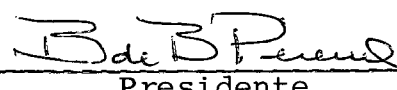


ANÁLISE RECURSIVA DE MODELOS LINEARES DE REGRESSÃO  
E SÉRIES TEMPORAIS: O MÉTODO BAYESIANO DE PREVISÃO

Luiz Eduardo Pazito Mendes

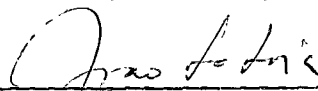
TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS  
DE PÓS-GRADUAÇÃO EM ENGENHARIA DA UNIVERSIDADE FEDERAL DO  
RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A  
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS (M.Sc.)

Aprovada por:

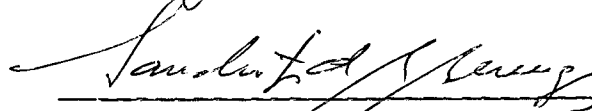


Presidente

BASÍLIO DE BRAGANÇA PEREIRA



JOÃO LUIZ MAURITY SABOIA



SANCHO EDUARDO B. BERENQUER

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 1978

## AGRADECIMENTOS

Ao Professor Basilio de Bragança Pereira, mais que orientador, um amigo.

Ao Professor Reinaldo de Souza, pelo incentivo, sugestões e colaboração.

Ao colega Roberto Pereira D'Araujo, pelas sugestões e críticas.

A Syléa de Figueiredo e Raquel Mellman, pela datilografia, paciência e boa-vontade.

## RESUMO

O trabalho trata da análise recursiva de modelos lineares de regressão e séries temporais.

A abordagem recursiva guarda íntima relação com a representação Markoviana de processos. Este fato permite estabelecer uma analogia entre os conceitos de parâmetro e de estado, justificando a utilização de um algoritmo de estimação de variáveis de estados: o filtro de Kalman.

Ênfase particular é dada à apresentação do método Bayesiano de previsão de Harrison e Stevens, cujas principais características são a rápida adaptabilidade a mudanças bruscas no comportamento da série, a aplicabilidade com número pequeno de observações e a possibilidade de comunicação direta sistema-analista em todos os instantes. Este método é aplicado a uma série de demanda mensal de energia elétrica.

## ABSTRACT

This work deals with the recursive analysis of regression and time series linear models.

The recursive approach is closely related to the Markovian representation of processes. Such fact leads to an analogy between the concepts of parameter and state, justifying the use of a state variable estimation algorithm: the Kalman Filter.

Special emphasis is given to the presentation of the Bayesian forecasting method developed by Harrison and Stevens which has as main features the quick adaptation to sudden changes in the series behavior, the capability of working out even with a small number of observations and the possibility of close communication forecaster-system at any moment. This method is applied to a monthly electrical energy consumption series.

## ÍNDICE

CAPÍTULO I	-	<u>INTRODUÇÃO</u>	
CAPÍTULO II	-	<u>REVISÃO DA LITERATURA</u>	
CAPÍTULO III	-	<u>ANÁLISE RECURSIVA DE MODELOS LINEARES DE REGRESSÃO POR MÍNIMOS QUADRADOS</u>	
III.1	-	Abordagem Clássica .....	13
III.2	-	Abordagem Recursiva: Parâmetros Constantes .....	17
III.2.1	-	Algoritmo Determinístico .....	17
III.2.2	-	Algoritmo Estocástico .....	20
III.2.3	-	Características do Algoritmo Estocástico ....	22
III.2.3.1	-	Vantagens .....	23
III.2.3.2	-	Interpretações .....	23
III.2.3.3	-	Inicialização .....	25
III.3	-	Abordagem Recursiva: Parâmetros Variáveis .....	26
III.3.1	-	Algoritmo de Ponderação Exponencial .....	26
III.3.2	-	Algoritmo Linear Dinâmico .....	28
III.3.3	-	Casos Particulares do Modelo Linear Dinâmico.	33
III.3.1.1	-	O Passeio Aleatório Multidimensional ..	33
III.3.2.2	-	O Modelo de Gauss-Markov de Primeira Ordem .....	34
III.3.2.3	-	Consideração .....	35
III.3.4	-	Generalizações do Modelo Linear Dinâmico ....	35
III.3.4.1	-	Modelo ARIMA dos Parâmetros .....	35
III.3.4.2	-	Variâncias Não Constantes .....	37
III.3.5	-	Um Teste para a Invariância dos Parâmetros ..	39

## CAPÍTULO IV - O FILTRO DE KALMAN

IV.1	- O Conceito de Estado em Sistemas Determinísticos ....	46
IV.2	- O Conceito de Estado em Sistemas Estocásticos .....	48
IV.3	- O Problema de Estimação de Estados .....	49
IV.4	- Filtro de Kalman Discreto: Solução Bayesiana .....	50
IV.4.1	- Caso Geral: Estimação Estocástica de Múltiplos Estágios .....	51
IV.4.2	- Caso Particular: o filtro de Kalman .....	55

## CAPÍTULO V - ANÁLISE RECURSIVA DE MODELOS LINEARES DE SÉRIES TEMPORAIS

V.1	- Modelos Espaço de Estados: O Método Bayesiano de Previsão a Curto Prazo .....	63
V.1.1	- Características .....	63
V.1.2	- Modelagem do Processo: O Modelo Linear Dinâmico ..	66
V.1.3	- Estimação dos Parâmetros: o filtro de Kalman .....	68
V.1.4	- Previsão .....	71
V.1.5	- Superposição .....	75
V.1.6	- Representação de Modelos Convencionais em Espaço de Estados .....	76
V.1.6.1	- Regressão Linear .....	77
V.1.6.2	- Modelo de Crescimento Linear Sazonal .....	78
V.1.6.3	- Modelo de Funções Periódicas .....	80
V.1.6.4	- Modelo ARIMA: o filtro de Kalman Extendido ...	83
V.1.7	- Modelos de Múltiplos Processos .....	87
V.1.7.1	- Modelo Único Representativo do Processo .....	87
V.1.7.2	- Modelo Múltiplo Representativo do Processo ...	89
V.1.8	- Exemplos de Intervenção do Analista .....	96

V.2	- Modelos Convencionais: O Método VI-MVA .....	96
V.2.1	- Hipóteses do Sistema .....	97
V.2.2	- Relações entre as Variáveis .....	98
V.2.3	- Estimação dos Parâmetros .....	100
V.2.4	- O Algoritmo de Variáveis Instrumentais (VI) .....	102
V.2.5	- O Algoritmo de Máxima Verossimilhança Aproximada (MVA) .....	104

CAPÍTULO VI - APLICAÇÃO DO MÉTODO BAYESIANO

VI.1	- Introdução: Características da Série .....	107
VI.2	- Modelo Utilizado .....	108
VI.3	- Informação de Entrada: Critério de Escolha dos Parâmetros de Entrada .....	110
VI.3.1	- Especificação da Distribuição Inicial dos Parâmetros .....	112
VI.3.2	- Especificação dos Modelos de Estado .....	113
VI.4	- Lógica do Programa de Múltiplos Estados .....	116
VI.5	- Intervenções do Analista .....	118
VI.6	- Desempenho do Modelo .....	119
VI.7	- Considerações Finais .....	121

CAPÍTULO VII - CONCLUSÕES E SUGESTÕES

REFERÊNCIAS .....	127
ANEXO .....	130

## CAPÍTULO I

### INTRODUÇÃO

O objetivo principal deste trabalho é a apresentação e a aplicação, a uma série de demanda mensal de energia elétrica, do método Bayesiano de previsão a curto prazo de séries temporais desenvolvido por Harrison e Stevens. Entretanto, a modelagem Markoviana por ele utilizada propicia um tipo de estimação essencialmente recursiva que, a cada instante, atualiza a estimativa dos parâmetros combinando a estimativa anterior com a informação daquele instante. Em outras palavras, a estimativa anterior carrega consigo toda a informação acerca do passado suficiente para prever seu efeito no futuro. Isto estabelece uma analogia marcante com o conceito de estado, permitindo encará-lo como um método de estimação das variáveis de estado de um sistema linear dinâmico estocástico e justificando o uso, para a estimação, de uma relação de recorrência comum em Teoria de Controles: o filtro de Kalman. O enfoque do trabalho passa a ser, então, a análise recursiva de modelos lineares de regressão (por mínimos quadrados) e séries temporais, procurando lhe dar consistência teórica e estabelecer sua evolução cronológica e apontando, quando se julgar procedente, tópicos da Teoria de Controles a ela relacionados. Neste particular é dada especial atenção do conceito de estado, à derivação do filtro de Kalman e a representação de séries temporais em modelos de espaço de estados. Ressalte-se que o objetivo principal continua



sendo o método Bayesiano de previsão a curto prazo de séries temporais, apenas situado num contexto mais amplo.

Este tipo de interesse se justifica na medida em que a previsão é um subsídio relevante à tomada de decisões. Como decisões criteriosas geralmente proporcionam economias consideráveis, a análise de séries temporais objetivando métodos eficientes de previsão se reveste de grande importância.

A formulação recursiva anteriormente citada permite uma abordagem Bayesiana ao problema. A cada instante, pode-se prover informações, através de distribuições a priori, de duas maneiras diferentes: internamente, estabelecendo leis de variação e propriedades estatísticas para os parâmetros e externamente, refletindo a ocorrência de fatos que influenciam o comportamento futuro da série. E este talvez seja o ponto principal: a inclusão formal do analista, com sua experiência e sensibilidade, como parte integrante do sistema de previsão.

A maioria dos métodos baseia a previsão numa análise retrospectiva dos dados, não levando em conta o fato presente e suas consequências futuras. Existe uma tendência a previsões demasiadamente calcadas no passado e este, em determinado instante, pode deixar de ser significativo. Exemplificando, uma série de vendas de um produto pode sofrer sensível modificação quando um competidor entra no mercado, quando se efetua uma propaganda maciça ou quando se alteram os hábitos do consumidor. Os efeitos provenientes, difíceis de expressar quantitativamente, necessitam de distribuições de probabilidade para descrever sua incerteza.

Um modelo que opera somente com a série histórica, transformando mecanicamente dados de entrada em informações de saída, não pode prever acontecimentos desta natureza e corre o risco de, a partir de determinado instante, se tornar inadequado. Surge, então, a necessidade de métodos mais flexíveis que permitam incorporar à série histórica informações transmitidas pelo analista e/ou possuam uma estrutura adaptativa que possibilite uma percepção rápida de mudanças externas, expressando-as convenientemente em termos quantitativos.

Estas idéias, basicamente, motivaram a Harrison e Stevens o desenvolvimento do método Bayesiano de previsão, cujas características se prestam a este tipo de situações.

Em linhas gerais, a esquematização do trabalho é a seguinte: o capítulo I situa o problema e define os objetivos. O capítulo II faz uma revisão da literatura relacionada ao assunto procurando estabelecer uma seqüência cronológica. O capítulo III aborda a estimação por mínimos quadrados de modelos lineares de regressão, mostrando rapidamente os resultados clássicos para posteriormente se deter nos algoritmos recursivos, nos casos de parâmetros constantes e variáveis. O capítulo IV introduz os tópicos da Teoria de Controles que têm relação com a parte estatística do trabalho, basicamente, o conceito de estado de um sistema e a derivação Bayesiana do filtro de Kalman. O capítulo V se ocupa da análise recursiva de séries temporais, apresentando os fundamentos teóricos e as características do método Bayesiano de previsão a curto prazo, a representação de vários modelos usuais de

séries temporais em modelos de espaço de estados e um método recursivo de estimação dos parâmetros de um modelo dinâmico de função de transferência baseado em modificações por variáveis instrumentais e máxima verossimilhança aproximada das equações recursivas de mínimos quadrados. O Capítulo VI mostra a aplicação do método Bayesiano à série de demanda mensal de energia elétrica das Centrais Elétricas do Espírito Santo S.A. (ESCELSA). O Capítulo VII contém conclusões e sugestões para possíveis desenvolvimentos de determinados tópicos. Seguem-se a bibliografia e os anexos.

## CAPÍTULO II

### REVISÃO DA LITERATURA

A primeira aplicação, na análise de observações, do princípio de estimação por mínimos quadrados do modelo de regressão linear data do início do século XIX e deve-se a GAUSS em 1801 e LEGENDRE em 1806.

A partir daí, o método se desenvolve naturalmente, evoluindo de um mero processo mecânico de ajuste para um contexto estocástico, no qual as observações e os estimadores são considerados variáveis aleatórias com propriedades e características conhecidas. Isto proporciona maior consistência teórica, possibilitando a obtenção de uma série de resultados importantes.

Em 1950, PLACKETT [14] acrescenta algo realmente novo à teoria, apresentando modificações no método aplicáveis a duas situações:

- (i) - quando a matriz relacionando o valor esperado das observações com os parâmetros desconhecidos não for de rank pleno. Nela, formaliza resultados obtidos por YATES e HALE em 1939 que, supondo a existência de uma relação fixa entre os parâmetros desconhecidos, determinam, por substituição, sua matriz de covariância;
- (ii) - quando se deseja atualizar as estimativas dos parâmetros, da sua matriz de covariância e da soma dos quadrados dos

resíduos, devido a presença de novas observações. Nela, desenvolve uma elegante formulação matricial recursiva que, em presença das novas observações, atualiza as estimativas com um esforço computacional menor que o método clássico (no qual é necessária a inversão de uma matriz).

Apesar da potencialidade, da elegância e da simplificação, a análise recursiva de mínimos quadrados não causa no meio estatístico a devida repercussão. Na realidade é através de KALMAN [ 8 ], um engenheiro de Controles, que a idéia se populariza a partir de um trabalho publicado em 1960. Considerando os parâmetros como as variáveis de estado de um sistema dinâmico descritos por um conjunto de equações lineares estocásticas, a hipótese de invariância, implícita no trabalho de Plackett, pode ser generalizada: eles são supostos variáveis com uma componente determinística (caracterizando o processo evolutivo normal do sistema) e uma componente estocástica (representando a aleatoriedade inerente a todo sistema físico). Para derivar esta formulação geral, Kalman utiliza o conceito de projeções ortogonais. Deve-se ressaltar que o trabalho se destina basicamente à teoria de filtragem, não sendo aventada explicitamente pelo autor sua relação com a idéia de mínimos quadrados recursivos da Estatística. Apesar disto, as equações de Plackett constituem caso particular das de Kalman sem que possivelmente este tenha conhecimento do trabalho daquele. As diversas formas do algoritmo de filtragem de Kalman (discreta, contínua e discreta-contínua) desempenham um importante papel na Teoria de Controles sendo aplicadas, inclusive, nas estimações da trajetória e da órbita da missão Apolo.

Em 1964, HO e LEE [7] propõem uma solução Bayesiana para derivar as equações do filtro de Kalman num trabalho que visa formular uma classe geral de problemas de estimação estocástica sob o ponto de vista da Teoria de Decisão Bayesiana. Esta solução, mais detalhada, aparece no mesmo ano num livro de LEE [10] que dá um tratamento aprofundado aos problemas de estimação, identificação e controle ótimos de sistemas estocásticos. Nele, o critério de ajuste recursivo por mínimos quadrados é deduzido explicitamente, sendo pela primeira apontada a ligação entre estimação de parâmetro e de estado.

Em 1969, YOUNG [17] mostra como são obtidas as equações de estimação recursiva por mínimos quadrados do modelo de regressão linear e como elas podem ser modificadas para permitir a percepção de possíveis variações dos parâmetros. A seguir, discute sua aplicação na estimação e identificação de processos dinâmicos a partir dos dados normais de operação.

Em 1971, HARRISON e STEVENS [4] aplicam os princípios Bayesianos à previsão a curto prazo de séries temporais. Pela primeira vez, um método modela explicitamente a ocorrência de situações "anormais" como mudanças de nível e de inclinação ou transições, calculando as probabilidades a posteriori destes eventos a cada instante e fornecendo, com isto, distribuições do nível e da inclinação, relevantes à tomada de decisões baseada nas previsões.

Em 1972, MACGREGOR [12] discute a abordagem de variáveis de estado para modelar sistemas dinâmicos lineares discretos

estocásticos e a relaciona à que utiliza modelos funções de transferência e modelos ARIMA. Mostra que a forma padrão do modelo de variáveis de estado usando dois vetores normais independentes pode sempre ser escrita numa forma mais simplificada, empregando um único vetor normal, dando várias destas representações para o modelo ARIMA geral.

Em 1974, YOUNG [18] formaliza a análise recursiva de mínimos quadrados estabelecendo seu processo evolutivo e suas principais características e indicando sua utilidade no delineamento de métodos recursivos de análise de séries temporais para modelos de sistemas dinâmicos estocásticos. Comenta que a maioria da literatura aborda modelos convencionais tipo "caixa-preta", preocupados basicamente com as relações entre as entradas e as saídas, contrastando com a escassez de publicações a respeito de modelos tipo "espaço de estados", mais voltados à descrição interna do processo. Estes modelos são importantes por facilitarem o uso de informações a priori sobre o sistema e por possibilitarem que se evite a especificação de formas canônicas satisfatórias para as equações de modelos multivariados.

Em 1974, AKAIKE [1] discute a relação da representação Markoviana de um processo estocástico estacionário com a representação ARMA (auto-regressivo médias móveis), com um critério de máxima verossimilhança para o ajuste da representação Markoviana. Introduce um procedimento prático para encontrar uma forma inicial desta representação, demonstrando com exemplos numéricos sua validade.

Em 1974, BLIGHT |2| discute algumas aplicações das abordagens Bayesiana e de Wiener - Kolmogorov na construção de equações recursivas para a estimação de parâmetros estocásticos. Aponta soluções para os casos de previsão, filtragem e amortecimento quando o parâmetro segue um processo Markoviano e é observado com erro. Utiliza a solução Bayesiana para gerar uma fatorização canônica da função geradora da autocovariância da série observada, generalizando estas técnicas para outros modelos.

Em 1976, HARRISON e STEVENS |5| e |6| apresentam o sistema Bayesiano de previsão a curto prazo de séries temporais em dois artigos: um teórico, enfatizando como principal aspecto a inclusão formal do analista no sistema de previsão e introduzindo, através dos modelos de múltiplos processos, a noção de incerteza quanto ao verdadeiro modelo em curso num determinado instante, numa generalização dos conceitos apresentados no trabalho de 1971; outro, contendo a aplicação do método a quatro séries temporais.

Em 1977, LEONARD e HARRISON |11| desenvolvem novas técnicas para o método de Harrison e Stevens. Supõem distribuições a priori contínuas para as variâncias desconhecidas do modelo e descrevem uma maneira de atualizar as variâncias e os parâmetros do processo. Computam, através de integração unidimensional, distribuições das futuras observações, comparando com os resultados obtidos pela aplicação de um modelo ARIMA Box e Jenkins. Propõem uma generalização da abordagem Bayesiana para situações em que as observações não têm distribuição normal.



Em 1978, LEDOLTER [9] discute a estimação recursiva de parâmetros em modelos de regressão linear (parâmetros constantes e variáveis) e modelos ARIMA. Nos modelos ARIMA mostra que, no caso de haver termos de médias móveis, o modelo precisa ser linearizado e a atualização da estimativa feita pelo filtro de Kalman Extendido.

Em 1978, SOUZA e HARRISON [16] generalizam o trabalho de 1976 de Harrison e Stevens apresentando uma abordagem à previsão baseada em princípios Bayesianos e Teoria da Informação. A formulação do modelo utiliza a função de entropia de Shannon e o princípio de máxima entropia de Jaynes, possibilitando o relaxamento das hipóteses de normalidade e linearidade necessárias a todos os métodos de previsão e tornando o modelo linear dinâmico de 1976 um caso particular desta formulação.

Todas estas publicações contêm tópicos diretamente relacionados a este trabalho. Outras publicações sobre modelos de séries temporais citadas a seguir se relacionam à modelagem de sistemas lineares dinâmicos estocásticos, campo desenvolvido quase independentemente por estatísticos, com interesse principal na modelagem de séries temporais, e engenheiros de controles, com interesse principal no controle dos processos caracterizados pelos modelos dinâmicos estocásticos. Torna-se interessante apresentá-los por se relacionarem indiretamente a temas abordados neste trabalho.

Na área da Estatística, o interesse em modelos de séries temporais desta espécie começa com YULE (1927) e SLUTSKY (1927). Eles mostram que comportamentos pseudo-periódicos seme-

lhantes aos verdadeiramente ocorridos em algumas séries temporais podem resultar de modelos contendo somente componentes aleatórias. YULE introduz a noção de séries auto-regressivas, enquanto SLUTSKY considera um caso especial do processo de médias móveis. Os desenvolvimentos subsequentes de modelos estacionários auto-regressivos e de médias móveis devem-se grandemente a trabalhos de WALKER (1931), WOLD (1938), KOLMOGOROV (1939), que propõe uma solução para o problema de filtragem e previsão e BARTLETT (1946) e (1955). A hipótese de estacionaridade é relaxada por YAGLON (1955) que introduz a idéia de considerar estacionárias algumas diferenciações do processo. Num famoso trabalho, BOX e JENKINS [2] em 1970 mostram que estas diferenciações permitem à série original ser homogeneamente não estacionária quando certas condições de inversibilidade e estacionaridade são satisfeitas pelas partes auto-regressiva e de médias móveis do modelo. A partir daí propõem uma classe bastante geral de modelos denominados modelos integrados auto-regressivos médias móveis (ARIMA) que também permite a incorporação de efeitos sazonais.

Na área de Controles, o ponto de partida deve-se a WIENER (1949) que, em contraste com as equações paramétricas de diferenças preferidas pelos estatísticos, usa representações não paramétricas para caracterizar os distúrbios estocásticos estacionários (autocorrelação, correlação cruzada e função densidade espectral). Ele mostra que a especificação de um filtro linear para a previsão de mínimo erro quadrático de um sinal estocástico ou a separação entre o sinal e o ruído levam a uma equação integral denominada equação de Wiener - Hopf. Para resolvê-la, pro-

põe um método de fatorização espectral válida no caso de uma série temporal com espectro racional. De todas as extensões que se seguem as mais importantes devem-se a ZADEH e RAGAZINNI (1950) e a BODE e SHANNON (1950) que apresentam um método de solução simplificado baseado na idéia de considerar uma série temporal estacionária com um espectro racional, como a saída de um filtro linear com um ruído branco de entrada. Até que em 1960 KALMAN, conforme mencionado anteriormente, dá uma abordagem geral nesta área para o problema de filtragem linear e previsão. Ele volta-se para uma representação paramétrica de modelos dinâmicos estocásticos em termos de equações diferenciais de vetores de estados, tornando o problema da fatorização espectral de Wiener e fornecendo resultados gerais para os problemas de filtragem e previsão de processos Markovianos multivariados não-estacionários.

De um modo geral, estas são as publicações mais importantes direta e indiretamente relacionadas ao trabalho.

## CAPÍTULO III

### ANÁLISE RECURSIVA DE MODELOS LINEARES DE REGRESSÃO POR MÍNIMOS QUADRADOS

O princípio de estimação por mínimos quadrados dos parâmetros do modelo de regressão linear é uma técnica bastante utilizada em aplicações práticas e em desenvolvimento de pesquisas científicas. Sua vasta aplicabilidade e aceitação devem-se, basicamente, ao fato de aliar simplicidade matemática a resultados precisos e coerentes, além de possuir um forte apelo intuitivo.

Neste capítulo, após breve exposição do resultado clássico, são apresentados os algoritmos recursivos, para os casos de parâmetros constantes e variáveis. Como ao longo do tempo raramente se conhece o modelo estocástico geral, são feitas algumas simplificações que levam aos modelos de Gauss-Markov de Primeira Ordem e de Passeio Aleatório. Para este apresenta-se um teste para verificar a veracidade ou não da hipótese de parâmetros variáveis.

#### III.1 - Abordagem Clássica

Seja o problema de regressão linear no qual uma variável  $y$  se relaciona linearmente, por hipótese, com  $N$  variáveis independentes  $x_j$  ( $j = \overline{1, N}$ ) cujos valores são conhecidos. Efetuan-

do-se  $t$  observações da variável dependente  $y$ , denotadas por  $y_i$  ( $i = \overline{1, t}$ ), cada uma vem "contaminada" pela presença de um ruído aleatório  $v_i$  decorrente de erros de medida ou de erros estocásticos devidos, possivelmente, à influência de variáveis omitidas no modelo. Sendo  $x_{ij}$  ( $j = \overline{1, N}$ ) o conjunto de valores associados à observação  $y_i$ , pode-se escrever o seguinte sistema de equações lineares relacionando as variáveis:

$$(3.1) \quad y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_N x_{iN} + v_i \quad , \quad i = \overline{1, t}$$

onde  $\theta_1, \theta_2, \dots, \theta_N$  são parâmetros constantes e desconhecidos cujos valores se deseja de alguma forma estimar e os ruídos  $v_i$  têm características especificadas no item III.2.2.

Uma maneira mais funcional de representar o sistema é através da formulação matricial:

$$(3.2) \quad Y = X\theta + v$$

onde:

$Y$  - vetor de observações ( $t \times 1$ )

$X$  - matriz de variáveis independentes ( $t \times N$ )

$\theta$  - vetor de parâmetros ( $N \times 1$ )

$v$  - vetor de ruídos ( $t \times 1$ )

sendo  $Y$  e  $X$  conhecidos,  $\theta$  e  $v$  desconhecidos.

Assim colocado o problema se resume em estimar o vetor  $\theta$  tendo como informação a matriz  $X$  e o vetor  $Y$ . O método mais conhecido para isto é o dos mínimos quadrados que consiste em determinar o valor  $\hat{\theta}$  que minimiza a soma dos quadrados dos re

síduos  $e_i$  ( $i = \overline{1, t}$ ), com os resíduos sendo definidos pela seguinte relação:

$$(3.3) \quad Y = X\hat{\theta} + e$$

onde

$e$  - vetor de resíduos ( $t \times 1$ )

Assim, o ruído  $v$  indica o erro verdadeiro na relação entre  $X$  e  $Y$  (quando são relacionados através do valor desconhecido  $\theta$ ) e o resíduo  $e$ , o erro estimado (quando são relacionados através de  $\hat{\theta}$ ):

Definindo:

$$(3.4) \quad J = \sum_{i=1}^t e_i^2 = e'e$$

o problema de estimação pode ser enunciado da seguinte maneira: determinar o valor  $\hat{\theta}$  que minimiza a função de critério  $J$  satisfazendo a relação (3.3), isto é:

$$\text{Min } J = e'e$$

s.a

$$Y = X\hat{\theta} + e$$

Desenvolvendo (3.4):

$$J = e'e = (Y - X\hat{\theta})'(Y - X\hat{\theta}) = Y'Y - \hat{\theta}'X'Y - Y'X\hat{\theta} + \hat{\theta}'X'X\hat{\theta}$$

Observando que  $\hat{\theta}'X'Y$  é escalar e portanto igual a seu transposto  $Y'X\hat{\theta}$ , vem:

$$J = Y'Y - 2\hat{\theta}'X'Y + \hat{\theta}'X'X\hat{\theta}$$

Como  $J$  é uma função unimodal no espaço dos parâmetros, a condição necessária e suficiente para sua minimização é que o valor  $\hat{\theta}$  satisfaça as condições:

$$\frac{\partial J}{\partial \hat{\theta}} = 0 \quad \text{e} \quad \frac{\partial^2 J}{\partial \hat{\theta}^2} > 0$$

Tem-se:

$$\frac{\partial J}{\partial \hat{\theta}} = 2X'Y + X'X\hat{\theta} + \hat{\theta}'X'X = 0$$

$$- 2X'Y + 2 X'X\hat{\theta} = 0$$

$$X'X\hat{\theta} = X'Y$$

$$(3.5) \quad \hat{\theta} = (X'X)^{-1}X'Y$$

Além disso:

$$\frac{\partial^2 J}{\partial \hat{\theta}^2} = 2 X'X$$

Como  $X'X$  é positiva definida, o valor  $\hat{\theta}$  de (3.5) é ponto de mínimo da função  $J$ .

Observe-se que a hipótese de independência entre as variáveis  $x_j$  ( $j = \overline{1, N}$ ) garante a existência de  $(X'X)^{-1}$  em (3.5)

### III.2 - Abordagem Recursiva: Parâmetros Constantes

#### III.2.1 - Algoritmo Determinístico

Introduzindo em (3.5) o índice  $t$  para indicar que a estimativa foi realizada levando-se em conta  $t$  amostras, chega-se a:

$$(3.6) \quad \hat{\theta}_t = P_t b_t$$

onde:

$$P_t = (X'X)^{-1}$$

$$b_t = (X'Y)$$

Definindo-se  $x_i$  como a  $i$ -ésima linha da matriz  $X$  ( $x_i = [X_{i1}, \dots, X_{iN}]$ ) e  $y_i$  como o  $i$ -ésimo elemento do vetor  $Y$ , as seguintes relações decorrem da definição de multiplicação matricial:

$$X'X = \sum_{i=1}^t x_i' x_i$$

$$X'Y = \sum_{i=1}^t x_i' y_i$$



Em consequência:

$$(3.7) \quad P_t = (X'X)^{-1} = \left[ \sum_{i=1}^t x_i'x_i \right]^{-1}$$

$$(3.8) \quad b_t = X'Y = \sum_{i=1}^t x_i'y_i$$

De (3.7) tira-se que:

$$(3.9) \quad P_t^{-1} = \sum_{i=1}^t x_i'x_i = \sum_{i=1}^{t-1} x_i'x_i + x_t'x_t = P_{t-1}^{-1} + x_t'x_t$$

Pré-multiplicando (3.9) por  $P_t$  e pós-multiplicando por  $P_{t-1}$ :

$$P_t P_t^{-1} P_{t-1} = P_t P_{t-1}^{-1} P_{t-1} + P_t x_t' x_t P_{t-1}$$

$$(3.10) \quad P_{t-1} = P_t + P_t x_t' x_t P_{t-1}$$

Pós-multiplicando por  $x_t'$ :

$$(3.11) \quad P_{t-1} x_t' = P_t x_t' + P_t x_t' x_t P_{t-1} x_t' = P_t x_t' (1 + x_t P_{t-1} x_t')$$

pós-multiplicando por  $(1 + x_t P_{t-1} x_t')^{-1} x_t P_{t-1}$ :

$$(3.12) \quad P_{t-1} x_t' (1 + x_t P_{t-1} x_t')^{-1} x_t P_{t-1} = P_t x_t' x_t P_{t-1}$$

Substituindo (3.12) em (3.10):

$$(3.13) \quad P_t = P_{t-1} - P_{t-1} x_t' (1 + x_t P_{t-1} x_t')^{-1} x_t P_{t-1}$$

Por outro lado, de (3.8) tira-se que:

$$(3.14) \quad b_t = \sum_{i=1}^{t-1} x_i' y_i + x_t' y_t = b_{t-1} + x_t' y_t$$

As equações (3.13) e (3.14) fornecem valores de  $P_t$  e  $b_t$  em função de  $P_{t-1}$  e  $b_{t-1}$  já estimados e da informação obtida na  $t$ -ésima amostra  $\{x_t, y_t\}$ . Isto sugere a tentativa de uma fórmula recursiva para determinação de  $\hat{\theta}_t$  que utilize a estimativa  $\hat{\theta}_{t-1}$  feita após  $(t-1)$  amostras e um termo corretivo baseado na informação contida na  $t$ -ésima amostra, representada por  $x_t$  e  $y_t$ . Para tal basta substituir (3.13) e (3.14) em (3.6), obtendo-se:

$$\hat{\theta}_t = \left[ P_{t-1} - P_{t-1} x_t' (1 + x_t' P_{t-1} x_t')^{-1} x_t' P_{t-1} \right] \left[ b_{t-1} + x_t' y_t \right]$$

$$(3.15) \quad \hat{\theta}_t = \hat{\theta}_{t-1} - P_{t-1} x_t' (1 + x_t' P_{t-1} x_t')^{-1} (x_t' \hat{\theta}_{t-1} - y_t)$$

Mas de (3.11) tira-se que:

$$P_{t-1} x_t' (1 + x_t' P_{t-1} x_t')^{-1} = P_t x_t' (1 + x_t' P_{t-1} x_t') (1 + x_t' P_{t-1} x_t')^{-1} = P_t x_t'$$

Substituindo esta expressão em (3.15):

$$\hat{\theta}_t = \hat{\theta}_{t-1} - P_t x_t' (x_t' \hat{\theta}_{t-1} - y_t)$$

$$(3.16) \quad \hat{\theta}_t = \hat{\theta}_{t-1} - P_t (x_t' x_t \hat{\theta}_{t-1} - x_t' y_t)$$

Portanto as equações (3.13) e (3.16):

$$P_t = P_{t-1} - P_{t-1} x_t' (1 + x_t P_{t-1} x_t')^{-1} x_t P_{t-1}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} - P_t (x_t' x_t \hat{\theta}_{t-1} - x_t' y_t)$$

forneem a partir de valores iniciais arbitrários  $\hat{\theta}_0$  e  $P_0$  uma solução recursiva para a equação (3.6):

$$\hat{\theta}_t = P_t b_t$$

### III.2.2 - Algoritmo Estocástico

O algoritmo recursivo apresentado é um método puramente determinístico de estimação, não levando em conta as hipóteses a respeito da natureza estocástica dos ruídos e não fornecendo nenhuma informação estatística acerca das estimativas. A consideração das hipóteses estatísticas clássicas da regressão linear proporciona uma melhora no seu desempenho, na medida em que permite estabelecer propriedades relativas à "qualidade" das estimativas (como por exemplo, não tendenciosidade, consistência, etc). Assim, tem-se:

#### Hipóteses

- 1 -  $E[v] = 0$ . Os ruídos têm média nula
- 2 -  $E[vv'] = \sigma^2 I_t$ . Os ruídos têm variância constante  $\sigma^2$  e são serialmente não correlacionados
- 3 - Os ruídos  $v$  e cada variável independente  $X_i$  ( $i=1, t$ ) são independentes entre si.

## Propriedades

1 -  $E[\hat{\theta}_t] = \theta_t$ . O estimador  $\hat{\theta}_t$  de (3.16) é não tendencioso

Prova:

De (3.2) e (3.5) vem (considerando o índice t):

$$\begin{aligned}\hat{\theta}_t &= (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\theta_t + v) = (X'X)^{-1}X'X\theta_t + (X'X)^{-1}X'v = \\ &= \theta_t + (X'X)^{-1}X'v\end{aligned}$$

Tomando o valor esperado:

$$E[\hat{\theta}_t] = E[\theta_t] + (X'X)^{-1}X'E[v]$$

Pela hipótese (1):

$$E[\hat{\theta}_t] = E[\theta_t] = \theta_t$$

2 -  $P_t^* \triangleq \text{Cov}(\hat{\theta}_t) = \sigma^2 P_t$ . A matriz de covariância do erro da estimativa igual ao produto da variância dos ruídos  $\sigma^2$  pela matriz  $P_t$  de (3.13).

Prova:

De (3.2), (3.5) e (3.7) vem:

$$\hat{\theta}_t - \theta_t = (X'X)^{-1}X'v = P_t X'v$$

Logo:

$$\begin{aligned}P_t^* &= \text{Cov}[\hat{\theta}_t] = E\{(\hat{\theta}_t - E[\hat{\theta}_t])(\hat{\theta}_t - E[\hat{\theta}_t])'\} = E[(\hat{\theta}_t - \theta_t)(\hat{\theta}_t - \theta_t)'] = \\ &= E[(P_t X'v)(P_t X'v)']\end{aligned}$$

Mas  $P_t$  é simétrica e portanto  $P_t = P_t'$ . Logo:

$$P_t^* = E \left[ P_t X' v v' X P_t \right] = P_t X' E \left[ v v' \right] X P_t$$

pela hipótese (2):

$$P_t^* = P_t X' \sigma^2 I_t X P_t = \sigma^2 P_t X' X P_t$$

De (3.7) vem:

$$P_t^* = \sigma^2 P_t P_t^{-1} P_t = \sigma^2 P_t$$

De posse destes resultados, pode-se substituir  $P_t$  por  $P_t^*/\sigma^2$  nas equações (3.13) e (3.16), obtendo o algoritmo:

$$(3.17) \quad P_t^* = P_{t-1}^* - P_{t-1}^* x_t' \left[ \sigma^2 + x_t P_{t-1}^* x_t' \right]^{-1} x_t P_{t-1}^*$$

$$(3.18) \quad \hat{\theta}_t = \hat{\theta}_{t-1} - \frac{P_t^*}{\sigma^2} (x_t' x_t \hat{\theta}_{t-1} - x_t' y_t)$$

É interessante ressaltar que (3.17) e (3.18) correspondem às equações obtidas por Plackett.

### III.2.3 - Características do Algoritmo Estocástico

#### III.2.3.1 - Vantagens

- em relação ao algoritmo determinístico, o fato de incorporar, a través da matriz de covariância do erro da estimativa  $P_t^*$  uma in-

formação estatística que indica o grau de confiabilidade dessa estimativa. Portanto para cada amostra  $\{x_i, y_i\}$  ( $i = \overline{1, t}$ ), fornece uma estimativa  $\hat{\theta}_i$  dos parâmetros e uma medida de sua precisão, além de informar acerca de sua convergência.

- em relação à solução clássica, o fato de não necessitar inversão de matrizes por ser escalar o termo  $[\sigma^2 + x_t P_{t-1}^* x_t']$ . Esta vantagem é considerável principalmente para sistemas de processamento em tempo real ("on-line") onde os dados são recebidos continuamente e as estimativas precisam ser atualizadas rapidamente em função das novas informações. A solução clássica requer, a cada nova observação, a inversão de uma matriz  $N \times N$ .

### III.2.3.2 - Interpretações

- algoritmo de inferência Bayesiana

Considerando  $P_t^*$  como a matriz de covariância do erro da estimativa, a recursão é iniciada escolhendo-se um valor  $P_0^*$  consistente com o nível de crença no valor  $\hat{\theta}_0$ . Deste modo  $\hat{\theta}_0$  com matriz de covariância a priori  $P_0^*$  associada representa uma estimativa a priori que, após o conhecimento da primeira amostra  $\{x_1, y_1\}$  dá origem à estimativa a posteriori  $\hat{\theta}_1$  com matriz de covariância a posteriori  $P_1^*$  associada, que por sua vez representa uma estimativa a priori para a segunda estimativa. O processo prossegue desta maneira: com a chegada de novas informações, a estimativa a posteriori torna-se a priori para a estimativa subsequente, caracterizando perfeitamente um processo de inferência Bayesiana.

- algoritmo de aproximação estocástica multidimensional

Basta notar que por (3.3) e (3.4) e pela definição de multiplicação matricial decorre:

$$(3.19) \quad J = e'e = (Y - X\hat{\theta}_t)'(Y - X\hat{\theta}_t) = \sum_{i=1}^t (y_i - x_i'\hat{\theta}_t)^2 = \sum_{i=1}^t (x_i'\hat{\theta}_t - y_i)^2$$

O gradiente de J em relação a  $\hat{\theta}_t$  fica:

$$(3.20) \quad \begin{aligned} \nabla_{\hat{\theta}_t}(J) &= \frac{\partial J}{\partial \hat{\theta}_t} = \left( \sum_{i=1}^t x_i'x_i \right) \hat{\theta}_t - \sum_{i=1}^t x_i'y_i = \\ &= \left[ \left( \sum_{i=1}^{t-1} x_i'x_i \right) \hat{\theta}_t - \sum_{i=1}^{t-1} x_i'y_i \right] + \left[ x_t'x_t\hat{\theta}_t - x_t'y_t \right] \end{aligned}$$

A comparação da última parcela de (3.20) com a de (3.18), sugere que se considere o termo  $P_t^*/\sigma^2 (x_t'x_t\hat{\theta}_{t-1} - x_t'y_t)$  como uma medida instantânea do gradiente da função de critério J no t-ésimo "instante amostral", modulado pela matriz  $P_t^*/\sigma^2$ . Encarando desta forma (3.18) representa um método discreto multidimensional do gradiente e o algoritmo (3.16) e (3.17), um algoritmo de aproximação (ou de gradiente) estocástica multidimensional com o ganho escalar substituído pela matriz de ganho  $P_t^*/\sigma^2$ .

Considerando a consistência da estimativa, a medida que a estimação prossegue a confiança na estimativa aumenta e o gradiente observado passa a decorrer, provavelmente, da imprecisão introduzida pelo ruído. Como o peso associado ao termo corretivo diminui, a matriz  $P^*/\sigma^2$  atua no sentido de suavizar ou filtrar esta imprecisão.

A interpretação como um algoritmo de aproximação estocástica é importante por dois motivos:

- (i) permite uma visualização "física" do seu funcionamento.
- (ii) proporciona um elemento de unificação entre vários métodos de inferência estatística superficialmente diferentes como estimação do modelo de regressão linear por mínimos quadrados, estimação de variáveis de estado e teoria de filtragem, como será visto no capítulo IV.

### III.2.3.3 - Inicialização

Para inicializar o algoritmo estocástico, há duas maneiras possíveis:

- arbitrar um valor  $\hat{\theta}_0$  (por exemplo  $\hat{\theta}_0$  igual ao vetor nulo ( $N \times 1$ )) e fazer  $P_0^*$  uma matriz diagonal com valores bem grandes, indicando pouca credibilidade na estimativa inicial e nenhum conhecimento da covariância entre as componentes desta estimativa. Pode-se mostrar que, desta forma, os resultados fornecidos pelo algoritmo tendem aos da solução clássica obtidos resolvendo repetidamente a equação (3.5), para  $t > N$  (os resultados são assintoticamente equivalentes). Na realidade a convergência é bastante rápida, desde que os elementos da diagonal de  $P_0^*$  sejam suficientemente grandes.
- para um  $M > N$ , calcular  $P_M$  (e conseqüentemente  $P_M^*$ ) e  $\hat{\theta}_M$  por (3.6), (3.7) e (3.8). A seguir, para  $t > M$ , obter  $P_t^*$  e  $\hat{\theta}_t$  pelo algoritmo estocástico.



### III.3 - Abordagem Recursiva: Parâmetros Variáveis

A estimação por mínimos quadrados do modelo de regressão linear anteriormente efetuada considera implícita, tanto a clássica quanto a recursiva, a hipótese que os parâmetros permanecem constantes ao longo de todo o processo. Não se leva em conta que uma parcela de seu valor possa decorrer de fatores puramente aleatórios ou que eles possam evoluir no tempo segundo uma regra determinística independente dos valores assumidos pelas variáveis  $\{x_i, y_i\}$  ou que num determinado instante possam sofrer uma variação esperada. Quando isto ocorre a aplicação pura e simples dos conceitos anteriormente expostos torna-se perigosa na medida em que pode conduzir a erros sensíveis de estimação.

Dois tipos de algoritmos utilizados para superar este problema são expostos a seguir: o de ponderação exponencial e o linear dinâmico.

#### III.3.1 - Algoritmo de Ponderação Exponencial

Uma maneira de detetar variações de parâmetros é encurtar a memória do processo de estimação fazendo com que o efeito das observações mais remotas se reduza ante a presença das mais recentes. Isto pode ser feito ponderando-se os dados por uma função exponencial, o que na realidade representa o encurtamento do processo de estimação através de um filtro passa-baixo ("low-pass") discreto com função de ponderação exponencial. A função de critério a ser minimizada toma então a forma:

$$J = \sum_{i=1}^t (x_i \theta_i - y_i)^2 (1-\alpha)^{t-i} \alpha$$

Para esta função de critério (3.13) e (3.14) ficam:

$$(3.21) \quad P_t^{-1} = (1-\alpha)P_{t-1}^{-1} + \alpha (x_t' x_t)$$

$$(3.22) \quad b_t = (1-\alpha)b_{t-1} + \alpha (x_t' y_t)$$

representando uma operação de filtragem "passa-baixo" discreta de primeira ordem nos elementos  $x_t' x_t$  e  $x_t' y_t$  com o "comprimento" da memória especificado pelo fator escalar  $\alpha (0 < \alpha < 1)$ .

A partir de (3.21) e (3.22), repetindo o procedimento do item III.2.1, chega-se a (substituindo  $P_t$  por  $P_t^*/\sigma^2$ ):

$$(3.23) \quad P_t^* = \frac{1}{1-\alpha} P_{t-1}^* - \frac{\alpha}{(1-\alpha)^2} P_{t-1}^* x_t' (\sigma^2 + \frac{\alpha}{1-\alpha} x_t P_{t-1}^* x_t')^{-1} x_t P_{t-1}^*$$

$$(3.24) \quad \hat{\theta}_t = \hat{\theta}_{t-1} - \alpha \frac{P_t^*}{\sigma^2} (x_t' x_t \hat{\theta}_{t-1} - x_t' y_t)$$

O efeito físico do fator  $\alpha$  é impedir que os elementos da matriz de covariância  $P_t^*$  se tornem muito pequenos permitindo, com isto, que as novas observações continuem a influenciar as estimativas. Assim, qualquer modificação no gradiente causada por variação dos parâmetros pode ser detetada e utilizada para atualizar sua estimativa.

O algoritmo de ponderação exponencial possui duas principais desvantagens:

- os efeitos do ruído são detetados simultaneamente com as variações de parâmetros porventura existentes, modificando também o gradiente e conseqüentemente, a estimativa. Assim a estimação só é satisfatória se a ordem de grandeza da variação dos parâmetros for bastante superior a da flutuação aleatória devida ao ruído.
- não permite seletividade a priori dos parâmetros, isto é, todos os parâmetros são tratados igualmente. Assim caso se saiba a priori que somente alguns parâmetros devam variar, não existe uma maneira de incorporar esta informação ao algoritmo.

Na prática, Young [17] afirma que melhores resultados são obtidos trabalhando-se com a variável escalar  $\alpha_t$ , função estritamente decrescente do tamanho da amostra, assim definida:

$$\alpha_t = \frac{\alpha}{1 - (1-\alpha)^t}, 0 < \alpha < 0.1 \text{ e } t = 1, 2, \dots$$

Note-se que

$$\lim_{t \rightarrow \infty} \alpha_t = \alpha$$

### III.3.2 - Algoritmo Linear Dinâmico

As limitações do algoritmo de ponderação exponencial sugerem o desenvolvimento de um algoritmo mais flexível no qual as variações dos parâmetros não sejam detetadas simultaneamente com

o efeito do ruído e as informações a priori sobre os parâmetros possam ser incorporadas seletivamente. Por uma questão de coerência, deve se basear nos algoritmos recursivos anteriormente desenvolvidos uma vez que a recursão permite a atualização imediata da estimativa dos parâmetros à medida que novas observações vão sendo recebidas.

A interpretação estatística das equações de mínimos quadrados sugere um elegante método de modelagem das variações dos parâmetros em que os problemas supracitados desaparecem. Supõe-se que eles evoluem entre estágios subsequentes segundo uma regra determinística dinâmica inerente ao processo superposta a uma perturbação de caráter aleatório. Esta hipótese pode ser modelada pelo seguinte modelo linear dinâmico:

$$(3.25) \quad \theta_t = G_{t-1,t} \theta_{t-1} + \Gamma_{t-1,t} w_{t-1}$$

$$(3.26) \quad y_t = x_t \theta_t + v_t$$

$\theta_t, y_t, x_t, v_t$  - definidos anteriormente

$G_{t-1,t}$  - matriz de transição (N x N) conhecida

$\Gamma_{t-1,t}$  - matriz de entrada (N x M) conhecida

$w_{t-1}$  - vetor de perturbação dos parâmetros (M x 1)

As seguintes hipóteses são feitas acerca do vetor  $w_{t-1}$ : é um vetor de variáveis aleatórias com média nula, matriz de covariância W e serialmente independentes. Ele possibilita a variação estocástica dos parâmetros.

O conhecimento da forma de variação dos parâmetros entre estágios subsequentes (representada pelas matrizes  $G$  e  $\Gamma$  e pelas propriedades estatísticas do vetor  $w$  em (3.25)) proporciona uma informação adicional a priori que pode ser usada na estimação de  $\theta_t$ . A idéia é utilizá-lo para variar de maneira semelhante sua estimativa, antes da obtenção da nova amostra. Desta forma, no  $t$ -ésimo "instante" de estimação pode-se realizar atualizações a priori (previsões) a partir das estimativas  $\hat{\theta}_{t-1}$  e  $P_{t-1}^*$  obtidas no  $(t-1)$ -ésimo "instante". Representando por  $\hat{\theta}_{t/t-1}$  esta estimativa a priori de  $\theta_t$  baseada na estimativa  $\hat{\theta}_{t-1}$  e na lei de variação (3.25) e por  $P_{t/t-1}^*$  a matriz de covariância associada, tem-se (de agora em diante os índices são omitidos das matrizes  $G_{t-1,t}$  e  $\Gamma_{t-1,t}$  por simplificação):

$$(3.27) \quad \theta_{t/t-1} = G\hat{\theta}_{t-1}$$

$$(3.28) \quad P_{t/t-1}^* = GP_{t-1}^*G' + \Gamma W\Gamma'$$

Isto porque de (3.25) e (3.27) vem:

$$E[\theta_t] = E[G\theta_{t-1} + \Gamma w_{t-1}] = GE[\theta_{t-1}] + \Gamma E[w_{t-1}] = G\theta_{t-1}$$

$$P_{t/t-1}^* = E[(\hat{\theta}_{t/t-1} - \theta_t)(\hat{\theta}_{t/t-1} - \theta_t)']$$

Mas:

$$\hat{\theta}_{t/t-1} - \theta_t = G\hat{\theta}_{t-1} - G\theta_{t-1} - \Gamma w_{t-1} = G(\hat{\theta}_{t-1} - \theta_{t-1}) - \Gamma w_{t-1}$$

Logo:

$$\begin{aligned}
 P_{t/t-1}^* &= E \left\{ \left[ G(\hat{\theta}_{t-1} - \theta_{t-1}) - \Gamma w_{t-1} \right] \left[ G(\hat{\theta}_{t-1} - \theta_{t-1}) - \Gamma w_{t-1} \right]' \right\} = \\
 &= GE \left[ (\hat{\theta}_{t-1} - \theta_{t-1})(\hat{\theta}_{t-1} - \theta_{t-1})' \right] G' - GE \left[ (\hat{\theta}_{t-1} - \theta_{t-1}) w_{t-1}' \right] \Gamma' - \\
 &- \Gamma E \left[ w_{t-1} (\hat{\theta}_{t-1} - \theta_{t-1})' \right] G' + \Gamma E \left[ w_{t-1} w_{t-1}' \right] \Gamma'
 \end{aligned}$$

Como por (3.25)  $w_{t-1}$  e  $\theta_{t-1}$  são independentes, as 2a. e 3a. parcelas se anulam. Logo:

$$\begin{aligned}
 P_{t/t-1}^* &= GE \left[ (\hat{\theta}_{t-1} - \theta_{t-1})(\hat{\theta}_{t-1} - \theta_{t-1})' \right] G' + \Gamma E \left[ w_{t-1} w_{t-1}' \right] \Gamma' = \\
 &= GP_{t-1}^* G' + \Gamma \Gamma'
 \end{aligned}$$

(3.27) e (3.28) combinadas aos resultados obtidos para o caso de parâmetros constantes, permitem a formulação de um algoritmo para a estimação do parâmetro variável  $\theta_t$  composto de duas partes:

Parte I: Previsão (Estimação a priori)

$$\hat{\theta}_{t/t-1} = G \hat{\theta}_{t-1}$$

$$P_{t/t-1}^* = GP_{t-1}^* G' - \Gamma \Gamma'$$

Nesta parte é feita a previsão para o t-ésimo "instante" de estimação baseado no conhecimento das estimativas  $\hat{\theta}_{t-1}$  e  $P_{t-1}^*$  e da lei de variação dos parâmetros.

Parte II: Correção

$$(3.29) \quad P_t^* = P_{t/t-1}^* - P_{t/t-1}^* x_t' \left[ \sigma^2 + x_t P_{t/t-1}^* x_t' \right]^{-1} x_t P_{t/t-1}^*$$

$$(3.30) \quad \hat{\theta}_t = \hat{\theta}_{t/t-1} - P_{t/t-1}^* x_t' \left[ \sigma^2 + x_t P_{t/t-1}^* x_t' \right]^{-1} \left[ x_t \hat{\theta}_{t/t-1} - y_t \right]$$

Nesta parte o conhecimento da t-ésima amostra permite, à luz da nova informação, a correção das estimativas a priori  $\hat{\theta}_{t/t-1}$  e  $P_{t/t-1}^*$  para  $\hat{\theta}_t$  e  $P_t^*$ . Estas, por sua vez, servirão de base para estimativas a priori para o instante  $(t + 1)$  e assim sucessivamente.

Estes resultados podem ser obtidos através de uma abordagem Bayesiana:

se  $\theta_o \sim N(\theta_o, P_o^*)$  então (o índice superior indica que estão sendo considerados todas as observações até o "instante" correspondente ao valor do índice):

$$(3.31) \quad (\theta_t, y_t | y^{(t-1)}, x^t, W, G, \hat{\theta}_o, P_o^*) \sim N \left\{ \begin{bmatrix} \hat{\theta}_{t/t-1} \\ x_t \hat{\theta}_{t/t-1} \end{bmatrix}, \right.$$

$$\left. \begin{bmatrix} P_{t/t-1}^* & P_{t/t-1}^* x_t' \\ x_t P_{t/t-1}^* & \sigma^2 + x_t P_{t/t-1}^* x_t' \end{bmatrix} \right\}$$

$$(3.32) \quad (\theta_t | y^t, x^t, W, G, \hat{\theta}_0, P_0^*) \sim N(\hat{\theta}_t, P_t^*)$$

Estas distribuições (a priori e a posteriori) são obtidas pelo filtro de Kalman, a ser visto detalhadamente no capítulo IV.

Como no caso de parâmetros constantes,  $\hat{\theta}_0$  e  $P_0^*$  precisam ser escolhidos para se dar início à recursão. Da mesma forma,  $\theta_0$  pode ser o vetor nulo e  $P_0^*$  uma matriz diagonal com elementos bem grandes refletindo a ignorância a respeito do valor real  $\theta_0$ . Se  $t$  for relativamente grande, pode-se utilizar a informação dos dados para a escolha de  $P_0^*$ .

### III.3.3 - Casos Particulares do Modelo Linear Dinâmico

O algoritmo linear dinâmico é bastante geral, podendo ser aplicado para atualizar a estimativa de qualquer modelo ARIMA dos parâmetros, como será visto no item III.3.4. Na prática o modelo real raramente é conhecido, obrigando a hipóteses simplificadoras. Dois destes modelos particulares são importantes.

#### III.3.3.1 - O Passeio Aleatório Multidimensional

É definido pela seguinte lei de variação dos parâmetros.

$$\theta_t = \theta_{t-1} + w_t$$

Isto equivale a:

$$\theta_t = \theta_0 + \sum_{i=1}^t w_i$$



Portanto  $\theta_t$  é a soma de  $\theta_0$  com  $t$  variáveis aleatórias independentes de média nula, fazendo com que os parâmetros sofram flutuações aleatórias em torno do nível  $\theta_0$  com variâncias definidas pelos elementos diagonais da matriz  $W$ .

O algoritmo de previsão necessita somente da matriz  $W$  para funcionar:

$$(3.33) \quad \hat{\theta}_{t|t-1} = \hat{\theta}_{t-1}$$

$$(3.34) \quad P_{t|t-1}^* = P_{t-1}^* + W$$

Com isto, os elementos diagonais devem refletir a taxa de variação esperada dos parâmetros. Caso se saiba que alguns dos parâmetros devam permanecer constantes, basta anular os elementos correspondentes da diagonal de  $W$ .

O algoritmo apresenta bom desempenho para a estimação de parâmetros cuja variação é lenta.

### III.3.3.2 - O Modelo de Gauss-Markov de Primeira Ordem

É definido pela seguinte lei de variação dos parâmetros

$$\theta_t = G\theta_{t-1} + w_t$$

onde  $G$  é uma matriz diagonal em que os elementos  $g_i (i=\overline{1,N})$  são tais que  $|g_i| < 1$ . Eles funcionam como um coeficiente de correlação entre os  $i$ -ésimos elementos dos vetores  $\theta_t$  e  $\theta_{t-1}$ , fazendo com que o algoritmo apresente bom desempenho quando a variação dos parâmetros é lenta e correlacionada no tempo.

O algoritmo de previsão fica:

$$\hat{\theta}_{t|t-1} = G\hat{\theta}_{t-1}$$

$$P_{t|t-1}^* = GP_{t-1}^* G' + W$$

### III.3.3.3 - Consideração

Estes casos particulares são mais utilizados na prática porque a forma geral (3.25) requer muita informação a priori; entretanto ela pode ser aplicada em alguns sistemas físicos dinâmicos bem definidos e dos quais se conhece o funcionamento. Neste caso a determinação da matriz  $G$  fica facilitada como no caso, por exemplo, da estimação dos coeficientes de controle e estabilidade de um míssil. Logicamente, estes exemplos fogem ao escopo do trabalho, não sendo portanto abordados. Para tal ver [17].

### III.3.4 - Generalizações do Modelo Linear Dinâmico

#### III.3.4.1 - Modelo ARIMA dos Parâmetros

(3.25) especifica um processo autoregressivo como modelo para os parâmetros. Entretanto, a introdução de parâmetros adicionais auxiliares (pode-se considerá-los variáveis de estado) permite que se coloque nesta forma qualquer processo ARIMA para modelo dos parâmetros. Assim seja o processo ARIMA:

$$(3.35) \quad (I - \phi_1 B - \dots - \phi_p B^p) \theta_t = (I - \psi_1 B - \dots - \psi_q B^q) w_t$$

onde:

$\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q \rightarrow$  matrizes ( $N \times N$ ) conhecidas

$B \rightarrow$  operador retardo ("backshift"), isto é,  $B^i \theta_t = \theta_{t-i}$

Sejam:

$$k = \max \{p, q+1\}$$

$$\phi_j = 0 \text{ para } j > p$$

$$\psi_j = 0 \text{ para } j > q$$

$I_{(k-1)N}$  a matriz identidade de dimensão  $(k-1)N$

$$w_{j,t}^* = - \psi_{j-1} w_t, \text{ para } 2 \leq j \leq k$$

$\theta_{j,t}^*$  vetores de parâmetros auxiliares ( $N \times 1$ ) para  $2 \leq j \leq k$

$\underline{0}$  o vetor nulo ( $1 \times N$ )

O processo (3.35) pode ser escrito na forma:

$$\begin{bmatrix} \theta_t \\ * \\ \theta_{2,t} \\ \cdot \\ \cdot \\ \cdot \\ * \\ \theta_{k,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & & & & & & & & \\ & \phi_2 & & & & & & & \\ & & I_{(k-1)N} & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ \phi_k & & \tilde{0} & \dots & \tilde{0} & & & & \end{bmatrix} \begin{bmatrix} \theta_{t-1} \\ * \\ \theta_{2,t-1} \\ \cdot \\ \cdot \\ \cdot \\ * \\ \theta_{k,t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ * \\ w_{2,t} \\ \cdot \\ \cdot \\ \cdot \\ * \\ w_{k,t} \end{bmatrix}$$

(3.26) por sua vez toma a forma:

$$y_t = [x_t, \tilde{0}, \dots, \tilde{0}] \begin{bmatrix} \theta_t \\ * \\ \theta_{2,t} \\ \cdot \\ \cdot \\ \cdot \\ * \\ \theta_{k,t} \end{bmatrix} + \begin{bmatrix} w_t \\ * \\ w_{2,t} \\ \cdot \\ \cdot \\ \cdot \\ * \\ w_{k,t} \end{bmatrix}$$

#### III.3.4.2 - Variâncias Não Constantes

O modelo de regressão linear (3.25) e (3.26) supõe variâncias constantes do ruído e das perturbações dos parâmetros. Considerá-las função do "instante" de estimação, generaliza o modelo tornando-o muito mais representativo da realidade. Assim, fazendo  $\text{var}(v_t) = \sigma_t^2$  e  $w_t = 0$  chega-se a um modelo estático:

$$Y_t = x_t \theta_t + v_t \quad , \quad \text{var}(v_t) = \sigma_t^2$$

$$\theta_t = \theta_{t-1} = \dots = \theta_0$$

Ele não necessita de transformações estabilizadoras de variância e leva em conta diretamente as implicações de transformações na variância do ruído que objetivam uma forma conveniente do modelo linear.

Fazendo-se  $\text{var}(v_t) = \sigma_t^2$  e  $\text{var}(w_t) = W_t$  chega-se a um modelo dinâmico:

$$Y_t = x_t \theta_t + v_t \quad , \quad \text{var}(v_t) = \sigma_t^2$$

$$\theta_t = \theta_{t-1} + w_t \quad , \quad \text{var}(w_t) = W_t$$

Este modelo é importante em econometria onde normalmente se utiliza o modelo linear clássico estático ( $\sigma_t^2 = \sigma^2$  e  $W_t = 0$ ) no qual os parâmetros são estimados a partir dos dados históricos. Isto traz dificuldades porque eles sofrem frequentes alterações ao longo do tempo, devido ao dinamismo das condições sócio-econômicas. Surge então o problema: usando somente as últimas amostras o erro da estimativa é muito grande; usando todas, o erro é pequeno mas ela pode estar totalmente desatualizada. O modelo dinâmico oferece meios de acompanhar o "movimento" da relação entre as variáveis, pelo fato das variâncias não serem constantes. Isto proporciona grande flexibilidade a ser explorada pela habilidade do "modelador". Exemplificando,  $W_t$  pode ser

escolhida como uma função da informação no espaço das variáveis independentes: quando há uma mudança nunca antes verificada nas condições sócio-econômicas,  $W_t$  torna-se grande refletindo o aumento da incerteza quanto à situação resultante. Da mesma forma,  $W_t$  pode ser feita constante, refletindo uma queda constante com o tempo do valor das informações anteriores.

Logicamente, existe uma infinidade de maneiras para definir  $W_t$ . No método de previsão de Harrison e Stevens, a definição de  $W_t$  e  $V_t$  desempenha um papel de fundamental importância para sua flexibilidade e eficácia, conforme será visto no capítulo V.

### III.3.5 - Um Teste para a Invariância dos Parâmetros

Usando as  $t$  amostras disponíveis, pode-se desenvolver um teste de hipóteses para verificar se os parâmetros de um particular modelo são realmente variáveis ou constantes, isto é, se o modelo real é dinâmico ou estático. O modelo considerado para o desenvolvimento do teste é um passeio aleatório multidimensional em que a variância das perturbações dos parâmetros  $W$  é suposta linearmente relacionada à variância do ruído  $\sigma^2$ , isto é:

$$W = \sigma^2 \Omega$$

onde  $\Omega \rightarrow$  matriz (N x N)

Obviamente para este modelo  $\Omega \neq 0$  significa parâmetros

variáveis e  $\Omega \neq 0$ , parâmetros constantes. Pode-se então definir o teste.

$$H_0: \Omega \neq 0$$

$$H_1: \Omega = 0$$

De (3.31) tira-se que para  $G = I$  a distribuição a priori fica:

$$(3.36) \quad (y_t | y^{(t-1)}, x^t, \sigma^2, \Omega, \hat{\theta}_0, P_0^*) \sim N(x_t \hat{\theta}_t |_{t-1}; \sigma^2 + x_t P_t^* |_{t-1} x_t')$$

Como as sucessivas observações são independentes tem-se:

$$(3.37) \quad p(y_1, y_2, \dots, y_t | x^t, \sigma^2, \Omega, \hat{\theta}_0, P_0^*) = p(y_1 | x^1, \sigma^2, \Omega, \hat{\theta}_0, P_0^*) \\ \cdot \prod_{i=2}^t p(y_i | y^{(i-1)}, x^i, \sigma^2, \Omega, \hat{\theta}_0, P_0^*)$$

De (3.36) e (3.37) obtêm-se:

$$(3.38) \quad (y_1, y_2, \dots, y_t | x^t, \sigma^2, \Omega, \hat{\theta}_0, P_0^*) \propto \left[ \prod_{i=1}^t \pi(\sigma^2 + x_i P_i^* |_{i-1} x_i') \right]^{-\frac{1}{2}} \\ \exp\left\{ -\frac{1}{2} \frac{\sum_{i=1}^t (y_i - x_i P_i^* |_{i-1})^2}{\sigma^2 + x_i P_i^* |_{i-1} x_i'} \right\}$$

onde  $\hat{\theta}_0$  e  $P_0^*$  podem ser tratados como parâmetros conhecidos.

Mas de (3.34) tira-se:

$$P_{i|i-1}^* = \frac{P_{i|i-1}^*}{\sigma^2} = \frac{P_{i-1}^*}{\sigma^2} + \frac{W}{\sigma^2} = P_{i-1} + \Omega$$

Logo:

$$\sigma^2 + x_i P_{i|i-1}^* x_i' = \sigma^2 (1 + x_i P_{i|i-1} x_i') \triangleq \sigma^2 f_i$$

$$\text{onde } f_i = 1 + x_i P_{i|i-1} x_i'$$

Portanto (3.38) se transforma em:

$$(Y_1, Y_2, \dots, Y_t | x^t, \sigma^2, \Omega, \hat{\theta}_0, P_0^*) \propto \sigma^{-t} \left[ \prod_{i=1}^t f_i \right]^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^t \frac{(Y_i - x_i \hat{\theta}_{i|i-1})^2}{f_i}\right\}$$

O logaritmo da função de verossimilhança de  $\sigma^2$  e  $\Omega$  fica (tratando  $\hat{\theta}_0$  e  $P_0^*$  como parâmetros conhecidos):

$$(3.39) \quad \text{Log} [L(\sigma^2, \Omega | Y^t, x^t)] = -t \log \sigma - \frac{1}{2} \sum_{i=1}^t f_i - \frac{1}{2\sigma^2} \sum_{i=1}^t \frac{(Y_i - x_i \hat{\theta}_{i|i-1})^2}{f_i}$$

Seja  $\hat{\sigma}^2$  o estimador de máxima verossimilhança de  $\sigma^2$ . Derivando (3.39) em relação a  $\hat{\sigma}^2$  e igualando a zero vem:

$$-t \frac{1}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^t \frac{(Y_i - x_i \hat{\theta}_{i|i-1})^2}{f_i} = 0$$



$$(3.40) \quad \hat{\sigma}^2 = \frac{1}{t} \sum_{i=1}^t \frac{(y_i - x_i \hat{\theta}_{i|i-1})^2}{f_i}$$

Usando o valor  $\hat{\sigma}^2$  de (3.40) chega-se ao logaritmo da função de verossimilhança concentrada ou perfil de verossimilhança:

$$(3.41) \quad \text{Log} [L(\hat{\Omega} | y^t, x^t)] = t \log \hat{\sigma} - \frac{1}{2} \sum_{i=1}^t \log f_i$$

É difícil determinar analiticamente o estimador de máxima verossimilhança  $\hat{\Omega}$ , ponto de máximo de (3.41). Entretanto isto pode ser conseguido através de uma técnica qualquer de otimização não-linear.

Devido à simetria da matriz  $\Omega$ , necessita-se determinar intervalos de confiança para  $N(N+1)/2$  elementos que, segundo Ledolter [9], para um nível de confiança de aproximadamente  $(1 - \alpha)$  são dados por:

$$(3.42) \quad \text{Log} [L(\hat{\Omega} | y^t, x^t)] - \frac{1}{2} \chi_{\frac{N(N+1)}{2}, \alpha}^2 \leq \text{Log} [L(\Omega | y^t, x^t)] \leq$$

$$\text{Log} [L(\hat{\Omega} | y^t, x^t)] + \frac{1}{2} \chi_{\frac{N(N+1)}{2}, \alpha}^2$$

onde  $\chi_{k, \alpha}^2$  é a abscissa da distribuição qui-quadrado com  $k$  graus de liberdade que deixa  $\alpha$  da área à sua direita.

De posse de (3.42) pode-se testar  $H_0$  contra  $H_1$ .

## CAPÍTULO IV

### O FILTRO DE KALMAN

A formulação recursiva das equações de estimação por mínimos quadrados do modelo de regressão linear guarda íntima relação com uma estrutura de estimação recursiva mais geral: o filtro de Kalman. O problema de estimar um conjunto de parâmetros variando segundo uma forma conhecida a priori como (3.25) a partir da equação de observação (3.26) é análogo ao de estimar o vetor de estados de um sistema linear dinâmico estocástico discreto:

$$(4.1) \quad \theta_t = G\theta_{t-1} + \Gamma w_{t-1}$$

a partir de um vetor de observações linearmente relacionado ao vetor de estados através da seguinte equação de observação:

$$(4.2) \quad y_t = F\theta_t + v_t$$

onde:

t - índice de ordem

$y_t$  - vetor de observações (Mx1)

$\theta_t$  - vetor de estados (Nx1)

F - matriz de variáveis independentes (MxN)

G - matriz do sistema (NxN)

- $\Gamma$  - matriz de entrada ( $N \times N$ )  
 $v_t$  - vetor aleatório ( $M \times 1$ ) com média nula e matriz de covariância  $V$  representando o ruído da observação  
 $w_{t-1}$  - vetor aleatório ( $N \times 1$ ) com média nula e matriz de covariância  $W$  representando a perturbação do sistema.

Os vetores aleatórios, por hipótese, são independentes entre si e entre estágios subsequentes.

As formas (3.25) - (3.26) e (4.1) - (4.2) são matematicamente equivalentes, a menos dos fatos de  $x_t$  ser variável e  $F$  (seu correspondente) constante, da observação ser um vetor ( $M \times 1$ ) (o que transforma a variância dos ruídos escalar  $\sigma^2$  numa matriz  $V(M \times M)$ ) e das matrizes  $G$  e  $\Gamma$  serem constantes.

(4.1) e (4.2) podem ser representadas pelo seguinte diagrama de blocos:

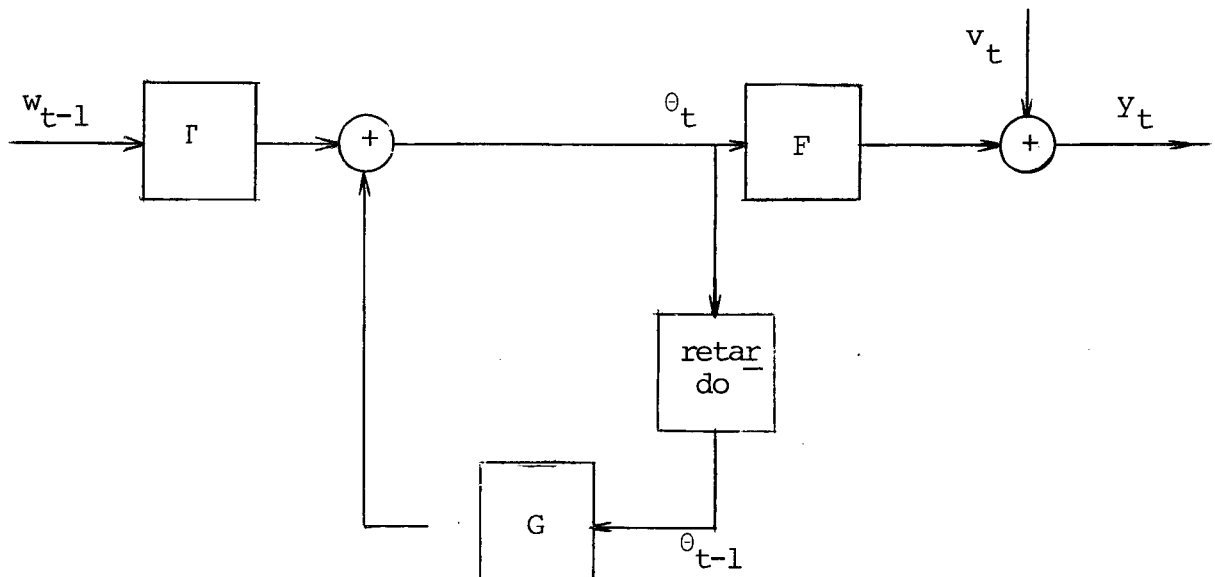


Figura 4.1

Torna-se patente a analogia entre estimação de parâmetro e de estado em modelos de sistemas dinâmicos, sugerindo uma tentativa de unificar estes dois conceitos. Fundamentalmente, neste tipo de modelos não existe diferença entre eles e a separação com que são normalmente tratados deve-se mais a razões históricas e a uma causa física: denomina-se parâmetro quando não há variação ou há uma variação lenta e estado quando a variação é relativamente rápida.

Nestes termos o filtro de Kalman, algoritmo de estimação do vetor  $\theta_t$ , assume particular importância neste trabalho. Suas equações se prestam não só à estimação recursiva de mínimos quadrados como permitem a formulação de métodos recursivos de análise e previsão de séries temporais. Torna-se, portanto, interessante mostrar como elas podem ser obtidas para uma melhor compreensão destes métodos e para justificar o algoritmo linear dinâmico de III.3.2. Das duas soluções mais conhecidas, a de Kalman [8] baseada em projeções ortogonais e a de Ho e Lee [7] baseada em inferência Bayesiana, optou-se pela segunda devido à sua maior relação com o método Bayesiano de previsão de Harrison e Stevens do Capítulo V.

Para situar o problema enuncia-se o conceito de estado em sistemas determinísticos e estocásticos, define-se o problema geral de estimação de estados (amortecimento, filtragem e previsão), introduz-se a versão discreta do filtro de Kalman e finalmente, deriva-se suas equações.

## IV.1 - O Conceito de Estado em Sistemas Determinísticos

Intuitivamente, o estado de um sistema determinístico pode ser considerado como a quantidade mínima de informação acerca de seu passado, suficiente para prever o efeito do passado no comportamento futuro do sistema. Chamando  $\theta_{t-1}$  o vetor representativo do estado no instante  $(t-1)$ , decorre da definição que o estado no instante  $t$  deve ser uma função somente de  $\theta_{t-1}$  e da entrada  $\mu_{t-1}$  no instante  $(t-1)$ , ou seja:

$$(4.3) \quad \theta_t = f_1(\theta_{t-1}) + f_2(\mu_{t-1})$$

De um modo geral, pode-se estabelecer um sistema de equações diferenciais (ou de diferenças) descrevendo o comportamento dinâmico de um processo físico para então colocá-lo na forma (4.3).

As variáveis de estado podem ter um significado físico. Entretanto como no caso de equações lineares de estado (caso deste trabalho) o sistema de entrada e saída não é afetado por transformações lineares não-singulares das variáveis de estado, existe uma infinidade de representações de variáveis de estado para um mesmo sistema de entrada e saída e, geralmente, a forma mais desejável de se trabalhar não é a forma original. Assim, ao contrário das entradas e das saídas que tem um significado físico concreto, o estado deve ser olhado, em princípio, como uma quantidade abstrata (no método Bayesiano de Harrison e Stevens, entretanto, ele possui significado físico).

A equação (4.3) pode geralmente ser linearizada para a forma:

$$(4.4) \quad \theta_t = G\theta_{t-1} + \Gamma\mu_{t-1}$$

onde:

$\theta_t$  - vetor de estados (Nx1)

$\mu_t$  - vetor de variáveis de entrada (Rx1)

G e  $\Gamma$  - matrizes (NxN) e (NxR)

A saída observada do sistema pode ser expressa como combinação linear das variáveis de estado:

$$(4.5) \quad y_t = F\theta_t$$

onde:

$y_t$  - vetor de saída (Mx1)

F - matriz (MxN)

Na forma (4.4), qualquer componente do vetor de estados pode ser escrito como uma equação de diferenças de N-ésima ordem e vice-versa, fazendo com que seja sempre possível escrever qualquer modelo linear dinâmico nesta forma.

Note-se que somente as entradas  $\mu$  e as saídas  $y$  são quantidades conhecidas diretamente. As variáveis de estado servem simplesmente para definir o comportamento interno do sistema. Existem meios de se obter informações sobre as variáveis de estado através das observações  $y$  ou de manipulá-las indiretamente pelo ajuste da entrada  $\mu$ . A manipulação envolve os conceitos de controlabilidade e observabilidade e foge aos objetivos do trabalho.

#### IV.2 - O Conceito de Estado em Sistemas Estocásticos

Num sistema estocástico existe a necessidade de se modificar o conceito de estado porque nele o comportamento futuro não pode ser determinado unicamente pelo estado presente. Uma extensão natural, então, é a exigência de que a distribuição de probabilidade futura do vetor de estados seja unicamente determinada pelo seu verdadeiro valor presente. Isto implica em que ele seja descrito por um processo Markoviano multivariado da forma (4.1).

Como no caso determinístico, a saída pode ser expressa como combinação linear das variáveis de estado e, portanto, por (4.5). Entretanto, a forma (4.1) e (4.5) não é a forma padrão do modelo linear estocástico de estados. Desde os trabalhos de Wiener tornou-se prática comum representar a observação do processo  $y_t$  como a soma de dois processos estocásticos: um relacionado ao estado  $\theta_t$  e um ruído indesejado  $v_t$ , originário de

erros de medida e erros estocásticos. Com isto, chega-se à forma (4.1) e (4.2) para representar o modelo linear estocástico de estados.

Supondo conhecidas as matrizes do modelo, o filtro de Kalman proporciona a estimativa ótima do vetor de estados  $\theta_t$  a partir do conjunto de observações  $y^t$ . Antes de derivá-lo, convém enunciar de uma maneira geral o problema de estimação de estados num sistema estocástico.

#### IV.3 - O Problema de Estimação de Estados

Seja o diagrama de bloco geral do sistema estocástico (4.1) e (4.2):

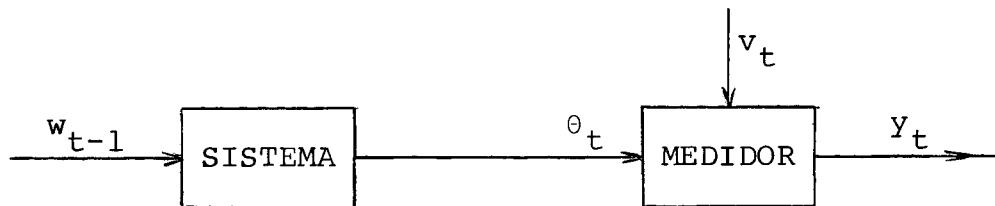


Figura 4.2

O problema geral de estimação de estados pode ser enunciado: conhecidas as relações físicas entre  $\theta_t$  e  $w_{t-1}$ , entre  $y_t$ ,  $\theta_t$  e  $v_t$ , as propriedades estatísticas de  $v_t$  e  $w_{t-1}$  e as observações  $y_1, y_2, \dots, y_t$ , determinar  $\hat{\theta}(\tau|t)$  a melhor estimativa de  $\theta_\tau$  segundo algum critério desejado.

Dependendo do valor de  $\tau$ , três tipos de problemas surgem:



- $\tau < t$  - problema de amortecimento
- $\tau = t$  - problema de filtragem
- $\tau > t$  - problema de previsão

De um modo geral, uma vez especificado o critério para a melhor estimativa, estes problemas são matematicamente bem definidos.

#### IV.4 - Filtro de Kalman Discreto: Solução Bayesiana

A versão discreta do filtro de Kalman foi inicialmente apresentada por Kalman em 1960. Em 1961, Battin derivou independentemente resultados semelhantes, onde conseguiu estabelecer o estimador linear ótimo tanto para caso de erros correlacionados, quanto para o de erros não correlacionados. Ele supôs para solução uma relação recursiva onde a nova estimativa é a soma de uma estimativa extrapolada, baseada em estimativas passadas, com uma matriz de "ganho" modificando os resíduos. Deste modo, o problema se resume na determinação da matriz de ganho ótima que minimiza o erro médio quadrático das estimativas.

O enfoque aqui utilizado enuncia o problema como um problema geral de estimação estocástica de múltiplos estágios, apresenta sua solução Bayesiana e depois a particulariza para o filtro de Kalman.

A idéia geral da solução Bayesiana é, conhecidas to-

das as observações, determinar a função densidade de probabilidade a posteriori do vetor de estados, já que ela contém todas as informações estatísticas de interesse. Com isto, o problema estocástico de estimação se transforma numa série de relações determinísticas que se pode visualizar e manipular.

#### IV.4.1 - Caso Geral: Estimação Estocástica de Múltiplos Estágios

Parte-se do pressuposto que no estágio de ordem  $t$  (em termos de parâmetros, no  $t$ -ésimo "instante" amostral) são conhecidos:

(i) - a equação governando a evolução do estado e a equação relacionando-o à observação

$$(4.6) \quad \theta_t = f(\theta_{t-1}, w_{t-1})$$

$$(4.7) \quad y_t = h(\theta_t, v_t) = g(\theta_{t-1}, w_{t-1}, v_t)$$

onde

$t$  - índice de ordem

$\theta_t$  - vetor de estados

$y_t$  - vetor de observações

$w_{t-1}$  - vetor de perturbações

$v_t$  - vetor de ruídos

(ii)- o conjunto de todos os vetores de observações

$$Y^t \triangleq \{Y_1, Y_2, \dots, Y_t\}$$

(iii)- as funções densidade de probabilidade

$$p(\theta_{t-1} | Y^{(t-1)})$$

$$p(w_{t-1}, v_t | \theta_{t-1})$$

De posse das informações (i), (ii), (iii), deseja-se estimar  $\theta_t$ , o vetor de estados no estágio  $t$ .

A solução deste problema pode ser desenvolvida nos seguintes passos:

(i)- obter  $p(\theta_t | \theta_{t-1})$ .

Isto pode ser feito a partir do conhecimento de  $p(w_{t-1}, v_t | \theta_{t-1})$ ,  $p(\theta_{t-1} | Y^{(t-1)})$  e de (4.6).

(ii)- obter  $p(Y_t | \theta_{t-1}, \theta_t)$

Isto pode ser feito a partir de  $p(w_{t-1}, v_t | \theta_{t-1})$ , (4.6) e (4.7).

$$(iii) - \text{ obter } p(\theta_t, y_t | Y^{(t-1)}) = \int p(y_t | Y^{(t-1)}, \theta_t) p(\theta_t | \theta_{t-1}) \cdot \\ \cdot p(\theta_{t-1} | Y^{(t-1)}) d\theta_{t-1}$$

A partir disto, as funções densidade marginais  $p(\theta_t | Y^{(t-1)})$  e  $p(y_t | Y^{(t-1)})$  podem ser obtidas diretamente.

$$(iv) - \text{ obter } p(\theta_t | Y^t) = \frac{p(\theta_t, y_t | Y^{(t-1)})}{p(y_t | Y^{(t-1)})} = \\ = \frac{\int p(y_t | Y^{(t-1)}, \theta_t) p(\theta_t | \theta_{t-1}) p(\theta_{t-1} | Y^{(t-1)}) d\theta_{t-1}}{\iint p(y_t | Y^{(t-1)}, \theta_t) p(\theta_t | \theta_{t-1}) p(\theta_{t-1} | Y^{(t-1)}) d\theta_t d\theta_{t-1}}$$

Esta equação representa a função densidade de probabilidade a posteriori do estado  $\theta_t$ , isto é, o conhecimento da sua natureza após o conjunto de todas observações  $Y^t$ . Por definição, ela contém todas as informações estatísticas necessárias à estimação.

(v) - obter, a partir de  $p(\theta_t | Y^t)$ , estimativas de  $\theta_t$ . Dependendo do objetivo desejado, várias funções de critério para a estimação podem ser definidas, originando várias estimativas, em algum sentido, ótimas. Exemplos típicos destas estimativas são os seguintes:

a) critério: maximizar a probabilidade de  $\theta_t$ .

solução:  $\hat{\theta}_t = \text{moda de } p(\theta_t | Y^t)$ , ou seja, a estimativa mais provável.

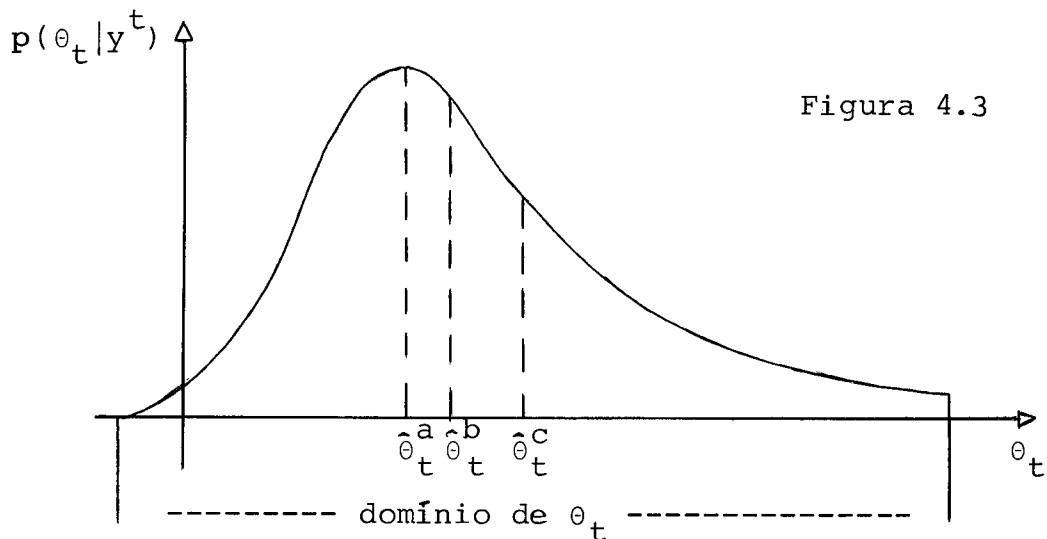
b) critério: minimizar  $\int_{-\infty}^{\infty} \|\theta_t - \hat{\theta}_t\|^2 p(\theta_t | y^t) d\theta$

solução:  $\hat{\theta}_t = E [\theta_t | y^t]$ , ou seja, a média condicional. Esta estimativa é também chamada de igual risco, porque a probabilidade de  $\theta_t$  ser maior ou menor que  $\hat{\theta}_t$  é a mesma.

c) critério: minimizar o máximo de  $|\theta_t - \hat{\theta}_t|$

solução:  $\hat{\theta}_t =$  mediana de  $p(\theta_t | y^t)$ , ou seja, a estimativa que minimiza o maior erro possível.

Estas três estimativas são representadas na figura seguinte, para o caso de variáveis escalares com  $p(\theta_t | y^t)$  genérica.



Logicamente, definindo-se outros critérios, novas estimativas podem ser obtidas. Além disso, pode-se determinar, a partir de  $p(\theta_t | y^t)$ , intervalos de confiança para cada caso. Deve-se notar que se  $p(\theta_t | y^t)$  for normal, as três estimativas ante

riormente definidas coincidem.

#### IV.4.2 - Caso Particular: O Filtro de Kalman

O filtro de Kalman é um caso de estimação particular no qual as relações funcionais (4.6) e (4.7) assumem a forma (4.1) e (4.2), reescritas a seguir:

$$\theta_t = G\theta_{t-1} + \Gamma w_{t-1}$$

$$y_t = F\theta_t + v_t$$

com as seguintes características:

(i)-  $v_t$  e  $w_{t-1}$  são ruídos brancos normais independentes com média nula e matrizes de covariância  $V$  e  $W$ , respectivamente.

(ii)-  $p(\theta_{t-1} | Y^{(t-1)}) \sim N(\hat{\theta}_{t-1}, P_{t-1}^*)$

(iii)-  $p(w_{t-1}, v_t | \theta_{t-1}, Y^{(t-1)}) = p(w_{t-1}, v_t) = p(w_{t-1}) p(v_t)$

Como  $w_{t-1}$  e  $v_t$  independem do estado, tem-se:

$$p(\theta_t, y_t | Y^{(t-1)}) = p(y_t | \theta_t) p(\theta_t | Y^{(t-1)})$$

Em consequência, a distribuição a posteriori  $p(\theta_t | Y^t)$  se transforma em:

$$(4.8) \quad p(\theta_t | Y^{(t)}) = \frac{p(y_t | \theta_t) p(\theta_t | Y^{(t-1)})}{p(y_t | Y^{(t-1)})}$$

Portanto, a solução do problema requer a determinação das três funções densidade de probabilidade à direita de (4.8). Assim:

$$(i) - p(\theta_t | Y^{(t-1)}) \sim N(G\hat{\theta}_{t-1}, P^*_{t|t-1}), \text{ onde:}$$

$$P^*_{t|t-1} \triangleq GP^*_{t-1}G' + \Gamma W \Gamma'$$

Isto porque:

$$(\theta_t | Y^{(t-1)}) = (G\theta_{t-1} + \Gamma w_{t-1} | Y^{(t-1)})$$

Como  $w_{t-1}$  independe de  $Y^{(t-1)}$ :

$$(\theta_t | Y^{(t-1)}) = (G\theta_{t-1} | Y^{(t-1)}) + \Gamma w_{t-1}$$

Logo:

$$E \left[ \theta_t | Y^{(t-1)} \right] = E \left[ G\theta_{t-1} | Y^{(t-1)} + \Gamma w_{t-1} \right] = GE \left[ \theta_{t-1} | Y^{(t-1)} \right] + \Gamma E \left[ w_{t-1} \right] = G\hat{\theta}_{t-1}$$

$$\text{Cov} \left[ \theta_t | Y^{(t-1)} \right] = \text{Cov} \left[ G\theta_{t-1} | Y^{(t-1)} \right] + \text{Cov} \left[ \Gamma w_{t-1} \right] = GP^*_{t-1}G' + \Gamma W \Gamma'$$

Finalmente,  $p(\theta_t | Y^{(t-1)})$  é normal por ser combinação linear de duas normais,  $p(\theta_{t-1} | Y^{(t-1)})$  e  $p(w_{t-1})$ .

$$(ii) - p(y_t | Y^{(t-1)}) \sim N(FG\hat{\theta}_{t-1}, FP^*_{t|t-1}F' + V)$$

Isto porque:

$$(y_t | Y^{(t-1)}) = (F\theta_t + v_t | Y^{(t-1)})$$

Como  $v_t$  independe de  $Y^{(t-1)}$ :

$$(y_t | Y^{(t-1)}) = (F\theta_t | Y^{(t-1)}) + v_t$$

Logo:

$$E [y_t | Y^{(t-1)}] = FE [\theta_t | Y^{(t-1)}] + E [v_t] = FG\hat{\theta}_{t-1}$$

$$\text{Cov} [y_t | Y^{(t-1)}] = F \text{Cov} [\theta_t | Y^{(t-1)}] F' + \text{Cov} [v_t] = FP^*_{t|t-1}F' + V$$

Finalmente,  $p(y_t | Y^{(t-1)})$  é normal por ser combinação linear de duas normais,  $p(\theta_t | Y^{(t-1)})$  e  $p(v_t)$ .

$$(iii) - p(y_t | \theta_t) \sim N(F\theta_t, V)$$

Isto porque:

$$(y_t | \theta_t) = (F\theta_t + v_t | \theta_t)$$



Como  $v_t$  independe de  $\theta_t$ :

$$(y_t | \theta_t) = (F\theta_t | \theta_t + v_t)$$

Logo:

$$E [y_t | \theta_t] = FE [\theta_t | \theta_t] + E [v_t] = F\theta_t$$

$$\text{Cov} [y_t | \theta_t] = F \text{Cov} [\theta_t | \theta_t] F' + \text{Cov} [v_t] = V$$

Finalmente,  $p(y_t | \theta_t)$  é normal porque  $p(v_t)$  é normal.

Substituindo as distribuições (i), (ii) e (iii) em (4.8), obtêm-se:

$$p(\theta_t | y^t) = \frac{1}{(2\pi)^{\frac{M}{2}} |V|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[ (y_t - F\hat{\theta}_t)' V^{-1} (y_t - F\hat{\theta}_t) \right] \right\} \cdot$$

$$\frac{1}{(2\pi)^{\frac{N}{2}} |P_{t|t-1}^*|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[ (\theta_t - G\hat{\theta}_{t-1})' P_{t|t-1}^{*-1} (\theta_t - G\hat{\theta}_{t-1}) \right] \right\} /$$

$$/ \frac{1}{(2\pi)^{\frac{M}{2}} |FP_{t|t-1}^* F' + V|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[ (y_t - FG\hat{\theta}_{t-1})' \cdot$$

$$\cdot (FP_{t|t-1}^* F' + V)^{-1} (y_t - FG\hat{\theta}_{t-1}) \right] \right\}$$

Logo:

$$\begin{aligned}
 (4.9) \quad p(\theta_t | y^t) &= \frac{|FP_t^*|_{t-1} F' + V|^{\frac{1}{2}}}{(2\pi)^{\frac{N}{2}} |V|^{\frac{1}{2}} |P_t^*|_{t-1}|^{\frac{1}{2}}} \cdot \\
 &\cdot \exp \left\{ -\frac{1}{2} \left[ (y_t - F\theta_t)' V^{-1} (y_t - F\theta_t) + \right. \right. \\
 &\quad + (\theta_t - G\hat{\theta}_{t-1})' P_t^*|_{t-1} (\theta_t - G\hat{\theta}_{t-1}) - \\
 &\quad - (y_t - FG\hat{\theta}_{t-1})' (FP_t^*|_{t-1} F' + V)^{-1} \cdot \\
 &\quad \left. \left. \cdot (y_t - FG\hat{\theta}_{t-1}) \right] \right\}
 \end{aligned}$$

Completando os quadrados em (4.9), chega-se a:

$$\begin{aligned}
 (4.10) \quad p(\hat{\theta}_t | y^t) &= \frac{|FP_t^*|_{t-1} F' + V|^{\frac{1}{2}}}{(2\pi)^{\frac{N}{2}} |V|^{\frac{1}{2}} |P_t^*|_{t-1}|^{\frac{1}{2}}} \cdot \\
 &\cdot \exp \left\{ -\frac{1}{2} (\theta_t - \hat{\theta}_t)' (P_t^*)^{-1} (\theta_t - \hat{\theta}_t) \right\}
 \end{aligned}$$

onde:

$$(4.11) \quad \hat{\theta}_t = G\hat{\theta}_{t-1} - P_t^*|_{t-1} F' (FP_t^*|_{t-1} F' + V)^{-1} (FG\hat{\theta}_{t-1} - y_t)$$

$$(4.12) \quad P_t^* = P_t^*|_{t-1} - P_t^*|_{t-1} F' (FP_t^*|_{t-1} F' + V)^{-1} F_t P_t^*|_{t-1}$$

Portanto, a distribuição a posteriori do vetor de parâmetros no instante  $t$  é normal com média  $\hat{\theta}_t$  e variância  $P_t^*$ , isto é:

$$(\theta_t | y^t) \sim N(\hat{\theta}_t, P_t^*)$$

(4.11) e (4.12) são matematicamente equivalentes a (3.29) e (3.30), a menos das diferenças apontadas na introdução do capítulo. Assim, estão demonstradas as equações recursivas do algoritmo linear dinâmico do item III.3.2.

Além disto, este resultado proporciona um subsídio teórico relevante ao desenvolvimento de métodos recursivos de análise e previsão de séries temporais.

## CAPÍTULO V

### ANÁLISE RECURSIVA DE MODELOS LINEARES

#### DE SÉRIES TEMPORAIS

Após a formulação recursiva das equações de mínimos quadrados do modelo de regressão linear e sua relação com o filtro de Kalman, resta a aplicação dos conceitos de estimação recursiva ao estudo de séries temporais. Os métodos recursivos para a estimação dos parâmetros de uma série temporal podem ser utilizados tanto nos modelos convencionais tipo "caixa-preta", que se preocupam basicamente com as relações entre as entradas e as saídas do sistema, quanto nos modelos tipo "espaço de estados", mais voltados à descrição interna do processo.

Para os modelos convencionais apresenta-se um método baseado em modificações por variáveis instrumentais e máxima verossimilhança aproximada (VI-MVA) das equações recursivas de mínimos quadrados que em relação aos tradicionais, embora não seja estatisticamente eficiente, possui a vantagem de produzir comparativamente estimativas de pequena variância com muito menor esforço computacional, além do potencial para aplicações em tempo real.

Para os modelos tipo "espaço de estados" apresenta-se um método baseado em inferência Bayesiana e desenvolvido por Harrison e Stevens, cujas principais características são as seguintes:

- capacidade do analista prover ao sistema, a cada instante, informações a priori, interna e externamente, conforme o exposto no Capítulo I. Isto possibilita uma intervenção humana baseada na teoria, estabelecendo a união da teoria com a prática.
- capacidade de detetar e se adaptar rapidamente, de maneira automática ou sob "intervenção" do analista, a situações "anormais" como mudanças bruscas de nível e inclinação, através de uma modelagem explícita destas situações. Os métodos não Bayesianos geralmente demoram a perceber estas mudanças, se adaptando de modo mais lento e produzindo erros sensíveis de previsão durante um tempo relativamente longo. Isto pode ser profundamente prejudicial à tomada de decisões e ao planejamento da operação do sistema em estudo.
- aplicabilidade quando se dispõe de um número pequeno de observações ou até mesmo no início do processo, quando ainda não se dispõe de observações (neste caso, baseada somente na expressão quantitativa da expectativa do analista). A importância deste fato ressalta quando se verifica ser comum a ocorrência de alterações na série histórica que tornam não significativos os dados anteriores a um determinado instante, fazendo com que somente uma pequena quantidade deles passe a conter informações relevantes. Um fato desta natureza pode impedir a aplicação, por exemplo, do método Box & Jenkins que geralmente só apresenta bom desempenho quando se dispõe pelo menos de cerca de 30 observações.

- modelagem da série por um modelo linear dinâmico com a estimativa dos parâmetros atualizada pelo filtro de Kalman, estabelecendo, implicitamente, a analogia parâmetro-estado. Este modelo inclui como casos particulares vários modelos convencionais como regressão linear, amortecimento exponencial e modelos ARIMA Box & Jenkins. Estes necessitam ser linearizados, utilizando-se na estimativa de seus parâmetros o filtro de Kalman Extendido.

Quase a totalidade do capítulo se constitui na apresentação deste método, sendo a parte final reservada ao método VI-MVA.

#### V.1 - Modelos Espaço de Estados: O Método Bayesiano de Previsão a Curto Prazo

##### V.1.1 - Características

###### (i) - Representação Paramétrica (ou de Variáveis de Estado)

Utiliza uma representação paramétrica do processo (em vez de uma representação funcional) que além de simplificar, possibilita a atribuição de um sentido "físico" (como nível, inclinação, etc.) a cada parâmetro, facilitando a visualização do mecanismo evolutivo do processo. Esta representação é a responsável pela possibilidade de comunicação sistema - analista em ambos os sentidos: o analista pode transmitir, através de informações sobre os parâmetros, sua opinião acerca da influência de fatos externos e o sistema pode responder, em termos de incerteza,

suas possíveis consequências. Assim, o analista pode entender e avaliar inferências sobre os parâmetros.

(ii) - Descrição Probabilística dos Parâmetros

O vetor de parâmetros é suposto um vetor aleatório multidimensional desconhecido cuja distribuição conjunta se deseja inferir para cada instante, a partir das observações até aquele instante. Esta informação é extrapolada para inferir os futuros valores da série. Deve-se ressaltar que a incerteza acerca do futuro provem não somente da grande aleatoriedade dos fatores que o determinam, mas também da incerteza acerca dos valores presentes dos parâmetros. A isto se alia a capacidade do modelo estabelecer corretamente a ligação do passado com o futuro.

No caso da observação ser um vetor, as distribuições conjuntas oferecem um meio simples de previsão sujeita a restrições e de uso dos dados de um produto para obtenção de informações acerca de produtos correlatos.

(iii) - Descrição Sequencial

Especifica a evolução do processo entre dois instantes subsequentes  $(t-1, t)$  descrevendo sua evolução sistemática e aleatória. Isto permite a inferência do vetor de parâmetros no instante  $t$  a partir:

- da distribuição a posteriori dos parâmetros no instante  $(t-1)$
- das distribuições do ruído da observação e da perturbação dos parâmetros no instante  $t$
- da observação no instante  $t$

Um aspecto importante do método é não haver exigência que o modelo representativo da transição  $(t-1, t)$  seja o mesmo para qualquer instante  $t$ . Isto proporciona grande flexibilidade já que diferentes modelos podem ser apropriados a determinadas transições (por exemplo, quando grandes alterações no comportamento da série são esperadas como decorrência de fatos externos).

#### (iv)- Modelos de Múltiplos Processos

Vários métodos estatísticos foram desenvolvidos para manipular a incerteza dos parâmetros de um determinado modelo, mas o problema da incerteza quanto ao próprio modelo não recebeu o mesmo tratamento. Pouca atenção tem sido dada ao problema da dúvida sobre que modelo está em curso determinado instante. Em séries temporais, além deste dúvida, é bastante lógico supor que diferentes modelos estejam em curso em diferentes instantes.



Existem dois tipos principais de incerteza quanto a modelos:

- entre um conjunto discreto de alternativas, um único modelo do processo está em curso em todos os instantes, não sendo conhecido qual (item V.1.7.1).
- o modelo do processo num determinado instante é uma combinação aleatória entre um número discreto de alternativas. Este tipo de abordagem possibilita a modelagem explícita das mudanças bruscas e transições da série (item V.1.7.2).

#### V.1.2 - Modelagem do Processo: O Modelo Linear Dinâmico

O processo é modelado por um modelo linear dinâmico (MLD) que especifica:

- a relação de dependência entre as observações e os parâmetros do processo.
- a evolução dos parâmetros no tempo, devido a uma regra sistemática inerente ao processo e a "choques" aleatórios.

O MLD utilizado é o seguinte:

$$(5.1) \quad Y_t = F_t \theta_t + v_t, \quad v_t \sim N(0, V_t) \text{ - equação de observação}$$

$$(5.2) \quad \theta_t = G\theta_{t-1} + w_t \quad , \quad w_t \sim N(0, W_t) \quad - \text{equação do sistema}$$

onde:

- $t$  - índice de tempo ( $t=1,2,3,\dots$ )
- $Y_t$  - vetor de observação ( $M \times 1$ ) no instante  $t$
- $F_t$  - matriz de variáveis independentes ( $M \times N$ ) conhecida no instante  $t$
- $\theta_t$  - vetor de parâmetros ( $N \times 1$ ) no instante  $t$
- $G$  - matriz do sistema ( $N \times N$ ) conhecida
- $v_t$  - vetor aleatório de ruído ( $M \times 1$ ) com distribuição normal de média nula e matriz de covariância  $V_t$  conhecida no instante  $t$
- $w_t$  - vetor aleatório de perturbação dos parâmetros ( $N \times 1$ ) com distribuição normal de média nula e matriz de covariância  $W_t$  conhecida no instante  $t$ .

A equação de observação especifica a dependência entre a observação e os parâmetros e, por hipótese, a distribuição de  $y_t$ , enquanto a equação do sistema especifica a evolução determinística (matriz  $G$ ) e a aleatória (vetor  $w_t$ ) dos parâmetros

A forma (5.1) e (5.2) possibilita duas interpretações principais para os parâmetros do processo:

- i) considerando  $F_t$  variável, expressam uma relação entre a variável dependente  $y_t$  e as variáveis independentes  $F_t$ , generalizando a interpretação paramétrica do modelo linear estático.

ii) considerando  $F_t$  constante, o MLD pode ser encarado como uma representação em espaço de estados de uma série temporal com os parâmetros admitindo interpretações físicas como nível e inclinação do processo. Neste sentido, estabelece-se uma ligação com a Teoria de Controles através da correspondência parâmetro-estado mencionada na introdução dos Capítulos IV e V.

Em relação ao MLD mais dois aspectos devem ser ressaltados:

- i) teoricamente nada impede que ele seja generalizado admitindo-se, por exemplo,  $G$  variável e distribuição conjunta para  $v_t$  e  $w_t$  com médias não nulas. Na prática, entretanto, estas modificações o tornam desnecessariamente complexo, uma vez que o modelo apresentado é suficiente para a grande maioria das aplicações.
- ii) ele pode representar combinações lineares de modelos lineares e modelos hierárquicos.

### V.1.3 - Estimação dos Parâmetros: O Filtro de Kalman

A forma (5.1) e (5.2) de representação do processo é análoga à forma (4.1) e (4.2) que representa a observação de um sistema linear dinâmico estocástico. Na realidade, considerando  $w_t = \Gamma w_{t-1}$  e a menos da hipótese das matrizes de covariância  $V_t$  e  $W_t$  e da matriz  $F_t$  variáveis no tempo, as duas formas

são matematicamente iguais. Isto sugere que se utilize na estimação dos parâmetros as equações de recorrência do filtro de Kalman (4.11) e (4.12) com estas alterações. Assim, partindo da suposição que no instante inicial  $t = 0$  o vetor de parâmetros  $\theta_0$  tem distribuição normal com média  $\hat{\theta}_0$  e matriz de covariância associada  $P_0^*$ , i é:

$$\theta_0 \sim N(\hat{\theta}_0, P_0^*)$$

e que o processo  $y_t$  é descrito pelo sistema (5.1) e (5.2), a distribuição a posteriori dos parâmetros no instante  $t$  é da forma:

$$(\theta_t | Y^t, F^t) \sim N(\hat{\theta}_t, P_t^*)$$

com a média  $\hat{\theta}_t$  e a matriz de covariância  $P_t^*$  sendo obtidas pelas relações:

$$(5.3) \quad \hat{\theta}_t = \hat{\theta}_{t|t-1} - P_{t|t-1}^* F_t' [V_t + F_t P_{t|t-1}^* F_t']^{-1} [F_t \hat{\theta}_{t|t-1} - Y_t]$$

$$(5.4) \quad P_t^* = P_{t|t-1}^* - P_{t|t-1}^* F_t' [V_t + F_t P_{t|t-1}^* F_t']^{-1} F_t P_{t|t-1}^*$$

onde:

$$(5.5) \quad \hat{\theta}_{t|t-1} = G \hat{\theta}_{t-1}$$

$$(5.6) \quad P_{t|t-1}^* = G P_{t-1}^* G' + W_t$$

(5.5) e (5.6) representam estimativas a priori da média e covariância dos parâmetros, atualizados após o conheci-

mento da informação relativa ao instante  $t$ , através de (5.3) e (5.4).

Alguns resultados intermediários destas relações merecem citação:

- $\hat{\theta}_{t|t-1}$  - estimativa a priori da média dos parâmetros  $\hat{\theta}_t$
- $P_{t|t-1}^*$  - matriz de covariância a priori associada à estimativa  $\hat{\theta}_{t|t-1}$
- $F_t \hat{\theta}_{t|t-1}$  - previsão de  $y_t$  um passo à frente feita no instante  $(t-1)$
- $(y_t - F_t \hat{\theta}_{t|t-1})$  - erro da previsão um passo à frente
- $(F_t P_{t|t-1}^* F_t' + V_t)$  - matriz de covariância associada a previsão de  $y_t$  um passo à frente
- $P_{t|t-1}^* F_t' [V_t + F_t P_{t|t-1}^* F_t']^{-1}$  - matriz que pode ser considerada uma extensão da constante de amortecimento escalar

Conforme o mencionado anteriormente esta formulação permite que se infira a distribuição a posteriori dos parâmetros no instante  $t$ ,  $(\theta_t | y^t, F^t)$  a partir:

- da distribuição a posteriori dos parâmetros no instante  $(t-1)$ ,  $(\theta_{t-1} | y^{(t-1)}, F^{(t-1)})$
- da observação no instante  $t$ ,  $y_t$
- das matrizes de covariância do ruído e da perturbação dos parâmetros no instante  $t$ ,  $V_t$  e  $W_t$

A grande vantagem disto é possibilitar:

- uma base teórica para previsões com poucas ou até nenhuma observação
- o desenvolvimento de estruturas que permitam a comunicação sistema-analista

#### V.1.4 - Previsão

Obtida a distribuição a posteriori do vetor de parâmetros no instante  $t$ , utiliza-se esta informação para inferir as distribuições das observações futuras do processo  $y_{t+k}$  ( $k=1,2,\dots$ ). Estas previsões decorrem da extrapolação da incerteza dos valores correntes dos parâmetros, combinada aos valores futuros das matrizes de covariância  $V_{t+k}$  e  $W_{t+k}$  sendo, portanto, fundamentalmente distribuições. Em conseqüência, pode-se escolher como previsão o parâmetro mais apropriado à situação em estudo, não havendo necessidade de associar previsão a valor esperado. Isto é importante quando as conseqüências de um erro de previsão cometido num determinado sentido são mais sérias que as de um erro de mesma magnitude cometido no sentido oposto. Exemplificando, num sistema hidrotérmico de geração de energia elétrica, a previsão de uma alta energia afluente ao sistema quando na realidade ocorre uma baixa afluência traz conseqüências muito mais desastrosas do que o caso inverso: pode-se gerar energia hidráulica em demasia e esvaziar os reservatórios. Assim, tendo

as previsões natureza de distribuições, pode-se fazê-las tendenciosas para levar em consideração fatos desta espécie.

A partir de (5.1) e (5.2) tira-se que o valor da observação no instante futuro ( $t+k$ ) pode ser obtido por:

$$(5.7) \quad Y_{t+k} = F_{t+k} \theta_{t+k} + v_{t+k}$$

$$(5.8) \quad \theta_{t+k} = G \theta_{t+k-1} + w_{t+k}$$

Supondo conhecidas as matrizes de covariância  $V_{t+k}, W_{t+1}, \dots, W_{t+k}$ , a previsão do valor futuro da observação  $Y_{t+k}$  requer a inferência, no instante  $t$ , dos valores futuros do vetor de parâmetros  $\theta_{t+k}$  e da matriz de variáveis independentes  $F_{t+k}$ . Sejam a média e a variância do vetor de parâmetros  $\theta_{t+k}$  definidos por:

$$(5.9) \quad \hat{\theta}_{k,t} = E \left[ \theta_{t+k} | Y^t, F^t \right]$$

$$(5.10) \quad P_{k,t}^* = \text{Var} \left[ \theta_{t+k} | Y^t, F^t \right]$$

Os valores  $\hat{\theta}_{0,t} = \hat{\theta}_t$  e  $P_{0,t}^* = P_t^*$  são conhecidos no instante  $t$  pelo filtro de Kalman. A partir deles, (5.9) e (5.10) podem ser obtidas pelas seguintes relações:

$$(5.11) \quad \hat{\theta}_{k,t} = G \hat{\theta}_{k-1,t} = G^2 \hat{\theta}_{k-2,t} = \dots = G^k \hat{\theta}_t$$

$$(5.12) \quad P_{k,t}^* = G P_{k-1,t}^* G' + W_{t+k} = \dots = G^k P_t^* G'^k + G^{k-1} W_{t+1} G'^{k-1} + \dots + G W_{t+k-1} G' + W_{t+k}$$

A partir de (5.7), (5.8), (5.11) e (5.12) pode-se obter a média e a variância da previsão  $k$  passos à frente da observação do processo:

$$(5.13) \quad \hat{Y}_{k,t} = E [Y_{t+k} | Y^t, F^t] = F_{t+k} \hat{\theta}_{k,t}$$

$$(5.14) \quad \hat{Y}_{k,t} = \text{Var} [Y_{t+k} | Y^t, F^t] = F_{t+k} P_{k,t}^* F_{t+k}' + V_{t+k}$$

(5.13) e (5.14) pressupõem o conhecimento de  $F_{t+k}$  no instante  $t$ . Caso isto não ocorra, elas se modificam para refletir a incerteza adicional na previsão decorrente da incerteza associada ao valor de  $F_{t+k}$ .

Supondo inicialmente que a observação do processo é escalar ( $M=1$ ) e que  $F_{t+k}$  possa ser estimado pelo vetor  $\hat{F}_{t+k}$  com matriz de covariância  $\hat{\Phi}_{t+k}$  associada, a média e a variância da previsão  $k$  passos à frente ficam (omitindo-se os índices  $t+k$  e  $k,t$  por simplificação):

$$(5.15) \quad \hat{y} = \hat{F} \hat{\theta}$$

$$(5.16) \quad \hat{Y} = E [(\hat{y} - y)(\hat{y} - y)'] = E [(\hat{F} \hat{\theta} - F \theta - v)(\hat{F} \hat{\theta} - F \theta - v)']$$



Da definição de multiplicação matricial decorre que (5.16) equivale a:

$$(5.17) \quad \hat{Y} = E \left[ \sum_{i=1}^N (\hat{F}_i \hat{\theta}_i - F_i \theta_i) \sum_{j=1}^N (\hat{F}_j \hat{\theta}_j - F_j \theta_j) \right] + V = \\ = \sum_{i=1}^N \sum_{j=1}^N \{ E[\hat{F}_i \hat{F}_j] E[\hat{\theta}_i \hat{\theta}_j] - F_i F_j \theta_i \theta_j \}$$

Fazendo  $E[\hat{F}_i \hat{F}_j] = \phi_{ij} + F_i F_j$  (onde  $\phi_{ij}$  é o elemento  $(i, j)$  de  $\Phi$ ) e  $E[\hat{\theta}_i \hat{\theta}_j] = p_{ij}^* + \theta_i \theta_j$  (onde  $p_{ij}^*$  é o elemento  $(i, j)$  de  $P^*$ ), pode-se escrever (5.17) na forma:

$$(5.18) \quad \hat{Y} = \sum_{i=1}^N \sum_{j=1}^N \{ (\phi_{ij} + F_i F_j) (p_{ij}^* + \theta_i \theta_j) - F_i F_j \theta_i \theta_j \} + V = \\ = \sum_{i=1}^N \sum_{j=1}^N (\phi_{ij} p_{ij}^* + \phi_{ij} \theta_i \theta_j + p_{ij}^* F_i F_j) + V$$

Colocando (5.18) na forma matricial:

$$(5.19) \quad \hat{Y} = \text{traço} (\Phi P^*) + \theta \Phi \theta' + F P^* F' + V$$

Para o caso de  $M > 1$ ,  $\phi_{ij}$  torna-se uma matriz  $\Phi_{ij}$ ,  $(M \times M)$ . Definindo a matriz  $Z$ ,  $(M \times M)$ , tal que seu elemento  $z_{ij} = \text{traço} [\Phi_{ij} (\hat{\theta}_{k,t} \hat{\theta}'_{k,t} + P_{k,t}^*)]$ ,  $(i, j = \overline{1, M})$  pode-se mostrar que (5.19) equivale a:

$$\hat{Y} = F P^* F' + V + Z$$

O termo  $Z$  representa a incerteza adicional na pre-

visão decorrente da incerteza associada ao valor de  $F_{t+k}$ .

Assim, para o caso de  $F_{t+k}$  desconhecido, as expressões da média e da variância da previsão  $k$  passos à frente ficam (reintroduzindo os índices):

$$(5.20) \quad \hat{Y}_{k,t} = \hat{F}_{k,t} \hat{\theta}_{k,t}$$

$$(5.21) \quad \hat{Y}_{k,t} = \hat{F}_{k,t} P_{k,t}^* \hat{F}_{k,t}' + V_{t+k} + Z$$

#### V.1.5 - Superposição

O princípio da superposição estabelece que uma combinação linear de modelos lineares é também um modelo linear. Será mostrado que para o caso de dois modelos lineares dinâmicos (a extensão para mais de dois é imediata) ele permanece válido. Sejam eles:

$$(5.22) \quad Y_{i,t} = F_{i,t} \theta_{i,t} \quad i = 1, 2$$

$$(5.23) \quad \theta_{i,t} = G_i \theta_{i,t-1} + w_{i,t} \quad i = 1, 2$$

As equações de observação são supostas, individualmente, sem ruído. A observação do processo é uma combinação das duas:

$$(5.24) \quad Y_t = Y_{1,t} + Y_{2,t} + v_t, v_t \sim N(0, V_t)$$

O sistema (5.22), (5.23) e (5.24) pode ser escrito na forma:

$$y_t = \begin{bmatrix} F_{1,t} & F_{2,t} \end{bmatrix} \begin{bmatrix} \theta_{1,t} \\ \theta_{2,t} \end{bmatrix} + v_t$$

$$\theta_t = \begin{bmatrix} \theta_{1,t} \\ \theta_{2,t} \end{bmatrix} = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} \theta_{1,t-1} \\ \theta_{2,t-1} \end{bmatrix} + \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix}$$

que é um MLD com  $\theta_t = (\theta_{1,t}, \theta_{2,t})'$

Com isto pode-se decompor um MLD de grande porte em MLDs mais simples e, portanto, mais fáceis de serem estudados individualmente. Exemplificando, num problema de previsão de vendas pode-se considerar o processo gerador da série como uma combinação de quatro MLDS: um representativo da dependência entre vendas e variáveis econométricas, um da variação sazonal, um do efeito de propaganda e um de fatores desconhecidos.

#### V.1.6 - Representação de Modelos Convencionais em Espaço de Estados

Uma série de modelos clássicos de séries temporais admite representação em espaço de estados, isto é, pode ser representada por um MLD. Alguns deles serão apresentados, primeiro na forma convencional e, posteriormente, na representação em es-

paço de estados.

A representação do modelo ARIMA requer sua linearização, conseguida através do desenvolvimento de uma série de Taylor, com o processo de estimação realizado pelo filtro de Kalman Extendido.

O símbolo  $\sigma$  denota uma perturbação aleatória no parâmetro correspondente.

#### V.1.6.1 - Regressão Linear

Tanto o modelo estático:

$$y_t = x_t \theta_t + v_t$$

$$\theta_t = \theta_{t-1} = \dots = \theta_0$$

quanto dinâmico

$$y_t = x_t \theta_t + v_t$$

$$\theta_t = \theta_{t-1} + \sigma \theta_t$$

são MLDS particulares em que  $M=1$ ,  $F_t=x_t$  e  $G=I$ , de acordo com o Capítulo III. Deve-se enfatizar a importância, em termos de flexibilidade, dos modelos não exigirem variâncias constantes, pe-

las razões expostas no item III.3.4.2.

#### V.1.6.2 - Modelo de Crescimento Linear Sazonal

Este modelo é bastante utilizado. Supõe que o processo é composto de um "nível" crescente (ou decrescente) ao longo do tempo superposto a um efeito sazonal, descrito por um vetor de fatores sazonais aditivos. Geralmente apresenta bons resultados em aplicações para séries de demanda e sua forma é a seguinte:

$$Y_t = \mu_t + \rho_{m(t),t} + v_t$$

$$\mu_t = \mu_{t-1} + \beta_t + \sigma\mu_t$$

$$\beta_t = \beta_{t-1} + \sigma\beta_t$$

$$\rho_{i,t} = \rho_{i,t-1} + \sigma\rho_{i,t} \quad (i = 1, \dots, T)$$

$$\sum_{i=1}^T \rho_{i,t} = \sum_{i=1}^T \sigma\rho_{i,t} = 0$$

onde os termos podem ser interpretados:

- $\mu_t$  - nível do processo no instante t
- $\beta_t$  - inclinação(crescimento) no instante t
- $\rho_{m(t),t}$  - fator sazonal no instante t associado ao "mês" m(t)
- T - período

Este modelo pode ser representado por um MLD com:

$$M=1$$

$$N=T + 2$$

$$F_t = (1, 0, \dots, 1, \dots, 0)$$

$$\theta_t = (\mu_t, \beta_t, \rho_{1,t}, \dots, \rho_{T,t})'$$

$$G = \left| \begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & \sim \\ \hline & 0 & I \end{array} \right|$$

$$w_t = (\sigma\mu_t + \sigma\beta_t, \sigma\beta_t, \sigma\rho_{1,t}, \dots, \sigma\rho_{T,t})'$$

$$W_t = E[\omega_t \omega_t'] = \left| \begin{array}{cc|cccc} \sigma\mu_t^2 + \sigma\beta_t^2 & \sigma\beta_t^2 & & & & \\ \sigma\beta_t^2 & \sigma\beta_t^2 & & & & \\ \hline & & 0 & & & \\ & & & \sigma\rho_{1,t}^2 & \dots & \sigma\rho_{1,t}\sigma\rho_{T,t} \\ & & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \\ & & & \sigma\rho_{T,t}\sigma\rho_{1,t} & \dots & \sigma\rho_{T,t}^2 \end{array} \right|$$

A presença da perturbação de nível  $\sigma\beta_t$  nos dois primeiros elementos do vetor  $w_t$  introduz uma covariância na matriz  $W_t$ . O vetor  $F_t$  é composto de zeros em todas as posições exceto três: a primeira, a segunda e a correspondente a  $m(t)$ . Com isto, apesar de não ser constante, varia sistematicamente, sendo conhecido em todos os instantes.

### V.1.6.3 - Modelo de Funções Periódicas

Existem situações, como o caso de dados semanais, que a representação da sazonalidade do processo através de  $T$  parâmetros pode não compensar o esforço computacional decorrente da melhoria do desempenho. As vezes, uma representação através de funções periódicas da forma:

$$f_i(t) = f_i(t+KT) \quad t=1,2,\dots, \quad K \in \mathbb{Z}$$

pode ser perfeitamente aceitável e introduzir uma simplificação computacional sensível.

Combinando esta representação sazonal com o modelo de crescimento linear chega-se a:

$$y_t = \mu_t + \sum_{i=1}^n \alpha_{i,t} f_i(t) + v_t$$

$$\mu_t = \mu_{t-1} + \beta_t + \sigma\beta_t$$

$$\beta_t = \beta_{t-1} + \sigma\beta_t$$

$$\alpha_{i,t} = \alpha_{i,t-1} + \sigma\alpha_{i,t}$$

Este modelo pode ser representado por um MLD com:

$$M = 1$$

$$N = n + 2$$

$$F_t = (1, 0, f_1(t), f_2(t), \dots, f_n(t))$$

$$\theta_t = (\mu_t, \beta_t, \alpha_{1,t}, \dots, \alpha_{n,t})$$

$$G = \left| \begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & \sim \\ \hline 0 & & I \end{array} \right|$$

$$w_t = (\sigma_{\mu_t}^2 + \sigma_{\beta_t}^2, \sigma_{\beta_t}^2, \sigma_{\alpha_{1,t}}^2, \dots, \sigma_{\alpha_{n,t}}^2)$$

$$W_t = \left| \begin{array}{cc|ccc} \sigma_{\mu_t}^2 + \sigma_{\beta_t}^2 & \sigma_{\beta_t}^2 & & & \\ \sigma_{\beta_t}^2 & \sigma_{\beta_t}^2 & & & \\ \hline & & \sigma_{\alpha_{1,t}}^2 & & \\ & & & \ddots & \\ & & & & \sigma_{\alpha_{T,t}}^2 \end{array} \right|$$

As funções  $f_i(t)$  mais usuais são da forma  $\sin \frac{(2\pi it)}{T}$  e  $\cos \frac{(2\pi it)}{T}$ ,  $i = 1, 2, \dots, T/2$ . Na prática, um número pequeno de harmônicos é em geral suficiente para uma representação adequada da sazonalidade do processo, fazendo com que o mo



delo tenha uma dimensão relativamente reduzida.

#### V.1.6.4 - Modelo ARIMA: O Filtro de Kalman Extendido

Pode-se definir um modelo integrado autoregressivo médias móveis (ARIMA) pela seguinte equação de diferenças:

$$(5.24a) \quad z_t - \mu = v_t - \phi_1(z_{t-1} - \mu) - \dots - \phi_p(z_{t-p} - \mu) - \psi_1 v_{t-1} - \dots - \psi_q v_{t-q}$$

$$(5.24b) \quad (1 - \phi_1 B - \dots - \phi_p B^p)(z_t - \mu) = (1 - \psi_1 B - \dots - \psi_q B^q) v_t$$

$$(5.24c) \quad \Phi(B)(z_t - \mu) = \Psi(B) v_t$$

onde

$\mu, \phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q$  - parâmetros

$B$  - operador retardo

$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  e  $\Psi(B) = 1 - \psi_1 B - \dots - \psi_q B^q$  - polinômios cujas raízes de  $\Phi(B)=0$  e  $\Psi(B)=0$  são supostas fora do círculo unitário, não tendo os dois raízes comuns.

$z_t$  - diferença estacionária da observação original

$v_t$  - seqüência de variáveis aleatórias independentes com distribuição  $N(0, \sigma^2)$

Seja  $\alpha = (\mu, \phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q)'$  o vetor de parâmetros. Conhecido o conjunto de observações até o instante  $t, z^t$ , pode-se derivar a estimativa de máxima verossimilhança de  $\alpha$  que, em geral, necessita de técnicas iterativas de maximização pelo fato da função de verossimilhança exata ser não linear nos parâmetros (ver |3|).

Este processo de estimação é essencialmente não recursivo e supõe parâmetros constantes. Para torná-lo recursivo e, portanto, permitir variações nos parâmetros basta colocar o modelo ARIMA (5.24) na forma de MLD (5.1) e (5.2). Entretanto, como o modelo ARIMA é não linear nos parâmetros, há necessidade de linearizá-lo e isto pode ser conseguido expandindo-o numa série de Taylor em torno de um valor de referência  $\bar{\alpha} = (\bar{\mu}, \bar{\phi}_1, \dots, \bar{\phi}_p, \bar{\psi}_1, \dots, \bar{\psi}_q)'$ .

(5.14) pode ser escrito (ver |3|):

$$z_t = \mu + \sum_{j=1}^{\infty} \pi_j(\phi, \psi) (z_{t-j} - \mu) + v_t$$

Definindo:

$$(5.25) \quad f(\alpha, z^{(t-1)}) \triangleq \mu + \sum_{j=1}^{\infty} \pi_j(\phi, \psi) (z_{t-j} - \mu)$$

chega-se a uma aproximação linear para o valor de  $z_t$  expandindo a função (5.25) numa série de Taylor em torno de  $\bar{\alpha}$ :

$$(5.26) \quad z_t \cong f(\bar{\alpha}, z^{(t-1)}) + (\mu - \bar{\mu}) a_t(\bar{\alpha}) + \sum_{i=1}^p (\phi_i - \bar{\phi}_i) b_{it}(\bar{\alpha}) + \\ + \sum_{j=1}^q (\psi_j - \bar{\psi}_j) c_{jt}(\bar{\alpha}) + v_t$$

onde:

$$(5.27) \quad a_t(\bar{\alpha}) = \frac{d}{d\mu} f(\alpha, z^{(t-1)}) \Big|_{\alpha=\bar{\alpha}} = \frac{\phi(B)}{\Psi(B)} \Big|_{\alpha=\bar{\alpha}} = \\ = \frac{1 - \bar{\phi}_1 B - \dots - \bar{\phi}_p B^p}{1 - \bar{\psi}_1 B - \dots - \bar{\psi}_q B^q}$$

$$(5.28) \quad b_{it}(\bar{\alpha}) = \frac{d}{d\phi_i} f(\alpha, z^{(t-1)}) \Big|_{\alpha=\bar{\alpha}} = \frac{B^i(z_t - \mu)}{\Psi(B)} \Big|_{\alpha=\bar{\alpha}} = \\ = \frac{(z_{t-i} - \mu)}{\Psi(B)} \Big|_{\alpha=\bar{\alpha}} = \frac{v_{t-i}(\alpha)}{\phi(B)} \Big|_{\alpha=\bar{\alpha}} = \frac{v_{t-i}(\bar{\alpha})}{1 - \bar{\phi}_1 B - \dots - \bar{\phi}_p B^p}$$

$$(5.29) \quad c_{jt}(\bar{\alpha}) = -\frac{d}{d\psi_j} f(\alpha, z^{(t-1)}) \Big|_{\alpha=\bar{\alpha}} = -\phi(B) \frac{B^j(z_t - \mu)}{[\Psi(B)]^2} \Big|_{\alpha=\bar{\alpha}} = \\ = -\frac{\phi(B)}{[\psi(B)]^2} (z_{t-j} - \mu) \Big|_{\alpha=\bar{\alpha}} = -\frac{v_{t-j}(\alpha)}{\Psi(B)} \Big|_{\alpha=\bar{\alpha}} = \\ = -\frac{v_{t-j}(\bar{\alpha})}{1 - \bar{\psi}_1 B - \dots - \bar{\psi}_q B^q}$$

As passagens intermediárias de (5.28) e (5.29) provêm de (5.24c) e de:

$$(5.30) \quad v_{t-i}(\alpha) = z_{t-i} - f(\alpha, z^{(t-i-1)})$$

Supondo os parâmetros variando segundo a expressão

$$\alpha_t = G\alpha_{t-1} + w_t, \quad w_t \sim N(0, W)$$

pode-se expandir o modelo ARIMA (5.24) em torno da trajetória  $\bar{\alpha}_t = G\bar{\alpha}_{t-1}$  e representá-lo na forma:

$$(5.30) \quad z_t - f(\bar{\alpha}, z^{(t-1)}) = \left[ a_t(\bar{\alpha}_t), b_{1t}(\bar{\alpha}_t), \dots, b_{pt}(\bar{\alpha}_t), c_{1t}(\bar{\alpha}_t), \dots, c_{qt}(\bar{\alpha}_t) \right] (\alpha_t - \bar{\alpha}_t) + v_t$$

$$(5.31) \quad \alpha_t - \bar{\alpha}_t = G(\alpha_{t-1} - \bar{\alpha}_{t-1}) + w_t$$

Este modelo pode ser encarado como um MLD com:

$$M = 1$$

$$N = p+q+1$$

$y_t = z_t - f(\bar{\alpha}, z^{(t-1)})$ : representa o desvio na medida da observação

$$F_t = \left[ a_t(\bar{\alpha}_t), b_{1t}(\bar{\alpha}_t), \dots, b_{pt}(\bar{\alpha}_t), c_{1t}(\bar{\alpha}_t), \dots, c_{qt}(\bar{\alpha}_t) \right]$$

$\theta_t = (\alpha_t - \bar{\alpha}_t)$ : representa o desvio na trajetória  $\bar{\alpha}_t$

$$V_t = \sigma^2$$

Assim dada uma trajetória  $\bar{\alpha}_t$  e o conjunto de observações  $z^t$ , pode-se calcular o desvio na medida da observação  $y_t = z_t - f(\bar{\alpha}_t, z^{(t-1)})$  e usar o filtro de Kalman para derivar uma estimativa  $\hat{\theta}_t$  do desvio na trajetória  $\theta_t = \alpha_t - \bar{\alpha}_t$ . De posse de  $\hat{\theta}_t$  chega-se imediatamente a  $\hat{\alpha}_t$ :

$$\hat{\alpha}_t = \hat{\theta}_t + \bar{\alpha}_t$$

A trajetória de referência  $\bar{\alpha}_0$  pode ser escolhida como a estimativa a priori de  $\alpha_0$ . Entretanto, se esta estimativa não for fidedigna, o desvio  $\theta_t$  pode se tornar demasiadamente grande, invalidando a linearização de (5.26). Uma maneira de contornar o problema é linearizar o modelo (5.30) e (5.31) em torno de cada nova estimativa  $\hat{\alpha}$  à medida que elas vão sendo obtidas, ou seja, fazer  $\bar{\alpha} = \hat{\alpha}$ . Este procedimento utiliza uma estimativa melhor para a trajetória assim que ela se torna disponível, impedindo a propagação de erros grandes em torno da trajetória devidos a más estimativas a priori.

O procedimento geral de linearizar recursivamente modelos não lineares é conhecido como uma filtragem de Kalman Extendida.

### V.1.7 - Modelos de Múltiplos Processos

Uma das características do método que lhe empresta grande flexibilidade é a incerteza em relação ao modelo do processo em cada instante. Dos dois tipos principais de incerteza comentados no item V.1.1, apresenta-se sucintamente o primeiro e detalhadamente o segundo. A razão disto é a seguinte: o primeiro guarda íntima relação com a teoria de decisão Bayesiana; sua descrição completa importa numa série de conceitos que além de fugir aos objetivos do trabalho, pouco subsídio traz à aplicação do método. Em consequência, somente se apresenta suas idéias. Quanto ao segundo, constitui uma parte bastante importante da aplicação do método.

#### V.1.7.1 - Modelo Único Representativo do Processo

Considera-se que existe um único modelo desconhecido que representa adequadamente o processo em todos os instantes. Como as matrizes  $F_t, G, V_t$  e  $W_t$  caracterizam um MLD no instante  $t$ , pode-se definir:

$$M_t = \{F_t, G, V_t, W_t\}$$

como a especificação de um modelo neste instante. De um modo geral, a própria situação em estudo sugere a escolha de  $F_t$  e  $G$ . Exemplificando, o modelo de crescimento linear sazonal frequente

mente representa muito bem séries de demanda. Por outro lado, a determinação das variâncias  $V_t$  e  $W_t$  é mais difícil.

Supondo que a incerteza acerca do modelo e seus parâmetros possa ser descrita por um conjunto discreto de modelos alternativos  $M^{(j)}$  ( $j=1,2,\dots,N$ ) com probabilidades associadas, o objetivo é inferir qual deles representa adequadamente o processo em todos os instantes. Neste tipo de incerteza pode-se distinguir duas espécies de problemas:

- a partir de um conjunto  $M_t^{(j)}$  escolher um único modelo para representar o processo e nele basear a previsão e as decisões.
- todo o conjunto  $M_t^{(j)}$  prove informações em que a previsão e as decisões se baseiam.

Como exemplo da segunda, pode-se imaginar uma situação onde se conhece a "forma" do modelo através das matrizes  $F_t$  e  $G$  (modelo de crescimento linear, por exemplo), mas não se conhece o valor de  $V_t$ . Neste caso o procedimento pode ser a especificação de um conjunto discreto de valores  $V_t^{(1)}, V_t^{(2)}, \dots, V_t^{(N)}$  à qual se atribui probabilidades a priori que são constantemente atualizadas e basear a previsão na informação de todo o conjunto  $V_t^{(i)}$ .

Os resultados concernentes à informação no instante  $t$  relacionando o vetor de parâmetros  $\theta_t$  às observações futu-

ras  $y_{t+k}$  foram estabelecidos condicionalmente a um MLD especificado. Considerando um conjunto de modelos  $M_t^{(j)}$ , as distribuições conjunta e condicional de  $\theta_t$  e  $M_t^{(j)}$  e a função de verossimilhança condicional de  $y_t$  são facilmente obtidas. Com isto, se chega à informação sobre as observações futuras através das distribuições:

$$(5.32) \quad (y_{t+k} | y^t, F^t, M^{(j)}) \sim N(\hat{y}_{k,t}^{(j)}, \hat{Y}_{k,t}^{(j)}) \quad , \quad j = \overline{1, N}$$

A partir de (5.32), introduzindo uma função de utilidade e utilizando o processo de decisão Bayesiano, pode-se escolher o modelo mais "apropriado". Por exemplo, o modelo que maximiza a esperança acumulada da função de utilidade.

#### V.1.7.2 - Modelo Múltiplo Representativo do Processo

##### V.1.7.2.1 - Introdução

A abordagem Bayesiana faz com que no instante  $(t-1)$  as únicas informações necessárias para prever a observação no instante  $t$  sejam a distribuição a posteriori dos parâmetros  $(\theta_{t-1} | y^{(t-1)}, F^{(t-1)})$  e o MLD que descreve a evolução do processo na transição  $(t-1, t)$ , não havendo necessidade do mesmo MLD descrever todas as transições.

Neste item se aborda a situação na qual nenhum MLD sozinho representa adequadamente o que acontecerá ao proces



so no próximo período. Isto possibilita um tratamento apropriado a ocorrências "anormais" súbitas, assumindo grande importância se for considerado que em processos sócio-econômicos estas ocorrências são até certo ponto normais (por exemplo em séries de vendas, observações esdrúxulas ocorrem com frequência, além de mudanças de nível do processo em decorrência de propaganda ou alteração de hábitos do consumidor). Estas situações podem ser convenientemente expressas como parte de um MLD, bastando relacionar, por exemplo, as mudanças nas variações do ruído e das perturbações aos parâmetros do processo (exceto quando as mudanças nas matrizes  $V$  e/ou  $W$  forem difíceis de antecipar e não forem associadas a quantidades mensuráveis à priori).

Para tratar este problema considera-se modelos que em qualquer instante  $t$  compreendam um conjunto de MLDs, cada qual com uma forma definida idêntica ( $F$  e  $G$  iguais) mas diferindo entre si pelas matrizes de covariância do ruído  $V$  e da perturbação  $W$ .

Como ilustração seja o modelo de crescimento linear do item V.1.6.2 sem a parte sazonal:

$$y_t = \mu_t + v_t$$

$$\mu_t = \mu_{t-1} + \beta_t + \sigma\mu_t$$

$$\beta_t = \beta_{t-1} + \sigma\beta_t$$

O ruído  $v_t$  afeta o processo somente no instante  $t$  não interferindo nos valores subsequentes. Já as perturbações de nível  $\sigma_{\mu_t}$  e inclinação  $\sigma_{\beta_t}$  se incorporam ao processo atuando sobre todos os valores futuros da série. As figuras 5.1, 5.2 e 5.3, nas quais se supõe um processo determinístico exceto no instante  $t$  quando  $v_t$ ,  $\sigma_{\mu_t}$  e  $\sigma_{\beta_t}$  são respectivamente não nulos, permitem visualizar o ocorrido. Um valor muito grande de  $v_t$  causa somente uma transiência instantânea, de  $\sigma_{\mu_t}$  uma mudança de nível permanente e de  $\sigma_{\beta_t}$  uma mudança de inclinação permanente.

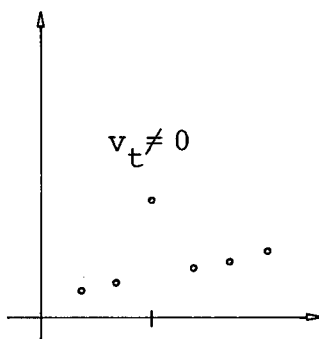


Figura 5.1

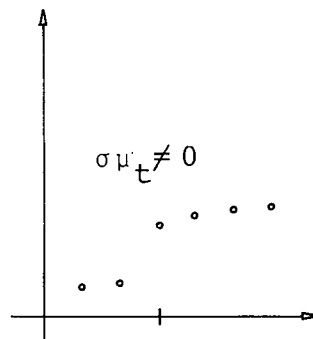


Figura 5.2

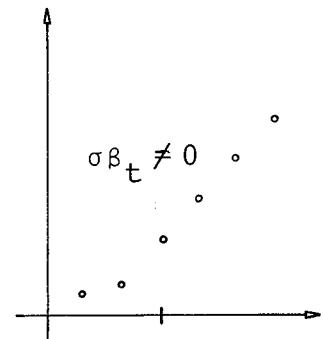


Figura 5.3

Do exemplo pode-se depreender o papel desempenhado por  $V_t$  e  $W_t$ . Um valor de  $V_t$  excepcionalmente grande indica uma observação "anormal" no instante  $t$  sem efeito no futuro. Já um valor grande de  $W_t$  indica uma mudança significativa no instante  $t$  que se incorpora ao processo afetando suas observações futuras. Isto permite que se modele explicitamente estas situações supondo a existência de um número de possíveis distribuições a partir das quais os valores do ruído e das perturbações dos parâmetros são gerados no instante  $t$ . Como  $w_t$  é  $(N \times 1)$  pelo menos

(N + 1) dessas "anormalidades" podem ser modeladas, especificando diferentes valores para  $V_t$  e  $W_t$ .

#### V.1.7.2.2 - Modelos de Múltiplos Estados

Esta abordagem dá origem ao modelo de múltiplos estados. Nela, um modelo  $M_t^{(j)} = \{F_t, G, V_t^{(j)}, W_t^{(j)}\}$ , todos conhecidos, define um estado  $j$  e caracteriza a evolução do processo no instante  $t$ . O modelo de múltiplos estados compreende todo o conjunto  $M^{(j)}$  e descreve o estado em curso no instante  $t$  por um processo Markoviano caracterizado por uma matriz de probabilidades de transição  $(M_{t-1}^{(i)}, M_t^{(j)})$ , onde  $(i, j = \overline{1, N})$ . Neste trabalho, entretanto, supõe-se por simplificação que cada modelo  $M^{(j)}$  tem uma probabilidade a priori constante  $\pi_j$  de estar em curso no instante  $t$ , independente do passado do processo.

Para exemplificar a definição de estados, seja o modelo de crescimento linear. Pode-se definir quatro estados básicos representados na figura 5.4:

- 1 - normal
- 2 - mudança de nível
- 3 - mudança de inclinação
- 4 - transiente

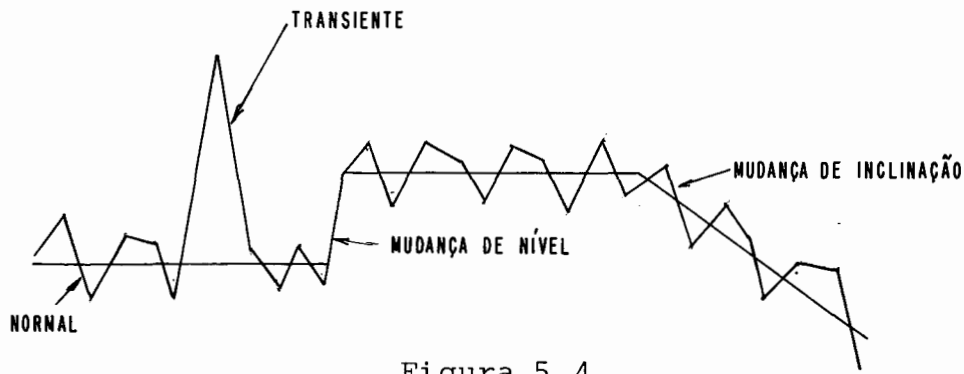


Figura 5.4

Cada um deles é caracterizado pelas variâncias do ruído e da perturbação (além das probabilidades de estado  $\pi_j, j = \overline{1,4}$ ) conforme mostra a tabela:

Estado	$V_t$	$\text{Var}(\sigma\mu_t)$	$\text{Var}(\sigma\beta_t)$
1	Normal	Muito Pequena	Muito Pequena
2	Normal	Grande	Muito Pequena
3	Normal	Muito Pequena	Grande
4	Grande	Muito Pequena	Muito Pequena

#### V.1.7.2.3 - Condensação de Posteriors

Dado um modelo  $M^{(j)}$ , a distribuição a posteriori de  $\theta_t$  pode ser obtida diretamente pelo filtro de Kalman usando os valores  $v_t^{(j)}$  e  $w_t^{(j)}$  apropriados. Supondo que as probabilidades a posteriori de cada estado  $j$  no instante  $t$  sejam conhecidas, a distribuição a posteriori de  $\theta_t$  é uma combinação linear, ponderada pelas probabilidades, de  $N$  distribuições normais distintas, onde  $N$  é o número de estados. No instante  $(t + 1)$ , a

distribuição a posteriori de  $\theta_{t+1}$  seria uma combinação de  $N^2$  normais distintas, cada uma correspondendo à aplicação do filtro de Kalman à transição entre os estados  $i$  e  $j$  nos instantes  $(t, t+1)$ . Prosseguindo se obteria  $N^3, N^4, \dots$  componentes num processo proliferativo que de algum modo deve ser contido. O seguinte procedimento é sugerido por Harrison e Stevens: supondo que no instante  $(t-1)$  toda a informação relevante ao futuro esteja resumida sob a forma de  $N$  distribuições normais e  $N$  probabilidades de estado a posteriori conhecidas

$$(\theta_{t-1} | M_{t-1}^{(j)}, Y^{(t-1)}) \sim N(\hat{\theta}_{t-1}^{(j)}, P_{t-1}^{*(j)}) \text{ e } P_{t-1}^{(j)}, \quad j = \overline{1, N}$$

onde  $P_{t-1}^{(j)}$  é a probabilidade do modelo  $M^{(j)}$  estar em curso no instante  $(t-1)$ , o filtro de Kalman gera no instante  $t$ , baseado nesta informação e na observação  $y_t, N^2$  distribuições a posteriori:

$$(\theta_t | Y^t, M_{t-1}^{(i)}, M_t^{(j)}) \sim N(\hat{\theta}^{(i,j)}, P_t^{*(i,j)})$$

e  $N^2$  probabilidades a posteriori

$$P_t^{(i,j)} = P\{M_{t-1}^{(i)}, M_t^{(j)} | Y^t\} \propto L(y_t | M_{t-1}^{(i)}, M_t^{(j)}, Y^{(t-1)}) \pi_j P_{t-1}^{(i)} =$$

$$= \frac{1}{(2\pi)^{\frac{M}{2}} |V(j)|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \left[ (y_t - F_t G \hat{\theta}_i)' (\hat{Y}^{(i,j)})^{-1} (y_t - F_t G \hat{\theta}_i) \right]\right\}.$$

$$\cdot \pi_j P_{t-1}^{(i)}$$

onde

$$\hat{Y}(i, j) = F_t P_{t-1}^*(i) F_t' + W(j)$$

que são grupadas em  $N$  distribuições e probabilidades a posteriori pelas seguintes equações de condensação:

$$P_t(j) = \sum_{i=1}^N P_t(i, j)$$

$$(\theta_t | Y^t, j) \sim N(\hat{\theta}_t(j), P_t^*(j))$$

onde:

$$\hat{\theta}_t(j) = \sum_{i=1}^N P_t(i, j) \hat{\theta}_t(i, j) / P_t(j)$$

$$P_t^*(j) = \sum_{i=1}^N P_t(i, j) \{ P_t^*(i, j) + (\hat{\theta}_t(i, j) - \hat{\theta}_t(j)) (\hat{\theta}_t(i, j) - \hat{\theta}_t(j))' \} / P_t(j)$$

Portanto o processo de condensação grupa toda a informação posteriori correspondente ao mesmo estado  $j$ , preservando as informações mais relevantes e reduzindo a dimensão do problema a posteriori ( $N^2$ ) para a dimensão a priori ( $N$ ).

### V.1.8 - Exemplos de Intervenção do Analista

Para encerrar a apresentação do método, um exemplo de como o analista e o sistema podem se comunicar. Seja a distribuição a priori do nível do processo no instante  $t$  dada pelo filtro de Kalman,  $(\mu_{t+1} | y^t, F^t) \sim N(100, 400)$ . O analista sabe que no mês seguinte um novo mercado se abre para o produto, estimando que as vendas aumentarão em cerca de 50%. Entretanto, está muito incerto em relação a este valor. Ele pode desprezar a estimativa do filtro de Kalman e estabelecer que  $(\mu_{t+1} | t) \sim N(150, 2500)$ . Após o conhecimento da observação  $y_{t+1}$ , o filtro de Kalman combina esta informação com a do analista e prossegue o processo estimativo. Da mesma forma o analista poderia ter aumentado a probabilidade de ocorrência do estado "mudança de nível".

Em outras palavras, a abordagem Bayesiana permite que se misture teoria e prática.

### V.2 - Modelos Convencionais: O Método VI-MVA

Este método se ocupa da estimação de parâmetros caracterizando uma representação de um sistema linear dinâmico através de um modelo de função de transferência discreto de séries temporais. A formulação recursiva das equações de estimação permite a estimação de parâmetros variáveis do sistema.

## V.2.1 - Hipóteses do Sistema

O sistema, por hipótese, possui duas entradas: um sinal determinístico  $u_t$  e um ruído branco  $v_t$  (seqüência de variáveis aleatórias com média nula, variância constante  $\sigma^2$ , serialmente não correlacionados e independentes de  $u_t$ ).

O sinal  $u_t$  passa pela função de transferência do sistema gerando uma saída sem ruído  $x_t$ . O ruído branco  $v_t$  passa pela função de transferência de ruído gerando uma perturbação  $\xi_t$ . Estes dois termos ( $x_t$  e  $\xi_t$ ) somados produzem a observação  $y_t$ , conforme a figura 5.5.

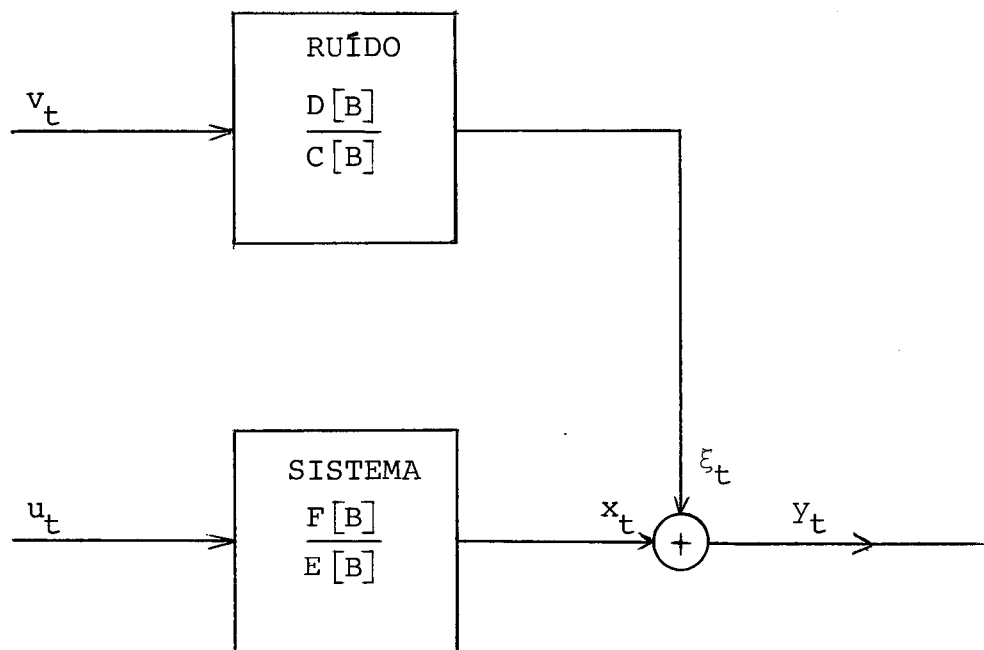


Figura 5.5

Em termos de análise convencional de séries temporais,  $x_t$  e  $\xi_t$  são gerados por um processo ARMA agindo sobre as entradas  $u_t$  e  $v_t$ .



## V.2.2 - Relações entre as Variáveis

Seja  $B$  o operador retardo definida no item III.3.4.1 e  $E[B]$ ,  $F[B]$ ,  $C[B]$  e  $D[B]$  polinômios em  $B$  da forma:

$$E[B] = 1 + e_1 B + \dots + e_N B^N$$

$$F[B] = f_0 + f_1 B + \dots + f_N B^N$$

$$C[B] = 1 + c_1 B + \dots + c_N B^N$$

$$D[B] = 1 + d_1 B + \dots + d_N B^N$$

As relações entre as variáveis podem ser escritas:

$$(5.33) \quad E[B] x_t = F[B] u_t$$

$$(5.34) \quad C[B] \xi_t = D[B] v_t$$

$$(5.35) \quad Y_t = x_t + \xi_t$$

Em termos das entradas:

$$(5.36) \quad Y_t = \frac{F[B]}{E[B]} u_t + \frac{D[B]}{C[B]} v_t$$

Para não introduzir complicações teóricas que nada de interessante acrescentariam ao problema, supõe-se que os pro-

cessos especificados pelas equações (5.34) e (5.35) são estáveis, o processo de ruído é inversível e que a entrada  $u_t$  proporciona excitação suficiente para permitir a completa identificação do sistema.

Combinando (5.33) e (5.35):

$$(5.37) \quad E[\bar{B}] y_t = F[\bar{B}] u_t + E[\bar{B}] \xi_t$$

(5.37) e (5.34) equivalem a:

$$y_t = -e_1 y_{t-1} - \dots - e_N y_{t-N} + f_0 u_t + f_1 u_{t-1} + \dots + f_N u_{t-N} + \xi_t + e_1 \xi_{t-1} + \dots + e_N \xi_{t-N}$$

$$\xi_t = -c_1 \xi_{t-1} - \dots - c_N \xi_{t-N} + d_1 v_{t-1} + \dots + d_N v_{t-N} + v_t$$

Definindo:

$$z_t = [-y_{t-1}, \dots, -y_{t-N}, u_t, \dots, u_{t-N}]$$

$$\theta^1 = [e_1, \dots, e_N, f_0, \dots, f_N]'$$

$$m_t = \xi_t + e_1 \xi_{t-1} + \dots + e_N \xi_{t-N}$$

$$n_t = [-\xi_{t-1}, \dots, -\xi_{t-N}, v_{t-1}, \dots, v_{t-N}]$$

$$\theta^2 = [c_1, \dots, c_N, d_1, \dots, d_N]'$$

chega-se às seguintes equações vetoriais para descrever o sistema global:

$$(5.38) \quad y_t = z_t \theta^1 + m_t$$

$$(5.39) \quad \xi_t = n_t \theta^2 + v_t$$

O vetor  $\theta^1$  contém os parâmetros que definem o sistema e vetor  $\theta^2$ , os parâmetros que definem a perturbação do sistema.

### V.2.3 - Estimação dos Parâmetros

Colocado nestes termos, o problema da estimação se resume em estimar os vetores de parâmetros  $\theta^1$  e  $\theta^2$  tendo como informação as entradas  $u_i$  e as observações  $y_i$ ,  $i = \overline{1, t}$ .

O algoritmo VI-MVA executa esta estimação em dois estágios conforme mostra a figura 5.6.

No primeiro estágio, o vetor  $\theta^1$  é estimado através de um algoritmo recursivo de mínimos quadrados modificado: utiliza-se variáveis instrumentais para assegurar consistência da estimativa. A partir da estimativa  $\hat{\theta}^1$ , estima-se a saída sem ruído  $\hat{x}_t$  através (5.33) e a perturbação  $\hat{\xi}_t$  através de (5.35).

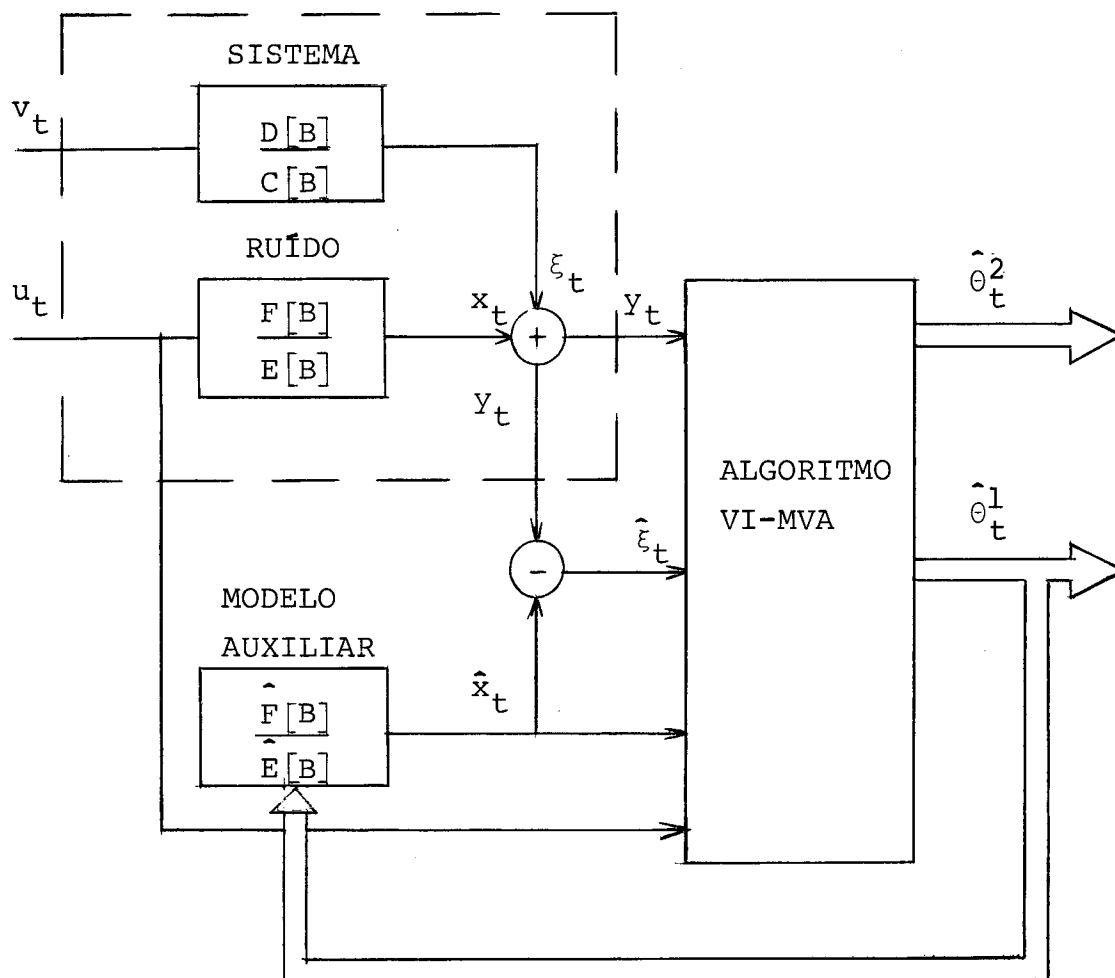


Figura 5.6

No segundo estágio da estimação, a estimativa  $\hat{\xi}_t$  anteriormente obtida serve como entrada para o algoritmo de máxima verossimilhança aproximada, uma outra modificação do algoritmo de mínimos quadrados recursivos, que proporciona uma estimativa consistente do vetor  $\theta^2$ .

## V.2.4 - O Algoritmo de Variáveis Instrumentais (VI)

A utilização de variáveis instrumentais no algoritmo recursivo de mínimos quadrados, prende-se ao fato de (5.38) representar um modelo estrutural em vez de um modelo de regressão linear, devido ao "ruído"  $m_t$  não ser independente do vetor  $z_t$ . Com isto, a aplicação pura e simples do algoritmo recursivo produz estimativas assintoticamente tendenciosas.

Seja o vetor sem ruído:

$$\tilde{x}_i \triangleq [x_{i-1}, \dots, x_{i-N}, u_t, \dots, u_{t-N}]$$

e  $\tilde{x}_i$  um vetor de variáveis instrumentais escolhido com as seguintes características: totalmente não correlacionado com a perturbação  $\xi_i$  e o mais altamente correlacionado possível com o vetor  $x_i$ . A estimativa não recursiva de  $\theta_1$  pode ser obtida de maneira similar a (3.6):

$$(5.40) \quad \hat{\theta}_t^1 = \hat{P}_t \hat{b}_t$$

onde:

$$\hat{P}_t = \left[ \begin{array}{c} t \\ \sum_{i=1} \tilde{x}_i' z_i \end{array} \right]^{-1}$$

$$\hat{b}_t = \frac{t}{\sum_{i=1} \tilde{x}_i' y_i}$$

A semelhança entre (5.40) e (3.6) sugere que se desenvolva um algoritmo recursivo para a estimação de  $\theta_t^1$  utilizando um raciocínio análogo ao do item III.2.1. Chega-se, então, ao algoritmo:

$$\hat{\theta}_t^1 = \hat{\theta}_{t-1}^1 - \hat{P}_{t-1} \hat{x}_t' \left[ 1 + z_t \hat{P}_{t-1} \hat{x}_t' \right]^{-1} \left[ z_t \hat{\theta}_{t-1}^1 - y_t \right]$$

$$\hat{P}_t = \hat{P}_{t-1} - \hat{P}_{t-1} \hat{x}_t' \left[ 1 + z_t \hat{P}_{t-1} \hat{x}_t' \right]^{-1} z_t \hat{P}_{t-1}$$

A escolha de  $\hat{x}_t$  é baseada nas características físicas do sistema. A partir de uma estimativa a priori do vetor  $\theta^1$ , constrói-se um modelo auxiliar do sistema que gera uma saída  $\hat{x}_t$  considerada uma estimativa inicial de  $x_t$ . O vetor de variáveis instrumentais é então definido:

$$\hat{x}_t = \left[ -\hat{x}_{t-1}, \dots, -\hat{x}_{t-N}, u_t, \dots, u_{t-N} \right]$$

À medida que a estimação prossegue, a estimativa / do vetor  $\theta^1$  se torna mais confiável, possibilitando a atualização do modelo auxiliar. Com isto, melhora-se a "qualidade" das variáveis instrumentais e, conseqüentemente, a eficiência das estimativas.

É interessante notar que o algoritmo VI recursivo pode ser considerado como uma extensão dos métodos de mínimos quadrados de dois estágios da Econometria.

## V.2.5 - O Algoritmo de Máxima Verossimilhança Aproximada (MVA)

A aplicação do método de variáveis instrumentais gera uma saída  $\hat{x}_t$  do modelo auxiliar que, após a convergência da estimativa  $\hat{\theta}^1$ , pode ser considerada uma estimativa da saída sem ruído  $x_t$ . Como consequência, obtém-se através de (5.35) uma estimativa  $\hat{\xi}_t$  do ruído do processo.

$$\hat{\xi}_t = y_t - \hat{x}_t$$

De posse de uma estimativa  $\hat{v}_t$  do ruído branco  $v_t$ , torna-se plausível a obtenção de uma estimativa consistente do vetor de perturbação do sistema  $\theta^2$  utilizando mínimos quadrados em (5.39) com  $\xi_t$  e  $v_t$  substituídos por  $\hat{\xi}_t$  e  $\hat{v}_t$ . Da mesma forma que no caso de variáveis instrumentais, uma versão recursiva pode ser obtida. O algoritmo de máxima verossimilhança aproximada produz uma estimativa  $\hat{\theta}^2$  através do seguinte algoritmo recursivo:

$$\hat{\theta}_t^2 = \hat{\theta}_{t-1}^2 - P_{t-1}^N \left[ 1 + \hat{n}_t P_{t-1}^N \hat{n}_t' \right]^{-1} \begin{bmatrix} \hat{n}_t & \hat{\theta}_{t-1}^2 - \hat{\xi}_t \end{bmatrix}$$

$$P_t^N = P_{t-1}^N - P_{t-1}^N \hat{n}_t' \left[ 1 + \hat{n}_t P_{t-1}^N \hat{n}_t' \right]^{-1} \hat{n}_t P_{t-1}^N$$

onde:

$$\hat{n}_t = \begin{bmatrix} -\hat{\xi}_{t-1}, \dots, -\hat{\xi}_{t-N}, \hat{v}_{t-1}, \dots, \hat{v}_{t-N} \end{bmatrix}$$

A partir da estimativa  $\hat{\theta}_t^2$ , pode-se estimar  $v_t$  por

(5.39):

$$\hat{v}_t = \hat{\xi}_t - \hat{n}_t \hat{\theta}_t^2$$

Este valor é utilizado na estimativa  $\hat{\theta}_{t+1}^2$  e assim sucessivamente.

O algoritmo VI-MVA constitui um método bastante flexível de análise de modelos de função de transferência de séries temporais, já tendo sido aplicado com sucesso em processos químicos, hidrológicos e macroeconômicos.

Young [17] o aplicou a um modelo de combustão de um gás em que a entrada  $u_t$  representa a taxa de gás utilizada e a saída  $y_t$ , a concentração de dióxido de carbono resultante. Box & Jenkins [3] identificou o seguinte modelo de função de transferência entre as duas variáveis:

$$y_t = \frac{f_3 B^3 + f_4 B^4 + f_5 B^5}{1 + e_1 B} u_t + \frac{1}{1 + c_1 B + c_2 B^2} v_t$$

O algoritmo recursivo possibilitou duas conclusões não reveladas pela análise tradicional:

- (i) a pequena diferença entre a saída do modelo auxiliar  $\hat{x}_t$  e a observação  $y_t$  sugere que o sistema é praticamente sem ruído.



- (ii) as estimativas obtidas indicam uma possível não estacionariedade dos parâmetros, particularmente no final dos dados.

Assim, o algoritmo recursivo permite a percepção do fenômeno, à medida que ele ocorre, enquanto na análise convencional sua ocorrência pode passar despercebida, conduzindo a conclusões erradas.

## CAPÍTULO VI

### APLICAÇÃO DO MÉTODO BAYESIANO

#### VI.1 - Introdução: Características da Série

A parte prática do trabalho consta da aplicação do método Bayesiano de previsão a curto prazo à série de consumo médio mensal de energia elétrica das Centrais Elétricas do Espírito Santo S.A. (ESCELSA). A escolha deve-se ao fato desta série apresentar algumas transiências, decorrentes de consumos anormais, e mudanças de nível e inclinação, decorrentes da entrada em operação de usinas de pelotização e siderurgias cujas cargas, devido ao relativo pequeno consumo de energia elétrica no estado, exercem sensível influência que se propaga no futuro: inúmeras indústrias de base se criam expandindo o número de empregos, desenvolvendo a região e melhorando o padrão de vida da população, tudo concorrendo para o aumento do consumo. Isto propicia a aplicação a pleno potencial do método: o conhecimento da data e do aumento de demanda esperados permite prover informações a priori e sentir seu desempenho nestas situações. Por outro lado, a ausência de informações a priori quando alguma mudança é esperada permite testar a percepção e a adaptabilidade do método diante de situações anormais. Em suma, a série em estudo possibilita avaliar o desempenho do método diante de alterações esperadas, inesperadas e, logicamente, durante o período de evolução normal do processo.

Desenvolveu-se um programa escrito na linguagem PL/I. Especificado o modelo representativo do processo, o programa gera a primeira previsão correspondente à primeira observação e prossegue ponto a ponto numa espécie de simulação do processo. A comparação entre os valores previstos e observados ao longo do tempo permite avaliar o desempenho do método. Isto pode ser feito através de representações gráficas e de critérios quantitativos como o erro médio quadrático, explicados no item IV.3.

## VI.2 - Modelo Utilizado

As características da série sugerem para representá-la um modelo paramétrico multiplicativo da forma:

$$Y_t = \mu_t \rho_{m(t),t} v_t$$

$$\mu_t = \mu_{t-1} \beta_t \sigma_{\mu_t}$$

$$\beta_t = \beta_{t-1} \sigma_{\beta_t}$$

$$\rho_{i,t} = \rho_{i,t-1} \sigma_{\rho_{i,t}} \quad , \quad i = \overline{1,12}$$

onde os parâmetros podem ser interpretados:

$\mu_t$  - nível do processo

$\beta_t$  - inclinação do processo

$\rho_{i,t}$  - fator sazonal do "mês"  $i$

$\sigma_{\mu_t}, \sigma_{\beta_t}, \sigma_{\rho_{i,t}}$  - distúrbios de nível, inclinação e sazonalidade com distribuição log-normal.

$v_t$  - ruído da observação com distribuição log-normal.

Este modelo não linear, composto de 14 parâmetros, pode ser transformado no modelo de crescimento linear sazonal do item V.1.6.2, trabalhando-se com o logaritmo de suas equações.

A incerteza acerca do modelo em curso é descrita através de um modelo de múltiplos processos composto de quatro estados, os mesmos do item V.1.7.2.2: normal, mudança de nível, mudança de inclinação e transiente. A ocorrência destes eventos, / conforme descrito naquele item, é reconhecida pelos modelos de estado que os representam. Opera-se, portanto, num contexto de múltiplos estados com probabilidades constantes a priori de ocorrência. Estas probabilidades, à medida que se recebe novas observações, são utilizadas em conjunto com outras informações para determinar as probabilidades de estado a posteriori, nas quais se baseiam as estimativas dos valores correntes dos parâmetros e as previsões.

A importância dos estados normal e transiente decorre do fato de que se o processo evolui normalmente, o estado normal assegura previsões sem flutuações violentas; quando surge uma observação "anormal" solitária, o estado transiente a reconhece e

o sistema não a leva em consideração.

### VI.3 - Informação de Entrada: Critério de Escolha dos Parâmetros

O sistema necessita de dois tipos de informações de entrada para funcionar:

- (i) a distribuição inicial dos parâmetros, especificada pela média  $\hat{\theta}_0$  e pela matriz de covariância associada  $P_0^*$ .
- (ii) os modelos de estado, especificados pelas matrizes de covariância do ruído da observação  $V_t$  e da perturbação dos parâmetros  $W_t$ .

Na modelagem utilizada (modelo de crescimento linear sazonal em combinação com um modelo de múltiplos processos de quatro estados), isto importa na especificação de 18 parâmetros de entrada (não confundir com os parâmetros do processo), que são devidamente manipulados pelo programa, dando origem às diversas distribuições necessárias.

É bastante lógico supor que alguns subconjuntos de parâmetros de entrada exerçam maior influência no desempenho do sistema, isto é, alguns parâmetros sejam mais significativos que outros. A fim de escolher o conjunto de valores mais representativo do processo, algum critério deve ser estabelecido; optou-se pelos dois apontados na introdução do capítulo, por parecerem os mais lógicos:

- (i) o erro médio quadrático entre a observação e a previsão um passo à frente, que fornece uma medida quantitativa dos desvios entre as duas, mas não permite sentir a adaptabilidade diante de mudanças bruscas.
- (ii) a representação gráfica da observação e da previsão um passo à frente, que permite visualizar o comportamento diante de mudanças bruscas.

Devido ao tempo proibitivo de computador, não se pode efetuar uma análise de sensibilidade completa, a fim de encontrar um conjunto de valores dos parâmetros de entrada que garantisse um ótimo global segundo estes critérios. Procedeu-se então da seguinte maneira: variava-se alguns parâmetros, enquanto os demais eram mantidos constantes. Após uma análise detalhada dos resultados, determinava-se uma região em que os valores assumidos por estes parâmetros eram mais significativos. Dentro desta região, escolhia-se 2 conjuntos de valores. Em seguida, variava-se outros parâmetros mantendo-se os demais constantes. Repetia-se o procedimento anterior com os parâmetros constantes assumindo os valores de cada conjunto anteriormente determinado e assim sucessivamente. Este tipo de busca obviamente não garante o ótimo global. Todavia, se a análise de sensibilidade for efetuada com atenção, atinge-se certamente um ponto de "quase" otimalidade a partir do qual talvez não compense mais investir esforços no sentido de melhorar o desempenho do sistema.

Convém salientar que todos os valores especificados a seguir se referem ao modelo multiplicativo do processo. As transformações para o modelo de crescimento linear sazonal são efetuadas automaticamente pelo programa.

### VI.3.1 - Especificação da Distribuição Inicial dos Parâmetros

Para descrever a incerteza inicial dos parâmetros, são fornecidos a mediana e os limites inferior e superior do intervalo de 95% de confiança. Tomando-se o logaritmo da mediana, determina-se a média  $\hat{\theta}_0$  e elevando-se ao quadrado o quociente entre a diferença dos logaritmos dos limites superior e inferior e a constante 3.96, determina-se de modo aproximado a matriz de covariância  $P_0^*$ . O fato dos limites serem arbitrados torna inconsequente a maneira aproximada da determinação de  $P_0^*$ .

Os valores utilizados foram:

PARÂMETRO	INFERIOR	MEDIANA	SUPERIOR
$\mu_0$	8500	9000	10000
$\beta_0$	0.95	1.022	1.111
$\rho_{i,0}$	0.5	1.0	2.0

A assimetria dos limites em relação à mediana decorre da forma assimétrica da distribuição log-normal, que possui uma cauda à direita.

## VI.3.2 - Especificação dos Modelos de Estado

O primeiro passo é a especificação do desvio padrão do ruído da observação (e conseqüentemente a variância). Conforme comentado na teoria, o fato de considerá-lo variável no tempo torna o modelo muito mais representativo da realidade. Como o parâmetro mais significativo do processo é seu nível, optou-se por expressá-lo como função do nível do processo em cada instante. Assim (em termos de variância):

$$V_t = (C \mu_t^P)^2$$

onde C e P são constantes. Esta formulação é bastante flexível. Por exemplo, P = 0 significa desvio padrão constante e P = 1, proporcional ao nível.

Harrison e Stevens sugerem que se especifique um conjunto de possíveis valores para a constante C, a cada um atribuído uma probabilidade a priori de ocorrência, determinando-se, desta forma, um valor mediano inicial  $\bar{C}_0$ . À medida que as novas observações vão sendo recebidas, atualiza-se as probabilidades e novos valores  $\bar{C}_t$  são obtidos. Isto caracteriza exatamente o modelo de múltiplos processos do primeiro tipo no qual um único modelo desconhecido representa o processo em todos os instantes. Entretanto, entendeu-se ser este procedimento desnecessário para a série estudada e optou-se pela determinação da constante segundo os critérios de análise expostos. Com isto, chegou-se à relação:



$$V_t = (0.08 \bar{\mu}_t)^2$$

Uma vez especificada a variância do ruído do processo, resta a especificação das variâncias das perturbações dos parâmetros e das probabilidades de estado, para a definição dos modelos de estado. As variâncias da perturbação de nível e do transiente são também consideradas funções do nível do processo, não só para torná-las mais representativas da realidade, como também para mantê-las coerentes com os valores da variância do ruído, facilitando o método na tarefa de identificar corretamente as diversas situações que surgem. Em relação à inclinação e aos fatores sazonais, a ordem de grandeza de seus valores é bastante inferior à do nível do processo. Não existe possibilidade do sistema confundir, por exemplo, o estado mudança de nível com o de inclinação, ao contrário dos estados mudança de nível, transiente e normal. Por estas razões, as variâncias das perturbações de inclinação e fatores sazonais não necessitam ser relacionadas ao nível do processo.

Por conveniência, especifica-se as perturbações para qualquer modelo de estado  $j (j=\overline{1, N})$  da seguinte forma:

$$\text{var } (v_t | j) = V_t + \left[ a(j) \bar{\mu}_t \right]^2$$

$$\text{var } (\mu_t | j) = \left[ b(j) \bar{\mu}_t \right]^2$$

$$\text{var } (\beta_t | j) = \left[ c(j) \right]^2$$

$$\text{var} (\rho_{i,t}|j) = \begin{cases} d^2(j) & , i = m(t) \\ d^2(j)/11 & , i \neq m(t) \end{cases}$$

A partir desta tabela e das probabilidades de estado, pode-se especificar os quatro modelos de estado utilizados:

MODELO DE ESTADO	a(j)	b(j)	c(j)	d(j)	$\pi(j)$
1- normal	0	0	0	0.01	0.893
2- mudança de nível	0	0.35	0	0	0.009
3- mudança de inclinação	0	0	0.008	0	0.009
4- transiente	0.32	0	0	0	0.089

As transformações das variâncias das perturbações do modelo multiplicativo para o aditivo são feitas através da relação entre as variâncias das distribuições log-normal e normal:

$$\sigma_N^2 = \log \left( \frac{1 + \sqrt{1 + 4\sigma_L^2}}{2} \right)$$

onde

$\sigma_N^2 \rightarrow$  variância da distribuição normal

$\sigma_L^2 \rightarrow$  variância da distribuição log-normal

Algumas considerações sobre os modelos de estado devem ser feitas:

- (i) é preferível tratar inicialmente uma mudança de nível como transiente do que o caso contrário, devido à mudança de nível influenciar os valores futuros da série. As probabilidades de estado utilizadas ( $\pi(4) \approx 10\pi(2)$ ) e as magnitudes das variâncias ( $0.32 > 0.25$ ) garantem que uma ocorrência anormal é tratada como um transiente até que os valores subsequentes da série confirmem a existência de uma mudança de nível.
  
- (ii) a variância da perturbação do fator sazonal deve ser pequena pelo fato da alteração do fator estar incluída na especificação do modelo do estado normal.
  
- (iii) a probabilidade de ocorrência do estado normal é cerca de 10 vezes maior que a soma dos outros, fazendo com que ele prevaleça na maior parte do tempo.

Com isto, ficam definidos os 18 parâmetros de entrada necessários ao funcionamento do método: 9 para a distribuição inicial dos parâmetros do processo, 2 para a variância do ruído, 3 para a variância das perturbações, 3 para as probabilidades de estado (a quarta é o complemento) e 1 para a variância do transiente.

#### VI.4 - Lógica do Programa de Múltiplos Estados

Para facilitar a compreensão da lógica do programa, ao lidar com o modelo de múltiplos estados, considere-se inicialmente a informação associada a um modelo de estado particular

num determinado instante  $(t-1)$  após a observação  $y_{t-1}$ . A partir da tabela do item VI.3.2, constrói-se 4 matrizes  $W_{t-1}$  e 4 escalares  $V_{t-1}$ , cada par correspondendo a um estado. O conhecimento acerca dos parâmetros é expresso através de 4 vetores  $\hat{\theta}_{t-1}$  com 4 matrizes de covariância associadas  $P_{t-1}^*$ , cada par correspondendo a um estado. Além disso, 4 probabilidades independentes do histórico expressam as possibilidades de uma transição para cada estado no instante seguinte.

Ao passar do instante  $(t-1)$  para o instante  $t$ , após a observação  $y_t$  se tornar disponível, o programa investiga as 16 probabilidades de transição  $(i,j)$  do processo estar no estado  $j$  no instante  $t$  tendo estado anteriormente no estado  $i$  (no instante  $t-1$ ). Para cada transição, o filtro de Kalman produz um vetor de média dos parâmetros e uma matriz de covariância associada, denotados por  $\hat{\theta}^{(i,j)}$  e  $P^{*(i,j)}$ , utilizando as matrizes correspondentes:  $\hat{\theta}_{t-1}^{(i)}$ ,  $P_{t-1}^{*(i)}$ ,  $V_t^{(j)}$  e  $W_t^{(j)}$  e a verossimilhança da ocorrência é calculada. Neste ponto se inicia o processo de condensação de informações: para cada estado  $j$ ,  $\hat{\theta}^{(1,j)}$ , ...,  $\hat{\theta}^{(4,j)}$ ,  $P^{*(1,j)}$ , ...,  $P^{*(4,j)}$  são condensados respectivamente em  $\hat{\theta}^{(j)}$  e  $P^{*(j)}$ , segundo as equações do item V.1.7.2.3. Da mesma forma, as verossimilhanças são normalizadas, transformadas em probabilidades de transição e condensadas em probabilidades de estado a posteriori.

Esta informação a posteriori condensada ( $\hat{\theta}^{(j)}$ ,  $P^{*(j)}$ ,  $p^{(j)}$ ,  $j = \overline{1,4}$ ) é utilizada para a previsão dos valores futuros da série, realizada segundo alguma regra estabelecida. A mais natural combina as previsões de cada modelo de estado, partindo do

pressuposto que o processo permanece no estado correspondente nos instantes subsequentes; a previsão é uma média das previsões dos modelos de estado ponderada pelas probabilidades de estado a posteriori. Outras duas regras possíveis são: considerar somente o estado de maior probabilidade a posteriori e considerar a média ponderada dos estados cujas probabilidades a posteriori excedam 0.1.

#### VI.5 - Intervenções do Analista

A série em estudo apresenta cinco mudanças sensíveis de nível acompanhadas de mudanças de inclinação. Elas se situam por volta de maio de 1970, agosto de 1970, abril e maio de 1973, setembro de 1975 e agosto a outubro de 1977. Cada uma corresponde à entrada em operação de siderurgia ou usina de pelotização cujas cargas são bastante representativas em relação ao consumo do estado. O conhecimento a priori destes fatos possibilitou a comunicação ao sistema das alterações esperadas nos parâmetros, explorando-se uma das características principais do método. Assim, nestes meses, acrescentou-se as estimativas fornecidas pelo filtro de Kalman os seguintes valores:

<u>OBS</u>	<u><math>\mu_t</math></u>	<u>Var(<math>\mu_t</math>)</u>	<u><math>\beta_t</math></u>
mai/70	5000	4000000	0.005
ago/70	5000	4000000	0
abr/73	5000	9000000	0.001
mai/73	7000	9000000	0
jun/73	6000	9000000	0
ago/73	5000	9000000	0
ago/77	5000	25000000	0.001
set/77	5000	25000000	0
out/77	5000	25000000	0

Em relação às mudanças de setembro de 1975, não foram fornecidas informações a priori, com o objetivo de testar a adaptabilidade automática do método.

#### VI.6 - Desempenho do Modelo

Para os valores dos parâmetros de entrada especificados no item VI.3, o erro médio quadrático entre a observação e a previsão um passo à frente foi de 2050, cerca de 4%, considerando que o valor médio da série se situa em torno de 50000. Este resultado foi obtido, utilizando a terceira regra de previsão: considerando somente os estados com probabilidade de estado a posteriori maior que 0.1.

A representação gráfica correspondente, em anexo, apresenta alguns aspectos interessantes:

- outubro e novembro de 1970 se constituem em dois transientes seguidos. A primeira observação o modelo interpreta como transiente. Entretanto, após a segunda, ele se convence que houve uma mudança de nível e fornece uma previsão para dezembro, adaptada ao novo "nível". Após a observação de dezembro, ele percebe o engano e retorna ao nível real da série.
  
- em novembro de 1977, a presença de um transiente origina uma previsão mais alta para dezembro, talvez antevendo uma mudança de nível. Entretanto, a observação de dezembro confirma o transiente anterior e a previsão para janeiro e fevereiro é feita dentro dos níveis normais da série. Mas em fevereiro, uma nova observação alta sugere que uma mudança de nível se avizinha, originando uma previsão mais elevada para março.
  
- por volta de agosto de 1975, ocorrem simultaneamente uma mudança de inclinação e uma ligeira mudança de nível. O sistema rapidamente percebe e se adapta à mudança de inclinação, mas não consegue acompanhar a contento a mudança de nível, originando previsões "paralelas" às observações. O mesmo fato ocorre por volta do meio de 1974.
  
- em março de 1977, uma observação baixa induziu o sistema a anular o crescimento, originando uma previsão para abril praticamente idêntica à anterior. Entretanto, a observação de abril sugere uma transiência ou uma mudança de nível. O sistema, então, faz as previsões subsequentes considerando este fato.

- as variações em torno dos níveis do processo não estão bem acompanhadas. Isto sugere duas interpretações: má especificação dos parâmetros sazonais ou o fato da série não apresentar características sazonais marcantes, como se supunha a priori.

#### VI.7 - Considerações Finais

- (i) embora o desempenho como um todo possa ser considerado bastante satisfatório, principalmente devido à irregularidade da série, alguns aspectos podem ser melhorados. O problema das variações em torno dos níveis do processo talvez possa ser superado pela introdução de um parâmetro adicional no modelo, relacionado à autocorrelação dos erros entre as observações e as previsões um passo à frente ou entre as observações e os níveis estimados.
- (ii) mudanças de nível e inclinação simultâneas não estão sendo percebidas prontamente. Geralmente, só a mudança de inclinação é rapidamente notada. Isto talvez possa ser superado através de uma modificação nas proporções relativas entre as probabilidades de estados e entre os fatores de variância.
- (iii) o efeito dos valores da média e da variância iniciais dos parâmetros, desde que coerentes, logo desaparece. Em pouco tempo, as previsões passam a independê-los e, quanto maior a variância, mais rapidamente isto acontece. Em função disto, estes parâmetros de entrada são os menos impor-



tantes no desempenho do sistema, quando se possui uma quantidade razoável de observações.

- (iv) das três regras de previsão do item VI.4, a primeira e a terceira apresentam melhores resultados, dependendo do conjunto de valores utilizado para os parâmetros de entrada. A segunda, entretanto, nenhuma vez supera as demais.
- (v) os quatro estados definidos parecem suficientes para representar a incerteza do processo. Poder-se-ia pensar num quinto estado, por exemplo, estado de mudança de sazonalidade. Para esta série, entretanto, nada acrescentaria de útil.
- (vi) em séries com padrões sazonais estáveis talvez seja mais proveitoso trabalhar com sazonalidade determinística, estimando-se os fatores sazonais a partir de um subconjunto da série histórica. No modelo utilizado, isto representa uma redução de 14 para 2 parâmetros, bastante significativa em termos de esforço computacional. Logicamente, o que deve orientar a escolha entre as duas abordagens é a característica sazonal da série em estudo.

## CAPÍTULO VII

### CONCLUSÕES E SUGESTÕES

O trabalho visa, basicamente, a formulação teórica unificada de modelos lineares de regressão e séries temporais que possuem como característica comum uma estrutura recursiva de estimação. Neste contexto, ressaltam também como características comuns a representação Markoviana dos processos e o algoritmo de estimação do filtro de Kalman. Em função disto, a apresentação separada dos modelos de regressão e de séries temporais deve-se muito mais a razões históricas e práticas do que teóricas; eles poderiam ser perfeitamente considerados casos particulares de um mesmo modelo linear geral.

A estimação recursiva dos parâmetros possibilita uma estruturação dinâmica que, ao contrário do tipo convencional de análise, admite a hipótese deles variarem no tempo e, quando convenientemente especificada, permite a rápida percepção destas variações. Este aspecto é importantíssimo: a vida em si é dinâmica e todo modelo que de alguma maneira a representa, deve também ser dinâmico, para não correr o risco de se tornar inadequado. Outro aspecto importante é o fato dos métodos não recursivos geralmente necessitarem de uma quantidade considerável de dados para operar, o que, na prática, muitas vezes, não se consegue. Neste particular, assume destaque o método Bayesiano de previsão de séries tem

porais de Harrison e Stevens, extremamente dinâmico, e que combina, em bases sólidas, teoria com prática. Ao contrário dos métodos convencionais, que só consideram as informações históricas, o método Bayesiano leva em conta uma fonte de informações preciosa e quase sempre negligenciada: o indivíduo que conhece o sistema e sabe da consequência de fatos externos no desenvolvimento futuro do processo. Outra característica interessante do método é a idéia da incerteza em relação ao modelo em curso em cada instante. Ela propicia grande flexibilidade e se constitui num campo de pesquisa a merecer atenção.

Fica como idéia geral do trabalho a ênfase aos modelos voltados à descrição da estrutura interna dos processos que representam, de certa forma análogos aos modelos de espaço de espaço de estado da Teoria de Controles. Eles permitem atribuir significado físico aos parâmetros, facilitando a visualização e a compreensão do mecanismo evolutivo do processo e a comunicação sistema-analista. Ultimamente, a linha de pesquisa em séries temporais tem-se orientado para o estudo de situações anormais e de mudanças de comportamento. Estes modelos se prestam perfeitamente a este tipo de enfoque.

Como sugestões para desenvolvimento prático do trabalho, ficam:

- (i) a insistência na aplicação do método Bayesiano à série de demanda de energia elétrica da Escelsa, visando a superar

os problemas apontados nos itens VI.6 e VI.7, principalmente o das variações em torno do nível do processo.

- (ii) aplicação de uma técnica de otimização não-linear para determinação do conjunto de valores ótimos dos parâmetros de entrada, isto é, dos modelos de estado e da distribuição inicial dos parâmetros do processo ótimos, segundo o critério de mínimo erro quadrático. Isto permite verificar se a diferença entre este procedimento e a análise de sensibilidade efetuada é significativa.
- (iii) sua aplicação a outras séries temporais, para uma conclusão mais definitiva sobre a sua real eficácia. Em séries sazonais, utilizar o modelo de funções periódicas do item VI.1.6.3 para avaliar se a simplificação introduzida compensa a possível perda de eficiência e se esta perda realmente se verifica.
- (iv) a comparação entre o seu desempenho e o de outros métodos, notadamente o Box & Jenkins. Isto deve ser efetuado em séries "bem" comportadas e "mal" comportadas.
- (v) aplicação do método VI-MVA a modelos de função transferência de séries temporais. Este método parece possuir grande potencial, tornando-se importante na previsão de séries cujas observações dependam dos valores assumidos por outra série em instantes anteriores. As vantagens desta abordagem sobre a de Box & Jenkins já foram apontadas e sua apli

cação a séries econômicas, cujo dinamismo torna frequentes variações nos parâmetros, deve ser feita. Ressalte-se que o método pode ser aplicado tanto com objetivos "on-line", para uma contínua atualização dos parâmetros, quanto "off-line", em que repetidas aplicações sobre subconjuntos de dados indicam o momento de se terminar a estimação: quando a diferença entre estimativas subsequentes for desprezível.

Como sugestões para desenvolvimento teórico do trabalho, ficam:

- (i) a generalização do método Bayesiano de Harrison e Stevens com o relaxamento das hipóteses de normalidade e linearidade. Isto pode ser feito, seguindo a linha de Souza e Harrison [16], que utiliza princípios de Teoria da Informação.
- (ii) a generalização do método VI-MVA para o caso de múltiplas entradas.

REFERÊNCIAS

- 1 - Akaike, H.- "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes, Ann.Inst.Statist.Maths, 26, 363-387 (1974)
- 2 - Blight, J.B.N.- "Recursive Solutions for the Estimation of a Stochastic Parameter" - Journal of American Statistical Association, vol. 69, nº 346, Theory and Methods Section, junho (1974).
- 3 - Box, G.E.P. e Jenkins, G.M., "Time Series Analysis, Forecasting and Control", Holden Day, San Francisco (1970).
- 4 - Harrison, P.J. e Stevens, C.F. - "Bayesian Approach to Short-Term Forecasting" - Oper Res. Quartely, 22, Nº 4, pg. 341-362 (1971)
- 5 - Harrison, P.J. e Stevens, C.F. - "Bayesian Forecasting" (with discussion), Journal of Statistical Society, Serie B, 38 (1976)
- 6 - Harrison, P.J. e Stevens, C.F. - Bayesian Forecasting in Action: Case Studies" - Warwick Statistical Report Nº 14 (1975).
- 7 - Ho, Y.C. e Lee, R.C.K. - "A Bayesian Approach to Problems in Stochastic Estimation and Control" - Trans IEEE, AC-9, pg.

- 8 - Kalman, R.E. - "New Methods and Results in Linear Prediction and Filtering Theory" - Research Institute for Advanced Studies, Baltimore, Md., nº 61-1 (1960).
- 9 - Ledolter, J. - "A Recursive Approach to Parameter Estimation in Regression and Time Series Models" - Technical Summary Report N. 1853, University of Wisconsin-Madison, Mathematics Research Center, maio (1978).
- 10 - Lee, R.C.K. - "Optimal Estimation, Identification and Control" - MIT Press, Cambridge, Massachusetts (1964).
- 11 - Leonard, T. e Harrison, P.J. - "Bayesian Updating for the Steady State Kalman Filter", Warwick Statistical Report (1976).
- 12 - MacGregor, J.F. - "State Variable Approach to Time Series Representation and Forecasting" - National Technical Information Service, AD 757 573, julho (1972).
- 13 - Mendes, L.E.P. e Pereira, B. de B. - "Método Bayesiano de Previsão de Séries Temporais: Aplicação a Uma Série de Demanda de Energia Elétrica", 3º Simpósio Nacional de Probabilidade e Estatística, USP (1978).
- 14 - Plackett, R.L. - "Some Theorems in Least Square" - Biometrika, pg. 149-157 (1950).

- 15 - Raifa, H. e Schlaifer, R. - "Applied Statistical Decision Theory" - Harvard University Press, Cambridge, Mass. (1961).
- 16 - Souza, R.C. e Harrison, P.J. - "Bayesian-Entropy Forecasting Approach", Warwick Statistical Report (1978).
- 17 - Young, P.C. - "Applying Parameter Estimation to Dynamic Systems", I e II, Journal of Control Engineering, vol. 16, nºs 10 e 11, pags. 119-125 e 118-124, outubro e novembro (1969).
- 18 - Young, P.C. - "Recursive Approaches to Time Series Analysis" Bull Inst. Maths. and Applicns., 10, pg. 209-224 (1974).
- 19 - Zellner, A. - "An Introduction to Bayesian Inference in Econometrics", New York, Wiley (1971).
- 20 - Feldstein, M.S. - "The Error of Forecasting Econometrics Models When the Forecasting-Period Exogenous Variables are Stochastic", Econometrica, 39, 55-60 (1971).



