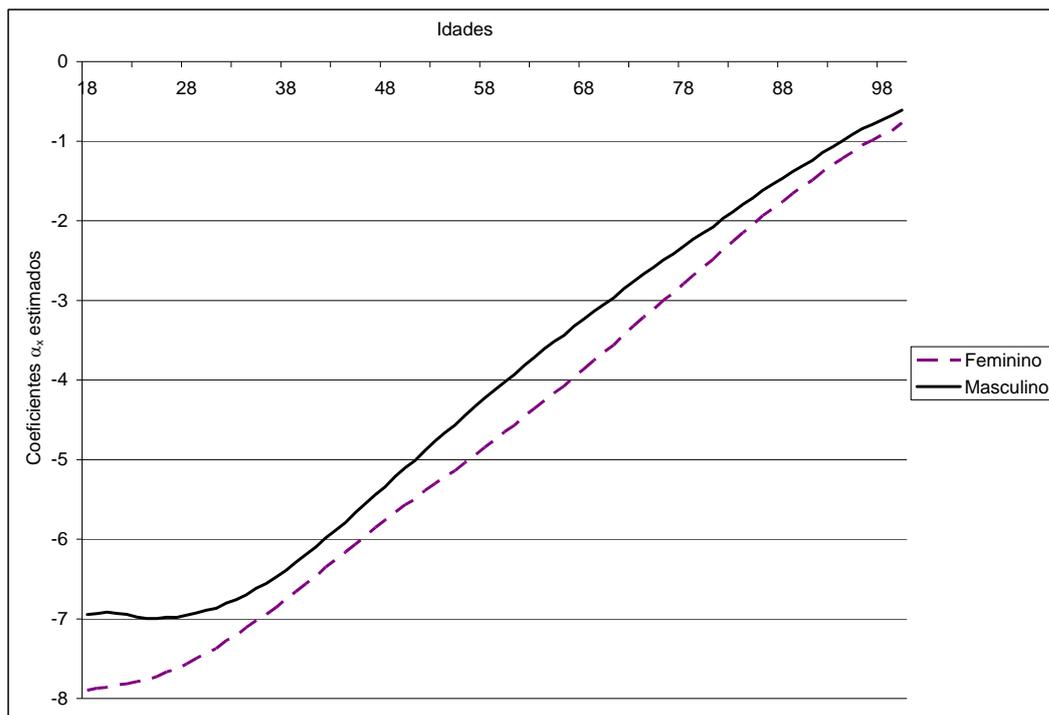


Figura 11 – Coeficientes  $\alpha_x$  estimados por sexo (1950-2003)

Com base nos valores  $\hat{\alpha}_x$ , foram obtidas as primeiras estimações de  $\kappa_t$  a partir da eq. (4). A tendência de queda do nível de mortalidade com o tempo é constatada para os dois sexos na Fig.12. Os parâmetros estimados  $\hat{\beta}_x$  conforme eq.(5), que avaliam a sensibilidade da variação nas taxas de morte quando o parâmetro  $\kappa_t$  varia, estão demonstrados na Fig. 13. Se  $\beta_x$  é pequeno, conclui-se que as taxas de morte variam pouco quando o nível geral de mortalidade se altera, o que é o caso da mortalidade nas idades mais avançadas. Para as idades acima de 75 anos aproximadamente, os coeficientes feminino e masculino apresentaram um comportamento similar, uma tendência decrescente. A curva observada para a população feminina, que mostra que quanto menor a idade, maior é a sensibilidade à variação no parâmetro  $\kappa_t$ , tem a mesma tendência da curva apontada por Lee (2000) ao aplicar a metodologia LC aos dados dos Estados Unidos e do Chile. Já a curva masculina apresentou valores menores para  $\hat{\beta}_x$  quando estes eram maiores na população feminina e vice-versa.

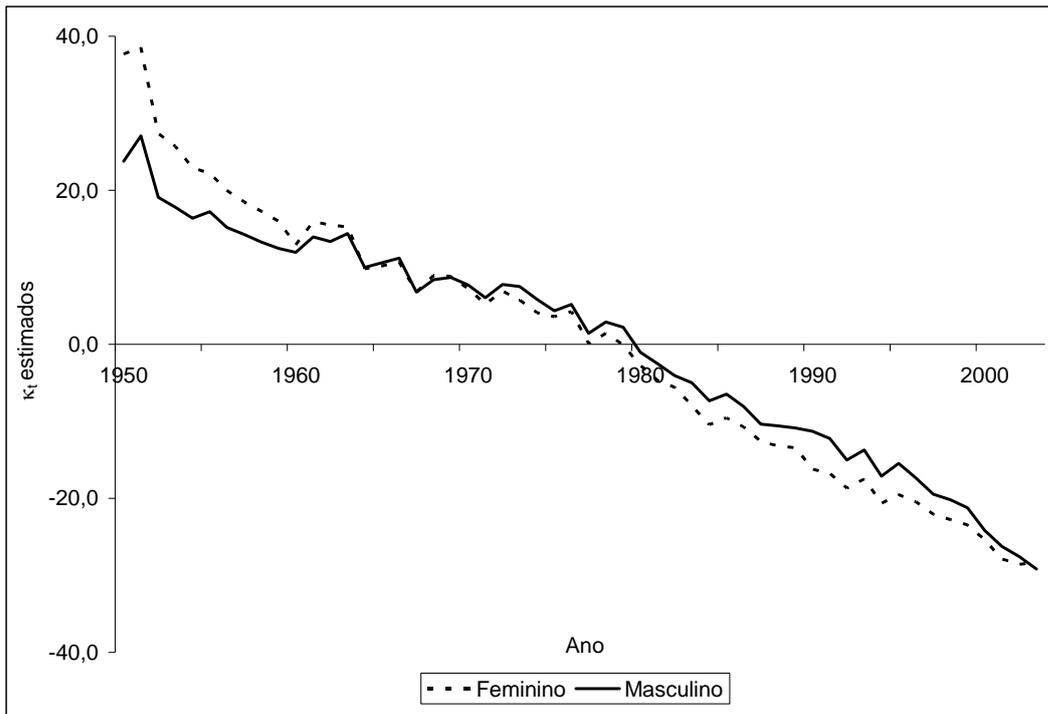


Figura 12 – Coeficientes  $\kappa_t$  estimados por sexo (1950-2003)

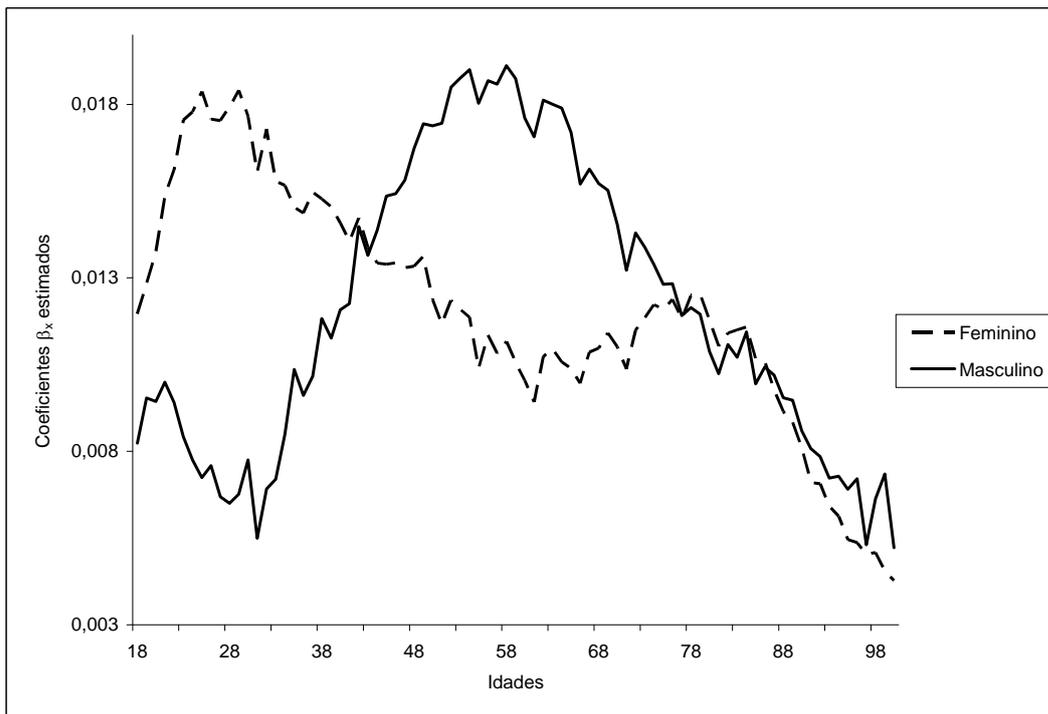


Figura 13 – Coeficientes  $\beta_x$  estimados, por sexo (1950-2003)

Para garantir que as taxas de morte ajustadas gerariam o número de mortes observado quando aplicadas à população analisada, foi realizada a reestimação do parâmetro  $\kappa_t$  segundo a eq.(6). Os parâmetros reestimados para a população feminina e masculina, bem como os estimados inicialmente, estão ilustrados nas Fig. 14. Os valores estimados para os parâmetros  $\alpha$ ,  $\beta$  e  $\kappa$  constam no Apêndice I deste trabalho.

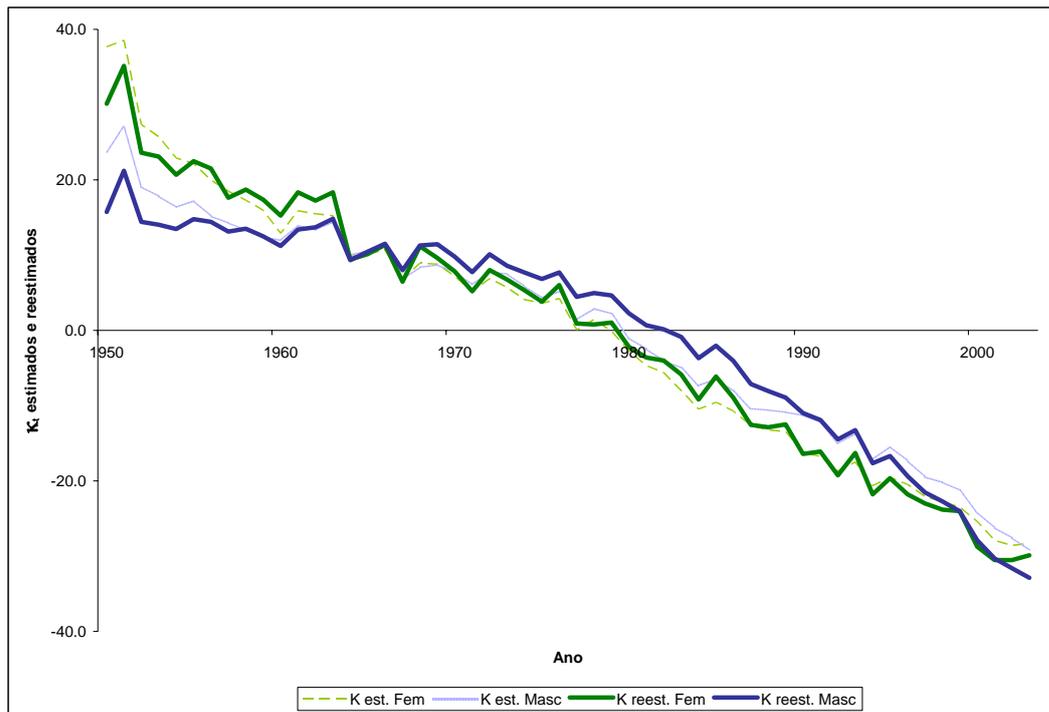


Figura 14 – Coeficientes  $\kappa_t$  reestimados (1950-2003)

Feita a reestimação do parâmetro  $\kappa_t$ , podemos escrever o modelo demográfico proposto na eq.(1). Uma vez que os parâmetros  $\alpha_x$  e  $\beta_x$  foram estimados e não dependem da variável tempo, a previsão das taxas de mortalidade futuras passam a depender exclusivamente da projeção dos valores  $\kappa_t$  futuros. Para tal projeção, as séries  $\kappa_t$  reestimadas feminina e masculina são modeladas como um processo de séries temporais. A Fig. 14 mostra um declínio dos valores de  $\kappa_t$  aproximadamente linear, maior para as mulheres do que para os homens, no período 1950-2003. A linearidade aproximada de  $\kappa_t$ , bem como a variação relativamente constante das séries no período são grandes vantagens para a construção de modelos de previsão. Os dados disponibilizados referem-se ao período de 1950 a 2003. No entanto, os 5 últimos anos disponíveis não foram

modelados, a fim de possibilitar uma comparação entre a previsibilidade dos modelos nos próximos 5 anos e o comportamento observado no período.

Uma simples inspeção nas séries mostra que as mesmas aparentam não-estacionariedade e apresentam uma tendência negativa sustentada. Os correlogramas ilustrados nas Fig. 15 e 16 também apresentam indícios de não-estacionariedade das séries, visto que observa-se um amortecimento lento da função de autocorrelação parcial (ACF). A não-estacionariedade das séries foi superada modelando-as em primeira diferença ou com ajuste de tendência, como será mostrado.

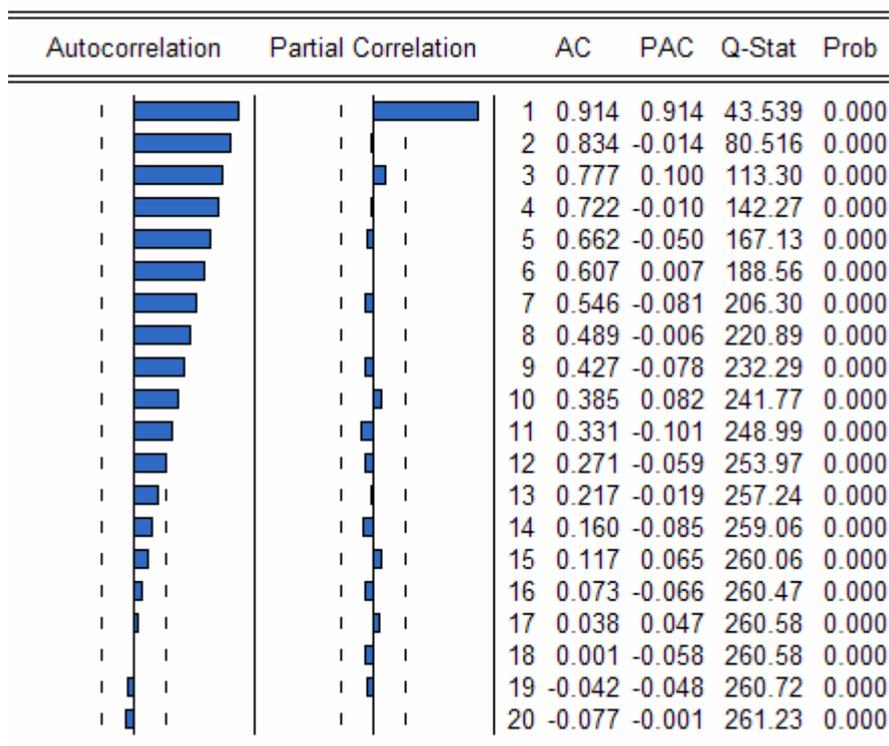


Figura 15 – Correlograma da série  $\kappa_t$  feminina

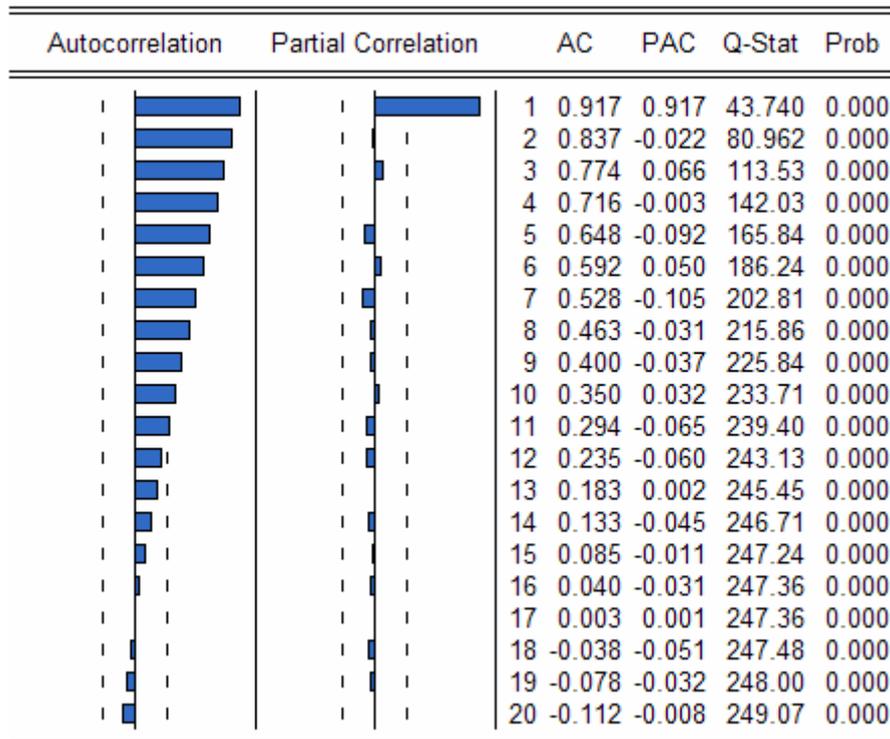


Figura 16 – Correlograma da série  $\kappa_t$  masculina

Devido à tendência decrescente das séries, foi considerada a presença de uma constante quando da estimação dos modelos. Em relação à série feminina, doravante denominada  $\kappa_{fem_t}$ , dois modelos testados apresentaram bons ajustes e uma análise de resíduos satisfatória. O primeiro modelo é um processo ARIMA, onde a série é modelada em primeira diferença, uma vez que já foi verificado que ela não é estacionária em nível. Os parâmetros dos modelos foram estimados através do pacote estatístico EViews, utilizando o método de Mínimos Quadrados, e estão apresentados na Tabela 3. O índice  $\kappa_{fem_t}$  modelado como um processo ARIMA(0,1,1) é representado algebricamente na forma:

$$\Delta \kappa_{fem_t} = c + \theta_1 \varepsilon_{t-1} + \varepsilon_t,$$

$$\varepsilon_t \sim N(0, \sigma^2),$$

onde  $\Delta$  representa o operador de diferença. Assumindo os valores estimados para os parâmetros, a equação estimada se torna:

$$\hat{\Delta \kappa_{fem_t}} = -1,101191 - 0,778194 \varepsilon_{t-1} + \varepsilon_t;$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

Tabela 3 – Parâmetros estimados - processo ARIMA(0,1,1), série  $\kappa fem_t$ 

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1,101191	0,086957	-12,66362	0,0000
MA(1)	-0,778194	0,088414	-8,801654	0,0000
$R^2 = 0,3825$				
AIC = 4,7029				

É importante destacar que o coeficiente  $R^2$  refere-se a  $\Delta \kappa fem_t$ , e não a  $\kappa fem_t$ . O segundo modelo estimado para  $\kappa fem_t$  foi ajustado com uma tendência linear e um processo média móvel (MA) 4. Neste ajuste, o  $R^2$  observado ficou próximo de 1, conforme exposto na Tabela 4. A representação algébrica para o modelo é:

$$\kappa fem_t = \alpha + \beta t + \varepsilon_t;$$

$$\varepsilon_t = \eta_t + \theta_4 \eta_{t-4};$$

$$\eta_t \sim N(0, \sigma^2)$$

Assumindo os valores estimados para os parâmetros, a equação proposta se torna:

$$\hat{\kappa fem}_t = 28,86891 - 1,077052 \times t + \varepsilon_t;$$

$$\varepsilon_t = \eta_t + 0,508999 \eta_{t-4}, \text{ onde } \eta_t \sim N(0, \sigma^2)$$

Tabela 4 – Parâmetros estimados - processo com tendência linear, série  $\kappa fem_t$ 

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	28,86891	0,993553	29,05622	0,0000
@TREND	-1,077052	0,034769	-30,97754	0,0000
MA(4)	0,508999	0,119627	4,254865	0,0001
$R^2 = 0,9770$				
AIC = 4,6498				

Comparando os dois modelos pelo método de Akaike (AIC) e observando o coeficiente  $R^2$ , optou-se por modelar a série feminina considerando a tendência linear e o processo MA(4). No caso da série masculina,  $\kappa masc_t$ , quando da tentativa de ajuste dos modelos, foi observado um comportamento atípico para o resíduo da observação referente ao ano 1951. De fato, os gráficos das séries masculina e feminina possuem um pico no ano 1951. Foi possível a estimação de

um modelo com uma análise de resíduos satisfatória para a série feminina, sem que fosse necessária a anulação do impacto da observação do ano 1951. No entanto, para a série masculina, o modelo preferido inclui uma variável dummy para remover a influência do referido ano. Observando a série masculina, nota-se que a tendência do declínio de  $\kappa_{\text{masc}_t}$  está entre uma tendência linear e uma tendência quadrática. O modelo utilizado para a previsão do índice masculino se baseia nas duas tendências, linear e quadrática, bem como na utilização de uma variável dummy referente ao ano 1951 e em um processo MA(12). As estimações para os parâmetros, obtidas via EViews, são apresentados na Tabela 5. A forma algébrica para o modelo é da forma:

$$\kappa_{\text{masc}_t} = \alpha + \beta_1 t + \beta_2 t^2 + \beta_3 D + \varepsilon_t ;$$

$$\varepsilon_t = \eta_t - \theta_{12} \eta_{t-12} ; \eta_t \sim N(0, \sigma^2),$$

onde D é a variável dummy assumindo valor 1 se o ano observado é 1951.

Inserindo os valores estimados para os parâmetros, observa-se:

$$\hat{\kappa}_{\text{masc}_t} = 13,43730 + 0,227291 \times t - 0,020347 \times t^2 + 9,345870 \times D + \varepsilon_t ;$$

$$\varepsilon_t = \eta_t - 0,907975 \eta_{t-12}$$

Tabela 5 – Estimação dos parâmetros - modelo ajustado para a série  $\kappa_{\text{masc}_t}$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13,43730	1,173386	11,45173	0,0000
@TREND	0,227291	0,090119	2,522115	0,0154
(@TREND)^2	-0,020347	0,001582	-12,86183	0,0000
D1951	9,345870	3,330911	2,805800	0,0074
MA(12)	-0,907975	0,030975	-29,31363	0,0000
$R^2 = 0,9940$				

O coeficiente em D indica que  $\kappa_{\text{masc}_t}$  foi 9,345870 maior do que o esperado em 1951. A análise de resíduos é necessária para verificação dos pressupostos adotados quando da modelagem de uma série temporal. Ao ajustar os modelos às séries, assume-se que os resíduos não são correlacionados e possuem distribuição normal. Cada uma dessas condições foi verificada a partir da análise de *outputs* do EViews.

As Fig. 17 e 18 exibem os resíduos dos modelos ajustados, bem como um comparativo dos valores observados x valores previstos:

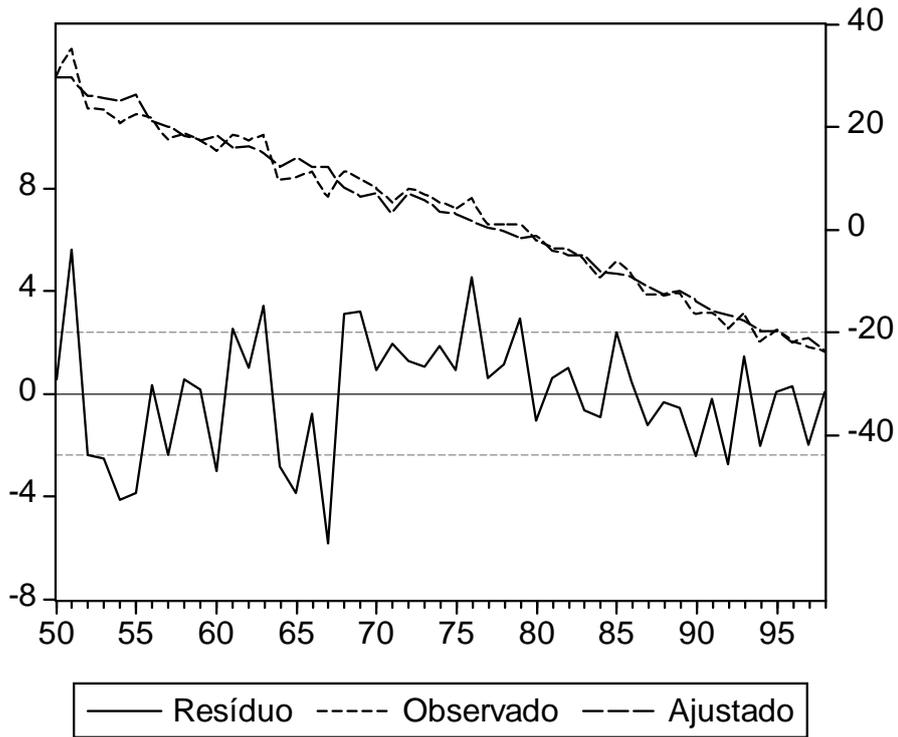


Figura 17 – Valores observados x ajustados x resíduos – Série  $\kappa_{fem_t}$

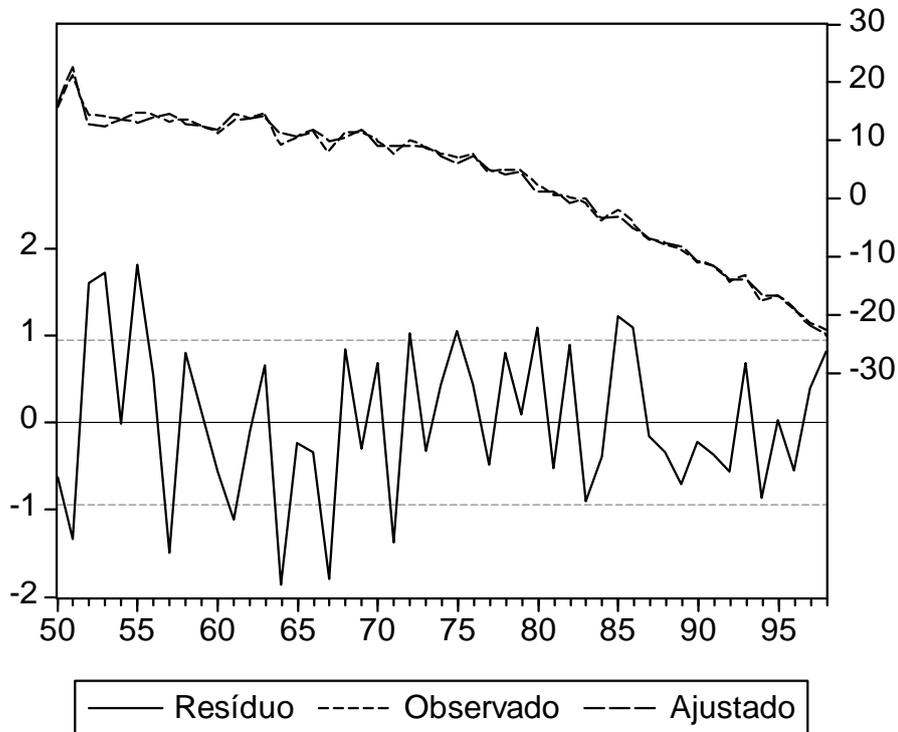


Figura 18 – Valores observados x ajustados x resíduos – Série  $\kappa_{masc_t}$

## 1. Autocorrelação dos resíduos

Tabela 6 – Teste de autocorrelação dos resíduos – Série  $\kappa fem_t$

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	1.617961	Probability	0.123771
Obs*R-squared	27.16501	Probability	0.130685

Tabela 7 – Teste de autocorrelação dos resíduos – Série  $\kappa masc_t$

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	0.688240	Probability	0.800222
Obs*R-squared	17.83708	Probability	0.598140

O teste LM para autocorrelação serial, conhecido como teste Breusch-Godfrey, foi utilizado para testar a correlação serial dos resíduos dos modelos avaliados. A hipótese nula do teste prevê que os erros são não-correlacionados até a ordem  $p$ . Os modelos foram testados com  $p = 20$  e concluiu-se que, ao nível de 5% de significância, aceita-se a hipótese de inexistência de correlação serial nos resíduos.

## 2. Normalidade dos resíduos

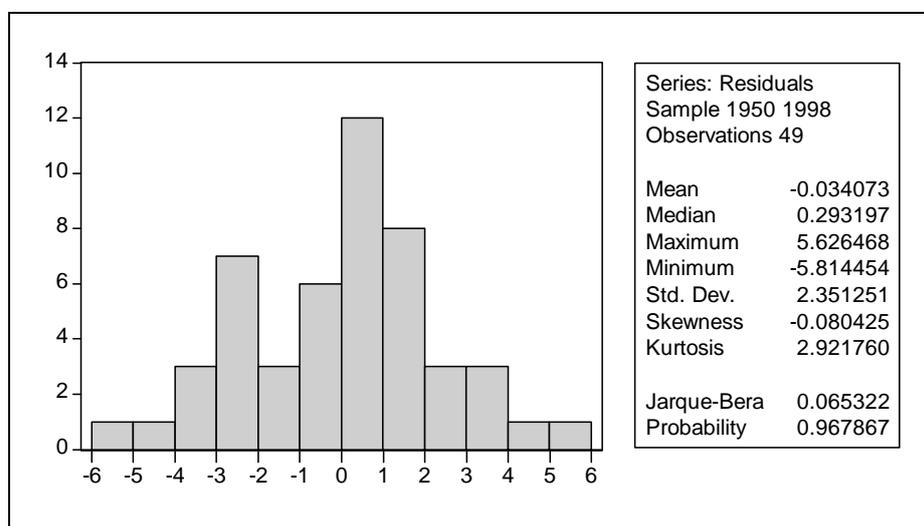


Figura 19 – Teste de normalidade dos resíduos – Série  $\kappa fem_t$

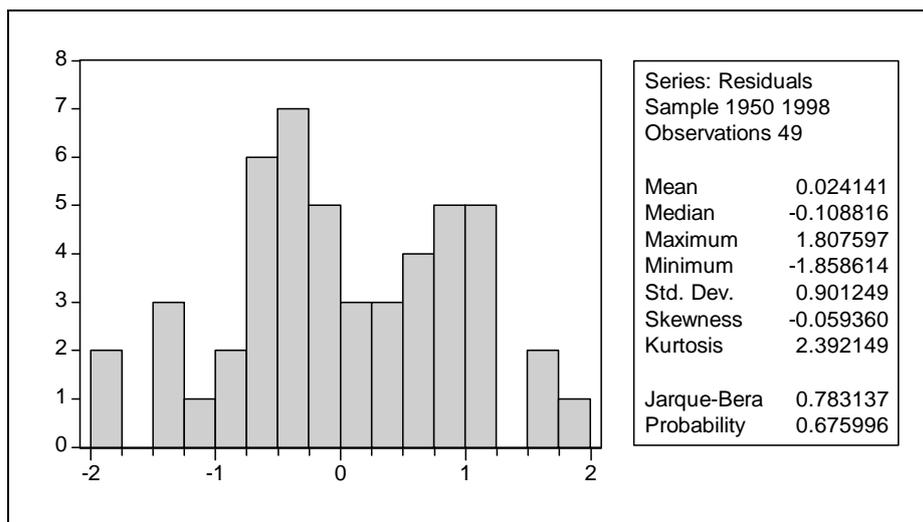


Figura 20 – Teste de normalidade dos resíduos – Série  $kmasc_t$

A normalidade dos resíduos de regressão foi investigada através do teste de Jarque-Bera. Neste teste, é investigada a hipótese nula de que os resíduos possuem distribuição normal. Com base nas probabilidades apresentadas, aceitamos a hipótese nula de normalidade dos resíduos ao nível de significância de 5%.

Partindo dos modelos validados, o passo seguinte é a previsão dos índices  $\kappa fem_t$  e  $\kappa masc_t$ . Para que fosse possível avaliar a capacidade preditiva do modelo, partiu-se do modelo estimado no período *in-sample* (1950-1998). Com base na previsão *out-of-sample*, ou *ex-post*<sup>3</sup>, é possível a comparação, através de medidas de aderência, entre as previsões e os valores realmente observados na série para o período 1999-2003 (*out-of-sample*). A Tabela 8 apresenta os valores das séries  $\kappa fem_t$  e  $\kappa masc_t$  no período *out-of-sample*. As Fig. 21 e 22 exibem as séries reais e os valores previstos com um intervalo de confiança de 95%. Todos os valores observados encontram-se dentro do intervalo de previsão de 95%, o que confirma uma boa capacidade preditiva dos modelos propostos.

<sup>3</sup> Previsão *ex-post*: utiliza-se o modelo estimado no período *in-sample* e extrapola-se o modelo no período *out-of-sample*, comparando o previsto com o realmente observado (Fernandes, 2006).

Tabela 8 – Medidas de aderência no período *out-of-sample*

Período	Observado	Previsão	Erro	% Erro
<b>Série <math>\kappa_{fem_t}</math></b>				
1999	-23.9949	-23.8747	-0.1202	0.50%
2000	-28.7352	-24.8344	-3.9008	13.57%
2001	-30.5103	-27.0666	-3.4437	11.29%
2002	-30.5146	-27.1110	-3.4036	11.15%
2003	-29.8683	-28.2148	-1.6535	5.54%
<b>Série <math>\kappa_{masc_t}</math></b>				
1999	-24.0245	-24.1382	0.1136	0.47%
2000	-27.8808	-25.7525	-2.1284	7.63%
2001	-29.2905	-27.2443	-2.0462	6.99%
2002	-31.6142	-29.5537	-2.0605	6.52%
2003	-32.8557	-31.3337	-1.5220	4.63%

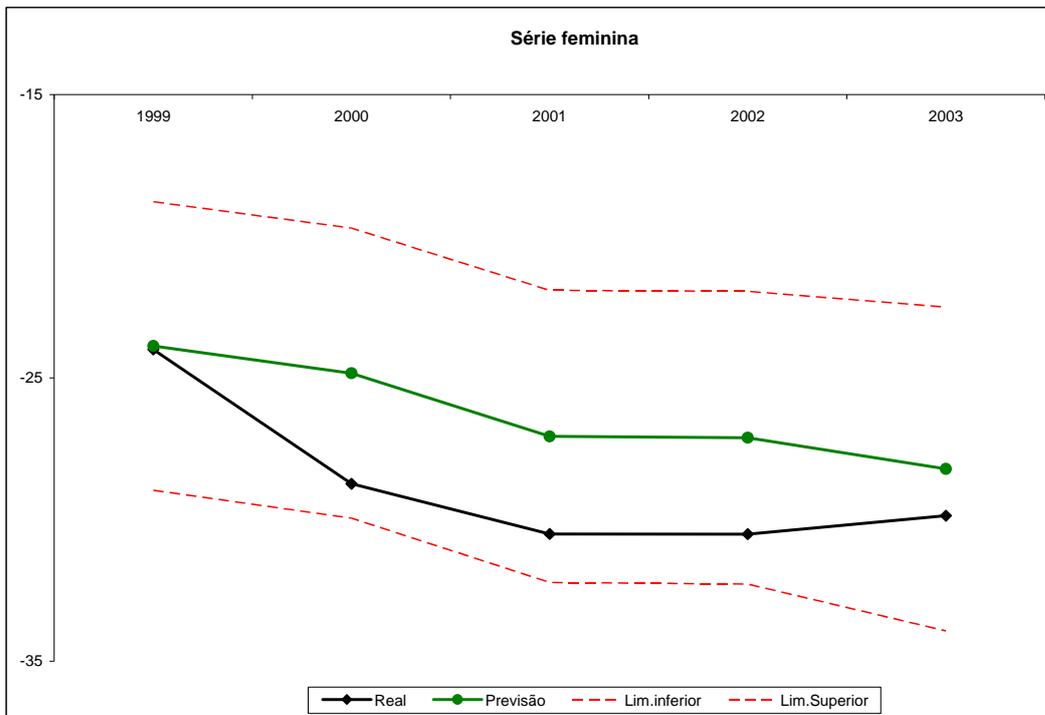


Figura 21 – Previsão out-of-sample da série  $\kappa_{fem_t}$

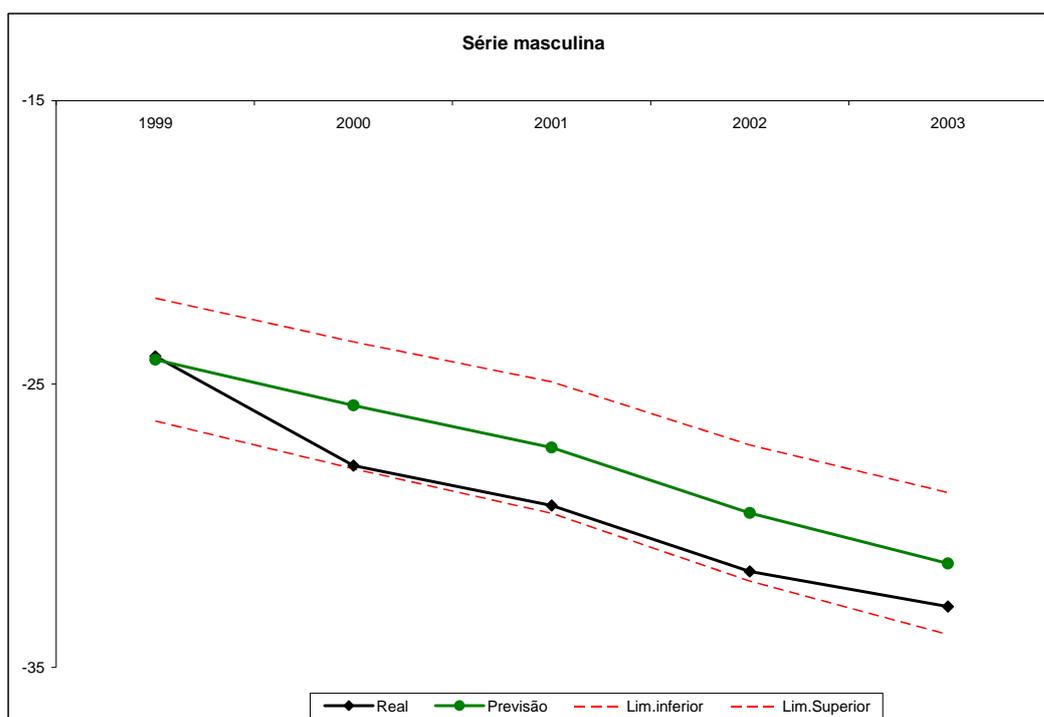


Figura 22 – Previsão out-of-sample da série  $\kappa_{masct}$

Após certificação da capacidade preditiva do modelo, as observações referentes ao período 1999-2003 foram incluídas no modelo para a obtenção da extrapolação das séries  $\kappa_t$  a partir da última observação disponível. Os novos parâmetros estimados para os modelos preditivos finais encontram-se na Tabela 9. As Fig. 23 e 24 ilustram as previsões das séries  $\kappa_{fem_t}$  e  $\kappa_{masct}$  para os 25 anos subsequentes à última observação (2004 – 2028). É válido notar que o método LC foi utilizado para calcular os intervalos de previsão concentrados somente na variabilidade de  $\kappa_t$ . Em Brouhns et al. (2005), mostra-se a possibilidade de inclusão de outras fontes de variação através de método de *bootstrap*.

Tabela 9 – Coeficientes estimados para os modelos preditivos finais

<b>Série feminina</b>				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	29.42971	0.959687	30.66593	0.0000
@TREND	-1.110437	0.030504	-36.40280	0.0000
MA(4)	0.519696	0.117248	4.432458	0.0000
$R^2 = 0,9816$				
<b>Série masculina</b>				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.06975	1.001806	12.04799	0.0000
@TREND	0.353629	0.073662	4.800670	0.0000
(@TREND)^2	-0.022815	0.001213	-18.81492	0.0000
D1951	8.359352	2.968737	2.815794	0.0070
MA(12)	-0.891080	0.026348	-33.81991	0.0000
$R^2 = 0,9957$				

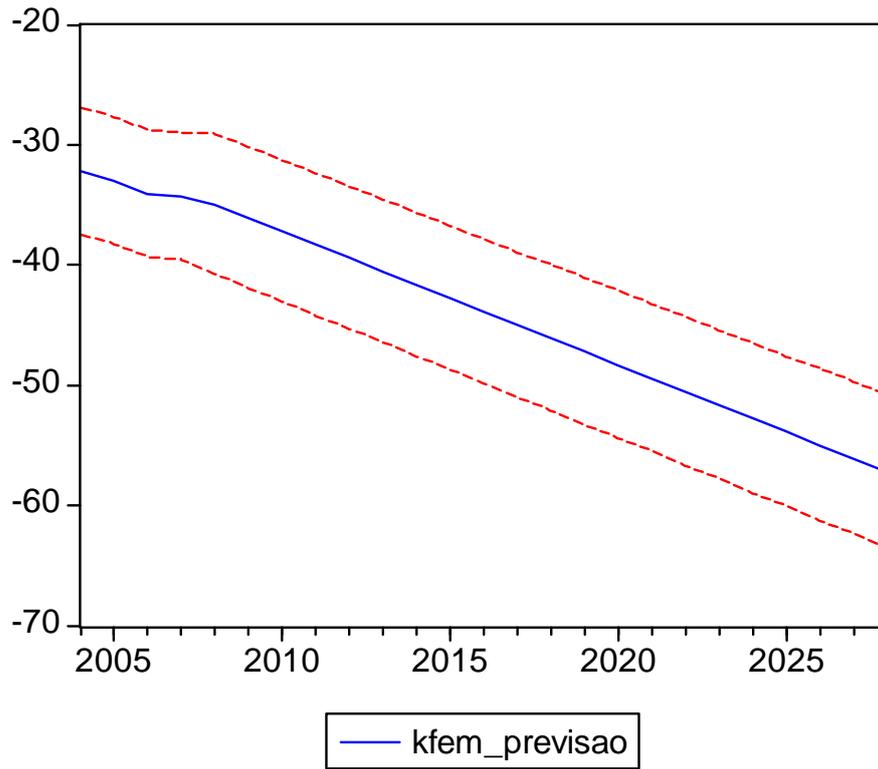


Figura 23 – Previsão da série  $\kappa_{fem_t}$  (2004 – 2028)

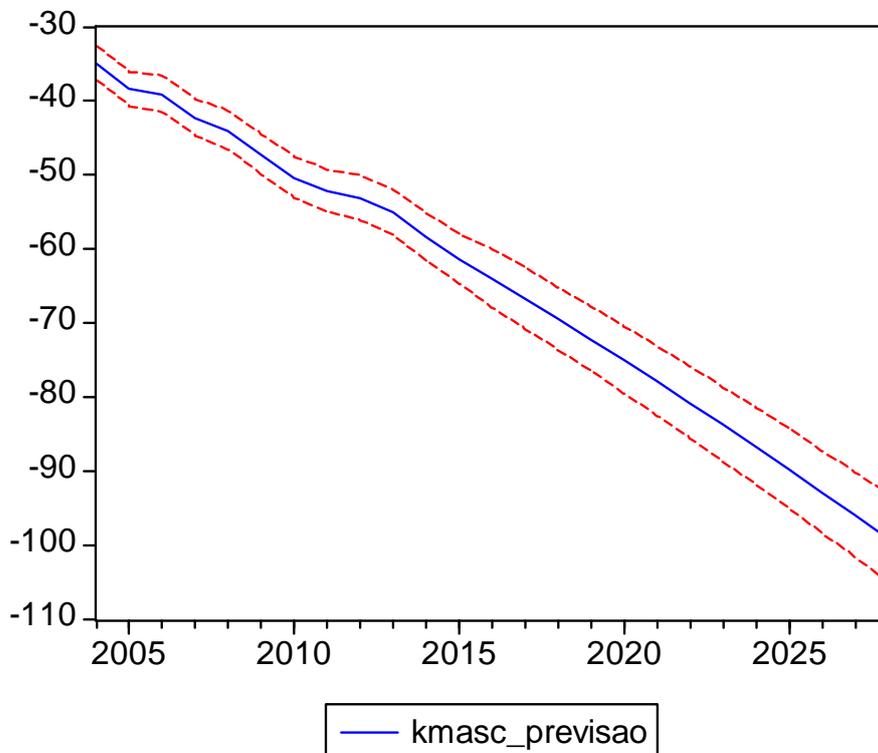


Figura 24 – Previsão da série  $\kappa_{masc_t}$  (2004 – 2028)

A partir dos valores de previsão do índice de mortalidade  $\kappa_t$ , as taxas de mortalidade podem ser projetadas e as tábuas de mortalidade construídas. Primeiramente, obtém-se as previsões das taxas de mortalidade a partir da última taxa de mortalidade empírica disponível,  $m_{x,2003}$ , através da fórmula:

$$\hat{m}_{x,2003+s} = m_{x,2003} \exp\left\{\hat{\beta}_x (\kappa_{2003+s} - \kappa_{2003})\right\}, s \text{ inteiro } \geq 0$$

A partir das taxas centrais de mortalidade projetadas, as probabilidades de morte a serem utilizadas nas tábuas de mortalidade são facilmente obtidas conforme relação exposta na seção 2.1:

$$q_x = \frac{2 m_x}{2 + m_x} \Rightarrow q_{x,t} = \frac{2 m_{x,t}}{2 + m_{x,t}}$$

O ideal é que seja feita a graduação das taxas cruas de mortalidade antes da obtenção das probabilidades de morte, para que as informações sejam suavizadas e produzam um comportamento de acordo com o esperado. Apesar da recomendação, o resultado apresentado para as probabilidades de morte calculadas via método LC neste trabalho não consideraram a graduação nas taxas. Uma referência sobre as formas de graduação mais utilizadas, especialmente nos casos brasileiros, pode ser obtida em Martins (2007).

A Fig. 25 exibe o logaritmo das probabilidades  $q_x$  projetadas através do método LC para as gerações femininas e masculinas nascidas nos anos 2004 e 2028. Percebemos a redução esperada para as probabilidades de morte entre os referidos anos. Nota-se que a curva do 2028 apresenta, a partir da idade 85 aproximadamente, probabilidades de morte masculinas esperadas menores do que as femininas. Os modelos feminino e masculino foram construídos de forma independente, não havendo a pressuposição de nenhuma relação entre eles. Portanto, a estimativa de probabilidades de morte menores para o sexo masculino é possível. Sabemos que as mulheres possuem uma sobrevida maior do que os homens em geral, mas não se pode afirmar se essa tendência vai permanecer, reduzir ou aumentar no futuro.

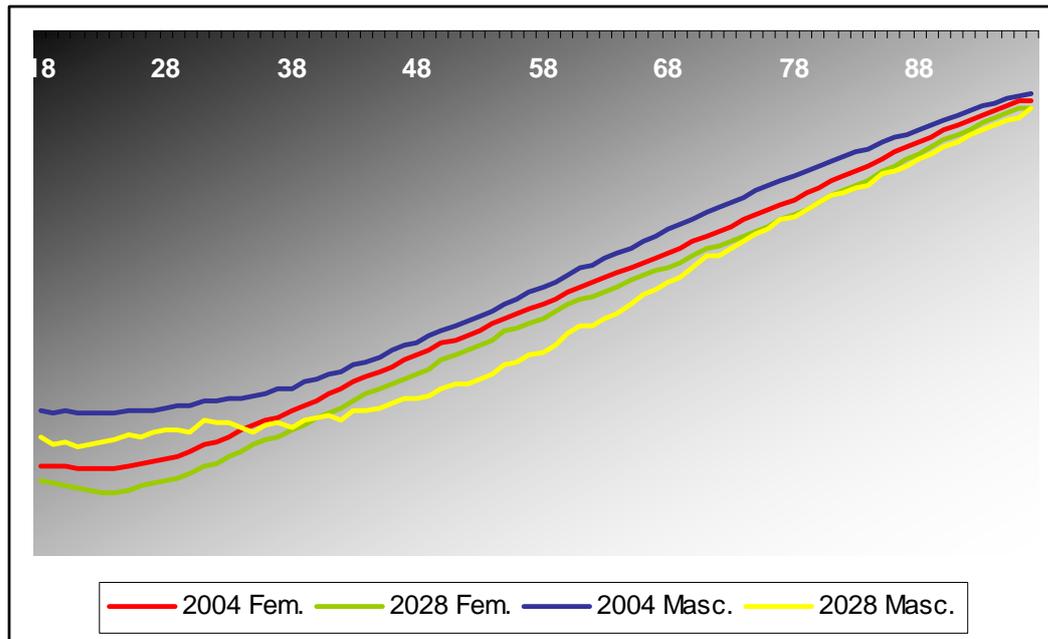


Figura 25 – Logaritmo dos  $q_x$  projetados via método LC

Com posse dos valores  $q_x$  e arbitrando um valor para  $l_0$ , raiz da tábua de mortalidade, é possível a obtenção das funções da tábua de mortalidade para todas as idades  $x$  através de fórmulas de recorrência. A raiz da tábua de mortalidade é geralmente arbitrada em 100.000 e representa o número de sobreviventes à idade 0 (recém-nascidos) da coorte utilizada na construção da tábua. Partindo dos valores  $q_{x,t}$  projetados, são definidas as funções da tábua de mortalidade dependentes da idade e do tempo através das relações descritas na seção 2.1:

$$l_{x+1,t} = l_{x,t} (1 - q_{x,t}) ;$$

$$d_{x,t} = l_{x,t} q_{x,t} ;$$

$$L_{x,t} = \frac{l_{x,t} - l_{x+1,t}}{2} ;$$

$$T_{x,t} = \sum_{x_i=x}^{\omega} L_{x_i,t} ; e$$

$$e_{x,t} = \frac{T_{x,t}}{l_{x,t}} ,$$

onde  $\omega$  é a idade limite da tábua de mortalidade.

## 5.2. Modelos Lineares Generalizados

Começamos nossa análise ajustando a eq.14 àquela que envolve a graduação de Gompertz  $G_x(s)$  dada pela eq.10, ou seja:

$$\eta_{x,t} = \log e_{xt} + \beta_0 + \sum_{j=1}^s \beta_j L_j(x') + \sum_{i=1}^r \alpha_i t^i \quad (16)$$

Os dados são compostos por 4482 informações acerca de  $d_{x,t}$  e  $e_{x,t}$  para cada sexo (83 idades x 54 anos de observação). A princípio, 10% da nossa amostra (449 observações) foi selecionada aleatoriamente e retirada do banco de dados original para a verificação do ajuste do modelo.

A seleção de variáveis explicativas a serem incluídas no modelo de regressão foi feita através do método *stepwise forward*, onde geramos uma sequência de modelos com um ou mais parâmetros, partindo do modelo nulo (contendo somente o intercepto). Os valores ótimos de  $r$  e  $s$  foram determinados monitorando-se a redução nas medidas Deviance e Akaike (AIC) à medida que os valores dos parâmetros aumentavam. Os valores das Deviances e do AIC dos modelos feminino e masculino estão reproduzidas nas Tabelas 10 e 11. Em Haberman & Wong-Fupuy (2004), sugere-se que sejam usados valores baixos para  $r$  e  $s$ , pois isso assegura uma melhor suavização dos dados. Em adição, Renshaw et al. (1996) apontam que a dependência no tempo é expressa por polinômios de baixo valor ( $r=2$ ). Com base nessas informações, simulamos os valores de Deviance e AIC obtidos variando o parâmetro  $r$  de 0 a 4 e o parâmetro  $s$  de 0 a 6.

Tabela 10 – Deviance e AIC – polinômios preditivos femininos

s	Deviance				
	r = 0	r = 1	r = 2	r = 3	r = 4
0	32.004.888,00	32.000.543,00	31.969.504,00	31.968.125,00	31.968.079,00
1	499.293,50	86.362,80	85.954,02	85.636,96	84.232,67
2	491.754,40	62.837,97	62.181,69	61.859,72	60.443,83
3	473.203,60	<b>47.916,74</b>	47.441,07	47.199,74	45.767,34
4	472.380,40	46.763,27	46.277,62	46.043,48	44.588,71
5	468.792,60	43.070,08	42.587,71	42.337,80	40.883,72
6	467.613,60	42.226,83	41.739,68	41.488,84	40.017,40

s	AIC				
	r = 0	r = 1	r = 2	r = 3	r = 4
0	32.041.438,00	32.037.095,00	32.006.058,00	32.004.681,00	32.004.637,00
1	535.845,70	122.917,00	122.510,20	122.195,10	120.792,90
2	528.308,60	99.394,16	98.739,87	98.419,90	97.006,02
3	509.759,80	<b>84.474,92</b>	84.001,25	83.761,93	82.331,53
4	508.938,60	83.323,45	82.839,80	82.607,67	81.154,90
5	505.352,80	79.632,27	79.151,90	78.903,99	77.451,91
6	504.175,80	78.791,02	78.305,87	78.057,03	76.587,59

Tabela 11 – Deviance e AIC – polinômios preditivos masculinos

s	Deviance				
	r = 0	r = 1	r = 2	r = 3	r = 4
0	28.880.125,00	28.807.475,00	28.777.820,00	28.775.131,00	28.774.514,00
1	669.353,10	216.477,50	173.280,20	172.442,20	171.292,40
2	661.887,20	214.723,90	172.100,70	171.271,00	170.154,50
3	579.714,40	140.499,60	98.955,60	98.456,98	97.373,54
4	542.124,00	102.669,90	<b>63.241,65</b>	62.606,62	61.790,06
5	539.817,60	100.542,30	61.096,53	60.445,93	59.571,19
6	538.299,30	98.576,83	59.422,88	58.760,69	57.872,45

s	AIC				
	r = 0	r = 1	r = 2	r = 3	r = 4
0	28.917.213,00	28.844.565,00	28.814.912,00	28.812.225,00	28.811.610,00
1	706.443,30	253.569,70	210.374,50	209.538,40	208.390,60
2	698.979,40	251.818,10	209.197,00	208.369,30	207.254,70
3	616.808,60	177.595,90	136.053,80	135.557,20	134.475,80
4	579.220,30	139.768,10	<b>100.341,90</b>	99.708,85	98.894,30
5	576.915,80	137.642,50	98.198,76	97.550,16	96.677,42
6	575.399,50	135.679,10	96.527,12	95.866,92	94.980,68

Para ajudar a selecionar os valores ótimos dos parâmetros r e s dos modelos, assim como em Renshaw et.al. (1996), as diferenças nas deviances em cada linha e coluna da tabela testadas foram aproximadas a uma distribuição qui-quadrado. A deviance do modelo com p parâmetros (modelo mais simples) menos a deviance do modelo com p+q parâmetros (modelo mais complexo) possui uma distribuição qui-quadrado com q graus de liberdade. O teste foi feito através da comparação da estatística obtida com uma  $\chi_1^2$ , sujeita às seguintes hipóteses:

$H_0$ : o modelo com p parâmetros é adequado;

$H_a$ : o modelo com p+1 parâmetros é adequado.

Fixando-se uma probabilidade de erro tipo I em 0,10 para o teste de inclusão de variáveis, para o modelo feminino, a inclusão dos 10 parâmetros um a um foi

verificada e o número de parâmetros escolhidos buscando-se otimizar a modelagem foi  $r = 1$  e  $s = 3$ , enquanto para o modelo masculino escolheu-se  $r = 2$  e  $s = 4$ . Desta forma, temos os modelos propostos:

Feminino:  $\log m_{x,t} = \log e_{x,t} + \beta_0 + \beta_1 L_1(x') + \beta_2 L_2(x') + \beta_3 L_3(x') + \alpha_1 t^1 + u_{x,t}$

Masculino:  $\log m_{x,t} = \log e_{x,t} + \beta_0 + \beta_1 L_1(x') + \beta_2 L_2(x') + \beta_3 L_3(x') + \beta_4 L_4(x') + \alpha_1 t^1 + \alpha_2 t^2 + u_{x,t}$ ,

onde  $L$  é o coeficiente de lagrange e  $u_{xt}$  é o termo de erro. O próximo passo é a estimação dos parâmetros e avaliação do erro padrão para que possamos verificar a significância estatística de cada parâmetro proposto. A modelagem GLM e a estimação dos parâmetros foi feita utilizando-se o pacote estatístico R. As tabelas 12 e 13 exibem os parâmetros estimados para os modelos feminino e masculino. O parâmetro de escala  $\phi$  é considerado igual a 1 na nossa modelagem, uma vez que não trabalhamos com dados de apólices, e sim com dados individuais.

Tabela 12 – Estimação dos parâmetros GLM masculinos

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z )	
<i>Intercepto</i>	-4.6058	0.0006	-7259.8000	<2e-16 ***	
$\beta_1$	3.8793	0.0014	2734.5000	<2e-16 ***	
$\beta_2$	0.3012	0.0014	207.8000	<2e-16 ***	
$\beta_3$	-0.1645	0.0013	-123.4000	<2e-16 ***	
$\alpha_1$	-0.3194	0.0005	-654.1000	<2e-16 ***	
---					
Signif.	codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
(Dispersion parameter for poisson family taken to be 1					
Null deviance: 32004888 on 4034 degrees of freedom					
Residual deviance: 47917 on 4030 degrees of freedom					
AIC: 84475					

Tabela 13 – Estimação dos parâmetros GLM masculino

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z )
<i>Intercepto</i>	-4.0549	0.0006	-6746.1000	<2e-16 ***
$\beta_1$	3.7200	0.0012	3124.2000	<2e-16 ***
$\beta_2$	0.2462	0.0014	174.2000	<2e-16 ***
$\beta_3$	-0.4274	0.0013	-321.1000	<2e-16 ***
$\beta_4$	0.2894	0.0015	192.3000	<2e-16 ***
$\alpha_1$	-0.3141	0.0005	-649.6000	<2e-16 ***
$\alpha_2$	-0.1801	0.0009	-197.3000	<2e-16 ***
--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'				
(Dispersion parameter for poisson family taken to be 1				
Null deviance: 28880125 on 4034 degrees of freedom				
Residual deviance: 63242 on 4028 degrees of freedom				
AIC: 100342				

As estatísticas t calculadas dividindo-se o parâmetro por seu erro padrão, indicaram significância estatística. Com base nos valores-p obtidos, pode-se afirmar que todos os parâmetros estimados foram considerados significativos estatisticamente ao nível de 5% de significância ou bem menos. A partir dos modelos estimados, testamos as possíveis interações entre as covariáveis. A inclusão dos termos de interação foi realizada conforme previsto na eq. 15. O método *stepwise* resultou nos seguintes modelos finais e estimativas:

$$\begin{aligned} \text{Feminino: } \log m_{x,t} = & \log e_{x,t} + \beta_0 + \beta_1 L_1(x') + \beta_2 L_2(x') + \beta_3 L_3(x') + \\ & + \alpha_1 t^{\text{II}} + \gamma_{11} L_1(x') t^{\text{II}} + u_{x,t} \end{aligned}$$

Tabela 14 – Estimação dos parâmetros GLM feminino com interação

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z )	
<i>Intercepto</i>	-4.6125	0.0006	-7155.87	<2e-16 ***	
$\beta_1$	3.8800	0.0014	2736.37	<2e-16 ***	
$\beta_2$	0.2958	0.0015	203.95	<2e-16 ***	
$\beta_3$	-0.1761	0.0013	-131.03	<2e-16 ***	
$\alpha_1$	-0.3580	0.0008	-465.02	<2e-16 ***	
$\gamma_{11}$	0.0988	0.0015	64.96	<2e-16 ***	
---					
Signif.	codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
(Dispersion parameter for poisson family taken to be 1					
Null deviance: 32004888 on 4034 degrees of freedom					
Residual deviance: 43684 on 4029 degrees of freedom					
AIC: 80244					

Masculino:  $\log m_{x,t} = \log e_{x,t} + \beta_0 + \beta_1 L_1(x') + \beta_2 L_2(x') + \beta_3 L_3(x') + \beta_4 L_4(x') + \alpha_1 t^1 + \alpha_2 t^2 + \gamma_{11} L_1(x') t^1 + \gamma_{21} L_1(x') t^2 + \gamma_{12} L_2(x') t^1 + \gamma_{22} L_2(x') t^2 + u_{x,t}$

Tabela 15 – Estimação dos parâmetros GLM masculino com interação

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z )	
<i>Intercepto</i>	-4.0995	0.0008	-5268.33	<2e-16 ***	
$\beta_1$	3.7324	0.0016	2312.80	<2e-16 ***	
$\beta_2$	0.1061	0.0019	55.47	<2e-16 ***	
$\beta_3$	-0.4622	0.0014	-340.96	<2e-16 ***	
$\beta_4$	0.2658	0.0015	174.90	<2e-16 ***	
$\alpha_1$	-0.2890	0.0009	-329.11	<2e-16 ***	
$\alpha_2$	-0.0829	0.0016	-50.41	<2e-16 ***	
$\gamma_{11}$	0.0608	0.0016	37.98	<2e-16 ***	
$\gamma_{21}$	-0.1266	0.0030	-41.90	<2e-16 ***	
$\gamma_{12}$	0.2038	0.0020	104.05	<2e-16 ***	
$\gamma_{22}$	0.3445	0.0036	95.84	<2e-16 ***	
---					
Signif.	codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
(Dispersion parameter for poisson family taken to be 1					
Null deviance: 28880125 on 4034 degrees of freedom					
Residual deviance: 33033 on 4024 degrees of freedom					
AIC: 70141					

Calculamos um intervalo de confiança de 95% para a média das observações da nossa amostra de 10% dos dados. Para cada observação, verificamos se o valor observado estava dentro do intervalo gerado. O índice de acerto de predição foi de aproximadamente 85%, o que consideramos um bom ajuste.

Estando os parâmetros estimados e os modelos validados, é possível a estimação dos valores projetados das taxas centrais de mortalidade para os anos futuros. A Fig. 26 exhibe o logaritmo das probabilidades de morte projetadas através do método GLM para as gerações femininas e masculinas nascidas nos anos 2004 e 2028. Assim como no método LC, nota-se uma redução das probabilidades de morte entre os anos 2004 e 2028 para os dois sexos. Apesar de não haver relação estabelecida entre os sexos para estimação das probabilidades de morte, as estimativas apresentaram o comportamento esperado da curva feminina possuir valores menores do que a curva masculina.

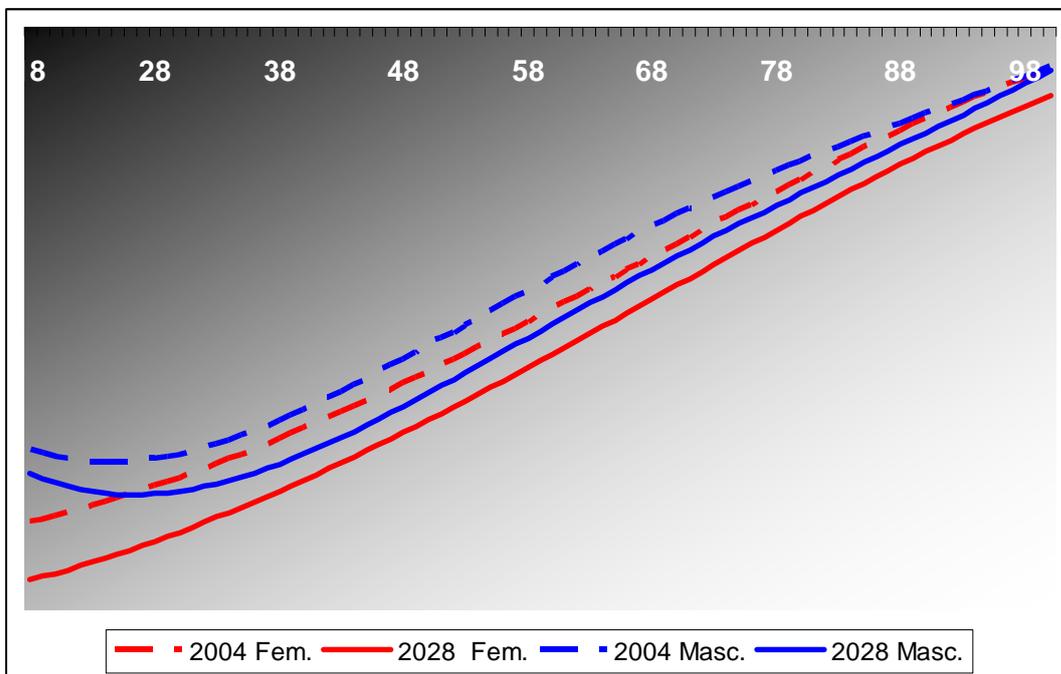


Figura 26 – Logaritmo dos  $q_x$  projetados via método GLM

Para fins de comparação entre os modelos, as Fig. 27 e 28 ilustram, respectivamente, o logaritmo das probabilidades de morte femininas e masculinas para as idades 30, 50 e 70, projetadas para os anos 2004 a 2028 pelos métodos LC e GLM. Nota-se o mesmo padrão de comportamento para as taxas, sendo que o método LC projeta sempre taxas menores do que o método GLM.

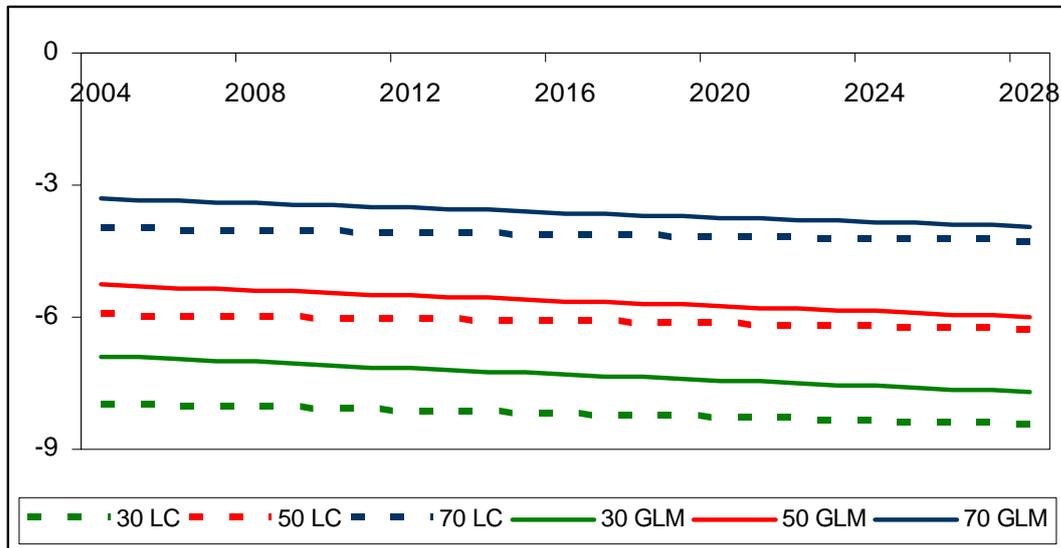


Figura 27 – Comparativo do logaritmo dos  $q_x$  projetados - Feminino

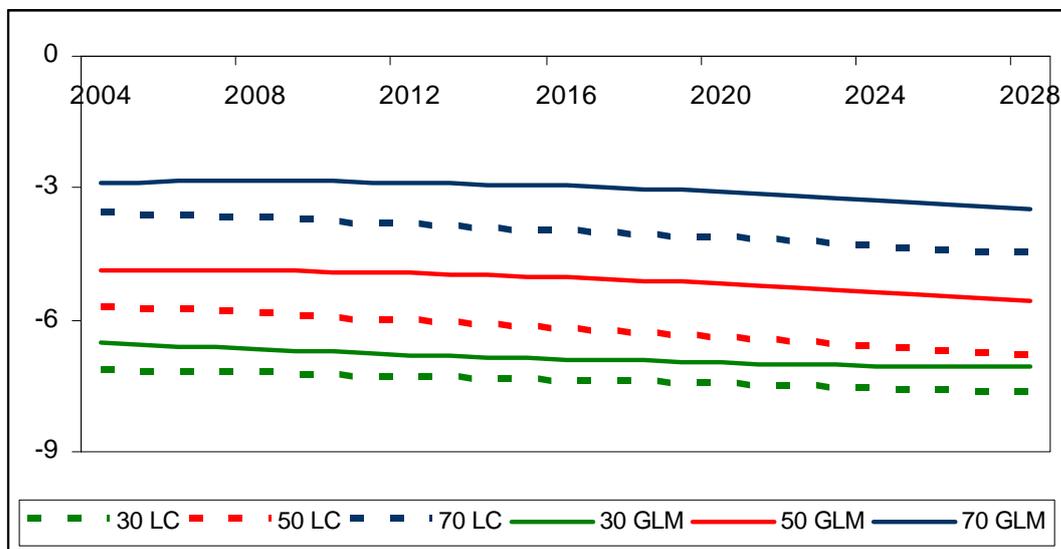


Figura 28 – Comparativo do logaritmo dos  $q_x$  projetados - Masculino

A partir dos  $q_{x,t}$  projetados, nascem as tábuas de mortalidade geracionais.