

# Modelos de parâmetros variáveis: uma resenha crítica\*

MARCELO S. PORTUGAL\*\*

*Este artigo apresenta uma resenha da literatura referente a modelos de parâmetros variáveis. Inclui não apenas as abordagens mais recentes, clássicas e bayesianas, baseadas no filtro de Kalman, mas oferece também alguma discussão de modelos mais antigos com variação paramétrica. Estas técnicas são muito úteis em economia, não apenas para fins de previsão, via modelos univariados de extração de sinal, mas ainda para a análise de modelos sujeitos a mudança estrutural.*

## 1 - Introdução

Há bastante tempo, vêm-se estudando modelos que permitem a variação dos parâmetros ao longo do período amostral. As primeiras tentativas de se construir modelos com coeficientes variáveis foram feitas nos anos 50 e 60 por Rubin (1950), Klein (1953), Kuh (1959), Rao (1965) e outros.

Várias histórias podem ser contadas para justificar a razão pela qual se deveria permitir que os coeficientes estimados variassem ao longo da amostra. A mais óbvia e razoável é que os próprios coeficientes verdadeiros são o resultado de um processo estocástico. Em termos de um modelo de série temporal, pode-se supor que o coeficiente seja autocorrelacionado, dependendo de alguma variável econômica, ou simplesmente aleatório com uma dada média e variância.

Por outro lado, quando o verdadeiro coeficiente for estável, o modelo de coeficientes variáveis pode ser visto como uma maneira de se tentar resolver erros de especificação, como a exclusão de variáveis. Assim, uma variação de parâmetro não explicada por alguma teoria pode ser vista como uma evidência de erro de especificação.

Um grande estímulo ao uso de modelos de parâmetros variáveis veio da crítica de Lucas, segundo o qual, quando os agentes econômicos têm expectativas racio-

---

\* Gostaria de agradecer a Ken Wallis, Alan Winters, David Collie e Sanjay Yadav pelos comentários a uma versão anterior deste trabalho.

\*\* Da Universidade Federal do Rio Grande do Sul.

nais, os parâmetros dos modelos econométricos já não serão constantes. Os agentes mudarão seus comportamentos em resposta a mudanças no regime de política econômica. De acordo com suas próprias palavras: “A respeito desse ponto, tenho argumentado simplesmente que a visão de parâmetro estável padrão da teoria econométrica e política quantitativa parece não casar com várias características importantes da prática econômica, enquanto uma estrutura geral alternativa, incorporando parâmetro estocástico com *drift*, combina mais exatamente com estas características” [cf. Lucas (1976)].

Note-se que, se o parâmetro mudar de algum modo paramétrico, existirão outros parâmetros (*deep parameters*) que ainda são constantes, isto é, a média e a variância do processo estocástico subjacente.

O principal objetivo deste trabalho é discutir modelos de parâmetros variáveis desenvolvidos durante as últimas décadas, dando especial atenção ao filtro de Kalman, cuja comparação com modelos anteriores parece evidenciar sua superioridade no tratamento de modelos de parâmetros variáveis.

Este artigo tem mais três seções: na seção seguinte apresentaremos os diversos modelos não-bayesianos introduzidos durante os anos 70, explorando também a relação entre eles; a terceira seção é dedicada ao filtro de Kalman, e contém não apenas a derivação das equações recursivas para filtragem (*filtering*) e suavização (*smoothing*), mas também uma discussão de outros tópicos, tais como diferentes esquemas de variação de parâmetro e a abordagem bayesiana para parâmetros variáveis; por fim, a última seção apresenta as conclusões e observações.

## 2 - Os diversos modelos propostos

### 2.1 - O modelo aleatório puro

O modelo de coeficientes aleatórios puro foi um dos primeiros a tentar resolver o problema da variação temporal [cf. Rao (1965), Hildreth e Houck (1968) e Swamy (1970)]. A idéia básica é criar um modelo em que os coeficientes fixos estimados representem a média da distribuição dos coeficientes variáveis. Assim, embora incorpore explicitamente a possibilidade de os coeficientes não serem fixos ao longo do período amostral, não permite a estimação de diversos coeficientes para cada ponto no tempo.

O modelo básico pode ser escrito do seguinte modo:

$$y_t = X_t' \beta_t + v_t \quad (1)$$

$$\beta_t = \mu + \varepsilon_t \quad (2)$$

onde:  $v_t \sim N(0, \sigma_v^2)$  e  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ .

Adicionalmente, supõe-se que ambos os erros são independentes e não-autocorrelacionados. Combinando (1) e (2):

$$y_t = X_t' \mu + u_t \quad (3)$$

onde:  $u_t = X_t' \varepsilon_t + v_t$ .

Dever-se-ia notar que, após esta substituição, o modelo torna-se um modelo de coeficientes fixos, com erro heterocedástico. A equação (3) pode, então, ser estimada por mínimos quadrados generalizados (MQG), e o coeficiente fixo  $\mu$  é a média estimada de  $\beta_t$ . Em outras palavras, ele mostra a variação média na variável dependente devido à variação de uma unidade nas variáveis exógenas. O estimador de MQG é:

$$b = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \quad (4)$$

O procedimento proposto para se obter  $\Omega$  estimado baseia-se nos resíduos da equação (3) calculada através de mínimos quadrados ordinários (MQO):

$$\begin{aligned} e &= y - X' b_{MQO} \\ &= Mu \end{aligned}$$

onde  $M = I - X(X'X)^{-1}X'$  é uma matriz conhecida.

Então  $E(ee') = E(Muu'M') = M\Omega M$ , da qual:

$$E(e^*) = M^* \text{Var}(u)$$

onde:  $e_{ij}^* = e_{ij}^2$  e  $m_{ij}^* = m_{ij}^2$ .

Finalmente, define-se:  $w = e^* - E(e^*)$  e  $G = M^*X^*$ .

Então:

$$\begin{aligned} e^* &= E(e^*) + w \\ &= M^* \text{Var}(u) + w \\ &= M^* (X^* \sigma_e^2 + \sigma_v^2) + w \\ e^* &= M^* X^* \sigma_e^2 + M^* \sigma_v^2 + w \end{aligned} \tag{5}$$

Uma vez que  $M$  e  $X$  são matrizes conhecidas, podemos estimar (5) através de MQO para se obterem estimativas das variâncias e, conseqüentemente, de  $\Omega$ .<sup>1</sup>

Conforme mostrado por Swamy (1970) e Hildreth e Houck (1968), este procedimento fornece um estimador não-tendencioso e consistente da matriz de variância-covariância além de consistente e assintoticamente eficiente da média.

## 2.2 - O modelo de regressão adaptativa

Desenvolvido inicialmente por Cooley e Prescott (1937a), este é basicamente um modelo de regressão padrão onde o intercepto pode variar. Supõe-se que o termo do erro aditivo é uma soma “não apenas de um elemento transitório que tem efeito no período corrente, mas também uma componente permanente cujo efeito persiste no futuro” [Cooley e Prescott (1973a, p. 364)].

Um exemplo dessa componente permanente poderia ser uma nova variável que, em algum ponto no tempo, se tornasse relevante para a explicação da variável dependente.<sup>2</sup> Neste modelo, o papel do intercepto variando no tempo é precisamente levar em consideração esta componente permanente do erro.

Normalmente, ao se lidar com este tipo de erro, especialmente quando a presença de variáveis omitidas é reconhecida, o procedimento comum é postular um erro auto-regressivo. É também bastante comum “ajustar” o termo constante sempre que as previsões *ex post* não apresentarem um desempenho muito bom.

Ao invés de se usar este procedimento *ad hoc*, Cooley e Prescott propuseram definir um modelo em que o termo constante pudesse variar:

---

1 Um procedimento similar é derivado em Theil (1971, p. 622-625). A diferença é que no caso de Theil o termo aleatório  $w$  é novamente heterocedástico com um padrão de heterocedasticidade conhecido. Assim, ele emprega MQG para obter as variâncias estimadas.

2 Outras possíveis fontes desta componente permanente são mudanças nos gostos e no desenvolvimento tecnológico.

$$y_t = \alpha_t + X_t' \beta + \varepsilon_t \quad (6)$$

$$\alpha_t = \alpha_{t-1} + u_t \quad (7)$$

onde:  $\varepsilon_t \sim N(0, (1 - \delta) \sigma^2)$ ;  $u_t \sim N(0, \delta \sigma^2)$  e  $0 \leq \delta \leq 1$ .

O parâmetro  $\delta$  é usado para medir a importância relativa da componente permanente. Note-se que, se não existir esta componente permanente no termo do erro aditivo, isto é,  $\delta = 0$ , então  $\alpha_t = \alpha_{t-1}$ , e estaremos de volta ao modelo de coeficientes fixos. Para se obter a equação a ser estimada, resolvemos (7) recursivamente e substituímos o resultado em (6):

$$\alpha_t = \alpha_0 + \sum_{i=1}^{t-1} u_{t-i}$$

Então:

$$y_t = \alpha_0 + X_t' \beta + \varepsilon_t + \sum_{i=0}^{t-1} u_{t-i}$$

Efetivamente, por estarem interessados em previsão, Cooley e Prescott resolvem os cálculos recursivos de modo inverso, obtendo  $\alpha_t$  em termos de  $\alpha_{t+1}$  e da soma do erro. Ao invés de se ajustar a constante, eles propuseram estimar seu valor mais recente:

$$y_t = \alpha_{t+1} + X_t' \beta + \varepsilon_t + \sum_{s=t}^T u_s \quad (8)$$

Em notação matricial:<sup>3</sup>

$$y = X' \beta^* + v \quad (9)$$

onde  $\beta^* = [\alpha_{t+1}, \beta_1, \beta_2, \dots, \beta_k]$  e  $v_t = \varepsilon_t + \sum_{s=t}^T u_s$ .

---

<sup>3</sup> Agora  $X'$  inclui uma coluna de 1's, como de costume.

A equação (9) é idêntica ao modelo de regressão padrão de coeficientes fixos. O único problema é o erro, que agora não é independente. Se  $\delta$  for conhecido, então o estimador de MQG de  $\beta^*$  e  $\sigma^2$  é dado por:

$$b_\delta = (X' \Omega_\delta^{-1} X)^{-1} X' \Omega_\delta^{-1} y$$

$$S_\delta^2 = T^{-1} (y - X' b_\delta)' \Omega_\delta^{-1} (y - X' b_\delta)$$

onde  $\Omega_\delta = (1 - \delta)I + \delta R$  e  $R$  é uma matriz  $T \times T$  com elementos  $r_{ij}$  tais que  $r_{ij} = \min [T - i + 1, T - j + 1]$ .<sup>4</sup>

Se não tivermos informações *a priori* acerca do valor de  $\delta$ , teremos que confiar em alguma técnica de localização numérica. Dada a função de verossimilhança das observações:

$$l(y; \beta^*, \sigma^2, \delta) = -T/2 \ln 2\pi - T/2 \ln \sigma^2 - 1/2 \ln |\Omega_\delta| - \\ - 1/2 \sigma^2 [(y - X\beta^*)' \Omega_\delta^{-1} (y - X\beta^*)]$$

podemos substituir os estimadores de  $\sigma^2$  e  $\beta^*$  para obter a função de verossimilhança concentrada:

$$l(y, \beta^*, \sigma^2; \delta) = -T/2 (\ln 2\pi + 1) - T/2 \ln S_\delta^2 - 1/2 \ln |\Omega_\delta|$$

e, então, procurar entre zero e 1 para obter um valor para  $\delta$  que maximize a verossimilhança concentrada.

Note-se que o custo da computação é um problema para esta rotina, visto que a matriz de variância-covariância tem que ser invertida várias vezes, dados os diversos valores de  $\delta$ . Cooley e Prescott (1973a) solucionam este problema transformando as variáveis de tal modo que  $\Omega_\delta$  seja diagonal, facilitando com isso os cálculos. Em outros trabalhos, eles também mostram que  $b_\delta$  é consistente, eficiente assintoticamente e robusto a erros de especificação que provocam mudança estrutural ao longo do tempo, relativamente a métodos de mínimos quadrados com e sem erros auto-regressivos de primeira ordem [Cooley e Prescott (1973b e 1973c)].

<sup>4</sup> Note-se que  $R$  é construída desse modo devido às somas cumulativas em (8) que fazem com que os  $u_t$  sejam autocorrelacionados.

Uma generalização desse modelo, bastante inconveniente, onde o coeficiente angular e o intercepto podem variar e ser estimado em vários pontos no tempo, é apresentada por Cooley e Prescott (1973c e 1976).

### 2.3 - Parâmetros de variação sistemática

Ao invés de postular coeficientes variando seqüencialmente, este modelo, proposto por Belsley (1973a, 1973b e 1973c), supõe que os coeficientes variarão de acordo com um conjunto de variáveis econômicas:

$$y_t = X_t' \beta_t + \varepsilon_t \quad (10)$$

$$\beta_t = \delta Z_t + u_t \quad (11)$$

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2 I)$$

$$u_t \sim N(0, \sigma_u^2 Q)$$

onde  $X_t$  e  $Z_t$  são vetores de dimensão  $k$  e  $r$ , respectivamente, e  $\beta$  é um vetor de dimensão  $k$ . Substituindo (11) em (10), obtemos:

$$\begin{aligned} y_t &= X_t' (\delta Z_t + u_t) + \varepsilon_t \\ &= (X_t' \otimes Z_t') \Gamma + v_t \\ &= w_t' \Gamma + v_t \end{aligned} \quad (12)$$

onde:

$$v_t = X_t' u_t + \varepsilon_t$$

$$w_t = X_t' \otimes Z_t'$$

$$E(v_t) = E(X_t' u_t + \varepsilon_t) = 0 \quad (13)$$

$$\text{Var}(v_t) = \text{Var}(X_t' u_t + \varepsilon_t) = \sigma_\varepsilon^2 + \sigma_u^2 X_t' Q X_t \quad (14)$$

Novamente, arranjando as observações em (12) para se obter uma forma matricial mais adequada, tem-se:

$$y = W\Gamma + v \quad (15)$$

onde  $y$  e  $v$  são vetores  $T \times 1$ ,  $W$  é uma matriz  $t \times k\tau$  e  $\Gamma$  é um vetor  $k\tau \times 1$  ( $\Gamma = \text{vec}(\delta)$ ).

O método de estimação vai depender das hipóteses que forem feitas a respeito de  $Z_t$  e  $u_t$ . No caso mais simples em que  $Z_t$  é conhecida, poder-se-ia aplicar MQO ou MQG dependendo da variância de  $u_t$ . Supondo que o modelo seja sistemático, isto é,  $\text{Var}(u_t) = 0$ , então  $v_t$  já não será heterocedástico, e podemos usar MQO em (15). Por outro lado, se tivermos uma componente aleatória em  $\beta_p$ , deveremos aplicar MQG. Uma solução neste caso, quando  $Q$  for conhecida, é seguir os mesmos passos propostos para o modelo aleatório puro, isto é, usar os resíduos estimados por MQO para se obter uma estimativa da variância do erro que será usada para corrigir a heterocedasticidade.

No modelo efetivamente proposto em Belsley (1973a),  $Z_t$  não é conhecida, mas a  $\text{Var}(u_t)$  é igual a zero. Por conseguinte, o problema central não é como estimar (15), mas sim como identificar o vetor  $Z_t$ .<sup>5</sup>

Belsley (1973a) propôs um método bastante complicado de se obter  $Z_t$ , cuja escolha poderia ser feita tentando diversos  $Z$ 's na regressão:

$$b_t = \sigma Z_t + h_t \quad (16)$$

onde  $b_t$  é um estimador de  $\beta$ , e  $h_t$  é um erro de ruído branco, tal que:

$$b_t = \beta_t + h_t \quad (17)$$

Isso corresponde a (11) no caso em que a  $\text{Var}(u_t)$  é zero. O problema é como obter  $b_t$ . Belsley (1973a) fornece um algoritmo para gerar uma série de  $b_t$ , através de regressões de  $y$  em  $X$  para diversos períodos, começando com o período  $t = 1$  a  $t = k + \tau$ , para se obter  $b_1$  e continuar mudando o período até  $t = T - k - \tau$  a  $t = T$ , para se obter  $b_T$ . Assim, essa técnica de "janela-móvel" (*moving-window*) pode fornecer uma série para  $b_t$ .<sup>6</sup>

5 Um modo mais simples e prático seria fazer uma regressão de MQO supondo parâmetros fixos e, então, representar graficamente os resultados contra diversas variáveis possíveis. Embora uma possível não-linearidade entre  $Z_t$  e os resíduos possa complicar as conclusões, uma simples inspeção do gráfico poderá oferecer uma boa compreensão.

6 Belsley (1973b) também fornece um teste- $F$  para a constância dos parâmetros quando  $Z$  não for conhecido.



Este procedimento não resolve completamente o problema, porque, conforme mostrado por Belsley (1973a),  $b_i$  é um estimador tendencioso de  $\beta_i$ . Este viés dependerá basicamente da taxa de variação de  $Z_i$ : caso  $Z_i$  se deslocasse lentamente no tempo, o viés seria menor.

Finalmente, dever-se-ia notar que Belsley efetivamente não oferece uma maneira nova para estimar parâmetros variáveis, mas um modelo para se escolher o conjunto de variáveis explicativas  $Z_i$ . Note-se ainda que o método proposto não é, de modo algum, teoricamente superior a uma simples tentativa de diversos  $Z_i$ 's diretamente em (14).<sup>7</sup> Uma vez que se conheça  $Z_i$ , a estimação de (14) é trivial.

## 2.4 - Regressões com troca de regime (*switching regressions*)

O modelo de regressão com troca de regime foi inicialmente desenvolvido nos anos 70 por Goldfeld e Quandt, e várias abordagens foram propostas para estimar e testar as hipóteses de regressão com troca de regime. Afirma-se basicamente que as observações amostrais podem ser divididas em um número, geralmente pequeno, de regimes distintos. Estes são diferenciados pelo fato de que — embora a forma da equação seja a mesma ao longo de todo o período — os valores dos parâmetros podem mudar entre regimes.

Assim, supondo que existam apenas dois regimes, o modelo de regressão com troca de regime pode ser expresso do seguinte modo:

$$y_t = X_t' \beta_1 + u_{1t} \text{ se } t \text{ pertence ao regime 1} \quad (18)$$

$$y_t = X_t' \beta_2 + u_{2t} \text{ se } t \text{ pertence ao regime 2} \quad (19)$$

Em geral, podemos ter dois conjuntos diferentes de parâmetros  $(\beta_1, \sigma_1)$  e  $(\beta_2, \sigma_2)$ .

Observe-se que o ponto central é a impossibilidade de se usar algum tipo de conhecimento *a priori* para dividir a amostra. Se as observações pudessem ser classificadas como se pertencessem a um regime específico, o problema de estimação e execução de testes de estabilidade seria trivial.

A primeira e mais simples abordagem deste modelo deve-se a Quandt (1958), o qual supõe que, numa amostra de tamanho  $n = n_1 + n_2$ , existem apenas dois regimes, de modo que as  $n_1$  primeiras observações estão associadas ao primeiro

<sup>7</sup> Poder-se-ia argumentar a favor do procedimento de Belsley com base nos custos computacionais, visto que seu algoritmo envolve apenas inversão de matrizes  $2 \times 2$ .

regime e as últimas  $n_2$  ao segundo. Conseqüentemente, este problema consiste em encontrar o ponto de reversão (*break point*) ótimo ( $t'$ ) e então estimar os coeficientes nos dois períodos. A equação (20) define a função de verossimilhança das observações  $y_t$ , dado  $t'$ :

$$L(y_t; t') = (1/2\pi)^{-n/2} \sigma_1^{t'} \sigma_2^{-(n-t')} \exp \left[ 1/2 \sigma_1^2 \sum_{t=1}^{t'} (y_t - X_t' \beta_1)^2 - 1/2 \sigma_2^2 \sum_{t=t'+1}^n (y_t - X_t' \beta_2)^2 \right] \quad (20)$$

Dois aspectos deste método deveriam ser realçados: primeiro, apenas um salto é permitido no período  $t'$ ; e, segundo, este ponto de reversão (*break point*)  $t'$  não pode estar próximo a nenhuma das extremidades da amostra, pois caso contrário a função de verossimilhança poderia ser indefinida.

Este modelo foi expandido por Goldfeld e Quandt (1972) e Quandt (1972) para possibilitar várias mudanças entre regimes. Goldfeld e Quandt (1972) supõem que em cada ponto no tempo a natureza escolhe entre dois regimes de acordo com uma variável  $Z_t$ :

$$\text{se } \sum_{i=1}^p \delta_i Z_{it} \leq 0 \quad \text{então a natureza escolhe o regime 1}$$

$$\text{se } \sum_{i=1}^p \delta_i Z_{it} > 0 \quad \text{então a natureza escolhe o regime 2}$$

onde os  $\delta$ 's são coeficientes desconhecidos.<sup>8</sup>

Definamos agora uma matriz diagonal  $D$  com elementos  $d(Z_t)$  de modo que:

$$d(Z_t) = 0 \quad \text{se } \sum_{i=1}^p \delta_i Z_{it} \leq 0$$

8 É possível que  $Z_t$  inclua alguns ou até mesmo todos os regressores.

$$d(Z_i) = 1 \quad \text{se} \quad \sum_{i=1}^p \delta_i Z_{it} > 0$$

Arranjando as observações em (18) e (19) e usando a matriz  $D$  para combiná-las, temos:

$$y = (I - D)X\beta_1 + DX\beta_2 + w \quad (21)$$

onde  $W = (I - D)u_1 + Du_2$ .

A função logarítmica de verossimilhança pode, então, ser escrita do seguinte modo (excluindo a constante):

$$l = - (1/2) |\Omega| - (1/2) [(y - (I - D)X\beta_1 + DX\beta_2)' \Omega^{-1} (y - (I - D)X\beta_1 + DX\beta_2)] \quad (22)$$

O problema com a maximização de (22) é que isso implica procurar várias combinações diferentes da matriz diagonal  $D$ . Para resolver este problema, Goldfeld e Quandt sugerem aproximar a função discreta (*unit step function*)  $d(Z_i)$  por uma função contínua. Eles escolhem a função densidade cumulativa normal:

$$d(Z_i) = \int_{-\infty}^{\sum_{i=1}^p \delta_i Z_{it}} \left[ \frac{1}{\sigma \sqrt{2\pi}} \right] \exp \left[ - \Phi^2 / 2 \sigma^2 \right] d\Phi \quad (23)$$

onde  $\sigma$  mede a excelência do ajustamento da discriminação entre dois regimes. Para  $\sigma$  pequenos, a expressão (23) deveria fornecer uma boa aproximação, mas quando  $\sigma$  não é pequeno, isto é, quando os dois regimes não podem ser separados facilmente, alguns valores de  $d(Z_i)$  não estarão próximos de zero ou 1.

Desse modo, pode-se substituir (23) em (22) e, então, maximizá-la para se obterem os  $\beta$ 's,  $\delta$ 's e  $\sigma$ .<sup>9</sup>

<sup>9</sup> Uma abordagem mais intuitiva seria usar os  $\delta$ 's estimados ou os  $d(Z_i)$ 's que deles decorrem, para separar a amostra e, então, usar MQO; porém, o uso do procedimento de MQO para tirar inferência dependeria dessa separação amostral.

Outra possibilidade é supor que a escolha de regime pela natureza é aleatória. Quandt (1972) desenvolve este modelo com uma escolha estocástica do regime. Em cada ponto do tempo existe uma probabilidade  $\alpha$  de a observação ter vindo do regime 1 e uma probabilidade  $1 - \alpha$  de ter vindo do regime 2.

Dado o pressuposto de normalidade, a densidade condicional de  $y_t$  pode ser escrita como uma média ponderada das funções densidade de probabilidade:

$$\begin{aligned} g(y_t/X_t) &= \alpha f_1(y_t/X_t) + (1 - \alpha) f_2(y_t/X_t) = \\ &= (\alpha / \sigma_1 \sqrt{2\pi}) \exp[-(1/2 \sigma_1^2)(y_t - X_t' \beta_1)^2] + \\ &+ ((1 - \alpha) / \sigma_2 \sqrt{2\pi}) \exp[-(1/2 \sigma_2^2)(y_t - X_t' \beta_2)^2] \end{aligned} \quad (24)$$

das quais se segue a função de verossimilhança:

$$L = [g_1 \times g_2 \times g_3 \times \dots \times g_n] \quad (25)$$

Note-se que novamente nos deparamos com um problema potencial idêntico ao encontrado em (20), ou seja, (25) poderá ser indefinida. Um modo de superar este problema é impor *a priori* que  $\sigma_1 = k \sigma_2$ , onde  $k$ , é qualquer constante positiva, para fazer com que (25) seja definida. Mesmo que esta suposição *a priori* não fosse feita, Kiefer (1978) [ver também Goldfeld e Quandt (1973b)] mostrou que existiria uma solução consistente.<sup>10</sup>

A generalização para o caso de mais de dois regimes é absolutamente direta. Se existirem  $f$  regimes diferentes, será necessário simplesmente redefinir a função logarítmica da verossimilhança do seguinte modo:

$$L = \sum_{i=1}^f \alpha_i f_i$$

onde  $\sum_{i=1}^f \alpha_i = 1$ .

---

10 O problema é mais complicado porque, conforme mostrado em Kiefer (1978), o próprio valor inicial deveria ser consistente. Uma alternativa seria usar uma função de geração de momentos, ao invés da função de verossimilhança. Este procedimento é sugerido por Quandt e Ramsey (1978), mas não há evidência de que seria melhor do que a estimação de verossimilhança.

Então, o sistema poderá mudar várias vezes não apenas entre dois regimes específicos, mas também para regimes mais novos.

Uma generalização final poderá levar em conta a possibilidade de se fazer uma mudança de um regime para outro, dependendo daquele que estiver ativo. Goldfeld e Quandt (1973a) apresentam esta generalização, colocando o modelo acima descrito dentro do arcabouço da cadeia de Markov.

Defina-se uma matriz de transição  $\Gamma$  onde o  $rs$ -ésimo elemento ( $r, s = 1, 2$ ) é a probabilidade de se deslocar para o regime  $s$  dado que o regime  $r$  estava ativo. Dadas as probabilidades no tempo zero, pode-se usar a matriz  $\Gamma$  para se obterem as probabilidades em todos os períodos subsequentes:

$$\Gamma = \begin{bmatrix} \tau_1 & 1 - \tau_1 \\ 1 - \tau_2 & \tau_2 \end{bmatrix}$$

Agora, o vetor de probabilidades poderá ser escrito como  $\alpha = (\alpha_{1t}, \alpha_{2t})$ , onde  $\alpha_{1t} = 1 - \alpha_{2t}$ . Então, a função densidade combinada pode ser reescrita assim:

$$g = \alpha_{1t} f_1(y_t/X_t) + (1 - \alpha_{1t}) f_2(y_t/X_t)$$

da qual podemos escrever a função de verossimilhança e maximizá-la para se obter  $\beta$ 's,  $\sigma$ 's,  $\tau$ 's e  $\alpha_{10}$ .<sup>11</sup>

Finalmente, dever-se-ia notar que o uso da função de verossimilhança em todos esses diferentes modelos fornece um teste natural para a constância dos parâmetros. O teste da razão da verossimilhança pode ser usado dentro da hipótese nula em que se impõe a condição de constância dos parâmetros.

## 2.5 - Algumas relações entre os modelos

Podemos, de certo modo, dividir os diversos modelos aqui apresentados em duas categorias: os dois primeiros — o aleatório puro (MAP) e o de regressão adaptativa (MRA) — podem ser classificados como modelos aleatórios, enquanto os outros dois — o de variação paramétrica sistemática (VPS) e o de regressão com troca do regime (RTR) — podem ser descritos como modelos sistemáticos. O ponto básico

<sup>11</sup> Seria simplesmente necessário estimar  $\alpha_{10}$  e  $\tau$  porque  $\alpha_t$  pode ser encontrado recursivamente como  $\alpha_t = \alpha_{t-1} \Gamma$ . Então,  $\alpha_t = \alpha_0 \Gamma^t$ .

é que nos MAP e MRA o vetor dos parâmetros é por si mesmo uma variável aleatória, enquanto nos VPS e RTR os parâmetros são funções determinísticas de alguma variável econômica específica.<sup>12</sup>

Vamos primeiro comparar os modelos aleatórios. Aqui a superioridade do MRA sobre o MAP é clara, não apenas por permitir que a trajetória do coeficiente seja estimada, mas ainda porque o MAP pressupõe uma média constante para  $\beta_t$ , enquanto no MRA a média condicional pode variar.

As ligações entre VPS e RTR são também fáceis de serem estabelecidas. No modelo VPS com apenas dois regimes ( $R = 2$ ), podemos definir  $Z_t$ :

$$\begin{aligned} Z_t &= (1 \ 0) \quad 1 \leq t < t' \\ &= (0 \ 1) \quad t' \leq t \leq T \end{aligned}$$

para se obter o RTR mais simples, e como:

$$\begin{aligned} Z_t &= (1 \ 0) \quad \text{com probabilidade } \alpha \\ &= (0 \ 1) \quad \text{com probabilidade } 1 - \alpha \end{aligned}$$

para se obter o método- $\alpha$  no RTR. Em ambos os casos supõe-se que a  $\text{Var}(u_t)$  em (13) é zero.

A conclusão de tudo isso é que o VPS não é um método para estimar parâmetros variáveis, mas sim um modo de se escolher o vetor  $Z_t$ , e, neste sentido, poderia ser visto como uma etapa preliminar ao RTR.

Outra comparação interessante pode ser feita entre o MRA e o VPS. Se for suposto que  $u_t$  é distribuído com incrementos independentes, (11) poderá ser reescrita como:

$$\beta_t = \beta_{t-1} + \delta \Delta Z_t + v_t \quad (26)$$

onde  $v_t = \Delta u_t$ .

Note-se que, para  $\Delta Z_t$  pequenos, (26) pode ser aproximada por (11). Porém,  $\Delta Z_t$  pequenos, isto é,  $Z_t$  movendo-se lentamente no tempo, é exatamente a exigência para que VPS funcione bem.

---

<sup>12</sup> Note-se que, mesmo quando a natureza escolhe estocasticamente entre regimes, isso pode ser verdadeiro. Substituindo  $\alpha$  por  $d(Z_t)$  em (24), a escolha de regime dependeria de  $Z_t$ .

### 3 - O filtro de Kalman

O filtro de Kalman foi desenvolvido originalmente na literatura de engenharia para se lidar com sinais aleatórios [ver Kalman (1960)]. Suas aplicações à econometria foram dificultadas provavelmente porque aborda o problema de regressão padrão de um modo bem diferente. Pelo modo que é usualmente derivado, o estimador  $b_t$  é o resultado da regressão de  $\beta_t$  na variável dependente  $y_t$ , ao invés de se fazer a regressão de  $y_t$  nas variáveis exógenas  $X_t$ . Isto é, o estimador  $b_t$  é a média condicional de  $\beta_t$  dado  $y_t, y_{t-1}, \dots, y_1$ .

Ao comparar o filtro de Kalman com a análise de regressão padrão, Duncan e Horn (1972) observaram o seguinte: “especificamente, o objetivo central em ambos os casos é estimar  $\beta$ , o vetor completo dos coeficientes incluídos na regressão”. Falando em termos de projeções de espaço vetorial, como faz Kalman, o regressionista convencional enfoca este problema projetando um vetor  $y$  de observações num vetor linear modulado pelas colunas de uma matriz  $X$ , isto é, “ajustando uma regressão de  $Y^*$  (a estimativa de  $y$ ) em  $X$ ”. As estimativas ótimas  $b$  de  $\beta$  são, então, os coeficientes da projeção ortogonal  $y^*$  assim obtido, isto é, o coeficiente de  $X$  na equação para  $y^*$ . A abordagem de Kalman, por outro lado, faz uso do fato de  $\beta_t$  ser agora uma variável aleatória. Através de uma redefinição adequada dos termos do espaço vetorial envolvidos, a própria estimativa ótima é encontrada diretamente como uma projeção ortogonal  $b_{t|s}$  de  $\beta_t$  no espaço vetorial modulado pelos  $\mu_1, y_1, \dots, y_t$ , dos dados incluídos. Em termos de regressão, isto significa “ajustar uma regressão de  $\beta_t$  em  $\mu_1, y_1, \dots, y_t$ ”.<sup>13</sup>

A derivação das equações do filtro de Kalman, apresentada nesta seção, será efetuada num contexto de regressão, visto que nosso objetivo principal é realçar seu uso na estimação de regressões de parâmetros variáveis. Deveria ficar claro, contudo, que o filtro de Kalman pode ter muitos outros usos. Diversos tipos de modelos podem ser arrançados dentro do arcabouço de espaço de estados, isto é, como nas equações (27) e (28) a seguir e, por conseguinte, estimados pelo filtro de Kalman. Exemplos disso são os modelos ARIMA e os modelos de componentes não-observáveis ou, como são às vezes denominados, modelos estruturais de séries temporais [ver Harvey (1989)]. Vamos explorar a representação de espaço de estados de Arima de modelos de componentes não-observáveis mais adiante nesta seção.

A idéia básica por trás do filtro de Kalman é de que a estimação é feita recursivamente em duas etapas: na primeira, encontra-se a “melhor” estimativa no período  $t$ , usando todas as informações disponíveis até o período  $t - 1$ ; e, na segunda, esta estimativa é então atualizada usando a informação nova que se tornou dispo-

---

<sup>13</sup> Ver Duncan e Horn (1972, p. 816). Note-se que  $\mu_1$  representa a média de  $\beta_t$  a priori e  $y^*$  o  $y$  estimado.

nível no período  $t$ . Esta estimativa é a melhor no sentido de que o erro quadrático médio é minimizado (EQMM).<sup>14</sup>

O modelo do filtro de Kalman é composto por equações de mensuração e de transição, que podem ser escritas como:

$$Y_t = X_t' \beta_t + \varepsilon_t \quad (27)$$

$$\beta_t = M_t \beta_{t-1} + S_t c_t + R_t u_t \quad (28)$$

onde:

$\beta$  = vetor de estados;

$Y_t$  = vetor das mensurações correntes;

$c_t$  = vetor de entrada (*input*); e

$X_t, M_t, S_t, R_t$  = matrizes fixas conhecidas com dimensões apropriadas.

Admite-se que  $\varepsilon_t$  e  $u_t$  têm média zero e matriz de covariância  $\sigma^2 H_t$  e  $\sigma^2 Q_t$ , respectivamente. Supõe-se ainda que  $\varepsilon_t$  e  $u_t$  são não-autocorrelacionados, não-correlacionados entre si e não-correlacionados com o vetor de estados no período zero, e que  $H_t$  e  $Q_t$  são conhecidas. Em geral, ambos os erros são, por suposição, distribuídos normalmente, embora a normalidade não seja crucial.<sup>15</sup>

### 3.1 - Equações de previsão

Em um ponto qualquer no tempo, conhecemos  $b_{t-1}$  e sua matriz de covariância:

$$\sigma^2 \Sigma_{t/t-1} = E[(b_{t/t-1} - \beta_t)(b_{t/t-1} - \beta_t)'] \quad (29)$$

Para se obter o melhor estimador de  $\beta_t$ , deveríamos usar a equação de transição (28), que mostra como o vetor de estados varia ao longo do tempo:

$$b_{t/t-1} = M_t b_{t-1} + S_t c_t \quad (30)$$

<sup>14</sup> É comum referir-se a esta estimativa como sendo a do EQMM, ao invés de estimativa não-tendenciosa de variância mínima, devido à natureza estocástica do parâmetro.

<sup>15</sup> Ocasionalmente, prefere-se trabalhar com uma distribuição mais geral. Neste caso, o estimador do filtro de Kalman seria o EQMM entre os estimadores lineares. Athans (1974), por exemplo, usa a normalidade, enquanto Harvey e Phillips (1982) trabalham com uma distribuição mais geral.



Para encontrar a equação de previsão da matriz de covariância no período  $t$ , dado seu valor no período  $t - 1$ , vamos definir o erro de previsão em (29) da seguinte forma:

$$b_{t/t-1} - \beta_t = M_t b_{t-1} + S_t c_t - \beta_t$$

mas, usando (28):

$$b_{t/t-1} - \beta_t = M_t (b_{t-1} - \beta_{t-1}) + R_t u_t \quad (31)$$

Substituindo (31) em (29) e usando o Lema de Projeção Ortogonal, tem-se [ver Jazwinski (1970, p. 202-203)]:

$$\Sigma_{t/t-1} = M_t \Sigma_{t-1} M_t' + R_t Q_t R_t' \quad (32)$$

Assim, (30) e (32) podem, então, ser utilizados para prever o vetor de estados  $\beta$  e a matriz de covariância  $\Sigma$  no período  $t$ , dadas suas estimativas no período  $t - 1$ .

### 3.2 - Equações de atualização

Vamos primeiramente definir o resíduo  $v_t$ :

$$\begin{aligned} v_t &= Y_t - \hat{Y}_{t/t-1} = X_t' \beta_t + \varepsilon_t - X_t' b_{t/t-1} \\ v_t &= X_t' (\beta_t - b_{t/t-1}) + \varepsilon_t \end{aligned} \quad (33)$$

Assim:

$$E(v_t) = X_t' E(\beta_t - b_{t/t-1}) + E(\varepsilon_t) = 0$$

Uma vez que  $b_t$  é não-tendencioso e:

$$\text{Var}(v_t) = E(v_t^2) = \sigma^2 F_t \quad (34)$$

onde:

$$F_t = X_t' \Sigma_{t/t-1} X_t + H_t \quad (35)$$

a derivação das equações de atualização para  $b$  e  $\Sigma$  pode ser feita através da aplicação de MQG a um modelo de parâmetros variáveis.<sup>16</sup> O procedimento para tornar a prova simples é redefinir o modelo do seguinte modo:

$$\begin{bmatrix} b_{t/t-1} \\ Y_t \end{bmatrix} = \begin{bmatrix} I \\ X_t \end{bmatrix} \beta_t + \begin{bmatrix} b_{t/t-1} - \beta_t \\ \varepsilon_t \end{bmatrix} \quad (36)$$

ou, em notação mais compacta:

$$Y_t^* = X_t^* \beta_t + \varepsilon_t^* \quad (36')$$

onde a matriz de covariância de  $\varepsilon_t^*$  é:

$$E(\varepsilon_t^* \varepsilon_t^{*'}) = \sigma^2 \Omega = \sigma^2 \begin{bmatrix} \Sigma_{t/t-1} & 0 \\ 0 & H_t \end{bmatrix} \quad (37)$$

Aplicando MQG em (36'), teremos:

$$b_{t/t} = (X_t^{*'} \Omega^{-1} X_t^*)^{-1} X_t^{*'} \Omega^{-1} Y_t^* \quad (38)$$

Para se obter uma fórmula mais adequada, podemos substituir (36) e (37) em (38):

---

<sup>16</sup> A prova para as equações de atualização usando o princípio de MQG pode ser encontrada em Duncan e Horn (1972), Sant (1977), Harvey (1981) e Diderrich (1985).

$$b_{t/t} = (\Sigma_{t/t-1}^{-1} + X_t H_t^{-1} X_t')^{-1} (\Sigma_{t/t-1}^{-1} b_{t/t-1} + X_t H_t^{-1} Y_t) \quad (39)$$

Usando um lema bem conhecido de inversão de matriz, a equação (39) pode ser reescrita assim:

$$b_{t/t} = b_{t/t-1} + K_t (Y_t - X_t' b_{t/t-1}) \quad (40)$$

onde o ganho  $K_t$  de Kalman é:

$$K_t = \Sigma_{t/t-1} X_t F_t^{-1}$$

Observe-se que o ganho do filtro é usado para melhorar as previsões do vetor de estados. Toda vez que este for tal que o  $Y_t$  corrente e o previsto sejam diferentes, o erro será incorporado na nova estimativa do vetor de estados para torná-la mais acurada. Em outras palavras, (40) diz que o valor esperado do vetor de estados, dadas todas as informações até o período  $t$ , pode ser dividido em dois componentes: o valor esperado deste vetor dadas as informações até  $t-1$ , mais um coeficiente multiplicado pelo erro de previsão de  $Y_t$  dadas as informações disponíveis até  $t-1$ . O ganho do filtro é exatamente este coeficiente que desconta o erro de previsão. Seguindo Chow (1983), podemos escrever (40) como:

$$E(\beta_t / Y_t, I_{t-1}) = E(\beta_t / I_{t-1}) + K_t [Y_t - E(Y_t / I_{t-1})]$$

onde  $Y_t$  representa a observação corrente no período  $t$  e  $I_{t-1}$  representa  $(Y_1, Y_2, \dots, Y_{t-1})$ .

Note-se que, se o erro de previsão em (40) for zero, então  $b_t = b_{t-1}$  e estaremos de volta ao caso de coeficientes fixos. Neste sentido, o modelo de coeficientes fixos pode ser visto como um caso do modelo variável, quando não usamos o erro de previsão no período  $t-1$  para melhorar a previsão para o período  $t$ .

Para se obter a equação de atualização para a matriz de covariância, podemos multiplicar (40) por  $-1$ , adicionar  $\beta_t$  e usar (27) para obter:

$$\beta_t - b_{t/t} = \beta_t - b_{t/t-1} - K_t [X_t' (\beta_t - b_{t/t-1}) + \varepsilon_t] \quad (41)$$

Tomando o valor esperado do produto cruzado de (41) e usando as definições de  $K_t$  e  $F_t$ , temos:

$$\Sigma_{t/t} = \Sigma_{t/t-1} - \Sigma_{t/t-1} X_t F_t^{-1} X_t' \Sigma_{t/t-1} \quad (42)$$

Como dissemos anteriormente, o modelo de parâmetros variáveis é simplesmente uma das aplicações possíveis do filtro de Kalman. Neste caso,  $y_t$  é um vetor com observações na variável endógena,  $\beta_t$  é um vetor de parâmetros a ser estimado,  $X_t$  é uma matriz fixa de variável exógena  $c_t = 0$  e  $H_t = M_t = R_t = I$ .

Para concluir a derivação, a única coisa que precisamos é um estimador para  $\sigma^2$ , o que pode ser obtido usando um enfoque de máxima verossimilhança.<sup>17</sup> Uma vez que  $v_t$  em (33) é normal e não-correlacionado serialmente, podemos escrever a função logarítmica da verossimilhança como:

$$\ln L = -(T/2) \ln 2\pi - 1/2 \sum_{t=1}^T \ln \sigma^2 (X_t' \Sigma_{t|t-1} X_t + H_t) - \\ - 1/2 \sigma^2 \sum_{t=1}^T [(Y_t - X_t' b_{t|t-1})^2 / (X_t' \Sigma_{t|t-1} X_t + H_t)]$$

Efetuada algumas manipulações e usando (33) e (35), obteremos:

$$\ln L = -(T/2) \ln 2\pi - (T/2) \ln \sigma^2 - 1/2 \sum_{t=1}^T \ln F_t - 1/2 \sigma^2 \sum_{t=1}^T v_t^2 / F_t \quad (43)$$

A equação (43) pode então ser maximizada com respeito a  $\sigma^2$ , fornecendo:

$$s^2 = 1/T \sum_{t=1}^T v_t^2 / F_t \quad (44)$$

que é o estimador de máxima verossimilhança de  $\sigma^2$  condicional a  $M_p, H_p, S_p, X_p, R_t$  e  $Q_r$ .

Note-se ainda que, se  $Q_r$  não fosse conhecida, poderia ser estimada maximizando a função logarítmica da verossimilhança concentrada. Substituindo (44) em (43), obteremos:

$$\ln L_c = -(T/2) (\ln 2\pi + 1) - (T/2) \ln s^2 - 1/2 \sum_{t=1}^T \ln F_t$$

<sup>17</sup> A variância das variáveis dos estados é, às vezes, denominada hiperparâmetro [ver Harvey (1987)].

Observe-se que poderíamos usar o procedimento da maximização acima descrito para obtermos outros elementos desconhecidos no modelo. Ocorre, porém, que, por vezes,  $M_t$  não é conhecida, como, por exemplo, quando estamos lidando com modelos ARIMA.

O problema, contudo, é ainda o mesmo, ou seja, encontrar  $\sigma^2$  e  $M_t$  que maximizem a função de verossimilhança, dados  $H_t$ ,  $X_t$ ,  $R_t$  e  $Q_t$ . Uma vez que o modelo está na forma de espaço de estados e os valores iniciais determinados, o filtro de Kalman é usado para gerar o erro de previsão  $v_t$  que é usado como entrada (*input*) para a função logarítmica da verossimilhança, a qual é, então, maximizada através da otimização numérica.

### 3.3 - Estimação recursiva

Em resumo, o filtro de Kalman pode ser sumariado pelas equações de previsão (30) e (32) e pelas equações de atualização (40) e (42). A estimação é feita recursivamente. Dado  $b_0$  e  $\Sigma_0$ , podemos usar (30) e (32) para obtermos uma estimativa inicial e, então, atualizá-la usando (40) e (42).

Na prática, podemos usar como valores iniciais uma matriz de covariância  $\Sigma_0 = kI$ , onde  $k$  é um número grande. Pode-se supor que o valor para  $b_0$  é zero. Note-se que  $b_0$  não tem muita importância porque, como  $k$  é grande,  $\Sigma_0$  dominará  $b_0$ . As estimativas iniciais provavelmente não serão muito precisas, mas, como o filtro de Kalman é invariável no tempo, pode-se suavizar os coeficientes após filtrá-los.

É possível também, conforme sugerido por Sant (1977), usar a equivalência entre MQG e o filtro de Kalman para se obterem as condições iniciais. Se existirem graus de liberdade suficientes, pode-se aplicar MQG às primeiras  $m$  observações para se obterem  $b_m$  e  $\Sigma_m$  e, então, usá-los como valores iniciais para filtrar as demais  $T - m$  observações.

Se, por outro lado, soubéssemos que o modelo é estacionário, poderíamos usar a média e a matriz de covariância da distribuição *a priori* de  $b_0$ . Visto que o modelo (27) e (28) não tem termo constante, deveríamos colocar  $b_0=0$ , e  $\Sigma_0$  seria encontrado resolvendo a equação (32) do seguinte modo [ver Harvey (1981, p. 49 e 109-110)]:

$$\Sigma_0 = M_t \Sigma_0 M_t' + R_t Q_t R_t'$$

Assim, a estimativa corrente consiste em, dados  $b_0$  e  $\Sigma_0$ , encontrar  $b_{t/t-1}$  e  $\Sigma_{t/t-1}$  e, em seguida,  $v_t$  e  $F_t$ . Uma vez que a nova observação  $Y_t$  estiver disponível, pode-se, então, obter as estimativas atualizadas  $b_{t/t}$  e  $\Sigma_{t/t}$ . Neste procedimento, as equações-chave são (30), (32), (33), (34), (40) e (42).

### 3.4 - Modelo de coeficientes convergentes estocasticamente

O modelo de coeficientes convergentes estocasticamente, também denominado “modelo de retorno à normalidade”, foi usado primeiramente por Rosenberg (1973) e mais recentemente por Harvey e Phillips (1982). Neste modelo, o parâmetro em cada ponto no tempo é uma combinação linear daquilo que ele era no último período e um valor fixo:

$$\alpha_t = \phi \alpha_{t-1} + (1 - \phi)\mu + u_t \quad (45)$$

ou, rearranjando os termos:

$$\alpha_t - \mu = \phi (\alpha_{t-1} - \mu) + u_t \quad (46)$$

onde  $u_t \sim N(0, \sigma^2 Q_t)$  e  $\phi$  é uma matriz diagonal que satisfaz a condição de estacionaridade, isto é, seus autovalores são menores que 1 em valor absoluto.

Neste modelo, o coeficiente  $\alpha_t$  é variável, mas muda lentamente no tempo em torno de uma média fixa  $\mu$ . Note-se que na equação de transição usual, onde  $\alpha$  segue um processo de passeio aleatório, seu valor no final da amostra pode ser substancialmente diferente daquele no início. Portanto, em casos onde se espera uma mudança estrutural substancial, poder-se-ia preferir usar a aproximação de passeio aleatório, enquanto no caso de mudanças suaves no tempo o modelo de coeficientes convergentes estocasticamente parece ser mais apropriado.

Dever-se-ia também notar que o modelo aleatório puro e a especificação de  $\beta_t$  no contexto de passeio aleatório podem ser encarados como casos especiais do modelo de coeficientes convergentes estocasticamente. Se  $\phi = 0$ , estaremos de volta ao modelo aleatório puro, enquanto que, se  $\phi = 1$ , obteremos (28), quando  $M_t = I$  e  $c_t = 0$ .

Para formular este modelo dentro do arcabouço do filtro de Kalman, podemos definir:

$$\beta_t = (\mu_t', \delta_t')'$$

onde  $\mu_t = \mu$  e  $\delta_t = \alpha_t - \mu$ . Então:

$$Y_t = X_t^* \beta_t \quad (47)$$

$$\begin{bmatrix} \mu_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \delta_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} u_t \quad (48)$$

onde  $X^* = (X_t', X_t')$ .

### 3.5 - O enfoque bayesiano

Até agora permanecemos dentro do arcabouço da estatística clássica. Contudo, o modelo de parâmetros variáveis das equações (27) e (28) pode também ser abordado do ponto de vista bayesiano. Tentaremos mostrar nesta seção que, no tocante a equações de previsão e de atualização, os enfoques clássico e bayesiano são equivalentes. A diferença está no tratamento das variâncias.

#### 3.5.1 - Interpretação bayesiana das equações de previsão e de atualização

As equações de atualização do filtro de Kalman foram derivadas acima usando o princípio de MQG. A literatura mostra que elas podem também ser derivadas sob o ponto de vista bayesiano [ver Meinhold e Singpurwalla (1983)].

O enfoque bayesiano da estatística preocupa-se principalmente com o modo pelo qual a informação muda a crença que se tem acerca das hipóteses e dos valores dos parâmetros. Nas próprias palavras de Lindley (1965, p. 30-31): "O tema principal da estatística é o estudo do modo pelo qual os dados (eventos) mudam os graus de convicção; de antes, pela observação de  $A$ , para depois." É esta a idéia que está por trás do Teorema de Bayes sobre probabilidade condicional:

$$P(H_n/A) \propto P(A/H_n)P(H_n)$$

onde  $n = 1, 2, \dots$ , e  $A$  é fixo.

Colocando de modo simples, este teorema estabelece que uma crença ou probabilidade *a priori* de um conjunto de hipóteses  $P(H_n)$  pode ser transformada numa crença ou probabilidade *a posteriori*  $P(H_n/A)$ , dada a verossimilhança de  $H_n$  em  $A$ ,  $P(A/H_n)$ . Este enfoque seqüencial para a estimação também se encontra dentro dos princípios básicos do filtro de Kalman.

Vamos tomar as equações (27) e (28) e reescrever o Teorema de Bayes do seguinte modo:

$$P(\beta_t/I_t) = P(Y_t/\beta_t, I_{t-1})P(\beta_t/I_{t-1}) \quad (49)$$

onde, como anteriormente,  $I_{t-1}$  representa  $(Y_1, Y_2, \dots, Y_{t-1})$ . No início, a distribuição *a priori*  $\beta_t/I_{t-1}$  é conhecida:

$$(\beta_t/I_{t-1}) \sim N(M_t b_{t-1}, M_t \Sigma_{t-1} M_t' + R_t Q_t R_t') \quad (50)$$

onde  $b_{t-1}$  e  $\Sigma_{t-1}$  são a média e a matriz de covariância de  $(\beta_{t-1}/I_{t-1})$ . Para se obter (50), simplesmente usamos (27), (28) e alguns resultados básicos da distribuição normal [ver Meinhold e Singpurwalla (1983)].

Observe-se ainda que, usando o resíduo  $v_t$ , definido em (33), podemos reescrever a probabilidade *a posteriori* e a verossimilhança de  $Y_t$  como a seguir:

$$P(\beta_t/I_t) = P(\beta_t/Y_t, I_{t-1}) = P(\beta_t/v_t, I_{t-1})$$

$$P(Y_t/\beta_t, I_{t-1}) = P(v_t/\beta_t, I_{t-1})$$

Isto é verdadeiro porque, uma vez que  $M_t, S_t, c_t$  e  $X_t$  são conhecidos, observar  $v_t$  é o mesmo que observar  $Y_t$ . Então, (49) pode ser reescrita do seguinte modo:

$$P(\beta_t/v_t, I_{t-1}) = P(v_t/\beta_t, I_{t-1})P(\beta_t/I_{t-1}) \quad (51)$$

A distribuição da verossimilhança  $P(v_t/\beta_t, I_{t-1})$  é normal com média e variância conforme definidas anteriormente.

Para, finalmente, se obter a distribuição *a posteriori* de (51), podem-se usar alguns resultados padrões em estatística multivariada para se obter (52) [ver Chow (1983, p. 8-14)]:

$$(\beta_t/v_t, I_{t-1}) \sim N [b_{t/t-1} + K_t(Y_t - X_t' b_{t/t-1}) \quad (52)$$

$$\Sigma_{t/t-1} - \Sigma_{t/t-1} X_t F_t^{-1} X_t' \Sigma_{t/t-1}$$

Uma comparação simples entre (52), (40) e (42) torna claro que as equações de atualização podem ser vistas como a média e variância da distribuição *a posteriori*. O filtro de Kalman pode, então, ser visto como um processo de atualização através do qual as convicções a respeito do vetor de estados são previstas, dependendo da excelência da suposição anterior.



Finalmente, dever-se-ia notar que a normalidade não é crucial para provar a equivalência entre os enfoques MQG e bayesiano. Na derivação de (50) e (52), a distribuição normal foi usada por motivos de simplificação.<sup>18</sup>

### 3.5.2 - Processo de aprendizagem pela variância (*variance learning*) e fatores de desconto

A diferença básica entre os enfoques clássico e bayesiano dos modelos de parâmetros variáveis está na estimação das variâncias observável e não-observável (*evolution variance*). No enfoque clássico, os hiperparâmetros são estimados por um procedimento de máxima verossimilhança, conforme descrito nas equações (43) e (44). Por outro lado, o enfoque bayesiano usa o processo de aprendizagem pela variância (*variance learning*) e fatores de desconto ao lidar com as variâncias.

Vamos considerar o modelo linear dinâmico mais simples representado pelas equações (53) a (56). A variância observável  $V$  é estimada seqüencialmente à medida que novas observações tornam-se disponíveis:

$$Y_t = \theta_t + v_t \quad (53)$$

$$\theta_t = \theta_{t-1} + \omega_t \quad (54)$$

$$v_t \sim N(0, V) \quad (55)$$

$$\omega_t \sim N(0, W_t) \quad (56)$$

A estimativa pontual de  $V$  é  $S_{t-1}$ , definida pelas equações (57), (58) e (59). As equações de atualização agora incluem a equação de atualização para  $S_{t-1}$ . À medida que mais observações tornam-se disponíveis, corrigimos nossa estimativa anterior de  $V$ . Note-se que  $n_t$  representa os graus de liberdade em cada período, mas  $d_t$  não é exatamente a soma dos quadrados dos resíduos. Se  $S_{t-1}/Q_t = 1$ , segue-se que  $d_t$  é a soma dos quadrados dos resíduos:

$$S_t = d_t/n_t \quad (57)$$

$$n_t = n_{t-1} + 1 \quad (58)$$

<sup>18</sup> A prova geral para as equações de atualização via Teorema de Bayes pode ser encontrada em West e Harrison (1989).

$$d_t = d_{t-1} + e_t^2 S_{t-1} / Q_t \quad (59)$$

onde  $e_t$  são os erros de previsão um passo à frente ( $Y_t - \hat{Y}_t$ ) e  $Q_t$  é a variância da previsão um passo à frente.

Uma correção é feita sempre que o resíduo ao quadrado for diferente de seu valor esperado, visto que  $E(e_t^2) = Q_t$ . Note-se que, por ser  $d_t = n_t S_t$ , a equação (59) pode ser reescrita como:

$$n_t S_t = n_{t-1} S_{t-1} + S_{t-1} e_t^2 / Q_t$$

$$n_t S_t = (n_t - 1) S_{t-1} + S_{t-1} e_t^2 / Q_t$$

$$S_t = S_{t-1} + S_{t-1} / n_t [(e_t^2 / Q_t) - 1] \quad (60)$$

Toda vez que  $e_t^2 = Q_t$ , não haverá correção na variância estimada. Se o resíduo ao quadrado for maior do que seu valor esperado, a expressão entre parênteses na equação (60) será positiva e, por conseguinte, a variância estimada será corrigida para cima, e vice-versa. Para iniciar as estimações, é necessário que algum valor para  $S_0$  esteja disponível.

A variância não-observável (*evolution variance*)  $W_t$  não é estimada diretamente. Ameen e Harrison (1984, 1985a e 1985b) mostram como a estimação de  $W_t$  pode ser substituída pela escolha de um fator de desconto  $0 < \delta \leq 1$ , o qual pode ser encarado como um indicador da velocidade de mudança do sistema. Ele nos diz quão informativa é uma observação à medida que ela vai envelhecendo. Se  $\delta = 1$  ou, equivalentemente, se  $W = 0$ , temos o modelo de parâmetros fixos.

### 3.6 - Resíduos auto-regressivos

Pode acontecer que, após a estimação através do filtro de Kalman, se encontre evidência de autocorrelação residual. O modelo deveria, então, ser ampliado para levar em conta a presença de correlação serial. Considere-se o seguinte modelo ampliado:<sup>19</sup>

<sup>19</sup> Para uma aplicação desse tipo de modelo, ver Wolff (1985).

$$Y_t = X_t' \beta_t + \varepsilon_t \quad (61)$$

$$\beta_t = \beta_{t-1} + u_t \quad (62)$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t \quad (63)$$

onde  $\eta_t$  é um erro de ruído branco.

Trata-se de uma versão simplificada do modelo de espaço de estados das equações (27) e (28) que incluem resíduos autocorrelacionados de primeira ordem. À exceção da autocorrelação residual, todas as demais hipóteses relacionadas ao modelo de espaço de estados são mantidas. Adicionalmente, supomos que:

$$E(\eta_i, u_j) = 0 \quad \text{para } i, j = 1, \dots, T$$

$$E(\eta_i, \beta_j) = 0 \quad \text{para } i, j = 1, \dots, T$$

$$E(\beta_0, \eta_t') = 0 \quad \text{para } t = 1, \dots, T$$

Usando (61) e (63), podemos, portanto, escrever as quase diferenças de primeira ordem:

$$Y_t + \rho Y_{t-1} + X_t' \beta_t - X_{t-1}' \rho \beta_{t-1} + \eta_t \quad (64)$$

Vamos definir:

$$\delta_t = \begin{bmatrix} \rho \\ \beta_t \\ -\rho \beta_{t-1} \end{bmatrix}, \quad Z_t' = [Y_{t-1} X_t' X_{t-1}'] \quad \text{e} \quad \xi_t = \begin{bmatrix} 0 \\ u_t \\ -\rho u_{t-1} \end{bmatrix}$$

Segue-se que o modelo de espaço de estados pode ser escrito do seguinte modo:

$$Y_t = Z_t' \delta_t + \eta_t \quad (65)$$

$$\delta_t = \delta_{t-1} + \xi_t \quad (66)$$

As equações (65) e (66) podem agora ser estimadas usando o filtro de Kalman. No modelo acima, supõe-se que  $\rho$  seja fixo, podendo ser estimado através de um procedimento recursivo pelos resíduos do filtro à maneira de Cochrane-Orcutt.

### 3.7 - Mínimos quadrados recursivos

A estimação de mínimos quadrados recursivos pode ainda ser efetuada pelo filtro de Kalman, uma vez que o modelo relevante está representado no arcabouço de espaço de estados. Note-se que, neste caso, o parâmetro verdadeiro não varia com o tempo. Desse modo, dever-se-ia esperar ver o vetor de estados tendendo para uma constante à medida que mais dados forem sendo usados. O filtro de Kalman é, neste caso, simplesmente um algoritmo para repetir a estimação de MQO à medida que o tamanho da amostra aumenta.

Considere-se o modelo de espaço de estados simplificado:

$$Y_t = X_t' \beta_t + e_t \quad (67)$$

$$\beta_t = \beta_{t-1} \quad (68)$$

onde  $e_t \sim N(0, \sigma^2)$  e  $\beta_0 = \beta_{t-m} \sim N(\beta_{t-m}, \Sigma_{t-m})$ .

Conseqüentemente, após usar as primeiras  $m$  observações para se obterem as condições iniciais, o filtro pode, então, ser iniciado. As equações de atualização são dadas por:

$$b_{t/t-1} = b_{t-1} = (X'X)_{t-1}^{-1} (X'Y)_{t-1}$$

$$\Sigma_{t/t-1} = \Sigma_{t-1} = \sigma^2 (X'X)_{t-1}^{-1}$$

enquanto as equações de previsão são:

$$b_t = b_{t-1} + k_t [Y_t - X_t' b_{t/t-1}]$$

$$\Sigma_t = \Sigma_{t/t-1} - \Sigma_{t/t-1} X_t f_t^{-1} X_t' \Sigma_{t/t-1}$$

onde:

$$k_t = (X'X)_{t-1}^{-1} X'_t f_t^{-1}$$

$$f_t = \sigma^2 [1 + X'_t (X'X)_{t-1}^{-1} X_t]$$

### 3.8 - Modelos ARIMA na forma de espaço de estados

Como dissemos anteriormente, os modelos Arima podem também ser facilmente representados no arcabouço de espaço de estados e, desse modo, estimados pelo filtro de Kalman.<sup>20</sup> Considere-se o modelo ARMA (2,1):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

Sua representação em espaço de estados é mostrada pelas equações (69) e (70). Neste caso, a matriz  $M_t$  não é conhecida e, por conseguinte, terá que ser estimada usando o enfoque de máxima verossimilhança descrito anteriormente:

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \end{bmatrix} \quad (69)$$

$$X'_t \quad \beta_t$$

$$\begin{bmatrix} \beta_{1t} \\ \beta_{2t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} \beta_{1t-1} \\ \beta_{2t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix} \varepsilon_t \quad (70)$$

$$\beta_t \quad M_t \quad \beta_{t-1} \quad R_t$$

Vamos agora considerar o caso geral de um ARMA (p,q):<sup>21</sup>

$$y_t = \phi_1 y_{t-1} + \dots + \phi_m y_{t-m} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_{m-1} \varepsilon_{t-m+1}$$

<sup>20</sup> Para mais detalhes sobre o modelo ARIMA em forma de espaço dos estados, ver Priestley (1981) e Akaike (1976).

<sup>21</sup> A representação do ARIMA pode ser obtida substituindo  $y_t$  por  $z_t$ , onde  $z_t = \Delta y_t$ .

onde  $m = \max(p, q + 1)$ . A representação de espaço de estados para o modelo geral ARMA pode ser assim escrita:

$$Y_t = \left[ \begin{array}{c|c} 1 & 0'_{m-1} \end{array} \right] \begin{bmatrix} \beta_{1t} \\ \vdots \\ \beta_{mt} \end{bmatrix}$$

$$\begin{bmatrix} \beta_{1t} \\ \vdots \\ \beta_{mt} \end{bmatrix} = \begin{bmatrix} \phi_1 & & & \\ & I_{m-1} & & \\ & & & \\ \hline & & & 0' \end{bmatrix} \begin{bmatrix} \beta_{1t-1} \\ \vdots \\ \beta_{mt-1} \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-1} \end{bmatrix} \varepsilon_t$$

### 3.9 - Modelos de componentes não-observáveis

Neste arcabouço, uma série temporal é modelada através da decomposição de seus elementos formadores básicos, ou seja, ela é decomposta em tendência ( $\mu_t$ ), ciclo ( $\psi_t$ ), sazonalidade ( $\gamma_t$ ) e componente irregular ( $\varepsilon_t$ ):

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t$$

Note-se que os próprios componentes variam no tempo, ou seja, o modelo leva em consideração uma tendência, um ciclo e um padrão sazonal mutáveis.

O modelo completo deveria especificar também o comportamento de cada um dos componentes individualmente. Um tratamento completo de modelos de componentes não-observáveis está além do escopo deste artigo. Portanto, vamos nos concentrar no caso do "modelo estrutural básico de séries temporais".<sup>22</sup>

<sup>22</sup> O modelo de ciclo e tendência é discutido em Portugal (1992). Uma discussão completa de modelos de componentes não-observáveis é encontrada em Harvey (1989).

No modelo estrutural básico de séries temporais, apenas três componentes são usadas, uma vez que a componente do ciclo é omitida. A tendência é modelada como um passeio aleatório com um *drift* variável, enquanto o próprio *drift* é um passeio aleatório. A componente sazonal é modelada através de uma combinação de senóides e co-senóides. O modelo pode ser expresso em:

$$\begin{aligned}
 y_t &= \mu_t + \gamma_t + \varepsilon_t \\
 \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\
 \beta_t &= \beta_{t-1} + \zeta_t \\
 \gamma_t &= \sum_{j=1}^{s/2} \gamma'_{jt}
 \end{aligned} \tag{71}$$

onde  $\gamma'_{jt}$  é formado por:

$$\begin{bmatrix} \gamma_{jt} \\ \gamma^*_{jt} \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma^*_{j,t-1} \end{bmatrix} + \begin{bmatrix} \omega_{jt} \\ \omega^*_{jt} \end{bmatrix} \tag{72}$$

e  $\varepsilon_t$ ,  $\eta_t$  e  $\zeta_t$  são todos normais com variâncias  $\sigma_\varepsilon^2$ ,  $\sigma_\eta^2$  e  $\sigma_\zeta^2$ ,  $\omega_t$  e  $\omega^*$  são normais com uma variância comum  $\sigma_\omega^2$ ,  $\lambda_j$  é a frequência e  $\gamma^*_{jt}$  aparece por construção.

Para simplificar, vamos supor que estamos lidando com dados trimestrais, de modo que  $s = 4$ . Assim, a frequência  $\lambda_j = 2\pi_j/4$  para  $j = 1, 2$ , o qual resulta  $\lambda_j = \pi, \pi/2$ . Usando alguns resultados trigonométricos básicos, podemos reescrever (72) do seguinte modo:

$$\begin{aligned}
 \gamma_{1t} &= \gamma^*_{1,t-1} + \omega_{1t} \\
 \gamma^*_{1t} &= -\gamma_{1,t-1} + \omega^*_{1t} \\
 \gamma_{2t} &= \gamma^*_{2,t-1} + \omega_{2t} \\
 \gamma^*_{2t} &= -\gamma_{2,t-1} + \omega^*_{2t}
 \end{aligned} \tag{73}$$

A representação de espaço de estados das equações (71) e (72) é:

$$y_t = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \end{bmatrix} \alpha_t + \varepsilon_t$$

$$\alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_{1t} \\ \gamma_{1t}^* \\ \gamma_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{1t-1} \\ \gamma_{1t-1}^* \\ \gamma_{2t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \omega_{1t} \\ \omega_{1t}^* \\ \omega_{2t} \end{bmatrix}$$

### 3.10 - Suavização (*smoothing*)

Até agora lidamos com filtragem. Nesta subseção vamos discutir o problema oposto, isto é, alisamento ou suavização. O filtro de Kalman é um algoritmo “para a frente” (*forward looking algorithm*), que fornece estimativas do vetor de estados no período  $t$ , usando todas as informações disponíveis até o período  $t - 1$ . As estimativas suavizadas, por outro lado, usam todas as informações da amostra. Elas começam no período  $T$  e voltam no tempo, fornecendo as melhores estimações para os períodos  $T - 1$ ,  $T - 2$ , e assim por diante. O algoritmo de suavização é iniciado através de  $b_{T/T}$  e  $\Sigma_{T/T}$  dados pelo filtro de Kalman e funciona para trás até  $t = 1$ .

Deixemos  $b_{t/T}$  e  $\Sigma_{t/T}$  representarem os estimadores de suavização do vetor de estados e da matriz de covariância. Então, as equações de suavização podem ser escritas do seguinte modo:

$$b_{t/T} = b_t + \Sigma_t^* (b_{t+1/T} - M_{t+1} b_t)$$

e:

$$\Sigma_{t/T} = \Sigma_t + \Sigma_t^* (\Sigma_{t+1/T} - \Sigma_{t+1/t}) \Sigma_t^{*'}$$

onde:

$$\Sigma_t^* = \Sigma_t M_{t+1}' \Sigma_{t+1/t}^{-1} \quad t = T-1, \dots, 1$$



Este é o suavizador de intervalo fixo ótimo. Existem outros algoritmos de suavização que podem ser usados, mas o suavizador pontual fixo parece ser o mais popular em econometria.<sup>23</sup>

#### 4 - Conclusões e observações

Neste trabalho apresentamos diversos procedimentos para estimar modelos de parâmetros variáveis no tempo. O filtro de Kalman tem um papel crucial e parece oferecer um arcabouço completo e flexível para tratar do problema da variação paramétrica. O filtro pode ser usado não apenas para estimar equações de regressão com parâmetros variáveis, mas também para modelos estruturais de séries temporais. Na realidade, qualquer modelo que puder ser representado em forma de espaço de estados pode ser estimado usando o filtro de Kalman.

O filtro pode ser usado tanto no enfoque clássico como no bayesiano. A principal diferença entre esses dois enfoques é o modo pelo qual as variâncias dos processos dos passeios aleatórios são estimadas. Na abordagem clássica, estes hiperparâmetros são estimados através de um método de máxima verossimilhança, enquanto que no caso bayesiano a variância observável é estimada por um processo de aprendizagem e o problema da estimação da variância não-observável (*evolution variance*) é resolvido escolhendo-se fatores de desconto para as várias partes do modelo. O enfoque bayesiano é muito útil em algumas situações específicas, particularmente em se tratando de mudanças estruturais, porque ele permite que se façam intervenções subjetivas e monitoramento do modelo.

#### Abstract

*This paper surveys the literature on time varying parameter models. It includes not only the more recent, classical and bayesian, approaches based in the Kalman filter, but also provides some discussion of earlier models of parameter variation. These techniques are very useful in economics not only for forecasting purposes, via univariate models of signal extraction, but also for analyzing models subject to structural change.*

#### Bibliografia

AMEEN, J. R. M. e HARRISON, P. J. Discount wheighted estimation. *Journal of Forecasting*, 3, 285-296, 1984.

---

<sup>23</sup> Ver Anderson e Moore (1979) para uma descrição de outros algoritmos de suavização.

- . Nominal discount Bayesian Models. In: BERNARDO, J. M. *et alii* (eds). *Bayesian Statistics*, 2. North-Holland, Amsterdam, 1985a.
- . Discount Bayesian multi-process modelling with CUSUMS. In: ANDERSON, D. O. (ed). *Times series analysis: theory and practice*, 5. North-Holland, Amsterdam, 1985b.
- ANDERSON, B. D. O., MOORE, J.B. *Optimal filtering*. Englewood Cliff: Prentice Hall, 1979.
- AKAIKE, H. Canonical correlation analysis of time series and the use of an information criterion. In: MEHRA, R., LAINIOTIS, D.C. (eds). *Advances and case studies in system identification*. New York: Academic Press, 1976.
- ATHANS, M. The importance of Kalman filtering methods for economic systems. *Annals of Economic and Social Measurement*, n.3, p. 49-64, 1974.
- BELSLEY, D. A. A test for systematic variation in regression coefficients. *Annals of Economic and Social Measurement*, n.2, p. 495-500, 1973a.
- . On the determination of systematic parameter variation in the linear regression model. *Annals of Economic and Social Measurement*, n.2, p. 487-494, 1973b.
- . The applications of Kalman filter in determination of systematic parameter variation. *Annals of Economic and Social Measurement*, n.2, p. 531-533, 1973c.
- BURRIGE, P., WALLIS, K. F. Prediction theory for autoregressive moving average processes. *Econometric Reviews*, v.7, p.65-95, 1988.
- CHOW, G. C. *Econometrics*. Singapore: McGraw-Hill, 1983.
- COOLEY, T. F., PRESCOTT, E. C. The adaptive regression model. *International Economic Review*, v.14, p.364-371, 1973a.
- . Tests of an adaptive regression model. *Review of Economics and Statistics*, v.55, p.248-256, 1973b.
- . Varying parameter regression: a theory and some applications. *Annals of Economic and Social Measurement*, n.2, p. 463-474, 1973c.
- . Estimation in presence of stochastic parameter variation. *Econometrica*, v.44, p.167-183, 1976.
- DIDERRICH, G. T. The Kalman filter from the perspective of Goldberg-Theil estimators. *The American Statistician*, v.39, p.193-198, 1985.

- DUNCAN, D. B., HORN, S. D. Linear dynamic estimation from the viewpoint of regression analysis. *Journal of the American Statistical Association*, v.67, p.815-821, 1972.
- GOLDFELD, S., QUANDT, R. *Nonlinear methods in econometrics*. Amsterdam: North-Holland, 1972.
- . A Markov model for switching regressions. *Journal of Econometrics*, v.1, p.3-16, 1973a.
- . The estimation of structural shifts by switching regressions. *Annals of Economic and Social Measurement*, n.2, p.475-486, 1973b.
- HARVEY, A. C. *Time series models*. Deddington: Philip Allan, 1981.
- . Applications of the Kalman filter in econometrics. In: BEWLEY, T. F. (ed.). *Advances in econometrics - fifth world congress*. Cambridge: Cambridge University Press, 1987, v. I.
- . *Forecasting structural time serie models and the Kalman filter*. Cambridge: Cambridge University Press, 1989.
- HARVEY, A. C., PHILLIPS, G. D. A. The estimation of regression models with time-varying parameters. In: DEISTLER, M., FÜRST, E., SCHWÖDIANER, G. (eds.). *Games, economic dynamics and time series analysis*. Wien-Würzburg: Physica-Verlag, 1982.
- HILDRETH, C., HOUCK, J. P. Some estimators for a linear model with Random coefficients. *Journal of the American Statistical Association*, v.63, p.584-595, 1968.
- JAZWINSKI, A. H. *Stochastic process and filtering theory*. New York: Academic Press, 1970.
- KALMAN, R. E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, Transactions of the ASME, series D, p.34-45, 1960.
- KIEFER, N. M. Discrete parameter variation: efficient estimation of a switching regression model. *Econometrica*, v.46, p.427-434, 1978.
- KLEIN, L. R. *A textbook in econometrics*. Evanston: Row Paterson, 1953.
- KUH, E. The validity of cross-sectionally estimated behavior equations in time series applications. *Econometrica*, v.27, p.197-214, 1959.
- LINDLEY, D. V. *Introduction to probability and statistics from a Bayesian point of view*. Cambridge: Cambridge University Press, 1965.
- LUCAS JR., R. E. Econometric policy evaluation: a critique. *Journal of Monetary Economics*, v.1, suppl., p.19-46, 1976.

- MEINHOLD, R. J., SINGPURWALLA, N. D. Understanding the Kalman filter. *The American Statistician*, v.37, p. 123-127, 1983.
- PORTUGAL, M. S. *Brazilian foreign trade: fixed and time varying parameter models*. University of Warwick, 1992 (Unpublish PhD Thesis).
- PRIESTLEY, M. B. *Spectral analysis and time series*. New York: Academic Press, 1981.
- QUANDT, R. E. The estimation of the parameters of a linear regression obeying two separate regimes. *Journal of the American Statistical Association*, v.53, p.873-880, 1958.
- . A new approach to estimating switching regressions. *Journal of the American Statistical Association*, v.67, p.306-310, 1972.
- QUANDT, R. E., RAMSEY, J.B. Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, v.73, p.730-738, 1978.
- RAO, C. R. The theory of least squares when the parameters are stochastic and its application of growth curves. *Biometrika*, v.52, p.447-458, 1965.
- ROSENBERG, B. The analysis of a cross section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement*, n.2, p.399-428, 1973.
- RUBIN, H. Note on Random coefficients. In: KOOPMANS, T. C. (ed.). *Statistical inference in dynamic models*. 1950, p. 419-421 (Cowels Commission Monograph, 10).
- SANT, D. Generalized least squares applied to time-varying parameters models. *Annals of Economic and Social Measurement*, n.6, p. 301-314, 1977.
- SWAMY, P. A. V. B. Efficient inference in a Random coefficient model. *Econometrica*, v.38, p.311-323, 1970.
- THEIL, H. *Principles of econometrics*. New York: John Wiley and Sons, 1971.
- WEST, M., HARRISON, J. *Bayesian forecasting and dynamic models*. New York: Spriger-Verlag, 1989.
- WOLFF, C. C. P. *Time varying parameters and the out of sample forecasting performance of structural exchange rate models*. London Business School, 1985 (Discussion Paper, 83/85).

(Originais recebidos em dezembro de 1992. Revisos em abril de 1993.)