

7

Estimação de Parâmetros Desconhecidos e a Questão dos Diagnósticos

Neste capítulo, são apresentadas e discutidas as expressões referentes às funções de log verossimilhança dos modelos em EE lineares (condicionalmente) Gaussianos e também da função de log verossimilhança aproximada relativa aos modelos em EE não lineares também Gaussianos (isto é, com erros Gaussianos) abordados via algum filtro não linear aproximado. Uma vez definido um modelo em EE linear (ou não linear, mas com tratamento aproximado), e conseqüentemente suas respectivas funções de log verossimilhança (ou de log verossimilhança aproximada, para modelos em EE não lineares com tratamento aproximado), torna-se possível a estimação de parâmetros desconhecidos, quantidades estas que, praticamente, sempre se fazem presentes em situações práticas contextualizadas no âmbito de uma análise estatística, por máxima verossimilhança (ou máxima verossimilhança aproximada). Naturalmente, as estimativas associadas a este método são obtidas via maximização das funções de log verossimilhança adotadas, procedimento que, em geral, exige e obriga o uso de métodos numéricos.

Também são discutidos alguns aspectos de estimação de máxima *quasi* verossimilhança, tecnologia geralmente adequada, principalmente, quando há indícios de violação do pressuposto de normalidade (condicional) do modelo proposto, por parte dos dados disponíveis ao ajuste.

Ao final, alguns procedimentos de diagnósticos são apresentados, como, por exemplo, gráficos e testes para resíduos, além de algumas referências adicionais sobre este tópico.

7.1

Função de Log Verossimilhança para Modelos Lineares

Suponha-se que esteja disponível uma série temporal de tamanho n (não necessariamente já observada¹) de um processo estocástico Y_t p -variado seguindo um modelo em EE linear (condicionalmente) Gaussiano, do tipo discutido no capítulo 2. Denote-se o vetor de parâmetros desconhecidos do modelo por ψ , seu espaço paramétrico por Θ e função de log verossimilhança referente à série disponível e a ao vetor ψ por $\log L(\psi)$, sendo que L é a função de verossimilhança, cuja expressão analítica genérica, avaliada em ψ , é usualmente dada por $L(\psi) \equiv p(Y_1, \dots, Y_n) = p(Y_1) \prod_{t=2}^n p(Y_t | Y_1, \dots, Y_{t-1})$. Será discutido mais adiante que a avaliação numérica desta última expressão depende sucintamente de quantidades associadas às recursões de Kalman discutidas no capítulo 3 e de qual inicialização das mesmas estaria sendo adotada.

7.1.1

Inicialização Não Difusa

Quando o modelo em EE promove meios para que se estabeleçam as quantidades iniciais $a_1 \equiv E(\alpha_1)$ e $P_1 \equiv V(\alpha_1)$, como é o caso de todos os modelos estacionários² [cf. Durbin e Koopman (2001a), páginas 111 e 112, para busca destas quantidades iniciais de modelos ARMA estacionários em forma de espaço de EE], então, sob normalidade, não é difícil deduzir que a log verossimilhança assumirá a forma específica

$$\log L(\psi) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \left(\log|F_t| + v_t' F_t^{-1} v_t \right), \quad (7.1)$$

¹ No entanto, a série há de ser observada no momento do cômputo das estimativas.

² Neste contexto de modelagem em EE linear, uma condição necessária e suficiente para estacionariedade é que o modelo seja tempo invariante e que a matriz T tenha todos os seus autovalores pertencentes ao círculo unitário [vide Anderson e Moore (1979) e Harvey (1989)].

na qual se nota a dependência de quantidades provenientes das recursões de Kalman, a saber, as inovações $v_t = Y_t - Z_t a_t$ e a sua matriz de covariâncias F_t^3 , para todo t (a_1 e P_1 já são inicialmente disponíveis!). Logo, para que se construa a função de verossimilhança avaliada em ψ , *deve-se passar o Filtro de Kalman aos dados uma vez com as matrizes do sistema avaliadas em ψ* . Para tanto, utilizam-se as equações de previsão dadas em (3.2) combinadas às de atualização dadas em (3.3), ou alternativamente o Filtro de Kalman “2 em 1”, cujas expressões estão em (3.4). Harvey (1989) popularizou a expressão de (7.1) pelo nome de *decomposição pelo erro de previsão*.

7.1.2

Inicialização Difusa

Para modelos que demandam inicializações difusas (conforme a Definição 3.4) – como é o caso de modelos em EE com alguns ou todos os elementos do vetor de estado seguindo processos não estacionários – a expressão para a log verossimilhança talvez precise de algumas modificações, e estas, por sua vez, dependem de qual tipo de inicialização difusa estaria sendo praticada.

7.1.2.1

Inicialização Difusa Aproximada

Quando se adota o procedimento descrito de inicialização aproximada descrito na subseção 3.6.1 desta Dissertação, há-de se conjecturar qual o momento $d+1$ em que a distribuição do vetor de estado passa a ficar própria. Para modelos univariados (isto é, $p = 1$), é possível provar [vide Harvey (1989)] que d coincide com o número de coordenadas não estacionárias do vetor de estado. Em outras situações mais gerais, sem a utilização de recursos como o Filtro de Kalman Exato (vide subseção 3.6.2), a busca de d , como já explanado previamente, assume caráter *ad hoc*.

³ Supõe-se, sem perda de generalidade, que esta matriz, cuja expressão se encontra no final da seção 3.1, é invertível. Caso isto não aconteça, pode-se redefinir o modelo em EE de forma a contornar o problema. Vide Koopman e Durbin (2000) ou o capítulo 6 de Durbin e Koopman (2001a) para maiores detalhes.

Uma vez estabelecido d , e tendo em mente de que as expressões para a inovação e para a sua matriz de variância e covariância não estão teoricamente bem definidos nos instantes $t = 1, \dots, d$, pode-se adotar a seguinte *função de log verossimilhança condicional* dada por

$$\log L(\psi) = -\frac{(n-d)p}{2} \log(2\pi) - \frac{1}{2} \sum_{t=d+1}^n (\log|F_t| + v_t' F_t^{-1} v_t). \quad (7.2)$$

Observe-se que esta versão da log verossimilhança faz com que se percam os d primeiros vetores de medidas, algo que pode significar falta de informação talvez relevante para o processo de estimação de parâmetros desconhecidos, principalmente para séries não tão “longas”. Além disso, as propagações de erro geradas pela adoção de um κ “grande” (mas não “infinito”!) podem gerar algum tipo de viés.

Koopman, Shephard e Doornik (1999) argumentam que não há muita diferença em se usar a expressão reproduzida em (7.1), em substituição àquela referida em (7.2), uma vez que F_t^{-1} , em geral, assume valores muito pequenos (ou seja, é próxima da matriz nula) nos instantes $t = 1, \dots, d$, na adoção de κ “grande”.

7.1.2.2

Inicialização Difusa Exata

Lembrando das vantagens relativas ao processo de inicialização difusa via Filtro de Kalman Exato, discutidas na subseção 3.6.2, destaca-se aquela que diz respeito à possibilidade de se construir uma função de log verossimilhança exata. Com efeito: é possível mostrar, sob esta tecnologia de inicialização, que expressões envolvendo os vetores iniciais das medidas são dedutíveis e podem ser inseridas no somatório de (7.2). Outra vantagem que se observa é a ausência de qualquer tipo de viés induzido por aproximações relativas à escolha de variâncias iniciais “grandes” (porém não “infinitas”).

A expressão para a função de log verossimilhança exata, denominada usualmente por *função de log verossimilhança difusa* e deduzida no capítulo 7 de Durbin e Koopman (2001a), se resume a

$$\log L(\psi) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^d \omega_t - \frac{1}{2} \sum_{t=d+1}^n (\log|F_t| + v_t' F_t^{-1} v_t)$$

na qual

$$\begin{aligned} \omega_t &= \log|F_{\infty,t}| & , F_{\infty,t} > 0 \\ &= \log|F_{*,t}| + v_t^{(0)'} F_{*,t}^{-1} v_t^{(0)} & , F_{\infty,t} = 0 \end{aligned} \quad (7.3)$$

sendo que expressões apropriadas para $F_{\infty,t}$, $F_{*,t}$ e $v_t^{(0)}$, referentes às recursões do Filtro de Kalman Exato, se encontram no capítulo 5 também de Durbin e Koopman (2001a).

Observe-se que esta log verossimilhança exata, assim como as anteriores dadas em (7.1) e (7.2), são diretamente adequadas ao se fazer uso da Metodologia de Doran (vide capítulos 4 e 5), pois na ocasião, embora se trabalhe com um vetor de medidas aumentado, ainda são ajustados modelos em EE lineares (condicionalmente) Gaussianos.

Em ajustes de modelos em EE lineares na parte aplicada desta Dissertação (capítulo 9), a função de log verossimilhança a ser adotada é a exata.

7.2

Função de *Quasi* Log Verossimilhança para Modelos Lineares

Se houver indícios de que o modelo ajustado não for regido por distribuições normais, as expressões dadas em (7.1), (7.2) e (7.3) ainda costumam ser usadas como representantes aproximadas da “verdadeira” função de log verossimilhança, em geral desconhecida. Neste contexto, tais expressões constituem as chamadas funções de *quasi* log verossimilhança, e os estimadores/estimativas, resultantes de maximização destas, são denominados *estimadores/estimativas de quasi verossimilhança*.

7.3

Propriedades Assintóticas dos Estimadores de (Quasi) Máxima Verossimilhança

O estimador de (*quasi*) máxima verossimilhança – o qual é o vetor aleatório definido por $\hat{\psi} \equiv \underset{\psi \in \Theta}{\operatorname{argmax}} \log L(\psi)$ – apresenta, eventualmente, propriedades assintóticas interessantes, como consistência forte e normalidade. No entanto, há que se postular algumas condições suficientes, *nem sempre tão diretas de serem checadas*, para que tais propriedades sejam garantidas. Tal tarefa é dependente do modelo em questão, no que diz respeito a três fatores: a) sua condição de estacionariedade; b) suas características tempo-invariantes; e c) de quais matrizes do sistema dependem ou não de elementos de ψ . Como as aplicações da Dissertação não concernem inferências estatísticas para parâmetros, não serão, portanto, expostos detalhes. O pesquisador mais interessado é convidado a estudar as seguintes referências: Pagan (1980), Harvey (1989), Ghosh (1989), Watson (1989), Bollerslev e Wooldridge (1992), Hamilton (1994) e White (1994).

7.4

Função de Log Verossimilhança Aproximada para Modelos Não Lineares Via Tratamento Aproximado

Suponha-se, agora, que esteja disponível uma série temporal extraída de um processo estocástico Y_t p-variado seguindo um modelo em EE não linear discutido na seção 6.1, com vetores de erro das equações das medidas e do estado Gaussianos e independentes entre si, e com os pressupostos adicionais requeridos pelo filtro não linear aproximado adotado. Denote-se novamente por ψ o vetor de parâmetros desconhecidos e por Θ o respectivo espaço paramétrico. Embora seja teoricamente difícil que se estabeleça uma log verossimilhança “correta” para ψ e para a série em questão, ainda resta uma alternativa matematicamente informal para a estimação de ψ . Esta constitui no método da máxima verossimilhança

aproximada, proposta, por exemplo, em Fuhrer (1992), Tanizaki (1996) e Tanizaki e Mariano (1996).

Em suma, o método é muito simples, dado que já se sabe o procedimento de construção de log verossimilhanças condicionais, como aquela descrita em (7.2). O que se deve fazer é aproximar as verdadeiras inovações e as respectivas matrizes de covariâncias condicionais por expressões derivadas dos filtros não lineares aproximados, gerando uma *função de log verossimilhança aproximada*. Por exemplo, se o Filtro de Kalman Estendido for o escolhido para o processo de estimação recursiva do estado, a função de log verossimilhança aproximada que se levantaria é dada pela expressão em (7.1) ou (7.2), mas com as seguintes expressões:

$$\begin{aligned} \tilde{u}_t &= Y_t - Z_t(a_t) \\ F_t &\equiv \dot{Z}_t P_t \dot{Z}_t' + H_t \end{aligned} \quad , \quad (7.4)$$

nas quais \dot{Z}_t , a_t e P_t estão definidas na subsecção 6.2.1. Construções análogas para os outros filtros não lineares aproximados podem ser estabelecidas.

A estimativa resultante da maximização numérica da função de log verossimilhança aproximada é chamada de *estimativa de máxima verossimilhança aproximada*.

Salientam-se como comentários finais desta seção os que seguem:

- Não há uma teoria assintótica desenvolvida para estes estimadores; e
- Todos os problemas, exaustivamente discutidos no capítulo 6, referentes ao filtro não linear adotado, são automaticamente herdados.

Portanto, não se pode dizer que o método da máxima verossimilhança aproximada é *bom*; porém, se constitui na única alternativa *vigente*.

7.5

Diagnósticos

Antes que se tomem decisões com base em um modelo em EE ajustado (ou qualquer outro modelo estatístico), deve-se verificar se este seria uma aproximação razoável para o verdadeiro mecanismo probabilístico gerador dos dados disponíveis para análise. A esta tarefa, dá-se o nome de *diagnósticos*, os quais, sob o enfoque freqüentista, se caracterizam, basicamente, pela análise gráfica e/ou inferencial de resíduos e pelo poder preditivo do modelo ajustado (isto é, se as expressões matemáticas estimadas conseguem “reproduzir” relativamente bem os dados).

7.5.1

Análise de Resíduos

No contexto de espaço de estado, a análise de resíduos se resume à análise da série temporal das inovações padronizadas $v_t^p \equiv (F_t^{1/2})^{-1}v_t$, na qual $F_t^{1/2}$ é qualquer matriz positiva definida, tal que $F_t = (F_t^{1/2})(F_t^{1/2})$, $t = 1, \dots, n^4$. Estas devem apresentar propriedades como: (1) média constante e aproximadamente nula; (2) homocedasticidade incondicional com variância próxima de 1; (3) ausência de correlação serial; e, de preferência (mas não indispensavelmente), (4) aspecto Gaussiano. Todas estas propriedades podem ser investigadas por procedimentos gráficos e descritivos, tais como os seguintes a serem usados na ordem abaixo⁵:

- 1) *plot* da série no tempo;
- 2) função de autocorrelação (não parcial e parcial);
- 3) função de densidade espectral suavizada;

⁴ Como F_t é matriz simétrica não negativa, então a existência de $F_t^{1/2}$ é assegurada pelo *Teorema da Decomposição da Raiz Quadrada* [cf. Kubrusly (2001), Teorema 5.85].

⁵ Esta ordem apresentada não é obrigatória de ser praticada; mas é recomendável pelo fato de ser *correta*. Outras possibilidades seriam eventualmente lícitas, desde que não transgridam nenhuma lógica estatística. Por exemplo, não seria razoável checar normalidade ou se calcular a variância amostral de uma série, antes que se estabeleça sua condição de estacionariedade.

- 4) média e variância amostrais;
- 5) histograma; e
- 6) QQ *plot* com base na distribuição normal.

Outra forma de se realizar este tipo de análise é caracterizada pela prática intensiva de testes de hipótese de especificações, de modo que se chequem os pressupostos em questão. Citam-se, já dispostos em uma ordem indicada para realização, os seguintes testes⁶:

- 1) Testes tradicionais de heterocedasticidade, como o de White e o do Multiplicador de Lagrange;
- 2) Testes de auto-correlação serial, como o de Ljung-Box e o do Multiplicador de Lagrange;
- 3) Testes de média e de variância baseados no Teorema Central do Limite para Processos Dependentes; e
- 4) Testes de normalidade, como o de Jarque e Bera e o de Anderson e Darling.

Por já estarem muito bem estabelecidos na literatura estatística, estes procedimentos (gráficos, medidas descritivas e testes de hipótese) não serão descritos pormenorizadamente nesta Dissertação – embora sejam eventualmente praticados no capítulo 9. O pesquisador interessado no assunto é convidado a estudar referências-padrões sobre análise de séries temporais; citam-se Hamilton (1994) e Brockwell e Davis (1996) como exemplos. No entanto, também são indicados textos que já direcionam a discussão sobre estes e outros procedimentos para a perspectiva da modelagem em EE, como Harvey (1989), de Jong e Penzer (1998) e Durbin e Koopman (2001a).

⁶ O mesmo conteúdo da nota anterior.

7.5.2

Análise do Poder Preditivo

O poder preditivo de um modelo em EE ajustado pode ser checado através da comparação de valores reais dos dados (no contexto presente, estes são as observações das medidas) com os valores previstos dos mesmos. Este procedimento pode ser realizado através da abordagem “dentro da amostra” (*in sample*) ou da abordagem “fora da amostra” (*out of sample*).

Indique-se por $Y_{t,j}$, $t = 1, \dots, n$, a série relativa a j -ésima coordenada do vetor de medidas, e por $\hat{Y}_{t,j/t-1}$ o correspondente previsto um passo à frente, sendo, obviamente, que $j = 1, \dots, p$. Para se checar a proximidade entre estas duas séries, as maneiras mais naturais seriam através de:

- Gráfico simultâneo das séries no tempo em um mesmo gráfico;
- Diagrama de dispersão entre as duas séries; e
- Cálculo de medidas descritivas, como:

$$\begin{aligned} \text{MAPE} &= 100 \frac{1}{n-d} \sum_{t=d+1}^n \frac{|Y_{t,j} - \hat{Y}_{t,j/t-1}|}{Y_{t,j}} \% \\ \text{MSE} &= \frac{1}{n-d} \sum_{t=d+1}^n (Y_{t,j} - \hat{Y}_{t,j/t-1})^2 \\ \text{PseudoR}^2 &= \left[\text{Corr}(Y_{t,j}, \hat{Y}_{t,j/t-1}) \right]^2. \end{aligned} \quad (7.5)$$

Para modelos adequados aos dados, espera-se que os procedimentos acima indiquem aderência relativamente satisfatória entre “observados” e “previstos” (*Scatter plot* com pontos em torno do gráfico da função identidade, Pseudo R^2 próximo de 1 etc.).