

MODELOS DE DEFASAGENS

DISTRIBUIDAS

SAMUEL HAZZAN

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR:

*Prof. Dr. CLÓVIS DE ARAUJO PERES*

- SÃO PAULO, DEZEMBRO DE 1978 -

## AGRADECIMENTOS

Ao Professor Doutor Clóvis de Araujo Peres, pela dedicação e incentivo durante a elaboração desta dissertação.

Aos Professores do Instituto de Matemática e Estatística da Universidade de São Paulo responsáveis pela minha formação.

Ao Professor Doutor Adroaldo Moura da Silva, da Fundação Instituto de Pesquisas Econômicas da Universidade de São Paulo pela sugestão e auxílio dados no exemplo de aplicação.

Ao Sr. João Baptista Esteves de Oliveira, pelo eficiente trabalho de datilografia.

## ÍNDICE

CAP. 1 - INTRODUÇÃO . . . . .	1
CAP. 2 - MODELOS DE DEFASAGENS DISTRIBUÍDAS . . . . .	8
2.1 - Conceitos Básicos . . . . .	8
2.2 - Modelos com Ponderação à Priori dos Parâmetros . . . . .	9
2.3 - Modelo Polinomial de Almon . . . . .	14
2.4 - Modelo de Koyck ou do declínio Geométrico . . . . .	22
2.5 - Modelo das Expectativas Adaptadas . . . . .	26
2.6 - Modelo dos Ajustamentos Parciais . . . . .	30
2.7 - Modelo de Solow ou das defasagens de Pascal . . . . .	38
CAP. 3 - MÉTODOS DE ESTIMAÇÃO . . . . .	41
3.1 - Introdução . . . . .	41
3.2 - Modelos de Koyck, Cagan e Nerlove . . . . .	42
3.3 - Modelo de Solow ou das Defasagens de Pascal . . . . .	63
3.4 - Observações Finais . . . . .	70
CAP. 4 - EXEMPLO DE APLICAÇÃO . . . . .	74
APÊNDICE 1 - PROPRIEDADES ASSINTÓTICAS DOS ESTIMADORES . . . . .	82
APÊNDICE 2 - O USO DE VARIÁVEIS INSTRUMENTAIS . . . . .	89
REFERÊNCIAS BIBLIOGRÁFICAS . . . . .	95

## CAPÍTULO 1

### INTRODUÇÃO

Em muitos problemas científicos e em particular em Economia, a variável em estudo no modelo (também chamada variável dependente ou endôgena) não reage imediatamente às variações das outras variáveis explicativas do modelo (também chamadas variáveis independentes ou exôgenas), mas a reação se processa de maneira gradual. Tal fato levou os cientistas a formularem modelos em cuja representação matemática as variáveis envolvidas comparecessem em diferentes períodos de tempo. Diz-se então que ocorrem defasagens entre as variáveis do modelo, e a expressão matemática do modelo é frequentemente denominada de "modelo de defasagens distribuídas".

Os primeiros trabalhos envolvendo modelos com defasagens distribuídas foram desenvolvidas em meados de 1930 sendo Irving Fisher um dos precursores, Além de G.C.Evans e C.F.Roos; todos eles utilizaram modelos com defasagens distribuídas em problemas ligados a Economia.

Em meados de 1940, trabalhos importantes foram realizados por T.Koopmans em estudos sobre ciclos econômicos. Outros trabalhos importantes surgiram também em meados de

1940 com Franz Alt [02] no estudo do consumo doméstico de óleo combustível.

Recentemente os modelos de defasagens distribuídas têm adquirido popularidade, como técnica de análise econômétrica, sobretudo graças aos trabalhos e estudo de Koyck [21], Cagan [--], Nerlove [30], Solow [36], Griliches [10] e Zellner [42] entre outros.

Vejamos a seguir alguns exemplos introdutórios onde ocorrem defasagens entre as variáveis dos modelos.

A função consumo

Há modelos segundo os quais admite-se que a função consumo no período  $t$  ( $C_t$ ) é explicada por valores passados da mesma ( $C_{t-1}, C_{t-2}, \dots$ ) devido ao que se denomina em Economia, "persistência de hábito", por valores passados e presente da renda disponível ( $Y_t, Y_{t-1}, Y_{t-2}, \dots$ ) e eventualmente, por outras variáveis ( $X_1, X_2, \dots$ ).

Simbolicamente

$$C_t = f(C_{t-1}, C_{t-2}, \dots, Y_t, Y_{t-1}, \dots, X_{1t}, X_{2t}, \dots).$$

A demanda por bens duráveis

Admite-se que a demanda de bens duráveis ( $D_d$ ) é explicada por valores passados e presente da renda disponível ( $Y_t, Y_{t-1}, \dots$ ), pelo preço presente do bem ( $P_t$ ) e pelo consumo passado do bem ( $Sd_{t-1}, Sd_{t-2}, \dots$ ).

Simbolicamente

$$D_d = f(Y_t, Y_{t-1}, \dots, P_t, Sd_{t-1}, Sd_{t-2}, \dots)$$

A demanda por bens não duráveis

Admite-se que a demanda por bens não duráveis (alimentos, bebidas, fumo, etc.) no período  $t$  ( $Q_t$ ), é explicada, pelo preço no período  $t$  ( $P_t$ ) do bem, pela renda disponível no período  $t$  ( $Y_t$ ) e por valores passados da demanda do bem ( $Q_{t-1}, Q_{t-2}, \dots$ ).

Simbolicamente

$$Q_t = f(Q_{t-1}, Q_{t-2}, \dots, Y_t, P_t).$$

As formas destas funções são em geral baseadas em hipóteses comportamentais, de fundo econômico.

Efetivamente, se quisermos pensar em fenômenos econômicos de uma forma mais realista devemos tornar os modelos mais dinâmicos; isto pode ser conseguido de certa forma tornando as variáveis envolvidas defasadas, de acordo com considerações de ordem econômica. De um modo geral, a teoria econômica tal qual é concebida hoje em dia é estática sendo os ajustamentos às solicitações feitos instantaneamente; raramente o processo de ajuste e as fases envolvidas são discutidas, mas a dinamização do modelo é fundamental.

Se tomarmos exemplos microeconômicos veremos a importância da dinâmica do modelo. É importante para o gover-

no, saber após quanto tempo os indicadores econômicos irão reagir à mudanças de variáveis sob seu controle. Outro aspecto importante é a rapidez com que os consumidores reagem às imposições de preços ou retração do crédito. Qual a rapidez com que as empresas reagem aos incentivos fiscais para investir. Qual a rapidez com que os investidores reagem às mudanças das taxas de juros.

Em resumo, podemos dizer que os modelos envolvendo variáveis defasadas constituem uma maneira de levar em consideração o tempo que se processa o ajuste entre variáveis em um fenômeno econômico e, talvez, a melhor maneira de torná-lo dinâmico.

Em geral costuma-se atribuir a três fatores principais a existência de modelos com variáveis defasadas.

#### 1 - FATORES PSICOLÓGICOS

Devido ao que se conhece por "persistência de hábito" (ou inércia) as pessoas não mudam imediatamente seus hábitos de consumo em razão da variação dos preços, ou da variação de suas rendas. Assim, quem se torna milionário repentinamente, por exemplo, por ganhar na loteria federal ou esportiva não muda imediatamente seus hábitos de vida, os quais perduraram em suas vidas por muito tempo, por não saber como reagir diante de um ganho inesperado. Ao passar o tempo, todavia, tal pessoa vai aprendendo a viver com a fortuna re

centemente conseguida.

Por outro lado, pessoas que auferem aumento de renda, ficam sem saber como reagir em relação ao consumo, por não saberem *à priori* se tal aumento é permanente, ou transitório. Alguns, por acreditarem que seja permanente, resolvem aumentar seu consumo, outros, por acreditarem que este evento seja transitório, preferem poupá-la; em grande parte os outros diante da incerteza desse aumento de renda, resolvem modificar paulatinamente seus hábitos de consumo.

## 2 - FATORES TECNOLÓGICOS

Suponhamos que o preço do capital em relação ao preço do trabalho decline, tornando a substituição do trabalho por capital, uma decisão vantajosa. Como tal troca se processa ao longo de um certo período de tempo, as firmas não tomam as vezes essa decisão, principalmente quando não sabem se essa diminuição do preço do capital é apenas temporária. Em virtude de tal fato a decisão da firma em trocar capital por trabalho se processa lentamente, daí a defasagem entre a diminuição do preço do capital e a decisão de substituir capital por trabalho.

Um exemplo interessante é o das calculadoras eletrônicas. Desde sua introdução no mercado na década de 60, seu preço real tem caído bastante; ao lado disso, as inovações tecnológicas introduzidas tem se processado rapidamen-



te. Isto origina no comprador, dúvidas com relação a sua compra, uma por não saberem se o preço futuro declinará ainda mais, e outra, por não conhecerem as futuras inovações que irão surgir.

### 3 - FATORES INSTITUCIONAIS

Tais fatores podem ser exemplificados como obrigações contratuais entre firmas com relação ao fornecimento de matéria-prima de uma à outra; ainda, pode-se citar, a aplicação de dinheiro em certas instituições à prazo fixo, que só após passar um certo período de tempo (pré-determinado) permite ao aplicar sacá-lo, mesmo que o emprego do dinheiro em outras formas de aplicação, redam maiores lucros.

Assim, de acordo com a legislação vigente, mudanças de comportamento muitas vezes devem esperar por algum tempo.

Para finalizar a introdução cabe observar que os modelos com variáveis defasadas envolvem ainda um certo grau de empirismo no que diz respeito ao tempo de ajuste às solicitações. Muitas vezes o número de períodos que comparece no modelo é determinado pelo tamanho da amostra, outras vezes é determinado por considerações à priori e finalmente adotam-se modelos com períodos distintos e escolhe-se aquele que a carreta o melhor ajuste do ponto de vista estatístico.

Nosso objetivo neste trabalho é descrever no capítulo 2, os diversos modelos concebidos, e suas aplicações.

No capítulo 3, veremos os principais métodos de estimação dos parâmetros dos modelos, bem como suas limitações. No capítulo 4, apresentaremos um exemplo de aplicação.

## CAPÍTULO 2

### MODELOS DE DEFASAGENS DISTRIBUÍDAS

#### 2.1 - CONCEITOS BÁSICOS

A expressão matemática do modelo de defasagens distribuídas é em geral concebida sob a forma

$$Y_t = \alpha + \beta_0 \cdot X_t + \beta_1 \cdot X_{t-1} + \beta_2 \cdot X_{t-2} \dots + \beta_m \cdot X_{t-m} + \varepsilon_t \quad (2.1)$$

onde  $X_t, X_{t-1}, \dots, X_{t-m}$  são variáveis não estocásticas, e  $\varepsilon_t$  uma variável aleatória (em geral admitida com média zero).

A expressão (2.1) também pode ser estendida sob a forma

$$Y_t = \alpha + \beta_0 \cdot X_t + \beta_1 \cdot X_{t-1} + \beta_2 \cdot X_{t-2} + \dots + \varepsilon_t \quad (2.2)$$

isto é

$$Y_t = \alpha + \sum_{S=0}^{\infty} \beta_S X_{t-S} + \varepsilon_t$$

onde, para que exista  $E(Y_t)$  se impõe que

$$\sum_{S=0}^{\infty} \beta_S < \infty$$

A princípio os parâmetros do modelo (2.1), poderiam

ser estimados, usando-se estimadores de mínimos quadrados ou outros estimadores com propriedades desejáveis, dependendo das especificações adotadas sobre  $\epsilon_t$ .

Todavia se  $m$  for grande, pode-se não ter observações suficientes para estimar os  $\beta_i$ 's. Mesmo tendo observações suficientes, em geral as variáveis  $X_t, X_{t-1}, \dots$  são altamente correlacionadas, advindo então problemas de multicolinearidade cujas conseqüências podem ser encontradas em [16] ou [19]. Tais problemas são contornados por modificações introduzidas no modelo (2.1).

Freqüentemente são sugeridas restrições aos parâmetros  $\beta_0, \beta_1, \beta_2, \dots$  em (2.1) e (2.2) de acordo com alguma hipótese decorrente do próprio fenômeno em estudo. Tais restrições reduzem o número de parâmetros de (2.1) e (2.2) e em geral, introduzem sérios problemas de estimação. Vejamos a seguir as diferentes formas propostas para se conseguir isto.

## 2.2 - MODELOS COM PONDERAÇÃO A PRIORI DOS PARÂMETROS

Esta é a forma mais simples de se reduzir o número de parâmetros de (2.1). O método consiste essencialmente em escrever (2.1) sob a forma

$$Y_t = \alpha + \beta \cdot \left( \frac{\beta_0}{\beta} X_t + \frac{\beta_1}{\beta} X_{t-1} + \dots + \frac{\beta_m}{\beta} X_{t-m} \right) + \epsilon_t$$

e fazer

$$w_i = \frac{\beta_i}{\beta} \quad i=0,1,2,\dots,m$$

onde tais valores  $w_i$ 's (chamados pesos) são escolhidos a priori. Assim fazendo

$$W_t = \sum_{i=0}^m w_i X_{t-i}$$

o modelo (2.1) fica sob a forma

$$Y_t = \alpha + \beta \cdot W_t + \varepsilon_t.$$

Desta forma o número de parâmetros a serem estimados se reduziu de  $m$  para apenas 2.

Embora a escolha dos  $w_i$ 's seja arbitrária, e admita uma infinidade de possibilidades, em geral ela segue certos critérios.

De Leeuw [04] foi um dos precursores na utilização de tal procedimento, em estudos econômicos sobre a função investimento.

Alguns critérios citados na literatura [20] e usados por De Leeuw são os seguintes:

a) Critério do Declínio dos Coeficientes

De acordo com este critério, os pesos  $w_i$ 's são assumidos como números decrescentes. Digamos que em (2.1)  $m$  seja igual a 4. Define-se então uma nova variável  $W_{1t}$  tal que

$$W_{1t} = w_0 \cdot X_t + w_1 \cdot X_{t-1} + w_2 \cdot X_{t-2} + w_3 \cdot X_{t-3} + w_4 \cdot X_{t-4}$$

onde

$$W_{1,t} = \frac{1}{2} X_t + \frac{1}{4} X_{t-1} + \frac{1}{8} X_{t-2} + \frac{1}{16} X_{t-3} + \frac{1}{32} X_{t-4}$$

Desta forma, o modelo (2.1) fica sob a forma

$$Y_t = \alpha + \beta \cdot W_{1,t} + \epsilon_t$$

e o número de parâmetros a serem estimados se reduziu de cinco para dois. De acordo com as suposições sobre  $\epsilon_t$  podem ser usados estimadores de mínimos quadrados ou outros estimadores com propriedades desejáveis.

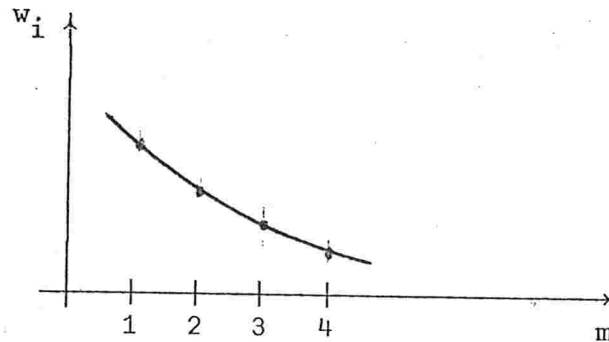


Figura 1

#### b) Critério da Uniformidade dos Coeficientes

Tal critério admite que os coeficientes de  $X_t, X_{t-1}, \dots, X_{t-m}$  são todos iguais. Define-se então uma nova variável  $W_{2t}$  tal que

$$W_{2t} = w_0 \cdot X_t + w_1 \cdot X_{t-1} + \dots + w_m \cdot X_{t-m}$$

onde  $w_0 = w_1 = \dots = w_m$ .

Desta forma o modelo original (2.1) pode ser escrito sob a forma

$$Y_t = \alpha + \beta \cdot W_{2t} + \varepsilon_t$$

Assim o número de parâmetros a serem estimados se reduz de  $\underline{m}$  para  $\underline{2}$ . Os estimadores de  $\alpha$  e  $\beta$  são obtidos, de acordo com as suposições feitas sobre  $\varepsilon_t$ .

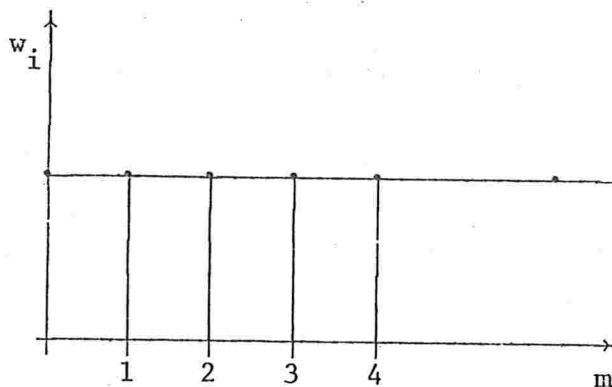


Figura 2

c) Critério do "V Invertido"

De acordo com tal critério, os pesos  $w_0, w_1, \dots, w_m$  inicialmente crescem monotonicamente para em seguida decrescerem monotonicamente.

Assim, admitindo-se  $m = 4$ , pode-se definir a variável  $W_{3t}$  como sendo

$$W_{3t} = \frac{1}{10} X_t + \frac{1}{4} X_{t-1} + \frac{1}{5} X_{t-2} + \frac{1}{7} X_{t-3} + \frac{1}{9} X_{t-4}$$

de modo que o modelo inicial (2.1) fica sob a forma

$$Y_t = \alpha + \beta \cdot W_{3t} + \varepsilon_t$$

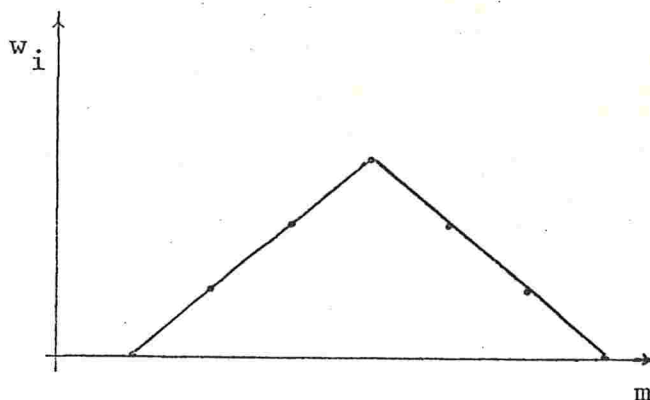


Figura 4

Embora tenha-se definido três critérios, para a ponderação â priori de  $X_t, X_{t-1}, \dots, X_{t-m}$ , o problema fica dependendo da escolha de  $m$  e dos valores numéricos dados aos pesos  $w_0, w_1, \dots, w_m$ . De Leeuw em seus estudos adotou os três critérios citados, para diferentes valores de  $m$ , adotando como critério de escolha o modelo que maximizava o coeficiente de determinação múltipla  $R^2$ .

Os valores de  $R^2$ , para os diferentes valores de  $m$  são dados abaixo, para os critérios de uniformidade, e do "V invertido".

$m$	critério da uniformidade	critério do "V invertido"
6	0,588	0,629
8	0,751	0,726
10	0,667	0,817
12	0,458	0,872
14	0,348	0,865



O modelo adotado por De Leeuw foi o do "V invertido" com  $m=12$ . Cumpre lembrar ainda que neste modelo, os pesos cresciam em progressão aritmética e também decresciam em progressão aritmética.

Antes de finalizarmos, convém lembrar que Irving Fisher, um dos precursores do uso de modelos de defasagens distribuídas, sugeriu o critério do declínio dos coeficientes, sendo que tal declínio se processava de acordo com os termos de uma progressão aritmética decrescente.

### 2.3 - MODELO POLINOMIAL DE ALMON

Shirley Almon [01] em um trabalho sobre a relação entre a apropriação de capital e gastos de capital sugeriu um método mais refinado para a redução do número de parâmetros do modelo expresso por (2.1).

Basicamente o método proposto, consiste em se admitir que no modelo

$$Y_t = \beta_0 \cdot X_t + \beta_1 \cdot X_{t-1} + \beta_2 \cdot X_{t-2} + \dots + \beta_m \cdot X_{t-m} + \epsilon_t \quad (2.3)$$

os coeficientes  $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ , podem ser aproximados por uma função polinomial  $f(z)$  de grau  $r$  (ver figura 4),

Em seu trabalho, Almon admitiu a possibilidade de tal aproximação baseado no Teorema de Weierstrass que afirma que uma função definida num intervalo fechado pode ser aproximada em todo domínio por uma função polinomial de grau

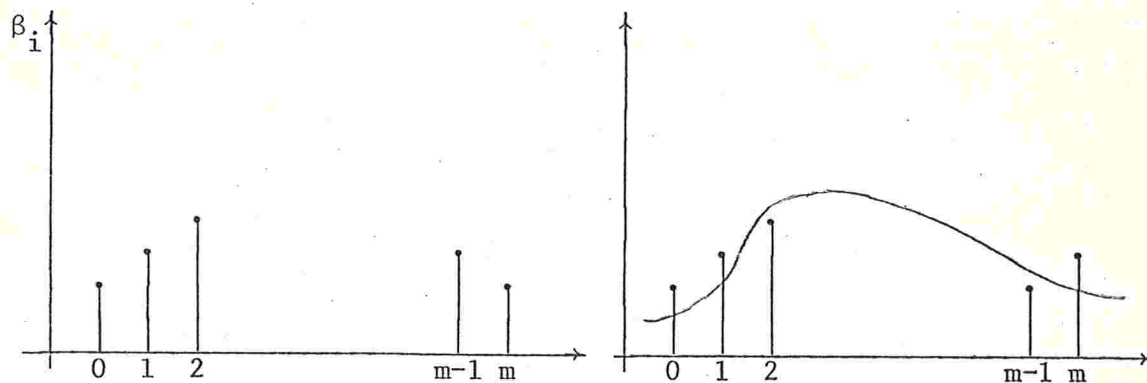


Figura 4

conveniente. Em seguida, usando a fórmula de interpolação polinomial de Lagrange e sucessivos e trabalhosos procedimentos algébricos, ele consegue estimar  $\beta_0, \beta_1, \dots, \beta_m$ .

A fim de contornar tais dificuldades de ordem prática, uma ligeira modificação costuma ser proposta e é citada em [11],[16],[19] e [20].

Tal modificação, consiste em se admitir diretamente que  $\beta_0, \beta_1, \dots, \beta_m$  são os valores de uma função polinomial de grau  $r$ , ( $r < m$ ), onde

$$f(i) = \beta_i \quad i=0,1,2,\dots,m.$$

e

$$f(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_r \cdot z^r.$$

Os seguintes passos são seguidos para a estimação dos  $\beta_i$ 's.

1º passo) Especifica-se o grau  $r$  do polinômio (em geral 3 ou 4).

2º passo) Expressa-se os  $\beta_i$ 's em função dos  $a_i$ 's de modo que

$$\beta_i = f(i) \quad \forall i, \in \{0, 1, 2, \dots, m\},$$

isto é

$$\begin{aligned} \beta_0 &= f(0) = a_0 \\ \beta_1 &= f(1) = a_0 + a_1 + a_2 + \dots + a_r \\ \beta_2 &= f(2) = a_0 + 2a_1 + 2^2 \cdot a_2 + \dots + 2^r \cdot a_r \\ \beta_m &= f(m) = a_0 + m \cdot a_1 + m^2 \cdot a_2 + \dots + m^r \cdot a_r \end{aligned} \tag{2.4}$$

Estimando-se  $a_0, a_1, \dots, a_r$  obtêm-se facilmente os estimadores de  $\beta_0, \beta_1, \dots, \beta_m$  através do sistema linear acima.

Retomando-se (2.3) e substituindo-se os  $\beta_i$ 's pelos valores dados em (2.4) tem-se

$$\begin{aligned} Y_t &= a_0 \cdot X_t + (a_0 + a_1 + a_2 + \dots + a_r) \cdot X_{t-1} + \\ &\quad + (a_0 + 2a_1 + 2^2 \cdot a_2 + \dots + 2^r \cdot a_r) \cdot X_{t-2} + \\ &\quad + (a_0 + 3a_1 + 3^2 \cdot a_2 + \dots + 3^r \cdot a_r) \cdot X_{t-3} + \dots \\ &\quad + (a_0 + ma_1 + m^2 \cdot a_2 + \dots + m^r \cdot a_r) \cdot X_{t-m} + \epsilon_t \end{aligned} \tag{2.5}$$

E portanto tem-se

$$\begin{aligned} Y_t &= a_0 (X_t + X_{t-1} + X_{t-2} + \dots + X_{t-m}) + \\ &\quad + a_1 (X_{t-1} + 2X_{t-2} + 3X_{t-3} + \dots + m X_{t-m}) + \\ &\quad + a_2 (X_{t-1} + 2^2 \cdot X_{t-2} + 3^2 \cdot X_{t-3} + \dots + m^2 X_{t-m}) + \dots \\ &\quad + a_r (X_{t-1} + 2^r \cdot X_{t-2} + \dots + m^r \cdot X_{t-m}) + \epsilon_t \end{aligned} \tag{2.6}$$

Fazendo-se

$$W_0 = X_t + X_{t-1} + X_{t-2} + \dots + X_{t-m}$$

$$W_1 = X_{t-1} + 2X_{t-2} + \dots + m \cdot X_{t-m}$$

$$W_2 = X_{t-1} + 2^2 \cdot X_{t-2} + \dots + m^2 \cdot X_{t-m}$$

.. . . . .

$$W_r = X_{t-1} + 2^r \cdot X_{t-2} + \dots + m^r \cdot X_{t-m}$$

O modelo inicial (2.3) fica transformado para

$$Y_t = a_0 \cdot W_0 + a_1 \cdot W_1 + a_2 \cdot W_2 + \dots + a_r \cdot W_r + \varepsilon_t \quad (2.7)$$

obtendo-se portanto uma redução no número de parâmetros.

3º passo) Admitindo-se que os  $\varepsilon_t$ 's tem média zero, mesma variância, e são não correlacionados, estimamos os  $a_i$ 's usando estimadores de mínimos quadrados em (2.7) que agora tem  $r$  ( $r < m$ ) parâmetros.

4º passo) Os estimadores dos  $\beta_i$ 's são agora facilmente obtidos a partir de (2.4).

As vantagens do uso deste enfoque do modelo polinomial de Almon pode ser enumerados.

1º) A manipulação é simples.

2º) O grau do polinômio pode ser escolhido empiricamente através de testes de significância sobre os coeficientes  $a_i$ 's. Por exemplo se  $a_3$  é significativamente diferentes de zero e  $a_4$  não, escolhe 3 como o grau do polinômio.

3º) Um aumento do grau do polinômio envolve a entrada de

uma variável adicional em (2.7) sem que haja necessidade de modificação das outras. Assim, o aumento do grau do polinômio consiste em acrescentar a (2.7) a variável

$$W_{r+1} = X_{t-1} + 2^{r+1} \cdot X_{t-2} + \dots + m^{r+1} \cdot X_{t-m}$$

sem necessidade de se alterar  $W_0, W_1, \dots, W_r$ .

Todavia críticas também podem ser feitas ao modelo, em primeiro lugar sobre a escolha empírica por tentativas do grau  $r$  do polinômio.

Outro fato relevante é a escolha de  $m$ . Embora Almon tenha seguido procedimentos praticamente iguais aos descritos, sua escolha de  $m$  baseava-se na utilização de diferentes valores para o mesmo, e escolhendo aquele valor de maximizava  $\bar{R}^2$ , isto é, o coeficiente de determinação múltipla, corrigido pelos graus de liberdade.

Antes de prosseguirmos o estudo do modelo polinomial de Almon, vamos ilustrar o que fizemos até aqui, com um exemplo extraído de [11].

Os dados abaixo fornecem os valores de estoques  $Y$  e vendas  $X$  do setor de manufatura no período 1955-1974 nos EUA. O modelo para o relacionamento entre as variáveis foi admitido como sendo. Dados tabela 1.

$$Y_t = \alpha + \beta_0 \cdot X_t + \beta_1 \cdot X_{t-1} + \beta_2 \cdot X_{t-2} + \beta_3 \cdot X_{t-3} + \varepsilon_t$$

onde  $\varepsilon_t \sim N(0, \sigma^2)$  e  $E(\varepsilon_i \varepsilon_j) = 0$   $i \neq j$

Os valores  $\beta_0, \beta_1, \beta_2, \beta_3$  foram admitidos como valo-

res de um polinômio do 2º grau

$$f(z) = a_0 + a_1 \cdot z + a_2 \cdot z^2$$

O modelo transformado, de acordo com o que vimos fi  
ca.

$$Y_t = \alpha + a_0 \cdot W_0 + a_1 \cdot W_1 + a_2 \cdot W_2 + \epsilon_t$$

onde

$$W_0 = X_t + X_{t-1} + X_{t-2} + X_{t-3}$$

$$W_1 = X_{t-1} + 2X_{t-2} + 3X_{t-3}$$

$$W_2 = X_{t-1} + 4X_{t-2} + 9X_{t-3}$$

TABELA 1

ANO	Y	X	W <sub>0</sub>	W <sub>1</sub>	W <sub>2</sub>
1955	45.069	26.480			
1956	50.642	27.740			
1957	51.871	28.736			
1958	50.070	27.280	110.236	163.656	378.016
1959	52.707	30.219	113.975	167.972	391.884
1960	53.814	30.796	117.031	170.987	397.963
1961	54.939	30.896	119.191	173.074	397.192
1962	58.213	33.113	125.024	183.145	426.051
1963	60.043	35.032	129.837	187.293	433.861
1964	63.383	37.335	136.376	193.246	445.548
1965	68.221	41.003	146.483	206.738	475.480
1966	77.965	44.863	158.239	220.769	505.631
1967	84.655	46.449	169.656	238.880	544.896
1968	90.875	50.282	182.603	259.196	594.952
1969	97.074	53.555	195.155	277.787	639.899
1970	101.645	52.859	203.145	293.466	672.724
1971	102.445	55.917	212.613	310.815	719.617
1972	107.719	62.017	224.348	322.300	749.348
1973	120.870	71.398	242.191	332.428	761.416
1974	147.135	82.078	271.410	363.183	822.719

X e Y são dados em milhões de dólares.

A equação estimada foi

$$Y_t = -8140,7564 + 0,6612 W_0 + 0,9020 W_1 - 0,4322 W_2$$

(1992,9809)	(0,1655)	(0,4831)	(0,1665)
(-4,0847)	(3,9960)	(1,8671)	(-2,5961)

$R^2 = 0,9961$       g.l. = 13

Os estimadores de  $\beta_0, \beta_1, \beta_2$  e  $\beta_3$  são dados por

$$\hat{\beta}_0 = \hat{a}_0 = 0,6612$$

$$\hat{\beta}_1 = (\hat{a}_0 + \hat{a}_1 + \hat{a}_2) = (0,6612 + 0,9020 - 0,4322) = 1,1310$$

$$\hat{\beta}_2 = (\hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2) = (0,6612 + 1,8040 - 1,7288) = 0,7364$$

$$\hat{\beta}_3 = (\hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2) = (0,6612 + 2,7060 - 3,8898) = -0,5226$$

Finalmente, a equação estimada do modelo original é

$$\hat{Y}_t = -8140,7564 + 0,6612 X_t + 1,1310 X_{t-1} + 0,7364 X_{t-2} - 0,5226 X_{t-3}$$

(1992,9803)	(0,1655)	(0,1798)	(0,1625)	(0,2307)
(-4,0847)	(3,9960)	(6,2903)	(4,5317)	(-2,3530)

O gráfico das estimativas dos  $\beta_i$ 's é dado na Figura 5.

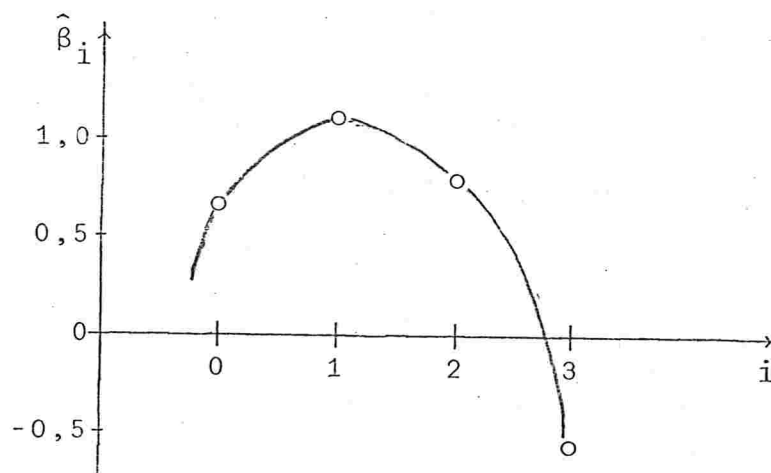


Figura 5

Recentemente Schmidt [35] tentando contornar a arbitrariedade do valor atribuído a  $m$  no modelo de Almon, propôs uma modificação que em última análise consistia em transformar o modelo

$$Y_t = \alpha + \beta_0 \cdot X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_m \cdot X_{t-m} + \varepsilon_t$$

em outro sob a forma

$$Y_t = \alpha + \sum_{i=0}^{\infty} \beta_i X_{t-i} + \varepsilon_t$$

de forma que  $\beta_i \rightarrow 0$  quando  $i \rightarrow \infty$ , da seguinte forma

$$\beta_i = \gamma^i (\lambda_0 + \lambda_1 \cdot i + \lambda_2 \cdot i^2 + \dots + \lambda_p \cdot i^p)$$

onde  $0 \leq \gamma < 1$  e  $i = 0, 1, 2, \dots$  ou seja os  $\beta_i$ 's são admitidos pertencentes a um polinômio de grau  $p$ ,  $f(z)$ , somente que cada  $\beta_i$  é "amortecido" pelo coeficiente  $\gamma^i$  que se torna cada vez menor, a medida que  $i$  aumenta.

Schmidt propôs estimadores para os parâmetros do modelo, e mostrou que eram consistentes e assintoticamente eficientes.

Ainda com relação ao problema da determinação de  $m$  e do grau do polinômio  $r$ , as seguintes considerações devem ser feitas.

Em primeiro lugar, se *a priori* se conhece  $m$  e o verdadeiro grau do polinômio  $r$  então os estimadores de  $\beta_0, \beta_1, \dots, \beta_m$  em (2.3) obtidos sob a forma descrita são normalmente



distribuídos e são os melhores estimadores lineares não viesados (desde que os  $\epsilon_t$  sejam normalmente distribuídos, homocedásticos e  $E(\epsilon_i \epsilon_j) = 0$  se  $i \neq j$  ver Dhrymes [05, pp.225-227].

Em segundo lugar Frost [09] em recente artigo, investigou as conseqüências do uso da técnica de Almon quando o grau do polinômio  $r$  e o valor de  $m$  são escolhidos por tentativas, no sentido de se descobrir qual a combinação de  $m$  e  $r$  que minimiza a variação residual (ou o coeficiente de determinação múltipla, corrigido pelos graus de liberdade  $\bar{R}^2$ ).

Frost demonstrou que os estimadores de  $\beta_i$ 's obtidos desta forma são viesados, não distribuídos normalmente e os estimadores de suas variâncias subestimam-nas.

Além disso ele aplicou simulação a um conjunto de observações e observou que o viés de estimação para os estimadores dos  $\beta_i$ 's era pequeno, na distribuição amostral, os estimadores dos  $\beta_i$ 's não se comportavam como variáveis normais e finalmente que as variâncias dos  $\hat{\beta}_i$ 's eram bastante viesadas. Tal fato levou a conclusão de que a análise de significância dos parâmetros  $\beta_i$ 's obtidos é um tanto perigosa e sujeita a erros.

#### 2.4 - MODELO DE KOYCK OU DO DECLÍNIO GEOMÉTRICO

Este é um dos mais conhecidos modelos de defasagens distribuídas, considerados na literatura econométrica. Foi introduzido por Koyck [21] em estudos sobre a resposta da ca

pacidade devida a mudanças na produção em certas indústrias norte-americanas.

Basicamente o modelo introduzido consiste em atribuir uma restrição aos coeficientes de (2.2). Tal restrição é a admitir que  $\beta_0, \beta_1, \beta_2, \dots$  constituem os termos de uma progressão geométrica decrescente. Tal restrição decorre da suposição de que em (2.2) valores mais recentes de  $X$  exercem maior influência sobre  $Y$  do que valores mais antigos, e que tal influência é suavizada exponencialmente ao longo do tempo. Assim, o modelo (2.2)

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \varepsilon_t$$

é tal que

$$\beta_i = \lambda^i \cdot \beta_0 \quad i=0,1,2,\dots, \quad (2.8)$$

onde  $0 < \lambda < 1$ .

O modelo original (2.2) de acordo com as suposições acima, fica sob a forma

$$Y_t = \alpha + \beta_0 X_t + (\lambda \cdot \beta_0) X_{t-1} + (\lambda^2 \cdot \beta_0) X_{t-2} + \dots + \varepsilon_t \quad (2.9)$$

O modelo (2.9) mesmo assim fica com um número infinito de parâmetros a serem estimados. Todavia Koyck contornou a situação da seguinte forma:

Tornou a equação (2.9) no período  $(t-1)$  obtendo

$$Y_{t-1} = \alpha + \beta_0 X_{t-1} + (\lambda \cdot \beta_0) X_{t-2} + (\lambda^2 \cdot \beta_0) X_{t-3} + \dots + \varepsilon_{t-1} \quad (2.10)$$

Multiplicou (2.10) por  $\lambda$ , obtendo

$$\lambda Y_{t-1} = \alpha\lambda + (\lambda\beta_0)X_{t-1} + (\lambda^2\beta_0)X_{t-2} + (\lambda^3\beta_0)X_{t-3} + \dots + \lambda\varepsilon_{t-1} \quad (2.11)$$

Em seguida fez (2.9) menos (2.11), obtendo

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta_0 X_t + (\varepsilon_t - \lambda\varepsilon_{t-1}) \quad (2.12)$$

ou ainda

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t \quad (2.13)$$

onde  $v_t = \varepsilon_t - \lambda\varepsilon_{t-1}$ .

O procedimento descrito é conhecido como "transformação de Koyck". Embora simples, ele reduziu o problema da estimação de infinitos parâmetros (2.9) para o problema da estimação de apenas três em (2.13),  $\alpha$ ,  $\lambda$  e  $\beta_0$ .

Uma ligeira modificação pode ser introduzida no modelo original de Koyck, se houver razões para admitir o declínio dos coeficientes a partir de  $\beta_k$  ( $k > 1$ ). Isto é

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_{k-1} X_{t-k+1} + \\ + [\beta_k X_{t-k} + (\lambda \cdot \beta_k) X_{t-k-1} + (\lambda^2 \cdot \beta_{k-1}) X_{t-k-2} + \dots] + \varepsilon_t \quad (2.14)$$

Aplicando a (2.14) a transformação de Koyck, obtém-se

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + (\beta_1 - \lambda\beta_0) X_{t-1} + (\beta_2 - \lambda\beta_1) X_{t-2} + \\ + \dots + (\beta_k - \lambda\beta_{k-1}) X_{t-k} + \lambda \cdot Y_{t-1} + v_t \quad (2.15)$$

onde  $v_t = \varepsilon_t - \lambda\varepsilon_{t-1}$ .

Além disso, o modelo de Koyck pode ser facilmente generalizado, se tivermos duas variáveis explicativas X e Z (ver [16, pp.299 e 300]).

Em resumo pode se dizer que o modelo de Koyck, em sua forma transformada contorna dois problemas básicos

- 1º) Aumenta o número de graus de liberdade do resíduo (devido ao pequeno número de parâmetros a serem estimados).
- 2º) Evita problemas de multicolinearidade entre os sucessivos valores de X ( $X_t, X_{t-1}, X_{t-2}, \text{etc.}$ ).

Todavia tais simplificações tem seu preço, pois ao lado delas nasceram outras conseqüências indesejáveis:

- 1º) No modelo sob a forma transformada (2.13), as perturbações aleatórias,  $v_t = \varepsilon_t - \lambda\varepsilon_{t-1}$  são autocorrelacionadas, se os  $\varepsilon_t$ 's não o forem.

De fato,

$$\begin{aligned} E(v_t \cdot v_{t+1}) &= E\{[\varepsilon_t - \lambda\varepsilon_{t-1}][\varepsilon_{t+1} - \lambda\varepsilon_t]\} = \\ &= E\{\varepsilon_t \cdot \varepsilon_{t+1} - \lambda\varepsilon_t^2 - \lambda\varepsilon_{t-1} \cdot \varepsilon_{t+1} + \lambda^2\varepsilon_t \cdot \varepsilon_{t-1}\} = \\ &= -\lambda E(\varepsilon_t^2) = -\lambda \cdot \sigma_\varepsilon^2 \neq 0. \end{aligned}$$

- 2º) A variável  $Y_{t-1}$  que aparece em (2.13) não é independente de  $v_t$ .

De fato,

$$E[v_t \cdot Y_{t-1}] = E[(\varepsilon_t - \lambda \varepsilon_{t-1}) \cdot [\alpha + \beta_0(X_{t-1} + \beta X_{t-2} + \dots) + \varepsilon_{t-1}]] = \lambda \sigma_\varepsilon^2 \neq 0.$$

A consequência de tal fato, é que os estimadores de mínimos quadrados de (2.13) são inconsistentes (demonstração no Apêndice 1).

Veremos no capítulo seguinte outros estimadores propostos a fim de contornarmos tal situação.

## 2.5 - MODELO DAS EXPECTATIVAS ADAPTADAS

Embora o modelo de Koyck seja uma forma elegante de impor restrições sobre os  $\beta_i$ 's de (2.2) ele é concebido de maneira puramente algébrica, sem sustentação teórica baseada em alguma hipótese comportamental.

O modelo que veremos a seguir, concebido por P. Cagan tem sustentação teórica baseada em hipóteses comportamentais.

O valor de Y no período t, é admitido como dependente não de  $X_t$ , mas do nível "esperado" ou "permanente" de X no período t (indica-se por  $X_t^*$ ), da seguinte forma

$$Y_t = \alpha + \beta \cdot X_t^* + \varepsilon_t \quad (2.16)$$

onde  $\varepsilon_t$  é em geral admitida como sendo uma variável aleatória com média zero.

Como  $X_t^*$  não é diretamente observável, postula-se que

$$X_t^* - X_{t-1}^* = (1 - \lambda) \cdot (X_t - X_{t-1}^*) \quad (2.17)$$

ou ainda

$$X_t^* = (1 - \lambda)X_t + \lambda \cdot X_{t-1}^*$$

onde  $0 \leq \lambda < 1$ .

Tal postulado admite que o valor "esperado" ou "permanente" de  $X$  no período  $t$  é uma média ponderada de  $X$  no período  $t$  e do valor "esperado" ou "permanente" de  $X$  no período  $(t-1)$ .

Tal postulado, baseia-se na pressuposição de que as expectativas do período  $t$  são dadas pela modificação das expectativas no período  $t-1$ , sob a influência da experiência do que sucede no período  $t$ .

Um exemplo de aplicação é a quantidade  $Q_t$  demandada de um bem (num período de alta inflação) ser admitida como função do preço "esperado"  $P_t^*$  do mesmo bem, sob a forma:

$$Q_t = \alpha + \beta \cdot P_t^* + \varepsilon_t$$

Outro exemplo decorre da "hipótese da renda permanente" concebida por M. Friedman, que postula o consumo  $C_t$  no período  $t$  como sendo determinado pela "renda esperada" ou "permanente"  $Y_t$ , da seguinte forma:

$$C_t = \beta \cdot Y_t^* + \varepsilon_t$$

onde

$$Y_t^* = (1 - \lambda) \cdot Y_t + \lambda \cdot Y_{t-1}^*$$

Um terceiro exemplo é o da demanda  $M_t$  da moeda ser função da taxa de juros "esperada"  $j_t^*$  da seguinte forma:

$$M_t = \alpha + \beta \cdot j_t^* + \varepsilon_t$$

onde

$$j_t^* = (1 - \lambda) \cdot j_t + \lambda \cdot j_{t-1}^*$$

De (2.17) é fácil concluir que

$$X_t^* = (1 - \lambda) (X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots). \quad (2.18)$$

De (2.17) e (2.18) obtêm-se

$$Y_t = \alpha + \beta(1 - \lambda) (X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots) + \varepsilon_t \quad (2.19)$$

que representa um modelo com defasagens distribuídas, análogo ao de Koyck, fazendo-se  $\beta(1 - \lambda) = \beta_0$ .

Da mesma forma que o modelo de Koyck, a estimação dos parâmetros de (2.19) nesta forma é impraticável. Todavia ela pode ser simplificada por meio da transformação de Koyck, isto é:

Defasando (2.19) de um período, e multiplicando membro a membro por  $\lambda$  e subtraindo o resultado de (2.19) chega-se facilmente a

$$Y_t = \alpha(1 - \lambda) + \beta(1 - \lambda)X_t + \lambda \cdot Y_{t-1} + v_t \quad (2.20)$$

onde

$$v_t = \varepsilon_t - \lambda\varepsilon_{t-1}$$

O modelo das expectativas adaptadas é, às vezes formulado de maneira ligeiramente diferente. A relação (2.16) é mantida, porém a (2.17) é postulada como

$$X_t^* = (1 - \lambda)X_{t-1} + \lambda \cdot X_{t-1}^* \quad (2.21)$$

Isto é,  $X_t$  de (2.17) é substituído por  $X_{t-1}$  em (2.21). Tal modificação pressupõe a hipótese de que as expectativas têm de ser revistas comparando-se  $X_{t-1}^*$  com  $X_{t-1}$ . Nesse caso, facilmente chega-se a

$$X_t^* = (1 - \lambda)(X_{t-1} + \lambda X_{t-2} + \lambda^2 X_{t-3} + \dots) \quad (2.22)$$

que substituída em (2.16) fornece

$$Y_t = \alpha + \beta(1 - \lambda)(X_{t-1} + \lambda X_{t-2} + \lambda^2 X_{t-3} + \dots) + \varepsilon_t \quad (2.23)$$

Finalmente, aplicando-se a (2.23) a transformação de Koyck, chega-se a

$$Y_t = \alpha(1 - \lambda) + \beta(1 - \lambda)X_{t-1} + \lambda Y_{t-1} + v_t \quad (2.24)$$

onde

$$v_t = \varepsilon_t - \lambda\varepsilon_{t-1}.$$

A diferença entre (2.24) e (2.20) é que na primei-



ra,  $X_t$  é substituída por  $X_{t-1}$ , o que não produz problemas de estimação dos parâmetros.

Da mesma forma que vimos no modelo de Koyck, o modelo (2.20) e (2.24) é tal que os  $v_t$ 's são correlacionados e  $v_t$  é correlacionado com  $Y_{t-1}$  (desde que os  $\varepsilon_t$ 's não sejam correlacionados). Tal fato introduz no modelo das expectativas adaptadas, os mesmos problemas de estimação que no modelo de Koyck.

## 2.6 - MODELO DOS AJUSTAMENTOS PARCIAIS

O modelo que veremos a seguir é outra forma de impor restrições sobre os  $\beta_i$ 's de (2.2) e que, da mesma forma que o modelo das expectativas adaptadas, tem sustentação teórica baseada em hipóteses comportamentais. Tal modelo foi introduzido por Nerlove [28].

Suponhamos que exista um "equilíbrio ótimo", "desejável" ou de "longo prazo" para uma variável  $Y$ , e que vamos indicá-lo por  $Y^*$ . Suponhamos ainda que tal valor  $Y$  seja descrito pela relação

$$Y_t^* = \alpha + \beta \cdot X_t + \varepsilon_t \quad (2.25)$$

Como  $Y^*$  não é observável diretamente, Nerlove postulou a seguinte hipótese comportamental, conhecida como a dos "ajustamentos parciais".

$$Y_t - Y_{t-1} = \gamma \cdot (Y_t^* - Y_{t-1}) \quad (2.26)$$

onde  $\gamma$  ( $0 < \gamma \leq 1$ ) é chamado coeficiente de ajustamento.

A equação (2.26) postula que a variação efetiva de  $Y$  entre dois períodos sucessivos ( $Y_t - Y_{t-1}$ ) é uma fração da variação desejável ( $Y_t^* - Y_{t-1}$ ). O fato de  $\gamma$  estar compreendido entre 0 e 1 pode decorrer de fatores de ordem tecnológica, institucional, ou de persistência de hábito.

Como visto no modelo das expectativas adaptadas, o modelo dos ajustamentos parciais, dado por (2.25) e (2.26) é as vezes ligeiramente modificado em (2.25), e expresso sob a forma

$$Y_t^* = \alpha + \beta \cdot X_{t-1} + \varepsilon_t \quad (2.27)$$

isto é, o nível "desejado" de  $Y$  no período  $t$  depende do nível de  $X$  no período  $t-1$  e não no período  $t$  como em (2.25).

Inicialmente vamos exemplificar o uso de tal modelo, admitindo que a quantidade de determinado produto agrícola que os produtores desejam ofertar no período  $t$  ( $Q_t^*$ ) dependa do preço desse produto no período anterior ( $P_{t-1}$ ). Expressamos tal fato através da relação

$$Q_t^* = \alpha + \beta \cdot P_{t-1} + \varepsilon_t$$

onde  $\varepsilon_t$  é uma variável aleatória com média 0. Todavia, devido à existência de uma rigidez a curto prazo ao nível dos fa

tores de produção, existe um descompasso entre a quantidade que os agricultores desejam oferecer e a quantidade efetivamente oferecida. Isto pode ser expresso pela relação

$$(Q_t - Q_{t-1}) = \gamma \cdot (Q_t^* - Q_{t-1}) \quad \text{onde } 0 < \gamma \leq 1$$

Um outro exemplo seria o de assumir que existe um equilíbrio ótimo, desejável ou de longo prazo de quantidade de capital ( $Y^*$ ) necessário à produção de um dado nível da produção ( $X$ ). Admitamos que  $Y_t^*$  é função de  $X_t$  da seguinte forma

$$Y_t^* = \alpha + \beta \cdot X_t + \varepsilon_t$$

Tipicamente, devido à rigidez de mudanças, decorrentes de razões tecnológicas, institucionais etc., a variação efetiva da quantidade de capital ( $Y_t - Y_{t-1}$ ) é apenas uma fração da variação desejável ( $Y_t^* - Y_{t-1}$ ), isto é

$$(Y_t - Y_{t-1}) = \gamma \cdot (Y_t^* - Y_{t-1})$$

Das equações (2.25) e (2.26) obtemos facilmente, por substituição

$$Y_t = \alpha + \beta \gamma X_t + (1 - \gamma) \cdot Y_{t-1} + \gamma \cdot \varepsilon_t \quad (2.28)$$

A equação (2.28) é formalmente a mesma do modelo de expectativas adaptadas, a menos da variação aleatória, que nesse caso é  $\gamma \cdot \varepsilon_t$ . Por outro lado, embora tendo a mesma forma, as hipóteses comportamentais que levaram à equação do modelo final são diferentes.

Do ponto de vista de estimação, os parâmetros de (2.28) não

apresentam sérios problemas como os do modelo de Koyck e de Cagan.

De fato, se admitirmos que os  $\epsilon_t$ 's de (2.25) são não correlacionados, é fácil concluir que  $\epsilon_t$  e  $Y_{t-1}$  de (2.28) são não correlacionados, o que torna os estimadores de mínimos quadrados dos parâmetros de (2.28) consistentes (ver Apêndice 1). Todavia para pequenas amostras os estimadores de mínimos quadrados são viesados, e isto é causado pela presença da variável aleatória  $Y_{t-1}$  no 2º membro de (2.28). Provemos tal afirmação no modelo simplificado

$$Y_t = \beta_0 + \beta_1 \cdot Y_{t-1} + \epsilon_t \quad (2.29)$$

onde

$$(1) \quad \epsilon_t \sim N(0, \sigma^2)$$

$$(2) \quad E(\epsilon_i \epsilon_j) = 0 \text{ para } i \neq j.$$

e provemos que o estimador de mínimos quadrados da  $\beta_i$  é viesado. Temos,

$$\hat{\beta}_i = \frac{\sum_{t=2}^n (Y_t - \bar{Y}) \cdot (Y_{t-1} - \bar{Y})}{\sum_{t=2}^n (Y_{t-1} - \bar{Y})^2} \quad (2.30)$$

onde

$$\bar{Y} = \frac{1}{n-1} \sum_{t=2}^n Y_t \quad \text{e} \quad \bar{\bar{Y}} = \frac{1}{n-1} \sum_{t=1}^{n-1} Y_t$$

Tomando (2.29), fazendo  $t = 2, 3, \dots, n$ , somando membro a membro e dividindo por  $(n-1)$  obteremos:

$$\bar{Y} = \beta_0 + \beta_1 \cdot \bar{Y} + \bar{\varepsilon} \quad (2.31)$$

onde

$$\bar{\varepsilon} = \frac{1}{n-1} \sum_{t=2}^n \varepsilon_t$$

Subtraindo (2.31) de (2.29), vem,

$$Y_t - \bar{Y} = \beta_1(Y_{t-1} - \bar{Y}) + (\varepsilon_t - \bar{\varepsilon}) \quad (2.32)$$

Multiplicando (2.32) por  $(Y_{t-1} - \bar{Y})$  e somando de 2 até  $n$ , temos:

$$\begin{aligned} \sum_{t=2}^n (Y_t - \bar{Y}) \cdot (Y_{t-1} - \bar{Y}) &= \beta_1 \sum_{t=2}^n (Y_{t-1} - \bar{Y})^2 + \\ &= \sum_{t=2}^n (\varepsilon_t - \bar{\varepsilon}) \cdot (Y_{t-1} - \bar{Y}) \end{aligned} \quad (2.33)$$

dividindo (2.33) por  $\sum_{t=2}^n (Y_{t-1} - \bar{Y})^2$ , teremos

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{t=2}^n (\varepsilon_t - \bar{\varepsilon}) \cdot (Y_{t-1} - \bar{Y})}{\sum_{t=2}^n (Y_{t-1} - \bar{Y})^2} \quad (2.34)$$

No cálculo de  $E(\hat{\beta}_1)$  notamos que é de difícil manipulação, em decorrência da 2.<sup>a</sup> parcela do 2.<sup>o</sup> membro de (2.24)

ser o quociente de variáveis aleatórias. Todavia o sinal da Esperança desta parcela depende do sinal do numerador, uma vez que o denominador é uma soma de quadrados.

Retomando a expressão (2.29) e fazendo sucessivamente  $t = 1, 2, 3, \dots, n-1$ , obteremos:

$$Y_1 = Y_1$$

$$Y_2 = \beta_0 + \beta_1 Y_1 + \epsilon_2$$

$$Y_3 = \beta_0 + \beta_0 \beta_1 + \beta_1^2 Y_1 + \beta_1 \epsilon_2 + \epsilon_3$$

$$Y_4 = \beta_0 + \beta_1 \beta_0 + \beta_1^2 \beta_0 + \beta_1^3 Y_1 + \beta_1^2 \epsilon_2 + \beta_1 \epsilon_3 + \epsilon_4$$

.. .. .

$$Y_{n-1} = \beta_0 + \beta_0 \beta_1 + \dots + \beta_0 \beta_1^{n-3} + \beta_1^{n-2} Y_1 + \beta_1^{n-3} \epsilon_2 + \beta_1^{n-4} \epsilon_3 + \dots + \epsilon_{n-1}$$

Somando membro a membro as relações acima e dividindo por  $(n-1)$ , teremos:

$$\bar{Y} = \phi(\beta_0, \beta_1, Y_1) + B_2 \cdot \epsilon_2 + B_3 \cdot \epsilon_3 + \dots + B_{n-1} \cdot \epsilon_{n-1} \quad (2.35)$$

onde

$$B_2 = (1 + \beta_1 + \beta_1^2 + \dots + \beta_1^{n-3}) / n-1$$

$$B_3 = (1 + \beta_1 + \beta_1^2 + \dots + \beta_1^{n-4}) / n-1$$

$$B_4 = (1 + \beta_1 + \beta_1^2 + \dots + \beta_1^{n-5}) / n-1$$

$$B_{n-1} = 1/n-1$$

Se  $\beta_1 > 0$ , teremos  $B_i > 0 \forall i \in \{2, 3, \dots, n-1\}$  pois cada

$B_1$  é média de potências de  $\beta_1$ .

Retomando a expressão (2.34)

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{t=2}^n (\varepsilon_t - \bar{\varepsilon}) \cdot (Y_{t-1} - \bar{Y})}{\sum_{t=2}^n (Y_{t-1} - \bar{Y})^2}$$

calculemos a esperança do numerador da 2.<sup>a</sup> parcela do 2.<sup>o</sup> membro, isto é, de

$$\sum_{t=2}^n (\varepsilon_t - \bar{\varepsilon}) \cdot (Y_{t-1} - \bar{Y}) = \sum_{t=2}^n (Y_{t-1} - \bar{Y}) \varepsilon_t - \sum_{t=2}^n (Y_{t-1} - \bar{Y}) \cdot \bar{\varepsilon}$$

Temos,

$$\sum_{t=2}^n (Y_{t-1} - \bar{Y}) \cdot \bar{\varepsilon} = \bar{\varepsilon} [(n-1) \cdot \bar{Y} - (n-1) \cdot \bar{Y}] = 0$$

Logo,

$$\sum_{t=2}^n (\varepsilon_t - \bar{\varepsilon}) \cdot (Y_{t-1} - \bar{Y}) = \sum_{t=2}^n (Y_{t-1} - \bar{Y}) \cdot \varepsilon_t$$

e

$$\begin{aligned} E \left\{ \sum_{t=2}^n (\varepsilon_t - \bar{\varepsilon}) \cdot (Y_{t-1} - \bar{Y}) \right\} &= E \left\{ \sum_{t=2}^n (Y_{t-1} - \bar{Y}) \varepsilon_t \right\} = \\ &= E \left\{ \sum_{t=2}^n Y_{t-1} \cdot \varepsilon_t - \sum_{t=2}^n \bar{Y} \varepsilon_t \right\} \end{aligned}$$

Mas,

$$\sum_{t=2}^n Y_{t-1} \cdot \varepsilon_t = Y_1 \varepsilon_2 + Y_2 \varepsilon_3 + Y_3 \varepsilon_4 + \dots + Y_{n-1} \cdot \varepsilon_n$$

e portanto

$$E \left\{ \sum_{t=2}^n Y_{t-1} \cdot \varepsilon_t \right\} = 0$$

Segue-se então que

$$E \left\{ \sum_{t=2}^n Y_{t-1} \cdot \varepsilon_t - \sum_{t=2}^n \bar{Y} \varepsilon_t \right\} = -E \left\{ \bar{Y} \sum_{t=2}^n \varepsilon_t \right\}$$

e lembrando (2.35)

$$E \left\{ \sum_{t=2}^n Y_{t-1} \cdot \varepsilon_t - \sum_{t=2}^n \bar{Y} \varepsilon_t \right\} = -E [ (B_2 \varepsilon_2 + B_3 \varepsilon_3 + \dots + B_{n-1} \varepsilon_{n-1}) \cdot (\varepsilon_2 + \varepsilon_3 + \dots + \varepsilon_{n-1}) ] = -\sigma^2 \cdot \sum_{i=2}^{n-1} B_i < 0.$$

Portanto se  $\beta_i > 0$ , o viés de  $\hat{\beta}_1$  é negativo.

Experiências realizadas por Malinvaud [26] sugerem que o uso prático de estimadores de mínimos quadrados em modelos de variáveis defasadas, como o que acabamos de ver, produzem resultados satisfatórios, apesar do viés existente.

Experiências realizadas por ele, mostraram que, em amostra com  $n = 20$  no modelo

$$Y_t = \beta_0 + \beta_1 \cdot Y_{t-1} + \beta_2 \cdot X_t + \varepsilon_t$$



onde  $\beta_1 = 0,6$ ,  $\hat{\beta}_1$  subestima  $\beta_1$ , sendo o viés da ordem de  $-0,08$ .

## 2.7 - MODELO DE SOLOW OU DAS DEFASAGENS DE PASCAL

Em certos problemas econômicos, um modelo de defasagem distribuídas (2.2) com restrições aos  $\beta_i$ 's de forma a decrescerem geometricamente, pode não ser adequado. Por exemplo, no relacionamento entre gastos de capital corrente e a apropriações de capital passado, é razoável admitir que os pesos ligados a apropriações de capital nos períodos  $t$ ,  $t-1$ ,  $t-2, \dots$  cresçam primeiro, e em seguida decresçam ao invés de decrescerem somente.

Solow [36] propôs que os  $\beta_i$ 's do modelo (2.2) seguissem um comportamento de um "V invertido", sendo tais  $\beta_i$ 's não números arbitrários, e sim números proporcionais às probabilidades de uma distribuição de Pascal. Isto é, considerando (2.2) como  $\alpha = 0$

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \dots + \varepsilon_t$$

Vamos escrevê-lo sob a forma

$$Y_t = \beta [w_0 \cdot X_t + w_1 \cdot X_{t-1} + \dots] + \varepsilon_t \quad (2.36)$$

onde

$$w_i = \frac{\beta_i}{\beta} \quad \forall_i \in \{0, 1, 2, \dots\}$$

Solow propôs que os  $w_i$ 's seguissem uma distribuição de Pas-

cal, isto é,

$$w_i = \binom{i+r-1}{i} (1-\lambda)^r \cdot \lambda^i \quad i=0,1,2,\dots$$

onde  $r$  é um inteiro positivo e  $\lambda$  um real entre 0 e 1 ( $r$  e  $\lambda$  são os parâmetros da distribuição de Pascal).

A figura 6 fornece o gráfico de  $w_i$  em função de  $i$  para  $\lambda = 0,6$  e  $r = 1, 2, 3$  e  $4$ .

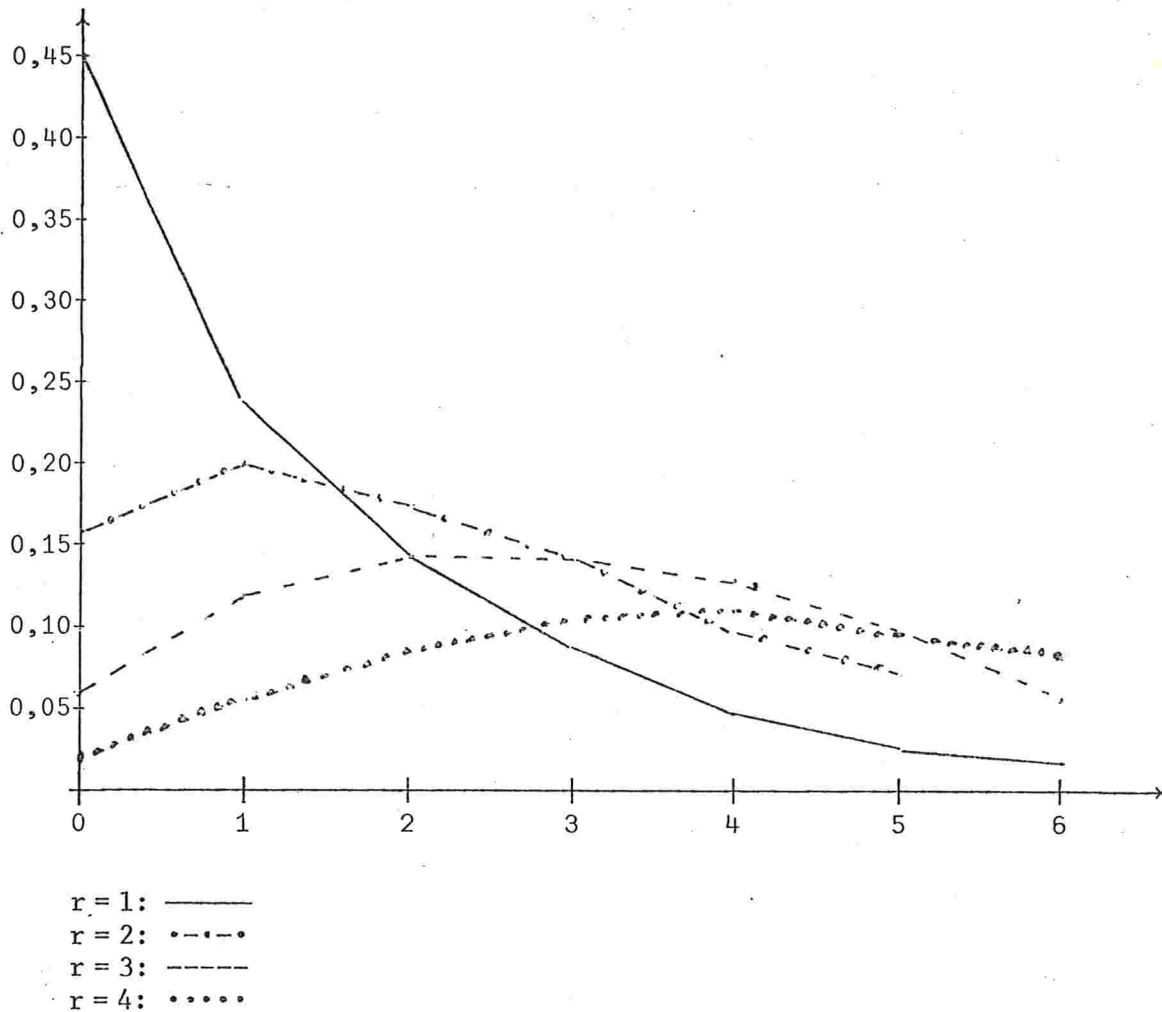


Figura 6

Observemos que a distribuição de Pascal se reduz à geométrica para  $r = 1$ , isto é

$$w_i = (1 - \lambda) \cdot \lambda^i$$

## CAPÍTULO 3

### MÉTODOS DE ESTIMAÇÃO

#### 3.1 - INTRODUÇÃO

Já vimos no Capítulo 2 que, se no conjunto das variáveis explicativas dos modelos compareceram apenas valores defasados de  $X$  ( $X_t, X_{t-1}$ , etc.) não há problemas para a estimação dos parâmetros; podemos aplicar diretamente os estimadores de mínimos quadrados ou ainda se houver um grande número de variáveis defasadas no modelo, podemos reduzir o número de parâmetros do modelo cuja expressão é (2.1) impondo a eles certas restrições, como por exemplo, atribuição de pesos arbitrários ou adotando o modelo polinomial de Almon.

Todavia, caso o modelo inclua valores defasados de variável dependente  $Y$  ( $Y_{t-1}, Y_{t-2}$ , etc.) no conjunto das variáveis explicativas ocorrem conforme já vimos, sérios problemas com respeito às propriedades dos estimadores, se eles forem de mínimos quadrados. A ocorrência de tais problemas é acarretada ou pela auto-correlação nos erros, ou então pela correlação entre o erro e a variável  $Y$  defasada que faz papel da variável explicativa, ou por ambos.

Passemos à análise dos métodos alternativos de estimação propostos em cada situação.

### 3.2 - MODELOS DE KOYCK, CAGAN E NERLOVE

Os três modelos já vistos no Capítulo 2, embora tenham sido concebidos sob hipóteses diferentes, recaem após algumas transformações em expressões do tipo

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + v_t \quad (3.1)$$

Os métodos de estimação de (3.1) vão depender essencialmente das hipóteses sobre o comportamento de  $v_t$ , as quais podem basicamente serem resumidas em três.

1º hipótese) os  $v_t$ 's são variáveis aleatórias independentes e normalmente distribuídas com média 0 e variância  $\sigma_v^2$ .

2º hipótese)  $v_t = \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$ ,  $0 < \lambda < 1$  onde, podem ser concebidos ainda os casos (a) e (b).

a) os  $\varepsilon_t$ 's são variáveis aleatórias independentes e normalmente distribuídas com média 0 e variância  $\sigma_\varepsilon^2$ ;

b)  $\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t$  onde  $|\rho| < 1$  e os  $u_t$ 's são variáveis aleatórias independentes e normalmente distribuídas com média 0 e variância  $\sigma_u^2$ .

3º hipótese)  $v_t = \rho \cdot v_{t-1} + u_t$  onde  $|\rho| < 1$  e os  $u_t$ 's são variáveis aleatórias independentes e normalmente dis

tribuídas com média 0 e variância  $\sigma_u^2$ .

### 3.2.1 - Estimação Sob a 1ª Hipótese

Esta é a hipótese frequentemente adotada no modelo dos ajustamentos parciais. Conforme vimos no Capítulo 2, o uso de estimadores de mínimos quadrados, nesse caso, nos levam a estimadores consistentes para os parâmetros, embora eles sejam ligeiramente viesados para amostras pequenas.

Assim, a estimação dos parâmetros do modelo (3.1) sujeito a erros cujo comportamento segue a 1ª hipótese deve ser feita da preferência com grande número de observações e com estimadores de mínimos quadrados, visto que os mesmos gozam da propriedade da consistência.

Por outro lado, no final deste capítulo veremos testes apropriados para detectar correlação nos erros e assim ter mais um instrumento para analisar a viabilidade da admissão desta 1ª hipótese.

### 3.2.2 - Estimação Sob a 2ª Hipótese - (a)

Esta é a hipótese frequentemente adotada nos modelos de Koyck e Cagan. Conforme vimos no Capítulo 2, o uso de estimadores de mínimos quadrados nos levam a estimadores inconsistentes dos parâmetros. Vejamos então os estimadores propostos, e que conduzem a estimadores consistentes dos parâmetros de (3.1).

### 3.2.2.1 - Método dos Mínimos Quadrados Generalizados

Nesse caso,  $v_t = \varepsilon_t - \lambda \varepsilon_{t-1}$   $0 < \lambda < 1$ , então  $E(v_t) = 0$ ,  
 $\forall t$

$$E(v_t^2) = \sigma_\varepsilon^2(1 + \lambda^2) \quad \forall t$$

$$E(v_t \cdot v_{t+s}) = \begin{cases} -\lambda \cdot \sigma_\varepsilon^2 \quad \forall t \text{ e } s = 1 \text{ ou } s = -1 \\ 0 \text{ para } |s| \geq 2, \quad \forall t. \end{cases}$$

Chamando de  $\underline{v}$  o vetor dos erros e de  $V$  a matriz de variância-covariância dos erros, teremos

$$V = E(\underline{v} \underline{v}') = \begin{bmatrix} (1 + \lambda^2) & -\lambda & 0 & \dots & 0 \\ -\lambda & (1 + \lambda^2) & -\lambda & \dots & 0 \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ 0 & 0 & 0 & (1 + \lambda^2) & \end{bmatrix} \quad (3.2)$$

Se  $\lambda$  for conhecido (por considerações *à priori*) então a matriz  $V$  também o será. Lembrando ainda que nos modelos de Koyck e Cagan o coeficiente de  $Y_{t-1}$  é  $\lambda$  (ver (2.13) e (2.20)), a expressão (3.1) pode ser escrita sob a forma

$$Y_t - \lambda \cdot Y_{t-1} = \beta_0 + \beta_2 \cdot X_t + v_t \quad (3.3)$$

O estimador de mínimos quadrados generalizados para o vetor de parâmetros de (3.3) é dado por

$$(X' \cdot V^{-1} \cdot X)^{-1} \cdot X' \cdot V^{-1} \cdot \underline{Y} \quad (3.4)$$

onde

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_n \end{bmatrix} \quad \text{e} \quad \underline{Y} = \begin{bmatrix} Y_1 - \lambda \cdot Y_0 \\ Y_2 - \lambda \cdot Y_1 \\ Y_3 - \lambda \cdot Y_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Y_n - \lambda \cdot Y_{n-1} \end{bmatrix}$$

O estimador dado por (3.4) é o estimador linear não viesado e de mínima variância para o vetor de parâmetros de (3.3) (para a demonstração, ver Johnston [16, pp. 209 e 210].

Para maiores detalhes sobre estimadores de mínimos quadrados generalizados ver [16, Cap. 7].

Entretanto, quando  $\lambda$  é não conhecido à priori outros métodos de estimação devem ser utilizados, que veremos a seguir.

#### 3.2.2.2 - Método de Zellner e Geisel

Quando  $\lambda$  dos modelos de Koyck e Cagan não é conhecido a priori, um método alternativo para estimar os parâmetros de (3.3) foi sugerido por Zellner e Geisel [42].

Basicamente, o método consiste em escrever (3.3) sob a forma



$$Y_t = \beta_0 + \lambda \cdot Y_{t-1} + \beta_2 \cdot X_t + (\varepsilon_t - \lambda \cdot \varepsilon_{t-1}) \quad (3.5)$$

Fazendo

$$w_t = Y_t - \varepsilon_t$$

teremos

$$w_t - \lambda \cdot w_{t-1} = Y_t - \lambda \cdot Y_{t-1} - (\varepsilon_t - \lambda \cdot \varepsilon_{t-1}) = \beta_0 + \beta_2 \cdot X_t$$

ou ainda

$$w_t = \lambda \cdot w_{t-1} + \beta_0 + \beta_2 \cdot X_t \quad (3.6)$$

Tomando (3.6) nos períodos  $t-1, t-2, \dots, 3, 2, 1$  e  $0$  e fazendo sucessivas substituições em (3.6) chega-se facilmente a

$$w_t = \lambda^t \cdot w_0 + \beta_0 (1 + \lambda + \lambda^2 + \dots + \lambda^{t-1}) + \beta_2 (X_t + \lambda \cdot X_{t-1} + \lambda^2 \cdot X_{t-2} + \dots + \lambda^{t-1} X_1) \quad (3.7)$$

onde  $w_0 = Y_0 - \varepsilon_0$  pode ser assumida como uma constante inicial.

Segue-se então que (3.5) fica sob a forma

$$Y_t = \lambda^t \cdot w_0 + \beta_0 \left( \sum_{i=0}^{t-1} \lambda^i \right) + \beta_2 \left( \sum_{i=0}^{t-1} \lambda^i \cdot X_{t-i} \right) + \varepsilon_t \quad (3.8)$$

Se  $\lambda$  fosse conhecido,  $Y_t$  seria uma função linear de  $w_0$ ,  $\beta_0$  e  $\beta_2$  com um erro  $\varepsilon_t$  normalmente distribuído, com média  $0$ , desvio padrão  $\sigma_\varepsilon$ , sendo os  $\varepsilon_t$ 's não correlacionados por hipótese. Assim, a aplicação de estimadores de mínimos

quadrados em (3.8) produziria estimadores com todas propriedades desejáveis.

Notemos ainda que a matriz das variáveis explicativas de 3.8 é

$$X = \begin{bmatrix} \lambda & 1 & X_1 \\ \lambda^2 & 1+\lambda & X_2 + \lambda \cdot X_1 \\ \lambda^3 & 1+\lambda+\lambda^2 & X_3 + \lambda \cdot X_2 + \lambda^2 \cdot X_1 \\ \vdots & \vdots & \vdots \\ \lambda^n & 1+\lambda^2+\dots+\lambda^{n-1} & X_n + \lambda \cdot X_{n-1} + \dots + \lambda^{n-1} \cdot X_1 \end{bmatrix} \quad (3.9)$$

Todavia, estamos admitindo desconhecer  $\lambda$ . O procedimento proposto por Zellner e Geisel, consiste em atribuir valores a  $\lambda$ , entre 0 e 1, obter para cada  $\lambda$  a matriz  $X$  e estimar  $w_0$ ,  $\beta_0$  e  $\beta_2$  de (3.8) usando estimadores de mínimos quadrados. Finalmente escolher o valor de  $\lambda$  (e conseqüentemente  $\beta_0$  e  $\beta_2$  que minimizem a variação residual.

Observemos que a atribuição de valores a  $\lambda$  se processa adotando-se zero como ponto de partida e um certo passo até atingir 1. Obtido  $\lambda$  que minimiza a variação residual, adota-se um passo menor e faz-se  $\lambda$  assumir valores em torno daquele ponto de mínimo e assim obtém-se uma melhor aproximação para o ponto de mínimo.

O método descrito foi aplicado por Zellner e Geisel para estimar os parâmetros de um modelo de função consumo nos

E.U.A. ã partir de dados trimestrais desde 1947 (I) até 1960 (IV). O modelo adotado foi o das expectativas adaptadas e derivado de

$$C_t = \lambda \cdot Y_t^* + \varepsilon_t$$

$$Y_t - Y_{t-1}^* = (1 - \lambda)(Y_t - Y_{t-1}^*)$$

onde

C = consumo real observado

Y = renda real observada

O modelo final sob a forma transformada ficou

$$C_t = \lambda \cdot C_{t-1} + \beta(1 - \lambda)Y_t + (\varepsilon_t - \lambda\varepsilon_{t-1})$$

Seguindo o procedimento descrito, o valor de  $\lambda$  que minimizava a variaçãõ residual foi 0,963 e o correspondente  $\hat{\beta}$  foi 1,129 (ver Zellner e Geisel [42, pp. 868 e 869]).

### 3.2.2.3 - Método de Regressãõ Ponderada

Tal método foi desenvolvido por Klein [18] e consiste em considerar (3.1) sob a forma

$$Y_t - \varepsilon_t = \beta_0 + \beta_2 X_t + \lambda(Y_{t-1} - \varepsilon_{t-1}) \quad (3.10)$$

que constitui formalmente a expressãõ dos modelos de Koyck Cagan.

A expressãõ (3.10) pode ser considerada como um mo

delo com erro nas variáveis, isto é,  $\varepsilon_t$  é o "erro" cometido na mensuração de  $E(Y_t)$  e  $\varepsilon_{t-1}$  é o "erro" cometido na mensuração de  $E(Y_{t-1})$ . Como os  $\varepsilon_t$ 's são homoscedásticos, podemos usar o método de regressão ponderada atribuído pesos iguais a cada um dos erros.

O procedimento de regressão ponderada leva a estimadores consistentes dos parâmetros de (3.10) e maiores detalhes sobre o método podem ser encontrados em Kmenta [19, no Capítulo 9].

Essencialmente o procedimento consiste em minimizar

$$S = \sum_{t=2}^n \varepsilon_t^2 + \sum_{t=2}^n \varepsilon_{t-1}^2 = \sum_{t=2}^n [Y_t - \beta_0 - \beta_2 X_t - \lambda \cdot \theta_{t-1}]^2 + \sum [Y_{t-1} - \theta_{t-1}]^2 \quad (3.11)$$

onde  $\theta_{t-1} = E(Y_{t-1})$ .

A minimização de  $S$  de (3.11) deve ser feita com relação a  $\beta_0$ ,  $\beta_2$ ,  $\lambda$  e cada um dos  $\theta_t$ 's.

Derivando (3.11) com respeito a  $\beta_0$ ,  $\beta_2$ ,  $\lambda$  e  $\theta_{t-1}$  e igualando a zero obteremos

$$\sum_{t=2}^n (Y_t - \beta_0 - \beta_2 X_t - \lambda \cdot \theta_{t-1}) = 0 \quad (3.12)$$

$$\sum_{t=2}^n X_t (Y_t - \beta_0 - \beta_2 X_t - \lambda \cdot \theta_{t-1}) = 0 \quad (3.13)$$

$$\sum_{t=2}^n \theta_{t-1} (Y_t - \beta_0 - \beta_2 X_t - \lambda \cdot \theta_{t-1}) = 0 \quad (3.14)$$

$$\lambda \cdot (Y_t - \beta_0 - \beta_2 X_t - \lambda \cdot \theta_{t-1}) + Y_{t-1} - \theta_{t-1} \quad (t=2,3,4,\dots,n) \quad (3.15)$$

De (3.12) obtemos

$$\beta_0 = \bar{Y} - \beta_2 \cdot \bar{X} - \lambda \left[ \frac{1}{n-1} \sum_{t=2}^n \theta_{t-1} \right] \quad (3.16)$$

onde

$$\bar{Y} = \frac{1}{n-1} \sum_{t=2}^n Y_t \quad \text{e} \quad \bar{X} = \frac{1}{n-1} \sum_{t=2}^n X_t$$

Substituindo (3.16) em (3.13), (3.14) e (3.15) podemos eliminar  $\beta_0$  expressar (3.16) sob a forma

$$\theta'_{t-1} = \frac{\lambda \cdot (Y'_t - \beta_2 \cdot X'_t) + Y'_{t-1}}{1 + \lambda^2} \quad (3.17)$$

onde,

$$\theta'_{t-1} = \theta_{t-1} - \frac{1}{n-1} \sum_{t=2}^n \theta_{t-1}$$

$$Y'_t = Y_t - \bar{Y}$$

$$X'_t = X_t - \bar{X}$$

$$Y'_{t-1} = Y_{t-1} - \frac{1}{n-1} \sum_{t=2}^n Y_{t-1}$$

Substituindo (3.17) em (3.13) e (3.14) e eliminando  $\beta_0$ , vem

$$\begin{aligned}
 & \lambda^2 \left[ \frac{\sum X'_t \cdot Y'_t \cdot \sum X'_t \cdot Y'_{t-1}}{\sum X'^2_t} - \sum Y'_t \cdot Y'_{t-1} \right] + \\
 & + \lambda \left[ \frac{\left[ \sum X'_t \cdot Y'_{t-1} \right]^2 - \left[ \sum X'_t \cdot Y'_t \right]^2}{\sum X'^2_t} + \sum Y'^2_t - \sum Y'^2_{t-1} \right] + \\
 & + \left[ \sum Y'_t Y'_{t-1} - \frac{\sum X'_t \cdot Y'_t \cdot \sum X'_t \cdot Y'_{t-1}}{\sum X'^2_t} \right] = 0 \quad (3.18)
 \end{aligned}$$

A expressão (3.18) é uma equação do 2º grau em  $\lambda$ , resolvendo-a, escolhemos a raiz que minimiza S, e ela será  $\hat{\lambda}$  o estimador de  $\lambda$ .

Finalmente usando  $\hat{\lambda}$  e as relações (3.12), (3.13), (3.14) e (3.15) obtemos os estimadores consistentes para  $\beta_0, \beta_2$  e os  $\theta$ 's.

Embora os estimadores assim obtidos sejam consistentes, eles não são assintoticamente os mais eficientes (ver Amemiya e Fuller [03, pp. 516 a 520]).

#### 3.2.2.4 - Método das Variáveis Instrumentais

Tal método foi desenvolvido por Liviatan [22] para o uso de modelos de defasagens distribuídas. No Apêndice 2 apresentamos um pequeno resumo sobre as variáveis instrumentais.

Essencialmente o método consiste em obter estimadores consistentes para os parâmetros de (3.1) quando ocorre correlação entre a variável  $Y_{t-1}$  e o erro (pois como já vi-

mos os estimadores de mínimos quadrados são inconsistentes).

Retomando

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + v_t \quad (3.18')$$

e considerando que  $X_t$  não é estocástica, Liviatan sugere as seguintes variáveis instrumentais

$$Z_{1t} = X_{t-1} \quad \text{e} \quad Z_{2t} = X_t$$

ou seja, a variável instrumental para  $Y_{t-1}$  é  $X_{t-1}$  e para  $X_t$  é a própria  $X_t$  (já que  $X_t$  é não estocástica).

As equações normais para obtenção dos estimadores de variáveis instrumentais são:

$$\sum_{t=2}^n Y_t = (n-1) \cdot \beta_0 + \sum_{t=2}^n Y_{t-1} \cdot \beta_1 + \sum_{t=2}^n X_t \cdot \beta_2$$

$$\sum_{t=2}^n X_t \cdot Y_t = \sum_{t=2}^n X_t \beta_0 + \sum_{t=2}^n X_t \cdot Y_{t-1} \cdot \beta_1 + \sum_{t=2}^n X_t^2 \cdot \beta_2 \quad (3.19)$$

$$\sum_{t=2}^n X_{t-1} \cdot Y_t = \sum_{t=2}^n X_{t-1} \cdot \beta_0 + \sum_{t=2}^n X_{t-1} \cdot Y_{t-1} \cdot \beta_1 + \sum_{t=2}^n X_t \cdot X_{t-1} \cdot \beta_2$$

O sistema (3.19) ainda pode ser expresso sob a forma matricial

$$Z' \cdot Y = (Z' \cdot X) \cdot \beta \quad (3.20)$$

onde

$$Y = \begin{bmatrix} Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \quad Z = \begin{bmatrix} 1 & X_2 & X_1 \\ 1 & X_3 & X_2 \\ \vdots & \vdots & \vdots \\ 1 & X_n & X_{n-1} \end{bmatrix} \quad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} e$$

$$X = \begin{bmatrix} 1 & Y_1 & X_2 \\ 1 & Y_2 & X_3 \\ \vdots & \vdots & \vdots \\ 1 & Y_{n-1} & X_n \end{bmatrix}$$

Sendo que a solução de (3.20) é

$$\hat{\beta} = (Z' \cdot X)^{-1} \cdot (Z' \cdot Y)$$

Finalmente cabe lembrar que os estimadores de variáveis instrumentais embora consistentes, não são necessariamente os assintoticamente mais eficientes (ver Hannan [12, pp. 219]).

### 3.2.3 - Estimação sob a 2ª Hipótese (b)

Nesse caso temos o modelo

$$Y_t = \beta_0 + \beta_1 \cdot Y_{t-1} + \beta_2 X_t + v_t \quad (3.21)$$

onde

$$v_t = \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$$



e

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t \quad |\rho| < 1$$

Nesse caso, notemos primeiramente que no caso particular de  $\rho = \lambda$ , teremos  $v_t = u_t$  isto é o erro de (3.1) é normalmente distribuído, com média 0 e os  $v_t$ 's não são correlacionados. Segue-se então que, nesse caso, os estimadores de mínimos quadrados aplicados a (3.1) são consistentes. Todavia a admissão a priori de que  $\rho = \lambda$  é difícil de ser justificada. Vamos então analisar o caso onde nada se conhece a respeito de  $\rho$  e  $\lambda$ .

### 3.2.3.1 - Método de Zellner e Geisel

O método consiste em tomar (3.1) com  $\beta_1 = \lambda$  e fazer

$$w_t = Y_t - \varepsilon_t \quad (3.21)$$

Assim,

$$w_t - \rho \cdot w_{t-1} = Y_t - \rho Y_{t-1} - (\varepsilon_t - \rho \varepsilon_{t-1}) \quad (3.22)$$

sendo que (3.22) pode ser escrita como,

$$w_t(\rho) = Y_t(\rho) - \varepsilon_t(\rho) \quad (3.23)$$

onde

$$w_t(\rho) = w_t - \rho \cdot w_{t-1}$$

$$Y_t(\rho) = Y_t - \rho \cdot Y_{t-1}$$

$$\varepsilon_t(\rho) = \varepsilon_t - \rho \cdot \varepsilon_{t-1}$$

$$X_t(\rho) = X_t - \rho \cdot X_{t-1}$$

De (3.1) obtemos

$$Y_t(\rho) = \beta_0(1 - \rho) + \lambda \cdot Y_{t-1}(\rho) + \beta_2 X_t(\rho) + \varepsilon_t(\rho) - \lambda \cdot \varepsilon_{t-1}(\rho) \quad (3.24)$$

Conseqüentemente, teremos:

$$w_t(\rho) = \beta_0(1 - \rho) + \lambda \cdot w_{t-1}(\rho) + \beta_2 \cdot X_t(\rho) \quad (3.25)$$

Tomando (3.25) nos períodos  $t-1, t-2, \dots, 2, 1, 0$  e substituindo sucessivamente em (3.25) teremos,

$$\begin{aligned} w_t(\rho) &= \lambda^t \cdot w_0(\rho) + \beta_0(1 - \rho)[1 + \lambda + \lambda^2 + \dots + \lambda^{t-1}] + \\ &+ [\beta_2 X_t(\rho) + \lambda \cdot X_{t-1}(\rho) + \dots + \lambda^{t-1} X_1(\rho)] \end{aligned} \quad (3.26)$$

e finalmente

$$\begin{aligned} Y_t(\rho) &= \lambda^t \cdot w_0(\rho) + \beta_0(1 - \rho)[1 + \lambda + \lambda^2 + \dots + \lambda^{t-1}] + \\ &+ \beta_2 [X_t(\rho) + \lambda X_{t-1}(\rho) + \dots + \lambda^{t-1} X_1(\rho)] + u_t \end{aligned} \quad (3.27)$$

Se  $\lambda$  e  $\rho$  forem conhecidos os parâmetros  $w_0(\rho)$ ,  $\beta_0(1 - \rho)$  e  $\beta_2$  podem ser estimados facilmente usando-se estimadores de mínimos quadrados.

A matriz  $X$  das observações é

$$X = \begin{bmatrix} \lambda & 1 & X_1(\rho) \\ \lambda^2 & 1 + \lambda & X_2(\rho) + \lambda X_1(\rho) \\ \lambda^3 & 1 + \lambda + \lambda^2 & X_3(\rho) + \lambda X_2(\rho) + \lambda^2 X_1(\rho) \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \lambda^n & 1 + \lambda + \lambda^2 + \dots + \lambda^{n-1} & X_n(\rho) + \lambda X_{n-1}(\rho) + \dots + \lambda^{n-1} X_1(\rho) \end{bmatrix}$$

Todavia, em geral  $\lambda$  e  $\rho$  são desconhecidos e Zellner e Geisel sugerem selecionar pares de valores para  $\lambda$  e  $\rho$  do produto cartesiano  $]0,1[X] - 1, 1[$  e em seguida, para cada par obter os estimadores de mínimos quadrados dos parâmetros de (3.27) e a correspondente variação residual.

Finalmente, sugerem que se escolha  $\lambda$  e  $\rho$  (e consequentemente  $\beta_0$  e  $\beta_2$ ) que minimizem tal variação residual.

Os autores aplicaram esse método ao problema já citado em (2.2.2) e obtiveram os seguintes resultados:

$$\hat{\lambda} = 0,66$$

$$\hat{\rho} = 0,69$$

$$\hat{\beta} = 0,94$$

Ver [42, pp. 870 e 871].

### 3.2.3.2 - Método das Variáveis Instrumentais

O método é essencialmente o mesmo que o já descrito em (2.2.4) e os estimadores obtidos gozam da propriedade

consistência.

### 3.2.4 - Estimação Sob a 3.<sup>a</sup> Hipótese

Neste caso temos

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + v_t \quad (3.1)$$

onde

$$v_t = \rho \cdot v_{t-1} + u_t \text{ onde } |\rho| < 1$$

e os  $u_t$ 's são variáveis aleatórias independentes e normalmente distribuídas com média 0 e variância  $\sigma_u^2$ .

Tal modelo costuma aparecer sob diferentes suposições teóricas baseadas em hipóteses comportamentais. Ele não decorre necessariamente do modelo transformado de Koyck, dos ajustamentos parciais ou das expectativas adaptadas.

Por outro lado, é fácil constatar que  $v_t$  e  $Y_{t-1}$  são correlacionados e isto faz com que os estimadores de mínimos quadrados sejam inconsistentes. Vejamos a seguir métodos para a obtenção de estimadores consistentes.

#### 3.2.4.1 - Método das Variáveis Instrumentais

O método é o descrito em (2.2.4) e os estimadores assim obtidos gozam da propriedade da consistência.

#### 3.2.4.2 - Método dos Mínimos Quadrados Após Transformação do Modelo

O modelo dado pela expressão (3.1) e sob a 3.<sup>a</sup> hi-

pótese, pode ser escrito sob a forma

$$Y_t = \beta_0 + \beta_1 \cdot Y_{t-1} + \beta_2 X_t + \rho Y_{t-1} + u_t \text{ onde } |\rho| < 1$$

ou ainda

$$Y_t - \rho Y_{t-1} = \beta_0(1 - \rho) + \beta_1(Y_{t-1} - \rho Y_{t-2}) + \beta_2(X_t - \rho X_{t-1}) + u_t \quad (3.28)$$

Se  $\rho$  for conhecido *à priori*, podemos aplicar os estimadores de mínimos quadrados *à* (3.28) e estimar facilmente  $\beta_0$ ,  $\beta_1$  e  $\beta_2$ . Os estimadores assim obtidos são consistentes e assintoticamente eficientes para os parâmetros.

Todavia, em geral, se desconhece *à priori* o valor de  $\rho$ . Nesse caso, vários procedimentos são sugeridos, e é o que veremos a seguir.

#### 3.2.4.3 - Métodos de Mínimos Quadrados com Iteração

Retomando a expressão (3.28), podemos reescrevê-la sob a forma

$$Y_t = \beta_0(1 - \rho) + (\beta_1 + \rho) \cdot Y_{t-1} - \beta_1 \rho Y_{t-2} + \beta_2 X_t - \beta_2 \rho X_{t-1} + u_t \quad (3.29)$$

A estimação de  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  e  $\rho$  via estimadores de mínimos quadrados em (3.29) levaria a um sistema não linear nos estimadores, cuja resolução analítica se torna extremamente complicada.

Tal dificuldade pode ser contornada, separando-se os parâmetros em dois grupos, um contendo  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  outro

contendo  $\rho$ . Em seguida, obtemos estimadores de mínimos quadrados condicionados ao valor (es) do (s) parâmetro (s) de cada grupo adotando os seguintes passos:

1º) Escolhemos um valor inicial para  $\rho$  ( $\hat{\rho}$ ) e encontramos os estimadores de mínimos quadrados de  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  em

$$(Y_t - \hat{\rho}Y_{t-1}) = \beta_0(1 - \hat{\rho}) + \beta_1(Y_{t-1} - \hat{\rho}Y_{t-2}) + \beta_2(X_t - \hat{\rho}X_{t-1}) + u_t \quad (3.30)$$

Tais estimadores serão designados por  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  e  $\hat{\beta}_2$ .

2º) De posse de  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  e  $\hat{\beta}_2$ , substituímos os mesmos nos lugares de  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  em (3.29) e obtemos o estimador de mínimos quadrados de  $\rho$  (que vamos indicar por  $\hat{\hat{\rho}}$ ). Isto é, encontramos o estimador de mínimos quadrados de  $\rho$  em

$$(Y_t - \hat{\beta}_0 - \hat{\beta}_1 Y_{t-1} - \hat{\beta}_2 X_t) = \rho (Y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 Y_{t-2} - \hat{\beta}_2 X_{t-1}) + u_t \quad (3.31)$$

ou seja

$$\hat{u}_t = \rho \cdot \hat{u}_{t-1} + u_t$$

onde os  $\hat{u}_t$ 's são os resíduos obtidos na estimação dos  $\beta_i$ 's em (3.30).

O valor  $\hat{\hat{\rho}}$  assim encontrado passa a fazer o papel do novo  $\hat{\rho}$  no 1º passo e os dois passos assim vão se repetindo até que se consiga para os  $\hat{\beta}_i$ 's um desejável grau de convergência numérica.

Observemos finalmente que o procedimento assim descrito é extremamente trabalhoso, a menos que se use compu

tadores.

### 3.2.4.4 - Método de Wallis

O método consiste essencialmente nos seguintes passos

1º) Estimar  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  do modelo

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + v_t \quad (3.1)$$

usando  $X_{t-1}$  como variável instrumental para  $Y_{t-1}$ . Isto é, obter  $\hat{\beta}$  através da relação

$$\hat{\beta} = (Z'X)^{-1} Z'Y$$

onde

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad Z = \begin{bmatrix} 1 & X_0 & X_1 \\ 1 & X_1 & X_2 \\ \vdots & \vdots & \vdots \\ 1 & X_{n-1} & X_n \end{bmatrix} \quad e$$

$$X = \begin{bmatrix} 1 & Y_0 & X_1 \\ 1 & Y_1 & X_2 \\ \vdots & \vdots & \vdots \\ 1 & Y_{n-1} & X_n \end{bmatrix}$$

2º) Calcular os resíduos da estimação dos parâmetros de (3.1)

$$\hat{\underline{Y}} = \underline{Y} - X \cdot \hat{\underline{\beta}}$$

onde

$$\hat{\underline{Y}} = \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \hat{v}_3 \\ \vdots \\ \hat{v}_n \end{bmatrix}$$

e assim, calcular r (estimador de  $\rho$ ) dado por

$$r = \frac{\sum_{t=2}^n \hat{v}_t \cdot \hat{v}_{t-1} / n-1}{\sum_{t=1}^n \hat{v}_t^2 / n} + \frac{3}{n}$$

onde o termo  $\frac{3}{n}$  é uma correção para o viés.

3º) Com a estimativa de  $\rho$  calculada anteriormente, obter a estimativa da matriz de variância-covariância dos erros  $\hat{V}$

$$\hat{V} = \begin{bmatrix} 1 & r & r^2 & \dots & r^{n-1} \\ r & 1 & r & \dots & r^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^{n-1} & r^{n-2} & r^{n-3} & \dots & 1 \end{bmatrix}$$



Finalmente estimar  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  usando o método dos mínimos quadrados generalizados, através da relação

$$\underline{\hat{\beta}}^* = (X' \hat{V} X)^{-1} \cdot X' \cdot \hat{V}^{-1} \cdot \underline{Y}$$

onde

$$\underline{\hat{\beta}}^* = \begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \hat{\beta}_2^* \end{bmatrix}$$

é o vetor dos estimadores, proposto por Wallis.

Com relação ao método proposto por Wallis, cabem as seguintes observações:

- a) O termo  $\frac{3}{n}$  usado para correção do viés na estimação de  $\rho$  foi escolhido entre outros, através de resultados experimentais obtidos por simulação (ver [40, pp.555]).
- b) Wallis demonstrou que os estimadores por ele propostos eram consistentes porém não assintoticamente eficientes.

Recentemente, Hatanaka [13] desenvolveu um método para estimar os parâmetros de modelo sob a 3.<sup>a</sup> hipótese, que é algo semelhante ao de Wallis, porém que conduz a estimadores consistentes e assintoticamente eficientes. Passemos a descrição do método.

Inicialmente, consideremos por simplicidade, o modelo dado pela expressão (3.1) sob a forma

$$Y_t = \beta_1 Y_{t-1} + \beta_2 X_t + v_t$$

com

$$v_t = \rho \cdot v_{t-1} + u_t$$

Hatanaka propôs os seguintes passos:

- (i) Usar o método das variáveis instrumentais para obter estimadores consistentes  $\tilde{\beta}_1$  e  $\tilde{\beta}_2$ . O estimador  $\tilde{\rho}$  é construído através da relação

$$\tilde{\rho} = \frac{\sum_t \tilde{v}_t \tilde{v}_{t-1}}{\sum_t \tilde{v}_{t-1}^2}$$

onde

$$\tilde{v}_t = Y_t - \tilde{\beta}_1 Y_{t-1} - \tilde{\beta}_2 X_t$$

- (ii) Aplicar o método dos mínimos quadrados, considerando  $Y_t - \tilde{\rho} Y_{t-1}$  como variável dependente e  $X_t - \tilde{\rho} X_{t-1}$ ,  $Y_{t-1} - \tilde{\rho} Y_{t-2}$  e  $\tilde{v}_{t-1}$  como variáveis independentes.

Consideremos  $\hat{\beta}_2$ ,  $\hat{\beta}_1$  e  $\hat{\rho}$  os coeficientes de  $X_t - \tilde{\rho} X_{t-1}$ ,  $Y_{t-1} - \tilde{\rho} Y_{t-2}$  e  $\tilde{v}_{t-1}$  respectivamente. Pondo  $\hat{\rho} = \tilde{\rho} - \hat{\rho}$  então  $\hat{\beta}_2$ ,  $\hat{\beta}_1$  e  $\hat{\rho}$  são os estimadores consistentes e assintoticamente eficientes propostos (para demonstração, ver [13, pp.203-205]).

### 3.3 - MODELO DE SOLOW OU DAS DEFASAGENS DE PASCAL

Retomando a expressão geral do modelo de Solow

$$Y_t = \beta[w_0 X_t + w_1 X_{t-1} + \dots] + \varepsilon_t \quad (2.36)$$

$$w_i = \binom{i+r-1}{i} (1-\lambda)^r \cdot \lambda^i \quad i=0,1,2,\dots$$

onde  $r$  é inteiro e positivo e  $\lambda$  um real entre 0 e 1 vejamos os métodos propostos para a estimação de  $\beta, \lambda$  e  $r$ .

Inicialmente a expressão (2.36) pode ser colocada em uma forma conveniente introduzindo-se o operador  $L$ , com a seguinte propriedade

- (i)  $LX_t = X_{t-1}$
- (ii)  $L^K X_t = L(L^{K-1} X_t) = X_{t-k}$

Assim, lembrando o Teorema binomial

$$\sum_{i=0}^{\infty} \binom{r+i-1}{i} X^i = (1-X)^{-r}$$

a expressão (2.36) pode ser escrita sob a forma

$$Y_t = \beta(1-\lambda)^r \sum_{i=0}^{\infty} \binom{r+i-1}{i} \lambda^i L^i X_t + \varepsilon_t = \beta(1-\lambda)^r (1-\lambda L)^{-r} X_t + \varepsilon_t \quad (3.32)$$

Finalmente de (3.32) obtém-se

$$(1-\lambda L)^r Y_t = \beta(1-\lambda)^r X_t + (1-\lambda L)^r \varepsilon_t \quad (3.33)$$

ou

$$\begin{aligned} Y_t - \binom{r}{1} \lambda \cdot Y_{t-1} + \binom{r}{2} \lambda^2 Y_{t-2} + \dots + (-1)^r \lambda^r \cdot Y_{t-r} &= \\ = \beta(1-\lambda)^r X_t + \sum_{K=0}^r \binom{r}{K} (-\lambda)^K \cdot u_{t-k} & \end{aligned} \quad (3.34)$$

Solow propôs estimadores para  $\beta$  e  $\lambda$  de (3.34) para cada valor específico  $\hat{a}$  priori de  $r$ , sob as suposições dos erros de (3.34) não serem ou serem correlacionados. Solow propôs ainda que se estimassem  $\lambda$  e  $\beta$  de (3.34) para diferentes valores de  $r$  e se escolhesse aquele que minimasse a variação residual.

Vejamos os estimadores propostos por Solow e ainda, estimadores de máxima verossimilhança propostos por Maddala e Rao.

### 3.3.1 - Método de Mínimos Quadrados

Tal método é aplicável nos parâmetros de (3.34) sob a suposição dos erros serem não correlacionados.

O método consiste em minimizar a expressão

$$\sum_{t=1}^n (Y_t - \binom{r}{1} \lambda Y_{t-1} + \binom{r}{2} \lambda^2 Y_{t-2} + \dots + (-\lambda)^r Y_{t-r} - \alpha X_t)^2$$

em relação a  $\lambda$  e  $\alpha$  (onde  $\alpha = \beta(1 - \lambda)^r$ ).

Para o caso de  $r = 3$  as equações normais serão

$$A - 3B\lambda + 3C\lambda^2 - D \cdot \lambda^3 - N \cdot \alpha = 0 \quad (3.35)$$

$$E - (3H + 2F)\lambda + (9I + G)\lambda^2 - (4J + 6K)\lambda^3 + 5L\lambda^4 - M\lambda^5 - \alpha(B - 2C\lambda + D \cdot \lambda^2) = 0 \quad (3.36)$$

onde:

$$\begin{aligned} A &= \sum X_t Y_t & P &= \sum Y_t^2 \\ B &= \sum X_t Y_{t-1} & E &= \sum Y_t Y_{t-1} \\ C &= \sum X_t Y_{t-2} & F &= \sum Y_t Y_{t-2} \\ D &= \sum X_t Y_{t-3} & G &= \sum Y_t Y_{t-3} \\ H &= \sum Y_{t-1}^2 & L &= \sum Y_{t-2} Y_{t-3} \\ I &= \sum Y_{t-1} Y_{t-2} & M &= \sum Y_{t-3}^2 \\ J &= \sum Y_{t-1} Y_{t-3} & N &= \sum X_t^2 \\ & & Q &= \sum X_t Y_{t-4} \\ K &= \sum Y_{t-2}^2 & R &= \sum Y_t Y_{t-4} \end{aligned}$$

Substituindo (3.35) em (3.36) e eliminando  $\alpha$  obtêm-se uma equação polinomial do 5º grau em  $\lambda$ , que deve ser resolvida (por métodos numéricos) em termos da raiz entre 0 e 1.

Observemos que os cálculos envolvidos são extensos e trabalhosos (para  $r = 3$ ). À medida que tomamos valores maiores para  $r$ , o cálculo se torna ainda mais complicado, necessitando para isso certamente o uso de computadores eletrônicos.

### 3.3.2 - Método dos Mínimos Quadrados Ponderados

Tal método é aplicável aos parâmetros de (3.34) sob

a suposição dos erros serem correlacionados.

Nesse caso, a aplicação dos estimadores de mínimos quadrados levaria a estimadores inconsistentes.

O método que veremos consiste em escrever (3.34) sob a forma

$$Y_t - u_t = \sum_{i=1}^r \begin{pmatrix} r \\ i \end{pmatrix} (-\lambda)^i (Y_{t-i} - u_{t-i}) + \alpha \cdot X_t \quad (3.38)$$

onde

$$\alpha = \beta(1 - \lambda)^r$$

Admitindo-se que os  $u_t$ 's são não correlacionados, a expressão (3.38) pode ser encarada como um modelo com erro nas variáveis (com excessão de X); admitindo-se ainda os  $u_t$ 's homoscedásticos, podemos aplicar a (3.38) o método dos mínimos quadrados ponderados que consiste em minimizar a soma dos quadrados ponderada por pesos proporcionais aos recíprocos de variâncias de

$$Y_t - \sum_{i=1}^r \begin{pmatrix} r \\ i \end{pmatrix} (-\lambda)^i Y_{t-i}$$

Isto é, devemos minimizar

$$\frac{\sum_{t=1}^n (Y_t - \sum_{i=1}^r \begin{pmatrix} r \\ i \end{pmatrix} (-\lambda)^i Y_{t-i} - \alpha X_t)^2}{1 + \sum_{i=1}^r \begin{pmatrix} r \\ i \end{pmatrix}^2 \lambda^{2i}} \quad (3.39)$$

com respeito a  $\alpha$  e  $\lambda$ .

Nos casos  $r = 2$  e  $r = 3$  a expressão (3.30) fica sendo

$$\frac{\sum_{t=1}^n (Y_t - 2\lambda Y_{t-1} + \lambda^2 Y_{t-2} - \alpha X_t)^2}{1 + 4\lambda^2 + \lambda^4}$$

e

$$\frac{\sum_{t=1}^n (Y_t - 3\lambda Y_{t-1} + 3\lambda^2 Y_{t-2} - \lambda^3 Y_{t-3} - \alpha X_t)^2}{1 + 9\lambda^2 + 9\lambda^4 + \lambda^6}$$

A manipulação algébrica deste método é extremamente complicada do ponto de vista numérico. Por exemplo, para  $r = 2$  as equações normais resultam (após a eliminação de  $\alpha$ ) em uma equação polinomial de 7º grau em  $\lambda$ , e que deve ser resolvida por métodos numéricos.

### 3.3.3 - Método da Máxima Verossimilhança

Recentemente Maddala e Rao [25] desenvolveram o método que veremos para estimar os parâmetros do modelo dado por

$$Y_t = (1 - \lambda)^r \sum_{i=0}^{\infty} \binom{r+i-1}{i} \lambda^i \cdot X_{t-i} + \varepsilon_t \quad (3.36)$$

Inicialmente definimos

$$n_i = E(Y_i) \quad i=0,1,2,\dots,r-1$$

Assim, (3.36) pode ser escrita sob a forma

$$Y_t = \beta Z_{1t} + \sum_{i=0}^{r-1} n_i Z_{i+2,t} + \varepsilon_t \quad (3.40)$$

onde

$$Z_{1t} = (1-\lambda)^r \sum_{i=0}^{t-r} \binom{r+i-1}{i} \lambda^i \cdot X_{t-i}$$

$$Z_{2t} = (-1)^{r-1} \binom{t-1}{r-1} \lambda^t$$

$$Z_{3t} = -\frac{r-1}{1} \frac{t}{t-1} \frac{1}{\lambda} Z_{2t}$$

$$Z_{4t} = -\frac{r-2}{2} \frac{t-1}{t-2} \frac{1}{\lambda} Z_{3t}$$

$$Z_{5t} = -\frac{r-3}{3} \frac{t-2}{t-3} \frac{1}{\lambda} Z_{4t}$$

$$\vdots$$

$$Z_{r+1,t} = \binom{t}{r-1} \cdot \lambda^{t-r+1}$$

onde

se  $t=1$ ,  $Z_3^t=1$  e os restantes  $Z$ 's são nulos;

se  $t=2$ ,  $Z_4^t=1$  e os restantes  $Z$ 's são nulos;

se  $t=r-1$ ,  $Z_{r+1,t}=1$  e os restantes  $Z$ 's são nulos.

Admitindo em (3.40) que os  $\varepsilon_t$ 's são independentes, e normalmente distribuídos, com média 0 e variância  $\sigma^2$ , a função de verossimilhança é tal que



$$L \propto \frac{1}{\sigma^N} \text{Exp} \left[ \frac{-1}{2\sigma^2} \sum_{t=1}^n (Y_t - \beta Z_{1t} - n_0 Z_{2t} - \dots - n_{r-1} \cdot Z_{r+1,t})^2 \right] \quad (3.41)$$

Para cada valor atribuído a  $\lambda$ , calculamos as variáveis sintéticas  $Z_{1t}, Z_{2t}, Z_{3t}, \dots, Z_{r-1,t}$  e obtemos os valores de  $\beta, n_0, n_1, \dots, n_{r-1}$  que minimizam  $L$  de (3.41). Finalmente escolhemos o valor de  $\lambda$  que torna mínima a variação residual (e conseqüentemente os valores de  $\beta, n_0, n_1, \dots, n_{r-1}$ ).

Observemos que os valores atribuídos a  $\lambda$  seguem um processo com valor inicial 0 e um certo passo até atingir o valor 1. Assim, obtido um valor de  $\lambda$  que minimize  $L$ , novos valores podem ser atribuídos a  $\lambda$ , em torno daquele, adotando passos menores, e assim por diante.

Observemos ainda que os estimadores assim obtidos, estão condicionados à escolha a priori de  $r$ . Maddala e Rao sugerem que o procedimento descrito seja adotado para diferentes valores de  $r$ , e finalmente adota-se o valor de  $r$  que minimize o coeficiente de determinação múltipla ( $\bar{R}^2$ ) corrigido pelos graus de liberdade em (3.40).

### 3.4 - OBSERVAÇÕES FINAIS

- a) Os métodos de estimação descritos muito embora produzam estimadores consistentes não são em sua maioria métodos que produzem os estimadores assintoticamente mais eficientes.
- b) A maioria dos métodos de estimação requerem cálculos

muito extensos o que exige o uso de computadores eletrônicos para sua manipulação. Uma exceção no caso, é o método das variáveis instrumentais que é extremamente simples, exigindo em geral a resolução de um sistema linear de 3 equações a 3 incógnitas.

- c) O teste de Durbin-Watson para detectar auto-correlação (de primeira ordem) nos erros é inapropriado em modelos onde  $Y_{t-1}, Y_{t-2}, \text{etc.}$  comparecem como variáveis explicativas (ver Jonhston [16, pp.309-312]).

Todavia, recentemente Durbin [07] elaborou um teste (usando a estatística  $h$ ) para detectar auto-correlação positiva (de primeira ordem) nos erros, desde que a amostra seja suficientemente grande.

Basicamente o procedimento consiste em estimar os parâmetros do modelo usando estimadores de mínimos quadrados. Em seguida, calcula-se os resíduos  $e_1, e_2, \dots, e_n$ .

Definindo-se

$$r = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^{n-1} e_t^2}$$

Durbin mostrou que para grandes amostras, a estatística

$$h = r \sqrt{\frac{n}{1 - n\hat{V}(b_1)}}$$

sob a hipótese de que  $\rho = 0$ , tem distribuição normal padronizada.

Na expressão de  $h$  temos

$n$  é o tamanho da amostra

$\hat{V}(b_1)$  é o estimador da variância de  $b_1$  (o coeficiente de  $Y_{t-1}$ ).

Por exemplo, se numa amostra de  $n = 100$ , tivermos  $r = 0,05$ ,  $\hat{V}(b_1) = 0,005$  então

$$h = 0,05 \sqrt{\frac{100}{1 - 100(0,005)}} = 0,7071$$

Assim, ao nível de significância 0,05 não rejeitamos a hipótese  $\rho = 0$ .

Observemos que o teste só depende da variância do estimador do coeficiente de  $Y_{t-1}$ , não importando os valores dos outros estimadores.

- d) Nos métodos de estimação procuramos seguir a metodologia da econometria clássica. Porém outros métodos existem, e que exigem conhecimentos mais especializados, como é o caso da Análise Espectral (um estudo completo desta teoria, aplicada aos modelos com defasagem distribuídas pode ser encontrado na monografia de Neves [31]).

- e) De acordo com a bibliografia consultada, até o recente momento, os problemas de determinação do grau do polinômio e de  $m$  no modelo de Almon, parecem estar em aberto. As alternativas propostas são passíveis de críticas, conforme já foi exposto.
- f) No método de Zellner e Geisel para estimação dos parâmetros dos modelos de defasagens distribuídas, tentamos uma maneira analítica de obter estimadores, a fim de contornar o método iterativo, porém não conseguimos chegar a resultados satisfatórios em virtude da complexidade de manipulação algébrica das expressões que comparecem no modelo.
- g) A utilização de um modelo de defasagem distribuídas, deve estar, presumimos, baseada em considerações teóricas, da ciência a qual está sendo aplicada. É necessário, antes de mais nada, o cientista raciocinar em termos de ciência da qual o mesmo faz parte, antes de utilizar um ou outro modelo. Isto é salientado em uma frase de Macedo [23, pp, 10] "Ao aplicarmos Econometria, devemos evitar o risco de adequar o problema ao modelo".

## CAPÍTULO 4

### EXEMPLO DE APLICAÇÃO

Neste capítulo, vamos exemplificar a teoria vista, na estimação dos parâmetros de um modelo que tenta analisar empiricamente o efeito de certas variáveis na taxa de juros anual, para os tomadores de letras de câmbio na cidade do Rio de Janeiro.

Sejam:

$j_t$ : taxa de juros anual de letras de câmbio em vigor no mês  $t$ .

$M_t$ : meios de pagamento no mês  $t$ .

$P_t$ : índice geral de preços no mês  $t$  (disponibilidade interna).

$Y_t$ : consumo industrial de energia elétrica na região Rio e São Paulo.

$\dot{M}_t$ : taxa de variação dos meios de pagamento

$$\dot{M}_t = \frac{M_t - M_{t-12}}{M_{t-12}}$$

$\dot{P}_t$ : taxa de variação do índice geral de preços

$$\dot{P}_t = \frac{P_t - P_{t-12}}{P_{t-12}}$$

$\dot{Y}_t$ : taxa de variação do consumo industrial de energia e elétrica

$$\dot{Y}_t = \frac{Y_t - Y_{t-12}}{Y_{t-12}}$$

O modelo a ser analisado é

$$j_t = C + \sum_{i=0}^s \alpha_i \dot{P}_{t-i} + \sum_{i=0}^s \beta_i \dot{M}_{t-i} + \sum_{i=0}^s \gamma_i \dot{Y}_{t-i} + \varepsilon_t$$

onde  $\varepsilon_t$  é admitida como variável normal, com média 0 e variância  $\sigma^2$  e  $E(\varepsilon_t \cdot \varepsilon_k) = 0$  para  $t \neq k$ .

Admitimos ainda no modelo que os  $\alpha_i$ 's,  $\beta_i$ 's e  $\gamma_i$ 's seguem um modelo polinomial de Almon.

Para a estimação dos parâmetros utilizamos um programa do "Econometric Software Package (ESP)", tal procedimento foi levado a efeito para diferentes valores de  $s$  e os graus do polinômio foram 2 e 3.

Os dados encontram-se na tabela 2 e a síntese dos resultados nas tabelas 3 e 4. Para a escolha do ajuste usamos a significância do teste F (nível de 5%) e o erro padrão da regressão.

TABELA 2

$t$	$j_t$	$\dot{P}_t$	$\dot{M}_t$	$\dot{Y}_t$
3/76	27,27	33,51	0,42	0,17
4/76	29,84	39,42	0,43	0,13

continua

5/76	31,99	41,80	0,41	0,14
6/76	31,68	42,48	0,43	0,12
7/76	32,92	45,60	0,45	0,16
8/76	32,76	45,98	0,36	0,15
9/76	33,54	48,03	0,37	0,18
10/76	35,43	48,15	0,41	0,12
11/76	35,43	48,05	0,36	0,10
12/76	37,03	48,23	0,37	0,17
1/77	39,61	49,04	0,34	0,13
2/77	38,80	47,76	0,34	0,17
3/77	37,51	49,36	0,36	0,06
4/77	39,29	50,18	0,42	0,18
5/77	39,94	49,88	0,41	0,09
6/77	38,64	48,05	0,35	0,12
7/77	42,08	43,74	0,37	0,09
8/77	43,08	40,06	0,39	0,06
9/77	42,58	37,35	0,41	0,11
10/77	41,75	37,89	0,40	0,12
11/77	41,91	39,10	0,37	0,09
12/77	46,44	38,55	0,37	0,12
1/78	41,91	37,04	0,40	0,12
2/78	42,08	36,87	0,41	0,12
3/78	42,24	34,92	0,41	0,13
4/78	41,91	31,09	0,38	0,08
5/78	39,94	33,36	0,38	0,09
6/78	40,10	35,79	0,38	0,11

Fontes: Boletim do Banco Central (agosto de 1978)  
 Conjuntura Econômica (agosto de 1978).

TABELA 3

Polinômio do 2º grau

s	R <sup>2</sup>	F	ERRO PADRÃO DA REGRESSÃO
3	0,8056	F(9,14) = 6,44579	1,98308

4	0,8074	F(9,13) = 6,05567	1,88022
5	0,8267	F(9,12) = 6,36252	1,64624
6	0,8217	F(9,11) = 5,63247	1,52368
7	0,8414	F(9,10) = 5,89347	1,36602
8	0,7139	F(9,9) = 2,49494	1,68543
9	0,7494	F(9,8) = 2,65756	1,51909
10	0,7198	F(9,7) = 1,99756	1,68956
12	0,8434	F(9,5) = 3,59004	1,30499
13	0,6715	F(9,4) = 0,90831	1,89964
14	0,7400	F(9,3) = 0,94880	1,87063
15	0,8168	F(9,2) = 0,99054	1,62683

TABELA 4 - POLINÔMIO DO 3º GRAU

s	R <sup>2</sup>	F	ERRO PADRÃO DA REGRESSÃO
4	0,9134	F(12,10) = 8,79448	1,43717
5	0,8497	F(12,9) = 4,23841	1,77081
6	0,8260	F(12,8) = 3,16576	1,76474
7	0,8925	F(12,7) = 4,84340	1,34403
8	0,8128	F(12,6) = 2,17076	1,66972
9	0,8609	F(12,5) = 2,57862	1,43150
10	0,8723	F(12,4) = 2,27733	1,50864
12	0,8079	F(12,2) = 0,70094	2,18302
13	0,9214	F(12,1) = 0,97464	1,85866

O resultado que nos pareceu satisfatório do ponto de vista de ajuste foi aquele onde o grau do polinômio é 3 e  $s = 7$ . Os coeficientes estimados nesse caso comparecem nas tabelas 2, 3 e 4.

A análise dos coeficientes deste modelo parecem indicar que



- a) Um aumento de  $\dot{P}$  nos períodos  $t-1$ ,  $t-2$ , e  $t-3$  tende a pressionar um aumento na taxa de juros no período  $t$ .
- b) Uma diminuição de  $\dot{M}$  nos períodos  $t$ ,  $t-1$ ,  $t-2$ ,  $t-3$ ,  $t-4$ , e  $t-5$  tende a aumentar a taxa de juros no período  $t$ ; isto provavelmente ocorre devido à diminuição da liquidez do sistema, que tende a aumentar a demanda de crédito, resultado num aumento da taxa de juros.
- c) Uma diminuição de  $\dot{Y}$  nos períodos  $t$ ,  $t-1$ ,  $t-2$ ,  $t-3$ ,  $t-4$ ,  $t-5$ ,  $t-6$  e  $t-7$  tende a aumentar a taxa de juros no período  $t$ ; isto provavelmente decorre do grande crescimento do produto nacional no período 73-74, e que, após 74, com a desaceleração econômica e queda no crescimento da demanda fez com que houvesse um aumento da demanda de crédito pelas empresas para fazer frente aos compromissos da acumulação passada.

TABELA 5 -  $\alpha_i$ 'S

LINE 9 ECONOMETRIC SOFTWARE PACKAGE \* FSP (VERSION 09/01/74) FSP \* 01/12/78

DISTRIBUTED LAG INTERPRETATION

COEFFICIENT ST'D. ERROR T-STATISTIC PLOT OF THE LAG DISTRIBUTION(\*) AND ST'D. ERROR BAND(+)

0	.1947E-01	0.4162	.4678E-01	+	*	+	+
1	0.3676	0.1940	1.895		*	+	+
2	0.4318	0.1706	2.531		*	*	+
3	0.3011	0.1173	2.586		*	*	+
4	.6461E-01	0.1802	0.3586		*	+	
5	-0.1885	0.2242	-0.8408		*	*	
6	-0.3691	0.1889	-1.954		*	+	
7	-0.3882	0.6082	-0.6382		*	+	

MEAN LAG = -14.5781

STANDARD ERROR = 17.3065

SUM OF LAG COEFFICIENTS = 0.238815

STANDARD ERROR = 0.242259

TABELA 6 -  $\beta_i$ 's

LINE 9 ECONOMETRIC SOFTWARE PACKAGE \* FSP (VERSION 09/01/74) ESP \* 01/12/78

DISTRIBUTED LAG INTERPRETATION

COEFFICIENT ST'D. ERROR T-STATISTIC PLOT OF THE LAG DISTRIBUTION(\*) AND ST'D. ERROR BAND(+)

0	-44.91	22.19	-2.023	+	*	+	.
1	-44.91	20.41	-2.200	+	*	+	.
2	-44.29	21.40	-2.070	+	*	+	.
3	-40.77	18.67	-2.184	+	*	+	.
4	-32.08	12.91	-2.485	+	*	+	.
5	-15.96	8.478	-1.883	+	*	+	.
6	9.863	13.78	0.7157	+	*	+	.
7	47.66	25.37	1.878	+	*	+	.

MEAN LAG = 0.430202

STANDARD ERROR = 1.27910

SUM OF LAG COEFFICIENTS = -165.407

STANDARD ERROR = 75.7386



## APÊNDICE 1

### PROPRIEDADES ASSINTÔTICAS DOS ESTIMADORES

Consideremos uma amostra de tamanho  $n$  e, associado a ela o estimador  $a_n$  do parâmetro  $\alpha$ . Observamos que, para um  $n$  fixado,  $a_n$  é uma variável aleatória; para cada  $n$  então vamos indicar a média e a variância de  $a_n$  por  $E(a_n)$  e  $\text{Var}(a_n)$ .

DEFINIÇÃO 1 - Chama-se esperança assintótica de  $a_n$  o valor  $\lim_{n \rightarrow \infty} E(a_n)$ . Diz-se que  $a_n$  é assintoticamente não viesado (ou não viciado) se  $\lim_{n \rightarrow \infty} E(a_n) = \alpha$ .

DEFINIÇÃO 2 - Chama-se variância assintótica de  $a_n$  o valor

$$n^{-1} \lim_{n \rightarrow \infty} E\{\sqrt{n}[a_n - E(a_n)]\}^2.$$

DEFINIÇÃO 3 - Diz-se que uma seqüência de variáveis aleatórias  $\{a_n\} = a_1, a_2, \dots, a_n, \dots$  converge em probabilidade para a constante  $\alpha$  se

$$\forall \epsilon > 0$$

$$\lim_{n \rightarrow \infty} P(|a_n - \alpha| > \epsilon) = 0$$

A notação usada é

$$a_n \xrightarrow{P} \alpha$$

ou

$$p \lim a_n = \alpha$$

DEFINIÇÃO 4 - Diz-se que  $a_n$  é um estimador consistente do parâmetro  $\alpha$ , se  $p \lim a_n = \alpha$ .

DEFINIÇÃO 5 - Diz-se que  $a_n$  é estimador assintoticamente eficiente para o parâmetro  $\alpha$  se as condições abaixo forem satisfeitas:

- a)  $a_n$  tem distribuição assintótica com média e variâncias finitas;
- b)  $a_n$  é consistente para  $\alpha$ ;
- c) nenhum outro estimador consistente de  $\alpha$  tem variância assintótica menor que  $a_n$ .

DEFINIÇÃO 6 - Dada uma seqüência de variáveis aleatórias  $\{a_n\} = a_1, a_2, \dots, a_n, \dots$  diz-se que a mesma converge em média quadrática para a constante  $\alpha$  se

$$\lim_{n \rightarrow \infty} E(a_n - \alpha)^2 = 0$$

A seguir, vamos enunciar alguns teoremas relacionados com tais definições. As demonstrações (em sua maioria) encontram-se em [14].

TEOREMA 1 - Uma condição suficiente para que o estimador  $a_n$

seja consistente para o parâmetro  $\alpha$ , é que

$$\lim_{n \rightarrow \infty} E(a_n - \alpha)^2 = 0$$

TEOREMA 2 - Se

$$\lim_{n \rightarrow \infty} \text{Var}(a_n) = 0 \quad \text{e} \quad E(a_n) = \alpha \quad (\text{ou} \quad \lim_{n \rightarrow \infty} E(a_n) = \alpha)$$

então  $a_n$  é consistente para o parâmetro  $\alpha$ .

TEOREMA 3 - Se  $p \lim a_n = \alpha$  e  $F(a_n)$  é uma função contínua de  $a_n$  então  $p \lim F(a_n) = F(\alpha)$ .

Em particular

$$(i) \quad p \lim (a_n^2) = (p \lim a_n)^2$$

$$(ii) \quad p \lim (a_n^{-1}) = (p \lim a_n)^{-1}$$

TEOREMA 4 - Se  $p \lim a_n = \alpha$  e  $p \lim b_n = \beta$  e  $F(a_n, b_n)$  é uma função contínua então  $\lim F(a_n, b_n) = F(\alpha, \beta)$ .

Em particular

$$(i) \quad p \lim (a_n + b_n) = p \lim a_n + p \lim b_n$$

$$(ii) \quad p \lim (a_n b_n) = (p \lim a_n) \cdot (p \lim b_n)$$

$$(iii) \quad p \lim \left( \frac{a_n}{b_n} \right) = (p \lim a_n / p \lim b_n) \quad (\text{se } p \lim b_n \neq 0)$$

DEFINIÇÃO 6 - Seja  $A$  uma matriz do tipo  $k \times p$ . Define-se limite em probabilidade da matriz  $A$  como sendo a matriz cujos elementos são ordenadamente os limites em probabilidades dos e

mentos de A. Isto é se  $A = (a_{ij})_{k \times p}$  então

$$p \lim A = (p \lim a_{ij})_{k \times p}$$

Desta forma são fáceis de concluir, por exemplo, as relações

$$(i) \quad p \lim (A \pm B) = p \lim A \pm p \lim B$$

$$(ii) \quad p \lim (A \cdot B) = (p \lim A) \cdot (p \lim B)$$

$$(iii) \quad p \lim (A^{-1}) = (p \lim A)^{-1}$$

TEOREMA 5 - Consideremos o modelo linear

$$Y_j = \alpha + \sum_{i=1}^K \beta_i X_{ij} + u_j \quad (j=1, 2, \dots, n)$$

que em forma matricial pode ser escrito

$$Y = X \cdot \beta + u$$

onde

$$(i) \quad E(u) = 0$$

$$(ii) \quad E(uu') = I \cdot \sigma^2$$

(iii) X é uma matriz cujos elementos são constantes e a característica de X é  $K+1 < n$

$$(iv) \quad \lim_{n \rightarrow \infty} \left( \frac{1}{n} X' X \right) = Q$$

onde Q é uma matriz não singular.



Nestas condições o estimador de mínimos quadrados de  $\beta$  é consistente.

Demonstração - Sabemos que o estimador  $b$  de mínimos quadrados para  $\beta$  é dado por

$$b = \beta + (X' \cdot X)^{-1} X' \cdot u$$

ou então

$$b = \beta + \left( \frac{1}{n} X' \cdot X \right)^{-1} \cdot \frac{1}{n} X' \cdot u$$

Tendo em conta a hipótese (iv)

$$p \lim b = \beta + Q^{-1} p \lim \left( \frac{1}{n} X' \cdot u \right)$$

Tendo em conta (i) e (ii)

$$E \left( \frac{1}{n} X' \cdot u \right) = 0$$

e a matriz de variância e covariância do vetor  $\frac{1}{n} E'u$  é

$$V \left( \frac{1}{n} X' \cdot u \right) = \frac{1}{n^2} E(X'u \cdot u'X) = \frac{\sigma^2}{n^2} X' \cdot X$$

Portanto

$$\lim_{n \rightarrow \infty} V \left( \frac{1}{n} X' \cdot u \right) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} \lim_{n \rightarrow \infty} \left( \frac{1}{n} X' \cdot X \right) = 0$$

Logo  $\frac{1}{n} X'u$  converge em média quadrática para a matriz nula e, de acordo com o Teorema 2

$$p \lim \left( \frac{1}{n} X' \cdot u \right) = 0$$

e como

$$p \lim b = \beta + Q^{-1} p \lim \left( \frac{1}{n} X'u \right)$$

Segue-se que

$$p \lim b = \beta$$

Isto é,  $b$  é estimador consistente de  $\beta$ .

TEOREMA 6 - Consideremos o modelo linear

$$Y_j = \alpha + \sum_{i=1}^K \beta_i X_{ij} + u_j \quad (j=1, 2, \dots, n)$$

ou ainda, na forma matricial

$$Y = X \cdot \beta + u$$

onde

- (i)  $E(u) = 0$
- (ii)  $E(uu') = I \cdot \sigma^2$
- (iii)  $X$  é uma matriz cujos elementos são variáveis aleatórias
- (iv)  $p \lim \left( \frac{1}{n} X' \cdot X \right) = Q$
- (v)  $p \lim \left( \frac{1}{n} X'u \right) \neq 0$ .

A pressuposição (v) pode ser expressa como: ao menos uma variável aleatória é assintoticamente correlaciona-

com o erro.

Nestas condições, o estimador  $b$  de mínimos quadrados para  $\beta$  é inconsistente.

Demonstração - Retomando a expressão

$$p \lim b = \beta + Q^{-1} p \left( \lim \frac{1}{n} X'u \right)$$

do Teorema 5, segue-se que

$$p \lim b \neq \beta$$

o que prova que  $b$  é inconsistente para  $\beta$ .

## APÊNDICE 2

### O USO DE VARIÁVEIS INSTRUMENTAIS

Inicialmente consideremos o modelo linear

$$Y_j = \beta \cdot X_j + u_j$$

onde

$$(i) \quad E(u_j) = 0$$

$$(ii) \quad E(u_j^2) = \sigma^2$$

$$(iii) \quad E(u_i u_j) = 0 \quad (i \neq j)$$

$$(iv) \quad p \lim \left( \frac{1}{n} \sum X_j^2 \right) = Q \quad (Q > 0)$$

$$(v) \quad p \lim \left( \frac{1}{n} \sum X_j u_j \right) \neq 0.$$

O estimador de mínimos quadrados de  $\beta$  é dado por

$$b = \frac{\sum X_j Y_j}{\sum X_j^2}$$

De acordo com o Teorema 6 do Apêndice 1

$$p \lim b - \beta = Q^{-1} p \lim \left( \frac{1}{n} \sum X_j u_j \right)$$

Se  $X_j$  e  $u_j$  tiverem correlação assintótica positiva

$$(p \lim \frac{1}{n} \sum X_j u_j > 0)$$

b tende a superestimar  $\beta$  pois  $p \lim b - \beta > 0$ , caso contrário

b tende a subestimar  $\beta$ , pois  $p \lim b - \beta < 0$ .

Admitamos que existam valores  $Z_j$ , de uma variável  $Z$  tal que

$$p \lim \left( \frac{1}{n} \sum Z_j X_j \right) \neq 0$$

e

$$p \lim \left( \frac{1}{n} \sum Z_j u_j \right) = 0$$

Tal variável se existir, é denominada variável instrumental.

Vamos provar que o estimador

$$\hat{\beta} = \frac{\sum Z_j Y_j}{\sum Z_j X_j}$$

chamado estimador de variável instrumental do parâmetro  $\beta$ , é consistente.

Temos

$$Y_j = \beta X_j + u_j$$

$$Z_j Y_j = \beta X_j Z_j + u_j Z_j$$

$$\sum Z_j Y_j = \beta \sum X_j Z_j + \sum Z_j u_j$$

como

$$\hat{\beta} = \frac{\sum Z_j Y_j}{\sum Z_j X_j} \implies \sum Z_j Y_j = \hat{\beta} \cdot \sum Z_j X_j$$

portanto,

$$\hat{\beta} \sum Z_j X_j = \beta \sum X_j Z_j + \sum Z_j \cdot u_j$$

$$\hat{\beta} = \beta + \frac{\sum Z_j u_j}{\sum Z_j X_j}$$

ou ainda

$$\hat{\beta} = \beta - \frac{\frac{1}{n} \sum Z_j u_j}{\frac{1}{n} \sum Z_j X_j}$$

portanto

$$p \lim \hat{\beta} = \beta + \frac{p \lim \left( \frac{1}{n} \sum Z_j u_j \right)}{p \lim \left( \frac{1}{n} \sum Z_j X_j \right)}$$

por definição de variável instrumental, segue-se que

$$p \lim \hat{\beta} = \beta$$

ou seja,  $\hat{\beta}$  é um estimador consistente de  $\beta$ .

Tomemos agora o caso, do modelo linear geral, escrito sob a forma matricial

$$Y = X \cdot \beta + u$$

onde

- (i)  $E(u) = 0$
- (ii)  $E(uu') = I \cdot \sigma^2$
- (iii)  $p \lim \left( \frac{1}{n} X'u \right) \neq 0$ .

De acordo com o Teorema 6 do Apêndice 1, sabemos que o estimador de mínimos quadrados para  $\beta$  ( $b = (X'X)^{-1}X'Y$ ) é inconsistente.

Admitindo que o vetor  $\beta$  tenha  $K$  parâmetros, suponhamos que exista uma matriz  $Z$ ,  $n \times K$  com as seguintes propriedades:

- (i)  $p \lim \left( \frac{1}{n} Z'u \right) = 0$
- (ii)  $p \lim \left( \frac{1}{n} Z'X \right) = Q$

onde  $Q$  é uma matriz não singular.

A matriz  $Z$ , se existir é chamada matriz de variáveis instrumentais.

O estimador

$$\hat{\beta} = (Z'X)^{-1}Z'Y$$

é chamado estimador de variáveis instrumentais do vetor  $\beta$ .

Vamos provar  $\hat{\beta}$  é estimador consistente para  $\beta$ . De fato, como

$$Y = X \cdot \beta + u$$

então

$$(Z'X)^{-1} \cdot Z'Y = (Z'X)^{-1} Z' \cdot X\beta + (Z'X)^{-1} Z'u$$

como

$$\hat{\beta} = (Z'X)^{-1} Z'Y$$

$$\hat{\beta} = X^{-1} (Z')^{-1} Z'X\beta + (Z'X)^{-1} Z'u$$

$$\hat{\beta} = \beta + (Z'X)^{-1} Z'u$$

portanto

$$p \lim \hat{\beta} = \beta + \left[ p \lim \left( \frac{1}{n} Z'X \right) \right]^{-1} p \lim \left( \frac{1}{n} Z'u \right)$$

e de acordo com a definição de matriz de variáveis instrumentais

$$p \lim \hat{\beta} = \beta$$

e conseqüentemente  $\hat{\beta}$  é estimador consistente para  $\beta$ .

#### OBSERVAÇÕES

- 1) Se todas as variáveis independentes forem assintoticamente não correlacionadas com o erro, a própria matriz  $X$  pode ser usada como matriz de variáveis instrumentais.
- 2) Para cada variável independente  $X_i$  assintoticamente correlacionada com o erro temos que encontrar outra



variável  $Z_i$  instrumental para substituir  $X_i$ . Por outro lado, se  $X_i$  não é assintoticamente correlacionado com o erro, ela própria pode ser a variável instrumental  $Z_i$ .

## REFERÊNCIAS BIBLIOGRÁFICAS

- [01] - Almon, S., "The distributed lag between capital appropriation and expenditures", *Econometrica*, vol. 33, nº 1, 1965, pp. 178-196.
- [02] - Alt, F., "Distributed Lags", *Econometrica*, vol. 10, nº 2, 1942, pp. 113-128.
- [03] - Amemiya, T. and Fuller, W.A., "A comparative study of alternative estimators in a distributed lag model", *Econometrica*, vol. 35, nº 3-4, 1967, pp. 509-529.
- [04] - De Leeuw, F., "The demand for capital goods by manufacturers; a study of quarterly time series", *Econometrica*, vol. 30, nº 3, 1962, pp. 407-423.
- [05] - Dhrymes, P.J., *Distributed lags, Problems of estimation and formulation*, San Francisco, Holden-Day, 1971.
- [06] - Doran, H.E. and Griffiths, W.E., "Inconsistency of the O.L.S. estimator of the partial adjustment adaptive expectations model", *Journal of Econometrics*, vol. 7, 1978, pp. 133-146.
- [07] - Durbin, J., "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables", *Econometrica*, vol. 38, nº 3, 1970, pp. 410-421.
- [08] - Ferguson, G.E., *Microeconomia*, Forense, Universitária, Rio de Janeiro, 1976.
- [09] - Frost, P.A., "Some properties of the Almon lag technique when one searches for degree of polynomial

- and lag", *Journal of the American Statistical Association*, vol. 70, nº 351, 1975, pp. 606-612.
- [10] - Griliches, Z., "Distributed lags: a survey", *Econometrica*, vol. 35, nº 1, 1967, pp. 16-49.
- [11] - Gujarati, D., *Basic Econometrics*, McGraw-Hill Book Company, 1978.
- [12] - Hannan, E.J., "The estimation of relationships involving distributed lags", *Econometrica*, vol. 33, nº 1, 1965.
- [13] - Hatanaka, M., "An efficient two-step estimator for the dynamic adjustment model with autoregressive error", *Journal of Econometrics*, vol. 2, 1974, pp. 199-220.
- [14] - Hoffmann, R e Vieira, S., *Análise de Regressão*, Editora Hucitec, Editora da USP, São Paulo, 1977.
- [15] - Johnston, J., *Métodos Econométricos*, Atlas, São Paulo, 1ª Edição, 1971.
- [16] - Johnston, J., *Econometric Methods*, McGraw-Hill, New York, Second Edition, 1972.
- [17] - Kelejian, H.H. e Oates, W.E., *Introdução à Econometria*, Editora Campus, 1978.
- [18] - Klein, L.R., "The estimation of distributed lags", *Econometrica*, vol. 26, nº 4, 1958, pp. 553-565.
- [19] - Kmenta, J. *Elements of Econometrics*, MacMillan Company, New York, 1971.
- [20] - Koutsouyiannis, A., *Theory of Econometrics*, MacMillan Company, London, 1973.
- [21] - Koyck, L.M., *Distributed lags and investment analysis*, Amsterdam, North-Holland Publishing Co., 1954.

- [22] - Liviatan, N., "Consistent estimation of distributed lags", *International Economic Review*, vol. 4, n°1, 1963, pp. 44-52.
- [23] - Macedo, R.B.M., *Cuidados lógicos e metodológicos na análise econométrica*, Apostila publica pela FEA-USP, 1978.
- [24] - Maddala, G.S. & Rao, A.S., "Test for serial correlation in regression models with lagged dependent variables and serially correlated errors", *Econometrica*, vol. 41, n° 4, 1973, pp. 761-774.
- [25] - Maddala, G.S. & Rao, A.S., "Maximum likelihood estimation of Solow's and Jorgenson's distributed lag models", *The Review of Economics and Statistics*, vol. 53, pp. 80-88.
- [26] - Malinvaud, E., *Statistical methods of Econometrics*, Rand McNally, Chicago, 1966.
- [27] - Nagar, A.L. & Gupta, Y.P., "The bias of Liviatan's consistent estimator in a distributed lag model", *Econometrica*, vol. 36, n° 2, 1968, pp. 337-342.
- [28] - Nerlove, M., "Estimates of the elasticities of supply of selected agricultural commodities", *Journal of Farm Economics*, vol. 38, 1956, pp. 496-509.
- [29] - Nerlove, M., "The dynamics of Supply: estimation of farmer's response to price", Baltimore, the *John Hopkins Press*, 1958.
- [30] - Nerlove, M., "lags in Economic behavior", *Econometrica*, vol. 40, n° 2, 1972, pp. 221-251.
- [31] - Neves, M.M.C., *Aplicação da análise, espectral ao estudo do modelo de defasagens distribuídas*, Dissertação de Mestrado, I.M.E. - U.S.P., 1977.

- [32] - Pastore, A.C., *A resposta da produção agrícola aos preços no Brasil*, APEC, São Paulo, 1973.
- [33] - Rocca, C.A. e outros, *Variáveis defasadas*, Apostila publicada pela FEA-USP, 1972.
- [34] - Rocca, C.A.; Pinto, M.B. e Martone, C.L., *Modelos Econométricos*, Apostila publicada pela FEA-USP, 1975.
- [35] - Schmidt, P., "A modification of the Almon distributed lag", *Journal of the American Statistical Association*, vol. 69, nº 347, 1974, pp. 679-681.
- [36] - Solow, R.M., "On a family of lag distributions", *Econometrica*, vol. 28, nº 2, 1960, pp. 393-406.
- [37] - Stewart, J., *Understanding Econometrics*, Hutchinson University Library, 1976.
- [38] - Taylor, L.D. & Wilson, T.A., "Three-pass least squares: a method for estimating models with a lagged dependent variable", *The Review of Economics and Statistics*, vol. 46, nº 4, 1964, pp. 329-346-
- [39] - Theil, H., *Principles of Econometrics*, Wiley, New York, 1971.
- [40] - Wallis, R.F., "Lagged dependent variables and serially correlated errors: a reappraisal of three-pass least squares", *The Review of Economics and Statistics*, vol. 49, 1967, pp. 555-567.
- [41] - Wonnacott, R.J. e Wonnacott, T.H., *Econometria*, Livros Técnicos e Científicos Editora, R. Janeiro, 1976.
- [42] - Zellner, A. & Geisel, M.S., "Analysis of distributed lag models with applications to consumption function estimation", *Econometrica*, vol. 38, nº 6, 1970, pp. 865-888.