

6 Estimação

6.1. Introdução

Várias técnicas foram utilizadas ao longo dos últimos anos na estimação de modelos da ETTJ. O trabalho de Gibbons e Ramaswamy (1993) foi pioneiro na estimação do modelo de CIR, os autores usaram momentos amostrais incondicionais dentro do arcabouço do método generalizado dos momentos – GMM – para estimar os parâmetros do modelo de CIR. Outro método com bastante tradição em estudos empíricos da ETTJ é a estimação por máxima verossimilhança. Chen e Scott (1993), Pearson e Sun (1994) e Duffie e Singleton (1997) usaram esse método na estimação de modelos do tipo CIR. Em linhas gerais, a idéia da estimação por quasi-máxima verossimilhança consiste de dois passos: 1. A partir de um número de taxas da ETTJ igual ao número de fatores/ estados do modelo, obtém-se o valor das variáveis de estado através da inversão da equação (25). 2. A partir daí, de posse do valor das variáveis de estado e da função de verossimilhança de tais variáveis⁴⁰, maximiza-se a função de verossimilhança em relação aos parâmetros do modelo, obtendo-se os estimadores. Outros métodos de estimação tem sido utilizados mais recentemente, Dai e Singleton (2000) utilizam o método dos momentos simulados (SMM⁴¹) que garante estimadores consistentes para qualquer modelo da família afim. Lamoureux e Witte (2002) utilizam técnicas de Monte Carlo e cadeias de markov para estimar o modelo de CIR, obtendo estimadores consistentes para os parâmetros do modelo.

⁴⁰ É importante salientar que dependendo da especificação para a dinâmica das variáveis de estado, a forma da função de verossimilhança de tais variáveis não é conhecida, exigindo o uso de valores aproximados dessa função (no caso de Vasicek e CIR, a função de verossimilhança tem forma analítica conhecida). Ait-Sahalia (1999) e Ait-Shalia e Kimmel (2002) desenvolvem métodos para a obtenção de aproximação analítica da função de verossimilhança no caso em que a especificação da dinâmica das variáveis de estado não fornece uma forma fechada para essa função de verossimilhança.

⁴¹ Simulated Method of Moments.

Neste trabalho, utilizamos o filtro de Kalman/ quasi-máxima verossimilhança como método de estimação para os quatro modelos que estudamos. Esse método de estimação já foi descrito e utilizado em trabalhos como Lund (1997), Duan e Simonato (1995) e Babbs e Nowman (1999).

Cada um dos métodos de estimação acima apresenta vantagens e desvantagens. Para citar os métodos mais modernos, Lamoureux et al. (2002) e Dai et al.(2000) apresentam a vantagem de fornecerem estimadores consistentes para qualquer modelo da família afim a ser estimado. Entretanto, ambos os métodos são computacionalmente bastante intensivos, o que dificulta bastante a sua implementação. O método de verossimilhança de Duffie e Singleton é computacionalmente bem menos intensivo, entretanto, para obter o valor da variável de estado a partir das taxas da ETTJ, o pesquisador precisa apontar quais taxas/ prazos da ETTJ seriam observadas sem erro, uma das críticas a esse método é que não parece claro porque a taxa de prazo de 1 ano seria observada sem erro e a taxa do prazo de 3 meses seria observada com erro, por exemplo.

No nosso caso, o grande atrativo do método que utilizamos foi sua boa performance em pequenas amostras, Duffee e Stanton (2001) realizaram, para um modelo da família afim, um estudo comparativo entre os vários métodos de estimação citados. O método do filtro de Kalman apresentou o melhor desempenho, inclusive em relação ao viés dos estimadores, em pequenas amostras.

Além disso, dada a flexibilidade do método, a partir da programação de um “código-mãe” em Matlab®⁴² para a estimação dos modelos de Vasicek, algumas modificações nos permitiram estimar também dois modelos à la CIR. Nenhum dos outros métodos nos permitiria atingir tais objetivos. O método do filtro de Kalman é utilizado a partir de sistemas de espaço de estado. Assim, nas próximas seções colocamos os modelos de Vasicek e de CIR na forma de sistemas de espaço de estado em tempo discreto.

⁴² Ver Apêndice A10 para o código para o modelo de Vasicek de um fator.

As próximas seções estão baseadas em Lund(1997 e 1997a), Duan e Simonato(1995), Gorieroux e Jasiak(2001) e Babbs e Nowman (1999).

6.2.

O modelo de Vasicek na forma de um sistema de espaço de estado em tempo discreto

Nessa seção, colocamos o modelo de Vasicek na forma de um sistema de espaço de estado em tempo discreto. Voltamos as duas principais equações do modelo de Vasicek de um fator:

$$dx_t = \kappa(\theta - x_t)dt + \sigma dW_t$$

$$y_{t,\tau} = A(\tau)/\tau + B(\tau)/\tau \cdot x_t$$

Como vemos, a equação que descreve a dinâmica da taxa livre de risco da economia está em tempo contínuo, utilizando o método de discretização de Euler descrito no capítulo 11 de Gourieroux e Jasiak (2001)⁴³ – obtemos como processo discreto para x_t , a equação dada abaixo:

$$x_{t+h} = \theta * (1 - e^{-\kappa * h}) + e^{-\kappa * h} * x_t + \sigma * \left(\frac{1 - e^{-2\kappa * h}}{2\kappa} \right)^{1/2} * \eta_t \quad (42)$$

onde:

η_t é um choque de distribuição normal $N \sim (0,1)$;

h é o intervalo de tempo decorrido entre duas observações consecutivas do processo discreto de x_t . No caso, por exemplo, de discretizarmos o processo de modo que tenhamos uma observação por semana do processo, teríamos para um calendário em dias corridos $h=0,1944$ ($\sim 7/360$) e para um calendário em dias úteis $h=0,01984$ ($\sim 5/252$). Como trabalhamos com observações semanais e calendário em dias úteis 252, o valor de nosso h é 0,01984.

⁴³ Esse método de discretização foi utilizado em Duan e Simonato (1995).

É importante notar que a discretização em (42) mantém o primeiro e o segundo momento do processo discretizado iguais aos primeiro e segundo momentos do processo contínuo. Tanto para o processo contínuo como para o processo discreto de x_t , temos como primeiro e segundo momentos:

$$E(x_{t+h} | x_t) = \theta * (1 - e^{-\kappa * h}) + e^{-\kappa * h} * x_t \quad (43)$$

$$Var(x_{t+h} | x_t) = \sigma^2 \cdot \left(\frac{1 - e^{-2\kappa * h}}{2\kappa} \right)$$

Como veremos, tal fato não ocorre na discretização do processo de CIR e é o motivo pelo qual a estimação do modelo de CIR é considerada uma estimação “aproximada”.

A equação (25) - $y_{t,\tau} = A(\tau) / \tau + B(\tau) / \tau \cdot x_t$ - determina as taxas da ETTJ a partir do valor para as variáveis de estado.

A equação (42) determina a dinâmica temporal da variável de estado⁴⁴ – por isso, tal equação é chamada equação de estado da economia. A equação (25) liga as variáveis de estado da economia às taxas observadas na ETTJ – por isso, tal equação é chamada equação das observações. A equação (42) também é conhecida como a dimensão temporal da ETTJ, pois ela determina como a ETTJ varia ao longo do tempo, afinal na equação (25) que determina a ETTJ, todos os parâmetros são invariantes no tempo, à exceção da variável de estado que tem sua dinâmica descrita pela equação (42). Além disso, a equação (25) é conhecida como a dimensão de corte da ETTJ, pois dado um período t , essa equação determina as várias taxas vigorando na ETTJ nesse momento.

Como vemos, a equação (25) implica uma correlação perfeita entre todas as taxas da ETTJ, afinal todas as taxas dependem da variável(is) de estado(s). É desnecessário dizer que tal correlação perfeita não se verifica na prática. As razões para isso podem ser diversas. Problemas em relação ao recolhimento do valor das

taxas observadas, por exemplo, podem ser interpretados como ruídos que nos impedem de observar a verdadeira ETTJ, sendo que aquilo que observamos é a ETTJ mais um ruído.

Assim, uma hipótese adicional muito utilizada no estudo empírico da ETTJ é que tais taxas são observadas com erro. Em nosso caso, seguindo a maior parte da literatura sobre o tema, utilizamos a hipótese de que **cada taxa da ETTJ** é observada com um erro normal idiossincrático de distribuição $N(0, \sigma(\varepsilon))$ ⁴⁵. Assim, a forma final do modelo de Vasicek na forma de um sistema de espaço de estado é dada por:

$$x_{t+h} = \theta * (1 - e^{-\kappa * h}) + e^{-\kappa * h} * x_t + \sigma * \left(\frac{1 - e^{-2\kappa * h}}{2\kappa} \right)^{1/2} * \eta_t$$

$$y_{t,\tau} = A(\tau) / \tau + B(\tau) / \tau \cdot x_t + \varepsilon_t$$
(44)

6.3. O filtro de Kalman linear

Considere o sistema de espaço de estado gaussiano dado abaixo:

$$S_t = C + F \cdot S_{t-1} + G \cdot v \quad (45)$$

$(n_x 1) \quad (n_x 1) \quad (n_x n) \quad (n_x 1) \quad (n_x n) \quad (n_x 1)$

$$y_t = A + H \cdot S_t + \omega \quad (46)$$

$(l_x 1) \quad (l_x 1) \quad (l_x n) \quad (n_x 1) \quad (l_x 1)$

As matrizes variância-covariância dos resíduos são dadas por:

$$R = E(\omega_t \omega_t) \text{ e } Q = E(v_t v_t) \quad (47)$$

$(l_x l) \quad (l_x l)$

⁴⁴ Em todo esse trabalho variável de estado, fatores ou fatores de risco são termos equivalentes.

⁴⁵ Em todo esse trabalho variável de estado, fatores ou fatores de risco são termos equivalentes.

A equação (45) é a equação de estado do sistema. O vetor S_t de variáveis de estado não é observado, sua dimensão é de um vetor $n \times 1$, indicando que esse é um sistema de n fatores. Em um modelo de Vasicek de 2 fatores, a dimensão do vetor S_t é 2.

A equação (46) é a equação das observações do sistema. As observações y_t são observadas, a dimensão do vetor de observações é de 1×1 . No nosso caso, vamos estimar nossos modelos com base na observação das taxas *swap* de 30, 60, 180 e 360 dias, logo a dimensão do nosso vetor de observações é 4.

Supondo conhecidas as matrizes C , F , G , A , H , R e Q , o filtro de Kalman linear nos fornece uma predição ótima para o valor das variáveis de estado (fatores) no instante t , dada a informação (ou seja, as observações) até o instante $t-1$. Além disso, o algoritmo do filtro de Kalman nos fornece a estimativa ótima para o vetor das variáveis de estado no instante t , dada toda a informação do sistema até t . Finalmente, o filtro de Kalman nos permite avaliar o valor da função de verossimilhança, a partir dessas estimativas para o vetor de variáveis de estado (fatores).

De forma reduzida, escrevemos abaixo as principais equações do algoritmo do filtro de Kalman:

$$\hat{S}_{t|t-1} = C + F \cdot \hat{S}_{t-1|t-1} \quad (48)$$

onde o valor do vetor inicial, $\hat{S}_{0|0}$, é dado pela esperança incondicional de S calculada a partir de (45);

$\hat{S}_{t|t-1}$ é a melhor predição para \hat{S}_t dado o conjunto de informação até $t-1$.

Referimo-nos a $\hat{S}_{t|t-1}$ como vetor de estados preditos.

$$P_{t|t-1} = F P_{t-1|t-1} F' + G Q G' \quad (49)$$

onde $P_{t|t-1}$ é a matriz de variância-covariância para o erro da predição, feita em t-1, do vetor de estado \hat{S}_t .

$$E(y_t | I_t) = A + H \cdot \hat{S}_{t|t-1} \quad (50)$$

onde $E(y_t | I_t)$ representa a melhor predição, em t-1, para as variáveis y_t . Referimo-nos a esse vetor como vetor de observações preditas.

Passamos a considerar agora para as estimativas do vetor de variáveis de estado e do vetor das variáveis de observação o conjunto de informação até t.

$$\hat{S}_{t|t} = \hat{S}_{t|t-1} + P_{t|t-1} H' (H P_{t|t-1} H' + R)^{-1} \left[y_t - (A + H \hat{S}_{t|t-1}) \right] \quad (51)$$

onde $\hat{S}_{t|t}$ é a melhor estimação para o vetor de variáveis de estado/ fatores, em t, dada a informação, ou seja, todas as observações até t, inclusive. Referimo-nos a $\hat{S}_{t|t}$ como o vetor de variáveis de estado filtrado.

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} H' (H P_{t|t-1} H' + R)^{-1} H P_{t|t-1} \quad (52)$$

onde $P_{t|t}$ é a matriz de erro associada ao vetor $\hat{S}_{t|t}$.

A partir das equações (48) a (52) é possível recursivamente obter estimativas para os vetores $\hat{S}_{t|t-1}$ e $\hat{S}_{t|t}$ para $0 < t < T$ (data final das observações).

A função de log-verossimilhança é dada por:

$$\log L(Y_t) = \sum_{t=1}^T \log f(y_t | I_{t-1})$$

$$f(y_t | I_{t-1}) = (2\pi)^{-1/2} |HP_{t|t-1}H' + R|^{-1/2} \cdot \exp \left[-\frac{1}{2} (y_t - A - HS_{t|t-1})'(HP_{t|t-1}H' + R)^{-1} (y_t - A - HS_{t|t-1}) \right]$$

(53)

Os valores filtrados e preditos para o vetor de variáveis de estado são ótimos entre todos os estimadores lineares e, dada a natureza gaussiana do sistema, tais estimadores são consistentes e ótimos entre todos os estimadores lineares ou não para o vetor de variáveis de estado⁴⁶.

Voltamos agora às equações principais do modelo de Vasicek de um fator sob a forma de um sistema de estado de espaço:

$$x_{t+h} = \theta * (1 - e^{-\kappa * h}) + e^{-\kappa * h} * x_t + \sigma * \left(\frac{1 - e^{-2\kappa * h}}{2\kappa} \right)^{1/2} * \eta_t$$

$$y_{t,\tau} = A(\tau) / \tau + B(\tau) / \tau \cdot x_t + \varepsilon_t, \text{ onde:}$$

$$A(\tau) = \ln \left[\frac{\exp(\gamma * (B(\tau) - (\tau)) - \sigma^2 * B(\tau)^2)}{4 * \kappa} \right]$$

$$B(\tau) = \left[\frac{(1 - \exp(-\kappa * \tau))}{\kappa} \right]$$

É fácil ver que esse é um sistema da forma do sistema dado pelas equações (45) e (46). Sendo que a equação da dinâmica da variável de estado ocupa a posição da equação (45) e a equação das observações das taxas da ETTJ ocupa a posição da equação (46).

⁴⁶ Para maiores informações sobre o algoritmo do filtro de Kalman ver o capítulo 13 de Hamilton (1994).

Nesse caso do Vasicek de um fator, as matrizes C , F , G , A^{47} , H , R e Q do algoritmo do filtro de Kalman são então dadas por:

$$C = \theta * (1 - e^{\kappa \cdot h}); F = e^{\kappa \cdot h} \text{ e } G = \sigma * \left(\frac{1 - e^{-2\kappa \cdot h}}{2\kappa} \right)^{1/2};$$

$$A = \frac{\ln \left[\frac{\exp(\gamma * (B(\tau) - (\tau)) - \sigma^2 * B(\tau)^2)}{4 * \kappa} \right]}{\tau} \text{ e } H = \frac{\left[\frac{(1 - \exp(-\kappa * \tau))}{\kappa} \right]}{\tau}; \text{ e}$$

$$R = \text{Var}(\varepsilon_t) \text{ e } Q = \text{Var}(\eta_t) = 1$$

O problema é que para calcular as predições e filtrações do vetor de variáveis de estado e o valor da função de logverossimilhança, é preciso conhecer C , F , G , A , H , R e Q , mas essas matrizes são funções dos parâmetros do modelo. Esse é o nosso trabalho de estimação, a partir de chutes iniciais para os parâmetros do modelo (σ , λ , θ , κ) e de um chute inicial para o valor da variância do erro da equação das observações ($\text{var}(\varepsilon)$), vamos maximizar o valor da função de logverossimilhança descrita no algoritmo de filtro de Kalman em relação aos parâmetros, através de um algoritmo de maximização de Newton-Raphson. Nossas estimativas serão os valores dos parâmetros que maximizam a função de logverossimilhança.

Em linhas gerais, a idéia da estimação é a seguinte:

1. Chuta-se um valor inicial para os parâmetros do modelo da ETTJ σ , λ , θ , κ e para a variância da equação das observações - $\text{var}(\varepsilon)$;
2. A partir dos parâmetros, definem-se as matrizes: C , F , G , A , H , R e Q ;
3. Calcula-se o valor filtrado e predito do vetor de variáveis de estado S_t para cada t , $0 < t < T$;

⁴⁷ Chamamos atenção para a notação aqui. Temos a matriz A referente ao sistema genérico de estado de espaço definida nas equações (45) e (46) e a matriz $A(\tau)$ que é solução da equação parcial diferencial de um modelo da ETTJ. As matrizes A e $A(\tau)$, *a priori*, não tem qualquer relação entre si.

4. calcula-se o valor da função de logverossimilhança. Se o máximo da função foi atingido (o máximo é atingido quando a diferença de valor entre duas iterações consecutivas da função de logverossimilhança é inferior a um valor de tolerância), o processo é encerrado. Caso o máximo não tenha sido atingido, volta-se para o passo 1, sendo que o novo chute para os parâmetros σ , λ , θ , κ e $\text{var}(\varepsilon)$ é escolhido através de uma função de otimização Newton-Raphson.

Ao fim desse processo, os valores finais de σ , λ , θ , κ e $\text{var}(\varepsilon)$ são as estimativas do filtro de Kalman para os parâmetros do modelo de Vasicek de um fator. O procedimento para a estimação do modelo de Vasicek de dois fatores é uma extensão trivial do procedimento descrito até aqui, já que o algoritmo do filtro de Kalman trata é multidimensional.

6.4.

O modelo de CIR na forma de um sistema de espaço de estado em tempo discreto

De forma análoga à seção 6.2, discretizamos o modelo de CIR de um fator, apresentando-o abaixo na forma de um sistema de espaço de estado:

$$x_{t+h} = \theta * (1 - e^{-\kappa * h}) + e^{-\kappa * h} * x_t + \left[\frac{\sigma^2}{\kappa} (e^{-\kappa * h} - e^{-2\kappa * h}) * x_t + \frac{\theta \sigma^2}{2\kappa} * (1 - e^{-\kappa * h})^2 \right]^{1/2} * \eta_t$$

$$y_{t,\tau} = A(\tau) / \tau + B(\tau) / \tau * x_t + \varepsilon_t$$

onde:

η_t é um choque de distribuição normal $N \sim (0,1)$;

h é o intervalo de tempo decorrido entre duas observações consecutivas do processo discreto de x_t ;

ε_t é um vetor-choque de componentes não correlacionados de distribuição normal $N \sim (0, \text{var}(\varepsilon))$;

A discretização de Euler no modelo de CIR não produz uma aproximação perfeita da dinâmica em tempo contínuo da(s) variável(is) de estado, no sentido de que o primeiro e segundo momento da variável x_t na forma discreta é somente

uma aproximação do primeiro e do segundo momento da variável x_t na forma contínua. O primeiro e segundo momento para a dinâmica em tempo contínuo de x_t podem ser encontrados no capítulo 11 de Gorieroux e Jasiak (2001). O primeiro e segundo momento do processo em tempo discreto podem ser facilmente inferidos das equações acima.

Outra modificação gerada pela discretização do processo de x_t é a mudança na distribuição condicional da variável x_t . Na forma contínua a distribuição condicional de x_t tem a forma de uma distribuição qui-quadrada, enquanto na forma discreta dada acima a distribuição condicional de x_t é uma distribuição normal.

A variável de estado x_t , como explicamos, no capítulo 4, nunca assume valores negativos. Entretanto, a discretização do processo abre a possibilidade de que x_t passe a assumir valores negativos.

Assim, a discretização do modelo de CIR gerou modificações na natureza do processo das variáveis de estado, quais sejam: 1. o primeiro e segundo momento são aproximações do primeiro e segundo momento originais; 2. a distribuição condicional de x_t , que era uma qui-quadrada, passou a ser uma distribuição normal; 3. a(s) variável(is) de estado passam a poder assumir valores negativos.

Essas modificações exigem algumas alterações no filtro de Kalman para que a otimalidade dos estimadores continue valendo.

6.5.

O filtro de Kalman não-linear

Duas modificações são introduzidas no algoritmo do filtro de Kalman linear para a estimação dos parâmetros do modelo de CIR.

1. A primeira modificação diz respeito à natureza heteroscedástica do modelo de CIR, onde a variância da variável de estado depende do nível da

variável de estado. A matriz G do filtro de Kalman passa em levar em conta a heterocedasticidade da variância da variável de estado;

2. Em relação à não-negatividade do valor das variáveis de estado, seguimos Lund (1997) que sugere que quaisquer valores preditos e/ ou filtrados para as variáveis de estado x_t que assumam valores negativos, em determinado instante t , devem ser substituídos por zeros. Esse procedimento é a razão da não-linearidade do filtro de Kalman.

Ainda não está claro se os estimadores do filtro de Kalman não-linear para os modelos afim não-gaussianos – como é o caso de CIR – são consistentes. Lund (1997a) discute amplamente essa questão, concluindo que os estimadores do filtro de Kalman não-linear são os melhores entre todos os estimadores lineares para o modelo de CIR, mas apresentando outros estimadores não-lineares possivelmente mais eficientes em grandes amostras.

É importante notar que Duffee e Stanton (2001) realizam um estudo comparativo entre vários métodos de estimação em amostras pequenas - similares em tamanho a usada por nós nesse trabalho. Os estimadores do filtro de Kalman não-linear apresentaram o menor viés entre todos os outros métodos econométricos testados, que incluíam o método da verossimilhança, o método simulado dos momentos e o método eficiente dos momentos.

6.6.

Experimento de Monte Carlo

Como vimos, os estimadores obtidos através do filtro de Kalman são ótimos e consistentes no caso de modelos da ETTJ gaussianos, como o de Vasicek. Os estimadores obtidos através do filtro de Kalman não-linear para modelos não-gaussianos, como o de CIR, podem ser inconsistentes, mas segundo Duffee e Stanton (2001) apresentam a melhor performance em pequenas amostras. Além disso, seu tempo de computação é de 30 a 60 vezes mais rápido do que métodos com os que utilizam cadeias de Markov e simulação de Monte Carlo (ver Lamoureux e Witte (2002)).

Com o intuito de testarmos as rotinas em MatLab® que construímos para nossas estimações, seguindo Duan e Simonato (1995), fizemos um pequeno experimento de Monte Carlo para estudar o viés em pequenas amostras de nossos estimadores, no caso do modelo de Vasicek.

Para tanto, escolhemos de antemão os parâmetros do modelo: σ , λ , θ , κ e $\text{var}(\varepsilon)$.

De posse do valor desses parâmetros, e, a partir da equação discretizada da variável de estado para o modelo de Vasicek, simulamos 500 observações, com espaçamento de uma semana⁴⁸ da variável de estado⁴⁹.

A partir daí, utilizando a equação da ETTJ obtivemos as taxas de juros para prazos de 30, 60, 180 e 360 dias. Para cada observação dessas taxas, simulamos uma observação normal com distribuição $N(0, \text{var}(\varepsilon))$ para o ruído de observação da taxa⁵⁰. Consideramos as taxas observadas as taxas inicialmente calculadas mais o ruído.

Simulamos assim uma amostra de 500 semanas em quatro prazos diferentes (30, 60, 180 e 360 dias). Utilizando o algoritmo de Kalman, estimamos os parâmetros com essa amostra.

Repetimos esse procedimento 50 vezes, realizando 50 estimações para 50 amostras diferentes.

Abaixo reportamos o valor correto do parâmetro e a média e o desvio-padrão das 50 estimativas realizadas para os parâmetros.

⁴⁸ Significando com isso que $h=0,01984(\sim(5/252))$.

⁴⁹ Simulamos as 500 observações normais necessárias para esse exercício, utilizando a função `randn` do MatLab®.

⁵⁰ Como temos quatro prazos diferentes (30, 60, 180 e 360) observados em 500 semanas diferentes, simulamos mais 2000 observações normais pela função `randn`.

	Valor Correto	Valor Estimado	
		Média	Desv.Pad.
K	0,5000	0,498	0,008
θ	0,2000	0,216	0,138
σ	0,3000	0,302	0,010
λ	0,5000	0,473	0,235
$\sigma(\varepsilon)$	0,0100	0,004	0,009

Tabela 2: Experimento de Monte Carlo

6.7.

Comentário sobre a função de otimização utilizada

Utilizamos duas funções diferentes para maximizar a função de verossimilhança em relação aos parâmetros do modelo. A primeira função que utilizamos era uma função de otimização côncava, à qual chamamos *optimal*⁵¹. Essa função implementa o algoritmo de Newton-Raphson, sendo que o passo ótimo que escolhe a distância que os parâmetros andam na direção do gradiente, não leva em conta o Hessiano da função. Escolhe-se o passo ótimo pelo algoritmo de Fibonacci.

A segunda função de otimização que utilizamos é função *fminsearch* que faz parte da biblioteca de funções do MatLab®. Essa função é uma função de otimização linear, fazendo uso de um método à la *grid search* multidimensional para otimizar a função em função dos parâmetros.

Fizemos todas as tentativas e combinações possíveis entre essas duas funções, e o procedimento que nos pareceu levar aos melhores resultados tanto quanto à precisão da estimação com quanto ao tempo utilizado, foi o de primeiro utilizar a função *optimal* e, depois, a partir dos resultados obtidos, utilizarmos a função *fminsearch*.

⁵¹ O código dessa função encontra-se no Apêndice A11.

Acreditamos que a razão para isso é que os resultados obtidos com a função `fminsearch` são muito dependentes do chute inicial para os parâmetros, mas, ao mesmo tempo, seu tempo de convergência é infinitamente menor que o da função `optimal`. Assim, utilizávamos a função `optimal` para chegarmos a um chute inicial bem próximo do que seria o valor correto para as estimativas – o tempo de convergência dessa chegava a levar até 7 horas por estimação. Utilizávamos, então, a função `fminsearch` para obtermos o valor das estimativas – o tempo de convergência nesse caso raramente ultrapassava 5 minutos.