

INVESTIGAÇÃO OPERACIONAL

DEZEMBRO 1984

NÚMERO 2

VOLUME 4

Publicação Científica da



Associação Portuguesa para o Desenvolvimento
da Investigação Operacional.

INVESTIGAÇÃO OPERACIONAL

Propriedade:

APDIO — Associação Portuguesa para o Desenvolvimento
da Investigação Operacional

ESTATUTO EDITORIAL

«Investigação Operacional», órgão oficial da APDIO cobre uma larga gama de assuntos reflectindo assim a grande diversidade de profissões e interesses dos sócios da Associação, bem como as muitas áreas de aplicação da I. O. O seu objectivo primordial é promover a aplicação do método e técnicas da I.O. aos problemas da Sociedade Portuguesa.

A publicação acolhe contribuições nos campos da metodologia, técnicas, e áreas de aplicação e software de I. O. sendo no entanto dada prioridade a bons casos de estudo de carácter eminentemente prático.

Distribuição gratuita aos sócios da APDIO



ÍNDICE

| | |
|---|-----|
| R. Castro Souza Identificação de Modelos ARIMA(p,d,q) através da Função de Autocorreção Extendida de Tiao & Tsay | 111 |
| M. Madden & P.V.D. Correia Interpretação Empírica do Dual de um pro- grama inteiro misto | 131 |
| L. Valadares Tavares, J.A. Antunes Ferreira & R. Carvalho Oliveira A Taxonomic Delimitation of Functional Urban Regions in a Heterogeneous Country | 149 |
| RESUMOS EM PORTUGUÊS | 177 |
| ABSTRACTS IN ENGLISH | 179 |



Associação Portuguesa para o Desenvolvimento
da Investigação Operacional.

INVESTIGAÇÃO OPERACIONAL

PUBLICAÇÃO SEMESTRAL

VOLUME 4 - Nº 2 DEZEMBRO 1984

ASSINATURA ANUAL: 3000\$00

DIRECTOR: JOSÉ MANUEL VIEGAS

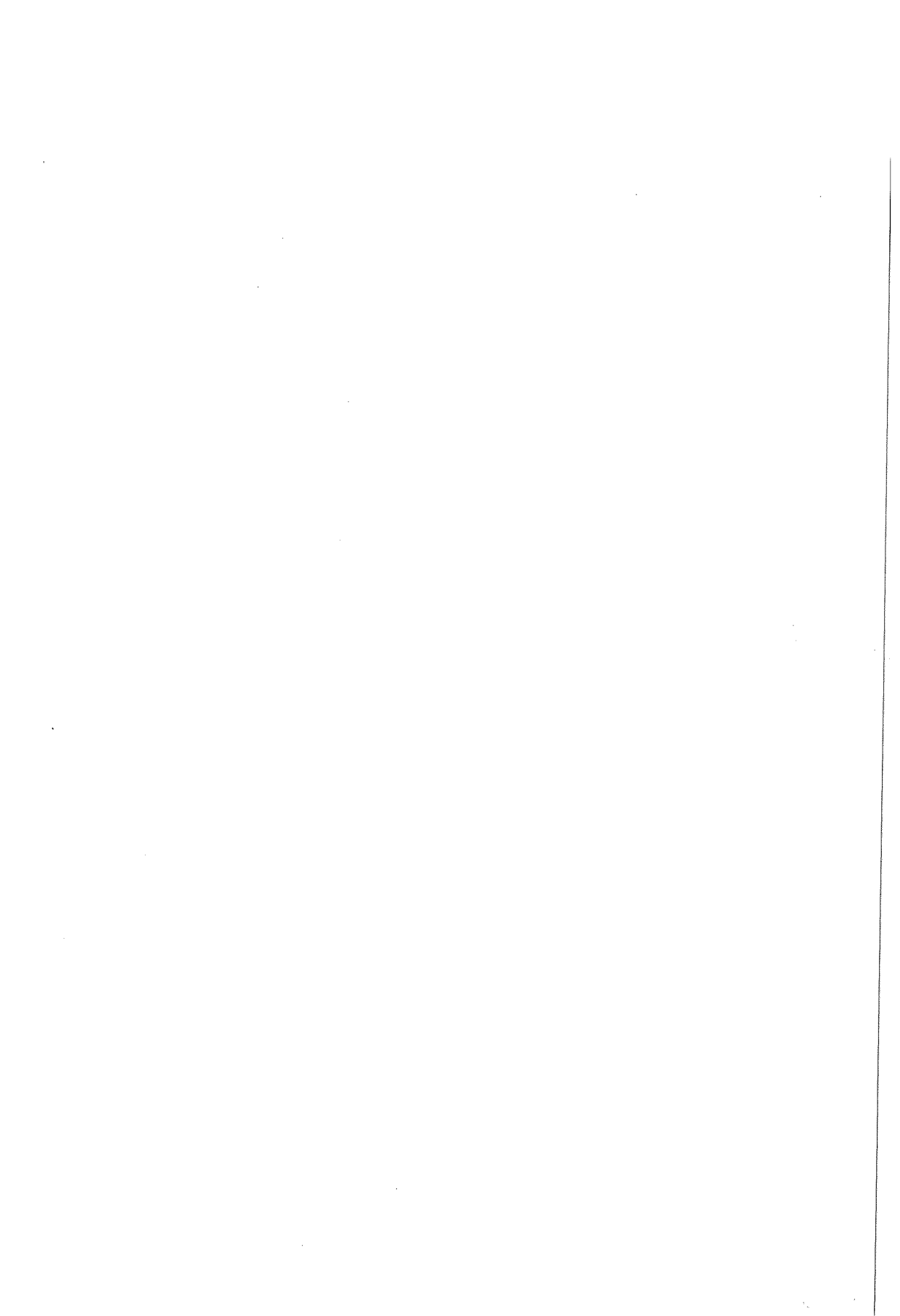
COMISSÃO EDITORIAL

Prof. João António Branco (IST)
Prof. J. Dias Coelho (UNL)
Prof. J. Borges Gouveia (FEUP)
Prof. J. João Júdice (U. Coimbra)
Prof. Mendes Mourão (JNICT)
Prof. Fernando Pacheco (U. Católica)
Prof. Gouveia Portela (IST)
Prof. Alberto Ralha (INIC)
Prof. L. Valadares Tavares (IST)
Prof.^a Isabel Hall Themido (IST)

Prof. Bernardo Calafate Vasconcelos (FEUP)
Prof. Mordecai Avriel (Israel)
Prof. Jaime Barceló (Espanha)
Prof. Josep Casanovas (Espanha)
Prof. D.V. Gokhale (USA)
Prof. M. Najim (Marrocos)
Prof. Alexander Rinnoy Kan (Holanda)
Prof. Masao Iri (Japan)
Prof. Bernard Roy (França)
Prof. Andrés Weintraub (Chile)

Este Número foi subsidiado pela Junta Nacional de Investigação Científica e Tecnológica e pelo Instituto Internacional de Engenharia de Sistemas e Desenvolvimento (IISED). A Revista "Investigação Operacional" está registada na Secretaria de Estado da Comunicação Social sob o nº 108 335.

Esta Revista é distribuída gratuitamente aos sócios da APDIO. As informações sobre inscrições na Associação, assim como a correspondência para a Revista, devem ser enviadas para a sede da APDIO - Associação Portuguesa para o Desenvolvimento da Investigação Operacional - CESUR, Instituto Superior Técnico, Av. Rovisco Pais, 1000 LISBOA.



IDENTIFICAÇÃO DE MODELOS ARIMA(p,d,q) ATRAVÉS DA
FUNÇÃO DE AUTOCORRECÇÃO EXTENDIDA DE TIAO & TSAY

REINALDO CASTRO SOUZA

GRUPO DE SISTEMAS, DEE
PUC/RJ, BRASIL

RESUMO

É apresentado o conceito de regressão AR(p) iterativa, a partir do qual podem ser obtidos estimadores consistentes da parte autoregressiva de modelos ARMA pelo método dos mínimos quadrados ordinários. A função de autocorrecção extendida a partir da regressão AR(p) iterativa é então obtida e seu estimador facilmente calculado.

Tal função permite a identificação directa dos parâmetros "p" e "q" dos modelos ARMA (p,q), apresentando-se como alternativa ao procedimento tradicional de reconhecimento de padrões das funções de autocorrecção e autocorrelação parcial proposto por Box & Jenkins.

1. INTRODUÇÃO

O modelo BJ [Box & Jenkins, 1970] conforme originalmente proposto pelos autores, consta basicamente de quatro estádios distintos para a sua aplicação metodológica, a saber: identificação, estimação, testes estatísticos e previsão. Destes, a identificação dos modelos é sem dúvida aquele que requer maior conhecimento, habilidade e experiência do utili-

zador, constituindo assim o estágio mais difícil da aplicação do método. Isto pode ser facilmente constatado se fizermos uma rápida revisão da literatura, onde verificaremos que várias foram as propostas alternativas para a identificação de modelos ARIMA, tais como: função de autocorrelação inversa [Cleveland,1972] & [Chatfield,1979]; Critério de Informação de Akaike AIC [Akaike,1974]; "R & S array" de Gray, Kelley & McIntire, (1978) e o "Corner Method" de Beguin, Gourieroux & Monfort, (1980). Entretanto, em todos eles não se conseguiu resolver de uma forma definitiva o problema de identificação de modelos ARIMA; especialmente quando estão presentes as duas componentes AR & MA no modelo para a série.

Consideraremos a seguinte notação para o modelo ARIMA (p,d,q) univariado à serie cronológica Z_t :

$$\phi(B) Z_t = \theta_0 + \theta(B) a_t$$

onde:

$$\phi(B) = U(B) \cdot \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$e: U(B) = 1 - U_1 B - \dots - U_d B^d$$

$$\theta(B) = 1 - \phi_1 B - \dots - \phi_{p-d} B^{p-d}$$

é a parte autoregressiva do modelo (AR)

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

é a parte médias móveis do modelo (MA)

- B é o operador de atraso, t.q., $B^k Z_t = Z_{t-k}$
- θ_0 é o termo constante
- a_t é o ruído branco Gaussiano, média nula e variância constante σ_a^2 .

Em relação aos polinômios AR e MA, supõe-se que todos os zeros de $U(B) \equiv 0$ estão no círculo unitário (equivalente aos operadores de diferença), os de $\phi(B) \equiv 0$ estão fora do círculo unitário e os de $\theta(B) \equiv 0$ sobre ou fora desse círculo. Supõe-se também que $\phi(B)$ e $\theta(B)$ não tem factores em comum.

O método de identificação proposto pelos autores Box & Jenkins, ou seja, aquele baseado no uso dos estimadores da FAC (Fundação de AutoCorrelação) e FACP (Fundação de AutoCorrelação Parcial) não apresentam problemas de maior no caso específico de modelos AR (corte brusco de PACF) e MA (corte brusco da FAC). Entretanto, as dificuldades aumentam sobremaneira no caso da presença simultânea de ambas componentes, i.e., para os modelos mixtos.

Os problemas acima citados constituem, na realidade, a motivação principal que levou Tiao & Tsay a investigarem uma forma alternativa de identificação. E noutras palavras, para modelos MA puros ou seja:

$$Z_t = \theta(B) \cdot a_t \text{ — FAC é função simples e útil na identificação}$$

porém, no caso de $\phi(B) \neq 1$ (modelo mixto), ou seja, para o modelo:

$$\phi(B) Z_t = \theta(B) \cdot a_t \text{ — tanto a FAC como a FACP são de difícil interpretação}$$

Entretanto, poder-se-ia contornar tal problema se fosse possível obter um estimador consistente da parte AR, ou seja:

$$w_t = \hat{\phi}(B) \cdot Z_t \simeq MA(q) \text{ — FAC de } w_t \text{ permite a identificação}$$

de "q" se $\hat{\phi}(B)$ for estimador consistente de $\phi(B)$.

Este trabalho considera a solução do problema acima exposto, em duas fases distintas:

1. Obtenção de estimadores consistentes da parte AR dos modelos BJ.
2. Método alternativo de identificação de modelos ARIMA.

A primeira parte será tratada através de um estudo detalhado das propriedades dos estimadores de mínimos quadrados ordinários (MQO) dos parâmetros AR que conduzirá ao conceito de regressão AR(p) iterativa o qual, conforme veremos, produzirá estimadores assintoticamente consistentes de $\Phi(B)$.

A partir destes estimadores é definida a Função de Autocorrelação Extendida (FACE) que permitirá a identificação rápida e fácil de modelos mixtos.

Correspondendo a estas ideias está sendo desenvolvido um programa de computador a ser implementado ao Centro de Processamento de Dados da PUC/RJ, que é basicamente composto de 2 estádios:

- obtenção de estimadores consistentes da parte AR do modelo através de uma rotina simples de MQO.
- cálculo da FACE

O programa fornece como saída além dos estimadores da FACE, uma análise destes acompanhada de sugestão de um possível modelo ARMA para a série em estudo.

2. REGRESSÃO AR(p) ITERATIVA

Seja o modelo ARMA(p,q) geral aplicado a Z_t :

$$Z_t = \sum_{\ell=1}^p \phi_{\ell} Z_{t-\ell} - \sum_{j=1}^q \theta_j a_{t-j} + a_t \quad (1)$$

O nosso objectivo consiste em obter estimadores consistentes dos parâmetros AR do modelo (i.e ϕ_{ℓ}). Estudemos inicialmente as propriedades dos estimadores MQO destes parâmetros para os possíveis casos particulares:

- Modelos AR puros (i.e $q = 0$)

. Estacionários: Estimadores MQO são consistentes,

$$U(B) \equiv 1$$

i.e:

$$\hat{\phi}_{\ell} = \frac{1}{P} \sum_{t=\ell}^P \phi_{\ell} ; \ell = 1, 2, \dots, p$$

converge em probabilidade

[Mann & Wald, 1943]

. Não estacionários: Estimadores MQO são consistentes

$$U(B) \neq 1$$

[Tiao & Tsay, 1981 and Lai & Wei, 1982]

- Modelos Mixtos ($p, q \neq 0$)

Estacionário: Estimadores MQO são não consistentes,

$$U(B) \equiv 1$$

como pode ser constatado pelo exemplo seguinte:

Exemplo:

$$(1 - \phi B)Z_t = (1 - \theta B)a_t \text{ — modelo verdadeiro}$$

$$Z_t = \beta Z_{t-1} + e_t \text{ modelo fixado para } Z_t;$$

onde:

$$MQO(\beta) = \hat{\beta} = \sum Z_t Z_{t-1} / \sum Z_{t-1}^2 = r_1$$

onde r_1 : autocorrelação de lag 1 de Z_t

Porém:

$\hat{\beta} = r_1 \frac{p}{1-p} \rho_1$, sendo ρ_1 para um ARMA (1,1) expresso por:

$$\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{(1+\theta^2 - 2\phi\theta)} \neq \phi,$$

a menos que $\phi = 1$ ou $\theta = 0$.

- . Não Estacionário: Pelo resultado acima, vemos que um AR(d) será consistente para U(B); porém um AR(k) (para $k > d$) não verificará esta propriedade.

Estes resultados são resumidos no quadro 1 abaixo:

| Modelos verd. | Modelos Fixado | Estimador MQO |
|------------------------|----------------|-----------------------|
| ARMA (p,0) estac. | AR(p) | Consistente |
| ARMA (p,0) n/estac. | AR(p) | Consistente |
| ARMA (p,q) estac. | AR(p) | Não Consistente |
| ARMA (p,q) n/estac. | AR(d) | Consistente p/U(B) |
| | AR(k); k > d | Não Consistente |

Quadro 1: Resumo dos estimadores MQO para a parte AR dos diversos modelos ARMA.

Vemos pelo quadro acima que ao fixar-se um modelo AR para a série Z_t e estimarmos os parâmetros correspondentes por MQO, tais estimadores serão sempre consistentes, excepto nos casos em que no modelo verdadeiro estiverem presentes componentes autoregressivas com zeros fora do círculo; i.e., componentes AR estacionárias.

Apresentamos seguidamente um procedimento geral a partir do qual será sempre possível obter estimadores consistentes, qualquer que seja o modelo.

Seja um modelo AR(p) adaptado ao mesmo conjunto de dados para os quais o modelo representado pela equação 1 é gerador:

$$Z_t = \sum_{\ell=1}^p \phi_{\ell}^{(0)} \cdot Z_{t-\ell} + e_{p,t}^{(0)}; \quad t = p + 1, \dots, n \quad (2)$$

Representando por $\hat{\phi}_{\ell(p)}^{(0)}$ o estimador MQO de $\phi_{\ell(p)}^{(0)}$, verificando-se que:

$$\hat{\phi}_{\ell(p)}^{(0)} \xrightarrow{P} \phi_{\ell}^{(0)}; \quad \ell = 1, \dots, p \text{ (i.e. } \hat{\phi}_{\ell(p)}^{(0)} \text{ é consistente)}$$

se e só se:

- Z_t segue um AR puro ($q = 0$)
- Z_t segue a ARMA (p,q) não estacionário puro ($U(B) \neq 1; \phi(B) \equiv 1$)

Sendo $\hat{e}_{p,t}^{(0)}$ o resíduo obtido do modelo acima, que no caso de não consistência do estimador, conterà informações sobre o processo Z_t , ou seja este não será um ruído branco.

Define-se como "1ª regressão AR(p) iterativa" aquela regressão linear dada pela seguinte equação:

$$Z_t = \sum_{\ell=1}^p \hat{\phi}_{\ell}^{(1)} Z_{t-\ell} + \beta_{1(p)}^{(1)} \cdot \hat{e}_{t,t-1}^{(0)} + e_{p,t}^{(1)} \quad (3)$$

Sendo $\hat{\phi}_{\ell}^{(1)}$ o estimador MQO de $\phi_{\ell}^{(1)}$, podendo-se verificar que:

$$\hat{\phi}_{\ell}^{(1)} \xrightarrow{P} \phi_{\ell}^{(1)}; \ell = 1, \dots, p \text{ (consistente)}$$

se e somente se $q \leq 1$.

De maneira análoga, representando por $\hat{e}_{p,k}^{(1)}$ o resíduo obtido da regressão (3), define-se a "2ª regressão AR(p) iterativa" pela equação:

$$Z_t = \sum_{\ell=1}^p \hat{\phi}_{\ell}^{(2)} Z_{t-\ell} + \beta_{1(p)}^{(2)} \hat{e}_{p,t-1}^{(1)} + \beta_{2(p)}^{(2)} \hat{e}_{p,t-2}^{(0)} + e_{p,t}^{(2)} \quad (4)$$

onde:

$$\hat{\phi}_{\ell}^{(2)} \xrightarrow{P} \phi_{\ell}^{(2)}; \ell = 1, \dots, p \text{ (consistente)}$$

se e somente se $q \leq 2$.

Prosseguindo desta forma, podemos definir a 3ª, 4ª, 5ª... regressão AE(p) iterativa que fornecerão estimadores MQO consistentes, dependendo do valor de $q \leq 3, \leq 4, \leq 5 \dots$

De uma forma geral, definimos a jésima regressão AR(k) iterativa pela equação:

$$Z_t = \sum_{\ell=1}^k \hat{\phi}_{\ell}^{(j)} \cdot Z_{t-\ell} + \sum_{i=1}^j \beta_{i(k)}^{(j)} \hat{e}_{k,t-1}^{(j-i)} + e_{k,t}^{(j)} \quad (5)$$

cujos estimadores obtidos por MQO, i.e., $\hat{\phi}_{\ell}^{(j)}$; $\ell = 1, 2, \dots, k$ & $\hat{\beta}_{i(k)}^{(j)}$; $i = 1, 2, \dots, j$; $j = 0, 1, 2, \dots$; são tais que:

$$\hat{\phi}_{\ell}^{(j)} \xrightarrow{P} \phi_{\ell}^{(j)}; \ell = 1, 2, \dots, k$$

se e somente se $q \leq j$.

Antes de concluirmos esta seção, é importante mencionar que o procedimento acima produz de facto estimadores consistentes da parte AR dentro das condições específicas. Entretanto, interpretando e. como estimadores de a., os estimadores de MQO dos coeficientes β 's que podem ser interpretados como estimadores dos parâmetros MA do modelo não são consistentes sob nenhuma hipótese, conforme mostra Tsay & Tiao (1982). Em relação à consistência dos estimadores da parte AR, os resultados apresentados podem ser resumidos nos teoremas a seguir enumerados:

Teorema 1:

$Z_t \sim \text{ARMA}(p,q)$; t.q. $U(B) \equiv 1$ (estacionário)

Então: $\hat{\phi}_{\ell}^{(j)}(k) = \phi_{\ell} + O(n^{-1/2})$; $\ell = 1, 2, \dots, k$ se:

(i) $k \geq p$ e $j = q$

(ii) $k = p$ e $j > q$

Teorema 2:

$Z_t \sim \text{ARMA}(p,q)$; t.q. $U(B) \neq 1$ (não estacionário)

Então: $\hat{\phi}_d^{(j)}(B) = U(B) + O(n^{-1})$

ou seja, converge mais rapidamente ou, reunindo os dois resultados num único teorema teremos:

Teorema 3:

$Z_t \sim \text{ARMA}(p,q)$; estacionário ou não

Então: $\hat{\phi}_{\ell}^{(j)}(k) = \phi_{\ell} + O(n^{-1/2})$; $\ell = 1, 2, \dots, k$

Se:

(i) $k \geq p$ e $j = q$

(ii) $k = p$ e $j > q$ (onde $\phi_{\ell} = 0$ para $\ell > p$)

além disto, $O(n^{-1/2}) \rightarrow O(n^{-1})$ se $\phi(B) = 1$; $k = p$ e $j \geq q$

Demonstramos assim que o procedimento de regressão AR iterativa produz estimadores consistentes para qualquer processo, estacionário ou não.

3. FUNÇÃO DE AUTOCORRELAÇÃO EXTENDIDA (FACE)

Mostraremos agora como os resultados da seção anterior poderão ser utilizados na definição de uma função dos dados que será particularmente útil na identificação de modelos ARMA.

Seja o ARMA (p,q) estacionário ou não, representado por:

$$\Phi(B) Z_t = \theta(B) \cdot a_t$$

e consideremos só diversos valores possíveis para os parâmetros p e q:

$$(i) \quad p = 0 \text{ — } Z_t \sim \text{MA}(q)$$

Neste caso a $\text{FAC}_j(Z_t)$ tem um corte brusco a partir do lag $j = q$

Define-se a FACE de ordem zero de Z_t , cujo estimador é representado por $r_j(0)$ por:

$$r_j(0) = \begin{cases} \neq 0 & j = 1, 2, \dots, q \\ = 0 & j < q \end{cases}$$

$$(ii) \quad p = 1 \text{ — } Z_t \sim \text{ARMA}(1, q)$$

Neste caso o estimador consistente de ϕ dependerá do valor verdadeiro de q, conforme se ilustra no quadro 2:

| | Estimador | Séries | Comportamento FAC | | | | | |
|--------------|----------------------|---|-------------------|----------|----------|----------|---|------|
| | | | da Série transf. | | | | | |
| q | consisten- | w | | | | | | |
| | | | lag | | | | | |
| te de ϕ | | Transformadas | | | | | | |
| | | | 0 1 2 3 4 5... | | | | | |
| 0 | $\hat{\phi}_1^{(0)}$ | $W_t^{(0)} = Z_t - \hat{\phi}_1^{(0)} Z_{t-1} \sim MA(0)$ | 1 | 0 | 0 | 0 | 0 | 0... |
| 1 | $\hat{\phi}_1^{(1)}$ | $W_t^{(1)} = Z_t - \hat{\phi}_1^{(1)} Z_{t-1} \sim MA(1)$ | 1 | ρ_1 | 0 | 0 | 0 | 0... |
| 2 | $\hat{\phi}_1^{(2)}$ | $W_t^{(2)} = Z_t - \hat{\phi}_1^{(2)} Z_{t-1} \sim MA(2)$ | 1 | ρ_1 | ρ_2 | 0 | 0 | 0... |
| 3 | $\hat{\phi}_1^{(3)}$ | $W_t^{(3)} = Z_t - \hat{\phi}_1^{(3)} Z_{t-1} \sim MA(3)$ | 1 | ρ_1 | ρ_2 | ρ_3 | 0 | 0... |

Quadro 2: Comportamento geral, modelo ARMA(1,q)

Consideramos agora o comportamento da FAC da série Transformada W_t na hipótese do modelo verdadeiro ser um ARMA (1,q) com $q = 1, 2, 3$ etc.. Estes resultados são apresentados nos quadros 3, 4.

| | $\hat{\phi}_1^{(1)}$ | FAC $w_t^{(1)}$ | | | | | |
|---|----------------------|-------------------|----------------|----------------|----------------|----------------|-------|
| | | lag | | | | | |
| | | | | | | | |
| | | 0 1 2 3 4 | | | | | |
| 0 | não consistente | 1 | $\rho_1^{(0)}$ | $\rho_2^{(0)}$ | $\rho_3^{(0)}$ | $\rho_4^{(0)}$ | |
| 1 | consistente | 1 | ρ_1 | 0 | 0 | 0 | |
| 2 | consistente | 1 | ρ_1 | 0 | 0 | 0 | |
| 3 | consistente | 1 | ρ_1 | 0 | 0 | 0 | |

Quadro 3: FAC $[w_t^{(i)}]$; modelo correcto ARMA (1,1)

| i | $\hat{\phi}_1^{(i)}$ | FAC [$w_t^{(i)}$] | | | | |
|---|----------------------|---------------------|----------------|----------------|----------------|----------------------|
| | | lag | | | | |
| | | 0 | 1 | 2 | 3 | 4 |
| 0 | não consistente | 1 | $\rho_1^{(0)}$ | $\rho_2^{(0)}$ | $\rho_3^{(0)}$ | $\rho_4^{(0)}$ |
| 1 | não consistente | 1 | $\rho_1^{(1)}$ | $\rho_2^{(1)}$ | $\rho_3^{(1)}$ | $\rho_4^{(1)}$ |
| 2 | consistente | 1 | ρ_1 | ρ_2 | 0 | 0 |
| 3 | consistente | 1 | ρ_1 | ρ_2 | 0 | 0 |

Quadro 4: FAC [$w_t^{(i)}$]; modelo correcto ARMA (1.2)

Seguindo o raciocínio dos quadros 3 e 4 acima, podemos facilmente determinar o comportamento da FAC das séries transformadas W_t para $q = 3, 4, 5 \dots$. Verificamos em todos eles a propriedade de corte brusco na FAC da série transformada a partir do momento em que o estimador do parâmetro AR se torna consistente; comportamento este que é caracterizado pela diagonal de cada quadro.

Define a FACE de ordem 1 de Z_t como sendo a FAC_j (i.e. de lag j) de $w_t^{(j)} = Z_t - \hat{\phi}_1^{(j)} Z_{t-1}$ cujo estimador amostral, $r_j(1)$, é tal que:

$$r_j(1) = 0 \text{ para } j > q \text{ e } p = 1$$

iii) $p = 2 - Z_t \sim ARMA(2, q)$

O equivalente ao quadro 2 seria:

| q | Estimadores consistentes de $\hat{\phi}_1$ & $\hat{\phi}_2$ | Séries $w_t^{(i)}$ transformadas | Comportamento da FAC da série transformada | | | | | | |
|---|---|---|--|----------|----------|----------|---|---|-----|
| | | $w_t^{(i)} = Z_t - \hat{\phi}_1^{(i)} Z_{t-1} - \hat{\phi}_2^{(i)} Z_{t-2}$ | lag | | | | | | |
| | | | 0 | 1 | 2 | 3 | 4 | 5 | ... |
| 0 | $\hat{\phi}_1^{(0)} ; \hat{\phi}_2^{(0)}$ | $w_t^{(0)} \sim MA(0)$ | 1 | 0 | 0 | 0 | 0 | 0 | ... |
| 1 | $\hat{\phi}_1^{(1)} ; \hat{\phi}_2^{(1)}$ | $w_t^{(1)} \sim MA(1)$ | 1 | ρ_1 | 0 | 0 | 0 | 0 | ... |
| 2 | $\hat{\phi}_1^{(2)} ; \hat{\phi}_2^{(2)}$ | $w_t^{(2)} \sim MA(2)$ | 1 | ρ_1 | ρ_2 | 0 | 0 | 0 | ... |
| 3 | $\hat{\phi}_1^{(3)} ; \hat{\phi}_2^{(3)}$ | $w_t^{(3)} \sim MA(3)$ | 1 | ρ_1 | ρ_2 | ρ_3 | 0 | 0 | ... |

Quadro 5: Comportamento geral, modelo ARMA (2,q)

A partir dos elementos do quadro 5 e das características do modelo, poderíamos construir, para cada valor de $q=1,2,\dots$ quadros similares aos quadros 3 e 4 que, conforme já verificamos, fornecem uma imagem do comportamento da FAC da série Transformada w_t ; $i = 0,1,2,\dots$

Define-se a FACE de ordem 2 de Z_t como a FAC_j (i.e. lag j) de $w_t^{(j)} = Z_t - \hat{\phi}_1^{(j)} Z_{t-1} -$

$\hat{\phi}_2^{(j)} Z_{t-2}$; cujo estimador amostral, representado por $r_{j(2)}$, é tal que:

$$r_{j(2)} = 0 \text{ para } j > q \text{ e } p = 2$$

iv) $p = 3, 4, 5 \dots$

De maneira análoga ao procedimento descrito em (ii) &

(iii) seguiríamos definindo a FACE de ordem $k = 3, 4, 5 \dots$ como sendo FAC_j de

$$w_t^{(j)} = Z_t - \sum_{i=1}^k \hat{\phi}_i^{(j)} Z_{t-i}$$

cujo estimador amostral $r_j(k)$ é tal que:

$$r_j(k) = 0 \text{ para } j > q \text{ e } p = k$$

Os valores $r_j(k)$ são normalmente apresentados como no quadro 6:

| AR | MA | q | | | | |
|-----|----|----------|----------|----------|----------|-------|
| | | 0 | 1 | 2 | 3 | |
| 0 | | $r_1(0)$ | $r_2(0)$ | $r_3(0)$ | $r_4(0)$ | |
| 1 | | $r_1(1)$ | $r_2(1)$ | $r_3(1)$ | $r_4(1)$ | |
| 2 | | $r_1(2)$ | $r_2(2)$ | $r_3(2)$ | $r_4(2)$ | |
| p 3 | | $r_1(3)$ | $r_2(3)$ | $r_3(3)$ | $r_4(3)$ | |
| . | | . | . | . | . | |
| . | | . | . | . | . | |
| . | | . | . | . | . | |

Quadro 6: Tabela das FACE estimadas

4. IDENTIFICAÇÃO DE MODELOS ARMA ATRAVÉS DA FACE

Demonstramos na seção anterior que a função FACE possui o mesmo comportamento de "corte brusco" idêntico à FAC dos modelos MA puros. Dessa forma, tal função pode constituir uma importante ferramenta na determinação de "q". Na prática, no entanto, desconhecemos ambos; assim sendo, é necessário que tenhamos "à priori" informações sobre o comportamento assintótico desta função (inclusive o efeito na FACE quando um número exagerado de termos AR são erroneamente fixados à

série ("overfitting problem").

É mostrado na seção 5 de Tsay & Tiao, (1982) uma série de resultados relativos a este problema de sobreajuste, o mais importante dos quais, é, inclusive usado na identificação de modelos ARMA conforme abaixo se apresenta:

Para um modelo ARMA (p,q) temos sempre:

$$r_j(k) = \begin{cases} c(k-p; j-q) & ; 0 \leq j-q \leq k-p \\ 0 & ; j-q > k-p \geq 0 \end{cases}$$

onde $c(k-p; j-q)$ é uma constante real não nula.

A identificação de p & q é feita utilizando-se o resultado da tabela FACE do quadro 6. Para ilustrar o emprego da mesma, consideremos a identificação de um modelo ARMA cujos parâmetros verdadeiros são $p = q = 1$. Então, pelo resultado acima teremos:

- (i) $r_j(0) = 0$ para todo $j = 1, 2, \dots$
- (ii) com $p=q=1$, então
 - $k=1 \implies r_j(1) = 0$; $j \geq 2$
 - $k=2 \implies r_j(2) = 0$; $j \geq 3$
 - .
 - .
 - .

Estes factos são mostrados no quadro 7, onde "x" denota um valor não nulo e "0" zero.

| AR \ MA | MA | | | | | | |
|---------|----|---|---|---|---|---|-------|
| | 0 | 1 | 2 | 3 | 4 | 5 | |
| 0 | x | x | x | x | x | x | |
| 1 | x | 0 | 0 | 0 | 0 | 0 | |
| 2 | x | x | 0 | 0 | 0 | 0 | |
| 3 | x | x | x | 0 | 0 | 0 | |

Quadro 7: Comportamento da Face de uma ARMA (1,1)

Assim, vemos claramente que os "0" formam uma configuração triangular com fronteiras nas linhas " $k = 1$ " & " $j-k = 1$ ". Os números correspondentes à linha e coluna do vértice superior esquerdo indicam as ordens AR & MA do modelo.

Generalizando, deveremos procurar numa tabela FACE o "triângulo de zeros", com fronteiras nas linhas $k = c_1 \geq 0$ e tentativamente identificar $p = c_1$ & $q = c_2$ (correspondente ao vértice superior esquerdo do referido triângulo).

Finalmente, para facilitar a análise da configuração da FACE, em vez de representar directamente os valores estimados de $r_{j(k)}$ nas posições correspondentes do quadro, utiliza-se a simbologia "x" para um valor significativo e "0" para um valor identicamente nulo. Para que o teste seja realizável, utilizamos a aproximação de Bartlett's para os $r_{j(k)}$ estimados, cujo estimador é dado por:

$$\text{Var} [r_{j(k)}] \approx \frac{1}{n-k-j}$$

Desta forma se:

$$|r_{j(k)}| < 2\sqrt{\frac{1}{n-k-j}} \Rightarrow r_{j(k)} - "0"$$

$$|r_{j(k)}| \geq 2\sqrt{\frac{1}{n-k-j}} \Rightarrow r_{j(k)} - "x"$$

5. EXEMPLOS

Para a aplicação do desenvolvimento teórico descrito nas seções anteriores desenvolveu-se um programa de computador em FORTRAN que possui basicamente duas rotinas de cálculo:

- estimação (por MQO) dos coeficientes das regressões AR(K) iterativas,
- estimação da FACE para diversas lags.

Apresentamos nesta seção a aplicação do referido programa (nossos agradecimentos ao Sr. Janio Mochdece Sales, do último período do curso de engenharia de Sistemas, DEE, PUC/RJ pelo empenho e dedicação na execução de parte do programa) a 3 séries simuladas correspondendo, respectivamente, aos modelos ARMA (0,2); ARMA (2,0) e ARMA (2,1) com as características especificadas no quadro 8.

| | Série 1 | Série 2 | Série 3 |
|-------------|---------------------------------|-----------------------------|-----------------------------|
| N | 500 | 500 | 500 |
| p | 0 | 2 | 2 |
| q | 2 | 0 | 1 |
| ϕ' s | - | $\phi_1 = .5; \phi_2 = -.5$ | $\phi_1 = .5; \phi_2 = -.5$ |
| θ' s | $\theta_1 = .5; \theta_2 = -.5$ | - | $\theta_1 = .4$ |

Quadro 8: Características das séries simuladas

O referido programa, além do cálculo do quadro das FACE estimadas possui uma rotina de reconhecimento de padrões a qual permite a identificação automática do modelo ARMA(p,q) para uma dada série. No caso presente, é feita a procura do triângulo de "zeros" e nele o vértice superior esquerdo que corresponde às ordens p e q do modelo. Nesta primeira versão do programa, os valores dos parâmetros (p e q) estão limitados a: $p \leq 5$ e $q \leq 4$. Dessa forma, não é possível a identificação de modelos para séries sazonais de período grande (por exemplo, séries mensais sazonais pois p e/ou q, ou ambos ≥ 12).

A saída do programa acompanhada da identificação automática é mostrada a baixo nos quadros 9, 10 e 11 para as 3 séries simuladas acima citadas.

Tabela Equivalente

| | | Q | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| P | 0 | x | x | o | o | o |
| | 1 | x | x | x | o | o |
| | 2 | x | x | x | x | o |
| | 3 | x | x | o | x | x |
| | 4 | x | x | o | o | x |
| | 5 | x | x | x | o | x |

MODELO SUGERIDO

ARIMA(0,0,2)

Quadro 9: Saída do programa (série 1)

Tabela Equivalente

| | | Q | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| P | 0 | x | x | x | o | x |
| | 1 | x | x | x | x | x |
| | 2 | o | o | o | o | o |
| | 3 | x | o | o | o | o |
| | 4 | o | x | o | o | o |
| | 5 | x | x | x | o | o |

MODELO SUGERIDO

ARIMA(2,0,0)

Quadro 10: saída do programa (série 2)

Vemos claramente que existe uma perfeita concordância entre os modelos identificados pelo programa e os modelos verdadeiros. Evidentemente é importante salientarmos que foram usadas as 500 observações geradas para cada série, número este exagerado se levarmos em conta que para a maioria das séries

reais encontradas na prática não se dispõe de tantas observações.

Tabela Equivalente

| | | Q | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| P | 0 | x | x | x | o | x |
| | 1 | x | x | x | x | x |
| | 2 | x | o | o | o | o |
| | 3 | x | x | o | o | o |
| | 4 | x | x | x | o | o |
| | 5 | x | x | o | o | o |

MODELO SUGERIDO

ARIMA (2.0.1)

Quadro 11: Saída do programa (Série 3)

Este aspecto está ser investigado através de séries simuladas, com o objectivo de determinar a validade do método para diversos tamanhos de séries, incluindo a versão sazonal. Além disso, estamos estudando a aplicação desta mesma metodologia à identificação de modelos ARMA(p,q) multivariados para vectores de séries cronológicas [Tiao & Tsay, 1982 e Sousa, 1982]. Estes resultados deverão ser publicados em breve.

BIBLIOGRAFIA

AKAIKE, H., 1974 - A New Look the Statistical Model Identification. IEEE Transsactions on Automatic Control, AC-15, 716-723

BEGUIN, J. M., GOURIEROUX, C. & MONFORT, A., 1980 - Identification of a mixed autoregressive-moving average process: the corner method. Time Serie, ed. O. D. Anderson, Amsterdam: North-Holland, 423-436

BOX, G. E. P. & JENKINS, G. M., 1970 - Time Series Analysis

Forecasting and Control, San Francisco: Holden-Day

CHATFIELD, C., 1979 - Inverse Autocorrelation. Journal of the Royal Statistical Society, A, 142, 363-377

CLEVELAND, W. S., 1972- The Inverse Autocorrelations of a Time Series and Their Applications. Technometrics, 14, 277-298

GRAY, H. L., KELLEY, G. D. & McINTIRE, D. D., 1978 - A New Approach to ARMA Modeling. Cummunications in Statistics, B7, 1-77

SOUZA, R. C., 1983 - Modelos Multivariados de Box & Tiao para Séries Temporais Vectoriais. Monografia GSM 09/83, Grupo de Sistemas, DEE, PUC/RJ, Brasil

TSAY R. S. & TIAO, G. C., 1982 - Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. Technical Report 683, Dep. of Statistics, University of Wisconsin, Madison, USA

TIAO, G. C. & TSAY, R. S., 1982 - Multiple Time Series Modeling and Extended Sample Cross Correlations. Technical Report 680, Dep. of Statistics, University of Wisconsin, Madison, USA.

INTERPRETAÇÃO EMPÍRICA DO DUAL DE UM PROGRAMA INTEIRO
MISTO

MOSS MADDEN

Department of Civic Design, University of Liverpool, UK

e

PAULO V. D. CORREIA

Secção de Urbanização e Sistemas, Departamento de
Engenharia Civil, Instituto Superior Técnico, Lisboa,
Portugal

RESUMO

O artigo apresenta as técnicas de programação matemática utilizadas na formulação do dual de um programa inteiro misto e sua interpretação.

Os autores desenvolvem a abordagem proposta por Balas (1970), e através da formulação de um problema simples sobre aquisição de solo por uma Autarquia Local, ilustrado por uma aplicação numérica, revelam como opera o sistema generalizado de preços-sombra e variáveis de subsídio/penalidade.

1. INTRODUÇÃO

Um programa linear contínuo possui as propriedades bem conhecidas de o valor óptimo máximo da função objectivo do primal ser igual ao valor óptimo mínimo da função do objectivo do dual respectivo, bem como de os valores das variáveis duais poderem ser interpretados como preços-sombra das restrições do

primal, e ainda os custos reduzidos representados pelas variáveis de folga das restrições do dual mostrarem quais os preços necessários à inclusão de variáveis primais na base.

Quando se introduzem no sistema variáveis inteiras, estas relações deixam de ser verdadeiras. Williams (1979), referido em Correia e Madden (1985) identifica as seguintes questões:

- 1) O valor óptimo da função objectivo do primal não é igual ao valor óptimo da função objectivo do dual;
- 2) Se uma restrição tem uma folga positiva, não representa necessariamente um bem gratuito, e pode portanto ter um valor económico positivo;
- 3) Em problemas inteiros mistos gerais não existem necessariamente atribuições de valor nas restrições que permitem que a solução óptima seja obtida de forma semelhante ao caso linear.

Uma abordagem possível para a resolução do problema gerado pelo dual de um programa inteiro misto (PIM) envolve a introdução de novas restrições tais que o PIM é redefinido como um programa linear (PL) referido à envolvente convexa do PIM original. Se isto for possível, as variáveis do dual do problema redefinido serão os preços-sombra do primal, mas as relações entre estas variáveis duais e o PIM original poderão ser pouco claras.

Este problema é abordado por Gomory e Baumol (1960), Alcaly e Klevorick (1966) e Williams (1979) entre outros, mas utilizando este método de acréscimo de novas restrições, permanecem os problemas relacionados com a questão de bens gratuitos.

Uma abordagem diferente é sugerida por Balas (1970), ao redefinir o ótimo dual/primal do PIM como sendo uma solução possível do primal com a propriedade de a solução associada do dual satisfazer todas as restrições do dual que correspondem a variáveis contínuas do primal, e que está "tão perto quanto possível" da satisfação das restrições do dual, que correspondam às variáveis do primal restringidas a valores inteiros. Este "tão perto quanto possível" deve ser interpretado no sentido de que a "lacuna", (definida como) a medida em que cada restrição do dual correspondente a uma variável primal restringida a valores inteiros é potencialmente violado é ponderado pela variável inteira respectiva, e a soma ponderada destas "lacunas" é minimizada em relação às variáveis de folga das restrições do dual em simultâneo com a maximização da função objectivo em relação às variáveis do primal.

Formalizando a abordagem de Balas:

$$\text{primal:} \quad \max \quad \underline{c} \cdot \underline{x} \quad (1)$$

$$\text{sujeito a} \quad \underline{A} \cdot \underline{x} \leq \underline{b} \quad (2)$$

$$\begin{aligned} x_j &\geq 0 \\ x_j &\text{ inteiro, } j \in N_1 \end{aligned} \quad (3)$$

$$\text{dual:} \quad \max_{\underline{x}^1} \quad \min_{\underline{u}, \underline{v}} \quad \underline{u} \cdot \underline{b} - \underline{v}^1 \cdot \underline{x}^1 \quad (4)$$

$$\text{sujeito a} \quad \underline{u} \cdot \underline{A} - \underline{v} = \underline{c} \quad (5)$$

$$\begin{aligned} u_i, x_j^1 &\geq 0 \\ x_j &\text{ inteiro, } j \in N_1 \end{aligned} \quad (6)$$

$$v_j \quad \left\{ \begin{array}{l} \text{não restringido, } j \in N_1 \\ \geq 0 \quad \quad \quad j \in N-N_1 \end{array} \right.$$

em que

\underline{x} é o vector das N variáveis do primal das quais N_1 são restringidas a valores inteiros.

\underline{u} é o vector das variáveis duais

\underline{v} é o vector com N elementos, das variáveis de folga das restrições do dual.

e \underline{A} , \underline{b} e \underline{c} são coeficientes

Balas prossegue com a identificação do PL que tem a mesma solução óptima do que o PIM original. Isto é:

$$\max \quad (\underline{c}^1 + \underline{s}^1) \underline{x}^1 + \underline{c}^2 \cdot \underline{x}^2 \quad (7)$$

$$\text{sujeito a} \quad \underline{A}_1 \cdot \underline{x}^1 + \underline{A}_2 \underline{x}^2 \leq \underline{b} \quad (8)$$

$$x_j^1, x_j^2 \geq 0 \quad (9)$$

em que os índices inferiores e superiores 1 e 2 se referem respectivamente às variáveis inteiras e às variáveis contínuas, e \underline{s} é um vector de subsídios ou penalidades.

\underline{s} é definido por:

$$s_j^2 = 0 \quad \left\{ \begin{array}{l} \bar{v}_j \quad \text{para } j \in N_1 \text{ e } \bar{x}_j > 0 \\ \min(0, \bar{v}_j) \quad \text{para } j \in N_1 \text{ e } \bar{x}_j = 0 \end{array} \right. \quad (10)$$

$$s_j^1 = \left\{ \begin{array}{l} \bar{v}_j \quad \text{para } j \in N_1 \text{ e } \bar{x}_j > 0 \\ \min(0, \bar{v}_j) \quad \text{para } j \in N_1 \text{ e } \bar{x}_j = 0 \end{array} \right. \quad (11)$$

O efeito de substituir \underline{c}^1 por $\underline{c}^1 + \underline{s}^1$ em (7) é o de mudar o declive do hiperplano da função objectivo de modo a

torná-lo paralelo à face do polítopo convexo, definido por (8) e (9), que contém o conjunto óptimo de \underline{x} . É portanto possível que existam soluções não inteiras para (7) a (9) que contenham a solução óptima \bar{x} , inteira para \bar{x}_1^1 , que são óptimas para (7).

Balas identifica casos especiais para os PIM's zero-um, em que (7) a (9) podem ser acrescidos de

$$\tilde{x}_j^1 \leq 1 \quad (12)$$

Balas sugere que a utilização deste método possibilita uma interpretação económica segundo a qual o habitual sistema de preços-sombra da programação linear é substituído por um sistema generalizado de preços-sombra u_i e subsídios ou penalidades v_j , em que os u_i são restringidos a não serem negativos e estão associados às restrições do primal, e os s_j não são restringidos, relacionando-se com as variáveis inteiras do primal.

Recapitulemos o processo necessário à formulação deste sistema:

- 1) Resolução do PIM original através de um método qualquer apropriado, restringindo as variáveis do primal a valores inteiros onde necessário;
- 2) Formulação do dual. Para todas as restrições do dual relacionadas com as variáveis inteiras do primal, introduzem-se variáveis de folga não restringidas, obtendo assim o efeito de remover cada restrição deste tipo do conjunto das restrições.

Introdução na função objectivo do dual de todas as variáveis de folga não restringidas, multiplicadas pelos

valores das variáveis do primal que lhes estão associadas, para o óptimo do primal, através da subtracção destes produtos da função objectivo do dual. Introdução de variáveis de folga restringidas nas outras restrições do dual. Resolução do dual através da minimização do conjunto variáveis duais e de folga não restringidas.

- 3) Formulação do PL equivalente com vista à interpretação dos subsídios/penalidades sobre cada variável primal restringida a valores inteiros.

2. UM PROBLEMA DE AQUISIÇÃO DE SOLO

Consideremos o seguinte problema, que representa uma versão simplificada de um outro apresentado por Correia e Madden (1985). Uma Câmara Municipal dispõe de um conjunto de critérios de mínimos e máximos de aprovisionamento de solo para diferentes tipos de equipamento social repartido por várias "áreas de gestão fundiária" (agf's). Dentro de cada agf existem vários terrenos à venda, alguns dos quais (no todo ou em parte) a Câmara deve adquirir para aí localizar o equipamento social. Alguns terrenos têm localizações isoladas enquanto outros se localizam em pares adjacentes, um urbano e outro rústico. A Câmara Municipal pretende minimizar o seu orçamento para estas aquisições.

Podemos apresentar o problema da seguinte forma:

maximizar

$$\begin{aligned}
 & - \sum_{ijk} c_{ik} \beta_j \bar{x}_{ijk} - \sum_{ijk} (c_{ik} \cdot \beta_j \cdot \hat{x}_{ijk} + \\
 & + c_{ik} \cdot \hat{JX}_{ij}^k) - \sum_{ijk} (c_{ik} \cdot \beta_j \cdot \bar{x}_{ijk} + c_{ik} \cdot \bar{JX}_{ij}^k)
 \end{aligned} \tag{13}$$

sujeito a

$$\sum_j \beta_{ij} \cdot X_{ijk} \leq A_{ik} \quad \text{para quaisquer } i \text{ e } k \quad (14)$$

$$\sum_j \beta_{ij} \cdot \hat{X}_{ijk} + \sum_j \hat{J}X_{ij}^k \leq \hat{A}_{ik} \quad \text{para quaisquer } i \text{ e } k \quad (15)$$

$$\sum_j \beta_{ij} \cdot \bar{X}_{ijk} + \sum_j \bar{J}X_{ij}^k \leq \bar{A}_{ik} \quad \text{para quaisquer } i \text{ e } k \quad (16)$$

$$-\sum_k X_{ijk} - \sum_k CX_{ij}^{\hat{k}\bar{k}} \leq -\text{MIN}_{ij} \quad \text{para quaisquer } i \text{ e } j \quad (17)$$

$$\sum_k X_{ijk} + \sum_k CX_{ij}^{\hat{k}\bar{k}} \leq \text{MAX}_{ij} \quad \text{para quaisquer } i \text{ e } j \quad (18)$$

$$\beta_{ij} \cdot CX_{ij}^{\hat{k}\bar{k}} - \hat{J}X_{ij}^k - \bar{J}X_{ij}^{\bar{k}} \leq 0 \quad \text{para quaisquer } i, j \text{ e } k \quad (19)$$

$$-\beta_{ij} \cdot CX_{ij}^{\hat{k}\bar{k}} + \hat{J}X_{ij}^k + \bar{J}X_{ij}^{\bar{k}} \leq 0 \quad \text{para quaisquer } i, j \text{ e } k \quad (20)$$

$$X_{ijk}, CX_{ij}^{\hat{k}\bar{k}} \leq 1 \text{ e inteiro} \quad (21)$$

em que

c_{ik} representa o custo do solo por hectare no terreno k na agf i

- β_j representa a área (em hectares) necessária a uma unidade de equipamento do tipo j , ou seja do uso do solo ;
- X_{ijk} representa a decisão de localizar ou não o uso do solo j no terreno k da agf i ;
- JX_{ij} representa a área de solo atribuída ao uso j no terreno da agf i quando k é adjacente a outro terreno;
- $\hat{C}X_{ij}^{kk}$ representa a decisão de localizar ou não o uso do solo j no terreno ou terrenos k e k na agf i ;
- A_{ik} representa a área do terreno k na agf i ;
- MIN_{ij} representa o requisito mínimo do uso do solo j para a agf i ;
- MAX_{ij} representa o requisito máximo do uso do solo j para a agf i ;

referindo-se $\hat{}$ e $-$ respectivamente aos terrenos urbanos e rústicos adjacentes. A função objectivo é uma função de minimização de custo. A restrição (14) assegura que a cada terreno isolado não serão atribuídos mais usos do que aqueles que a sua área permite, enquanto as restrições (15) e (16) asseguram o mesmo para terrenos urbanos e rústicos em parcelas adjacentes. As restrições (17) e (18) asseguram que em qualquer agf se farão as aquisições de solo suficientes para satisfazer pelo menos os requisitos mínimos mas não mais do que os requisitos máximos para cada uso do solo.

É evidente que num programa de minimização de custo deste tipo, a restrição (18) é redundante, mas decidimos mantê-la com vista a possibilitar quaisquer mudanças na função objetivo, como seja a maximização da área a adquirir. As restrições (19) e (20) asseguram que no caso de um uso atravessar a linha divisória entre dois terrenos adjacentes, a sua área total é igual à área necessária a esse uso do solo. A restrição (21) obriga as variáveis de decisão a tomarem os valores 0 ou 1. X_{ijk} é igual a 0 se o uso do solo j não for atribuído ao terreno isolado k na agf i , e é igual a 1 se o uso j for atribuído. $CX_{ij}^{\hat{k}\bar{k}}$ é igual a 1 se o uso do solo j for atribuído ao par de terrenos adjacentes k e \bar{k} da agf i . Este uso pode-
 confinar apenas a um dos terrenos, ou então estender-se através da divisória entre os dois terrenos. A variável toma o valor 0 no caso do uso j não ser atribuído.

O dual Balasiano deste problema tem a seguinte formulação:

$$\begin{aligned}
 & \max_{X_{ijk}, \hat{X}_{ijk}, X_{ijk}, CX_{ij}^{kk}} \min_{U, V, W} \sum_{ik} A_{ik} \cdot U^1 + \sum_{ik} \hat{A}_{ik} \cdot U^2 + \sum_{ik} \bar{A}_{ik} \cdot U^3 - \\
 & \sum_{ij} \text{MIN}_{ij} \cdot U^4 + \sum_{ij} \text{MAX}_{ij} \cdot U^5 - \\
 & \sum_{ijk} X_{ijk} \cdot V - \sum_{ijk} \hat{X}_{ijk} \cdot \hat{V} - \\
 & \sum_{ijk} \bar{X}_{ijk} \cdot \bar{V} - \sum_{ijk} CX_{ij}^{\hat{k}\bar{k}} \cdot W
 \end{aligned} \tag{22}$$

sujeito a:

$$\beta_j \cdot U_{ik}^1 - U_{ij}^4 + U_{ij}^5 - V_{ijk} = -c_{ik} \beta_j \quad \text{para quaisquer } i, j \text{ e } k \quad (23)$$

$$\beta_j \cdot U_{ik}^2 - \hat{V}_{ijk} = -c_{ik} \hat{\beta}_j \quad \text{para quaisquer } i, j \text{ e } k \quad (24)$$

$$\beta_j \cdot U_{ik}^3 - \bar{V}_{ijk} = -c_{ik} \bar{\beta}_j \quad \text{para quaisquer } i, j \text{ e } k \quad (25)$$

$$U_{ik}^2 - U_{ijk}^6 + U_{ijk}^7 - \hat{Z}_{ijk} = -c_{ik} \quad \text{para quaisquer } i, j \text{ e } \hat{k} \quad (26)$$

$$U_{ik}^3 - U_{ijk}^6 + U_{ijk}^7 - \bar{Z}_{ijk} = -c_{ik} \quad \text{para quaisquer } i, j \text{ e } \bar{k} \quad (27)$$

$$-U_{ij}^4 + U_{ij}^5 + \beta_j \cdot U_{ijk}^6 - U_{ijk}^7 \cdot W_{ijk} = 0 \quad \text{para quaisquer } i, j \text{ e } \hat{k} \quad (28)$$

com V_{ijk} , \hat{V}_{ijk} , \bar{V}_{ijk} e W_{ijk} não restringidos

U_{ik}^1 é o preço-sombra da área de um terreno isolado, e representa o decréscimo do orçamento se uma unidade de área adicional fosse tornada disponível para o uso j no terreno k da agf i . U_{ik}^2 e U_{ik}^3 são preços-sombra semelhantes ao anterior, mas para terrenos urbanos e rústicos adjacentes. U_{ij}^4 é o preço-sombra dos requisitos mínimos. Representa o decréscimo do orçamento se o requisito mínimo do uso do solo j na agf i fosse reduzido em uma unidade.

U_{ij}^5 é um preço-sombra semelhante relativo aos requisitos máximos, e deverá ser nulo para a função objectivo (13). U_{ijk}^6

e U_{ijk}^7 não parecem ter uma interpretação palpável neste problema.

O PL equivalente ao PIM original pode então ser formulado com a seguinte função objectivo:

$$\begin{aligned}
 \text{maximizar} & - \sum_{ijk} \sum_{ik} (c_{ik} \cdot \beta_j + s_{ijk}^1) X_{ijk} - \\
 & - \sum_{ijk} [(c_{ik}^{\wedge} - \beta_j + s_{ijk}^2) \hat{X}_{ijk}^{\wedge} + c_{ik}^{\wedge} \cdot \hat{J}X_{ij}^{\wedge k}] \\
 & - \sum_{ijk} [(c_{ik}^{\bar{}} \cdot \beta_j + s_{ijk}^3) \bar{X}_{ijk}^{\bar{}} + c_{ik}^{\bar{}} \cdot \bar{J}X_{ij}^{\bar{k}}] \\
 & - \sum_{ijk} s_{ijk}^4 \cdot CX_{ijk}^{\wedge k} \quad (29)
 \end{aligned}$$

sujeito às restrições (14) a (20)

$$e \quad X_{ijk} \leq 1 \quad \text{para quaisquer } i, j \text{ e } k \quad (30)$$

$$\hat{X}_{ijk}^{\wedge} \leq 1 \quad \text{para quaisquer } i, j \text{ e } \hat{k} \quad (31)$$

$$\bar{X}_{ijk}^{\bar{}} \leq 1 \quad \text{para quaisquer } i, j \text{ e } \bar{k} \quad (32)$$

$$CX_{ijk}^{\wedge k} \leq 1 \quad \text{para quaisquer } i, j \text{ e } \hat{k} \quad (33)$$

Nesta formulação os s^x são definidos como

$$s_{ijk}^1 = \begin{cases} V_{ijk}^* & \text{para } X_{ijk}^* = 1 \\ \min(0, V_{ijk}^*) & \text{para } X_{ijk}^* = 0 \end{cases} \quad (34)$$

$$s_{ijk}^2 = \begin{cases} \hat{V}_{ijk}^{\wedge} & \text{para } \hat{X}_{ijk}^{\wedge} = 1 \\ \min(0, \hat{V}_{ijk}^{\wedge}) & \text{para } \hat{X}_{ijk}^{\wedge} = 0 \end{cases} \quad (35)$$

$$s_{ijk}^3 = \begin{cases} \bar{v}_{ijk}^* & \text{para } \bar{x}_{ijk}^* = 1 \\ \min(0, \bar{v}_{ijk}^*) & \text{para } \bar{x}_{ijk}^* = 0 \end{cases} \quad (36)$$

$$s_{ijk}^4 = \begin{cases} w_{ijk}^{\wedge} & \text{para } CX_{ij}^{\wedge k} = 1 \\ \min(0, w_{ijk}^{\wedge}) & \text{para } CX_{ij}^{\wedge k} = 0 \end{cases} \quad (37)$$

Com vista à interpretação dos subsídios/penalidades, comecemos por examinar s_{ijk}^1 . Existem quatro resultados possíveis:

- a) $X_{ijk} = 1$ e $s_{ijk}^1 \geq 0$
- b) $X_{ijk} = 1$ e $s_{ijk}^1 \leq 0$
- c) $X_{ijk} = 0$ e $s_{ijk}^1 \leq 0$
- d) $X_{ijk} = 0$ e $s_{ijk}^1 = 0$

Em a) s_{ijk}^1 é positivo, e portanto actuando contra a maximização da função objectivo. Deve ser interpretado como uma penalidade pagável pelo agente de decisão "local" (adl) se quiser manter o uso do solo j na sua parcela k na agf i. Por outro lado, em b) s_{ijk}^1 é negativo, aumentando a maximização da função objectivo, e pode ser interpretado como um subsídio pagável ao adl pelo município com vista a "encorajar" a implementação do uso do solo j no terreno k da agf i. Em c) temos um subsídio destinado a impedir a implementação de um uso do solo num de terminado terreno e em d) não existe qualquer subsídio ou penalidade, uma vez que se $X_{ijk} = 0$, s_{ijk}^1 nunca poderá ser maior do que 0, em consequência de (34). Os outros s^x poderão ser interpretados de modo semelhante.

3. RESOLUÇÃO DO PROBLEMA - UMA APLICAÇÃO NUMÉRICA

Com vista à demonstração da resolução do dual Balasiano e à formulação do sistema de subsídios/penalidades, examinaremos uma aplicação numérica. Os quadros 1,2, 3 e 4 contêm os dados para um caso de estudo hipotético contendo duas agf's com três usos do solo

QUADRO 1: CARACTERÍSTICAS DO CASO DE ESTUDO

| agf | terrenos isolados | terrenos adjacentes | | usos do solo |
|-----|-------------------|---------------------|----------|--------------|
| | | urbano | rústico | |
| M1 | 3 (1,2,3) | 1(1u) | 1(1) | 3 (a,b, c) |
| M2 | 2 (1,2) | 2(1u,2u) | 2(1r,2r) | 3 (a,b, c) |

QUADRO 2: DESIGNAÇÃO DE CADA TERRENO, ÁREA E PREÇO RESPECTIVO

| TERRENO | ÁREA (HA) | PREÇO(10 ⁷ esc./ha) |
|---------|-----------|--------------------------------|
| M11 | 0.5 | 0.5 |
| M12 | 2.0 | 0.8 |
| M13 | 0.5 | 1.0 |
| M11u | 0.4 | 1.0 |
| M11r | 0.8 | 0.3 |
| M21 | 1.5 | 1.0 |
| M22 | 2.0 | 0.8 |
| M21u | 1.0 | 1.3 |
| M22u | 2.0 | 1.0 |
| M21r | 6.0 | 0.2 |
| M22r | 5.0 | 0.2 |

QUADRO 3: ÁREAS NECESSÁRIAS PARA CADA TIPO DE USO DO SOLO

| USO | ÁREA (ha) |
|-----|-----------|
| a | 0.5 |
| b | 1.0 |
| c | 2.0 |

QUADRO 4: REQUISITOS MÁXIMOS/MÍNIMOS

| agf | USOS | | | | | |
|-----|------|-----|-----|-----|-----|-----|
| | a | | b | | c | |
| | max | min | max | min | max | min |
| M1 | 3 | 1 | 2 | 1 | 2 | 1 |
| M2 | 4 | 2 | 3 | 1 | 4 | 2 |

O quadro 2 contém os valores dos coeficientes A e C, o quadro 3 os coeficientes β e o quadro 4 os valores de MAX e MIN. Substituindo estes coeficientes e parâmetros em (13) a (21) e formando as variáveis apropriadas, obtemos o primal do nosso problema. A resolução deste primal, utilizando um algoritmo de programação inteira mista, produziu o seguinte resultado:

Função objectivo : - 3.49

Variáveis inteiras

não nulas:

$$X_{111}^{11}, X_{132}^{11}, CX_{12}^{11}, CX_{21}^{11}, CX_{21}^{22}, CX_{32}^{11}, \\ CX_{22}^{22}, CX_{23}^{22}.$$

Variáveis contínuas $\hat{JX}_{12}^1=0.2, \bar{JX}_{12}^1=0.8, \bar{JX}_{21}^1=0.5, \bar{JX}_{23}^1=2.0,$

não nulas:

$$\bar{JX}_{21}^2=0.5, \bar{JX}_{22}^2=1.0, \bar{JX}_{23}^2=2.0$$

O impacto deste conjunto de resultados na área em estudo pode ser visualizado através da Figura 1. Podemos verificar que todos os requisitos mínimos são respeitados, e que na agf 1 o uso do solo b se estende por dois terrenos adjacentes.

Como seria de esperar num exercício de minimização de custos, os terrenos mais baratos são adquiridos dentro do possível. Na agf 2 todas as atribuições de uso se localizam em terrenos rústicos e baratos adjacentes a terrenos urbanos.

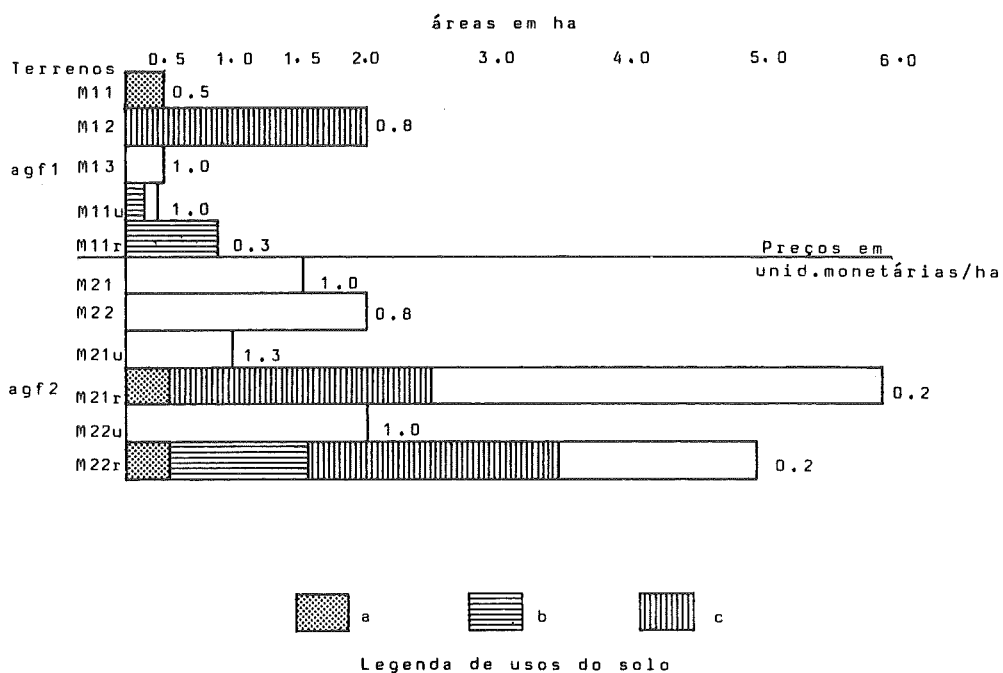


FIGURA 1

Com base na informação resultante da solução do primal, podemos prosseguir com a formulação do dual Balasiano.

A função objectivo do dual é obtida por substituição dos valores dos coeficientes em (22), notando que surgem nesta função apenas as variáveis de folga duais $V_{111}, V_{132}, W_{121}, W_{211}, W_{212}, W_{221}, W_{231}, W_{232}$.

As restrições são obtidas por substituição em (23) a (28), e o problema é resolvido utilizando um algoritmo de programação linear. A solução é a seguinte:

$$\begin{aligned} \text{Função objectivo:} & \quad - 3.49 \\ \text{Variáveis não nulas:} & \quad U_{11}^1=0.5, U_{12}^1=0.2, U_{11}^3=0.7, U_{11}^4=0.5, \\ & \quad U_{12}^4=1.0, U_{13}^4=2.0, U_{21}^4=0.4, U_{22}^4=0.2, \\ & \quad U_{23}^4=1.6 \end{aligned}$$

As variáveis U_{ik} podem ser interpretadas como rendas cobráveis sobre os terrenos k de i . Só são não nulas obviamente para terrenos completamente ocupados, e representam a diferença em unidades de custo/hectare entre cada terreno completamente ocupado (adquirido) e o terreno mais barato não completamente ocupado (adquirido). As variáveis U_{ij} representam os aumentos de custo que resultariam se os requisitos mínimos para cada j nas diferentes agf's fossem aumentados de uma unidade. Uma análise na Figura 1 confirma estas observações. No entanto, também devemos inspeccionar os valores de na solução do dual. Os valores não nulos dos s^x são obtidos a partir dos valores não nulos das variáveis de folga não restringidas V, \hat{V}, \bar{V} e W utilizando (34) a (37). Os s^x não nulos são os seguintes:

$$s_{211}^4 = -.3, s_{212}^4 = -.3, s_{231}^4 = -1.2 \text{ e } s_{232}^4 = -1.2$$

Estes valores são interpretados como subsídios pagáveis pelo município a qualquer adl com vista a encorajar a implementação do uso do solo pretendido para o terreno em causa.

Podemos agora analisar cada um destes subsídios: s_{211}^4 refere-se à manutenção do uso do solo a no par de terrenos adjacentes identificado por lu e lr da agf 2. O custo do terreno lr para o uso a é igual a $0.2 \times 0.5 = 0.1$ unidades monetárias. Se o uso do solo a fosse implementado no terreno menos caro seguinte, custaria $0.8 \times 0.5 = 0.4$ unidades. A diferença de 0.3 é o subsídio pagável por forma a garantir que a implementação terá lugar no terreno mais barato. O mesmo subsídio é aplicável à localização do uso do solo a no terreno 2r da agf 2.

s_{231}^4 e s_{232}^4 referem-se ao subsídio pagável com vista evitar a re-localização do uso e noutros terrenos da agf 2. O custo previsto para a implementação de c é de $2 \times 0.2 = 0.4$. O custo da sua localização no terreno menos caro seguinte seria de $2 \times 0.8 = 1.6$. A diferença é o subsídio calculado de 1.2. De notar que $s_{222}^4 = 0$ porque é possível relocar o uso do solo b do terreno 2r da agf 2 para o terreno lr na mesma agf.

Uma interpretação alternativa destes subsídios, uma vez que neste exercício estamos apenas a lidar com adl's hipotéticos, seria a de que o município estaria disposto a pagar mais pela área necessária em certos terrenos, portanto acima dos preços de mercado. A diferença de valor a pagar seria até ao limite indicado pelo subsídio.

4. CONCLUSÕES

Neste artigo aplicamos a formulação do dual de um PIM devida a Balas (1970) a um problema de aquisição de solo por uma autarquia local. Utilizando informação hipotética mas realista, avaliamos um conjunto de valores de variáveis duais e de subsídios/penalidades que sugerimos poderem ser interpretados como a diferença entre o preço de mercado de

terrenos e a soma de dinheiro que a autarquia local estaria disposta a pagar pela área necessária em cada terreno.

Utilizando mais informação sobre os níveis de utilidade associados aos usos potenciais possíveis em cada terreno de aquisição possível pelo município permitir-nos-á dar uma interpretação consistente ao significado das variáveis de subsídios/penalidades, e colocar ao dispor das autarquias, em particular dos seus agentes decisórios, uma verdadeira ferramenta de gestão em que poderão fundamentar a sua formulação de estratégias de gestão fundiária municipal.

5. REFERÊNCIAS

- Alcaly, R.E. and A.K. Klevorick 1966 "A Note on the Dual Prices of Integer Programs", *Econometrica* Vol. 34, No.1, pp.206-214.
- Balas, E. 1970 "Duality in Discrete Programming" in H.W.Kuhn (ed.) *Proceedings of the Symposium on Mathematical Programming*, Princeton, 1970 (P.U.P., Princeton), pp. 179-197.
- Correia, P.V.D. and M. Madden 1985 "Optimisation of Land Purchasing and Management using Mixed-Integer Programming: a Case Study in a Portuguese Municipal Authority". *Environment and Planning B, Planning and Design* (forthcoming).
- Gomory, R.E. and W. Baumol 1960 "Integer Programming and Pricing". *Econometrica* Vol.28, No.3, pp. 521-550.
- Williams, H.P. 1979 "Model Building in Mathematical Programming", John Wiley, Chichester.

A TAXONOMIC DELIMITATION OF FUNCTIONAL URBAN REGIONS IN A
HETEROGENEOUS COUNTRY

L.VALADARES TAVARES, J.A.ANTUNES FERREIRA and
R. CARVALHO OLIVEIRA

CESUR, Instituto Superior Técnico, Lisboa, Portugal

ABSTRACT

The identification of the interdependence between the urban centres and the surrounding hinterlands is a crucial issue for the processes of urban planning and policy making. Though there is not a unique criterion to define functional urban regions, a standardized procedure based on employment and commuting data has been applied in the regionalization of several european countries. However, reliable data on employment and commuting flows are not available in Portugal and indirect indicators of the interdistricts relationships had to be defined : the expected travel time by public transport modes. A numerical taxonomy method was then applied to generate clusters that can be used to define functional urban regions.

1. INTRODUCTION

The identification of the interdependence between the urban centres and the surrounding hinterlands is a crucial issue for the processes of urban planning and policy making. These plans and policies can not ignore the pressure that the population living in areas closely linked to the urban centre put upon this centre which has either to adjust its supply of

goods and services to the demand generated by these populations or to influence and divert this demand in some other more desirable direction. The urban centre can not therefore be dissociated from the surrounding areas with which it forms a complex system with multiple links and interactions that have to be identified and characterized.

Obviously, the stronger the links and the interactions the higher the pressure put upon the central city by the resident population in its sphere of influence. A very strong and important type of relationship can be described as the interaction between the place of residence and workplace, giving rise to the notion of functional urban region that can therefore be viewed as a daily urban system. The definition of the boundaries of such an urban system is therefore a crucial issue not only because of the required transportation infrastructure and facilities but also because of the demand for other goods and services that such a strong link with the central city tends to generate.

The purpose of this paper is to present a framework developed to delineate functional urban regions in Portugal. Most delimitation procedures develop the analysis in two stages :

- i) Identification of dominant urban centres (urban cores) on the grounds of a minimum required population and/or number of jobs.
- ii) Allocation of each hinterland spacial unit to the urban core to which it is more closely linked (usually using commuting patterns).

This procedure has been widely used ((1) to (5)) producing self-contained regions with minimum connections across region

boundaries.

However, when it comes to applying this framework to developing countries, two major difficulties can arise. First the scarcity of the required data and, second, the high heterogeneity that characterizes the socio-economic spatial arrangement illustrated by the large dispersion of the commonly accepted socio-economic indicators. This frequently dichotomized character of developing countries may render ineffective the procedures based on threshold analysis, and new approaches might be needed. The difficulties that Hall and Hay faced in their urban regionalization of Portugal, (3), serve to illustrate this situation.

Reliable data on the flows of capital, labour, services, passengers or goods between the spatial units are not available in Portugal. Therefore, indirect indicators of the inter-spatial units relationships had to be defined : the expected travel time between every two spatial units (by bus and train), used as a measure of accessibility (or potential commuting). A Numerical Taxonomy method based on the concept of minimum spanning tree and developed at CESUR, (6), was then applied to generate clusters that can be used to define functional urban regions (section 3).

In section 4 a discussion of the results is presented together with a comparison with the function urban regions delineated by Hall and Hay.

2. DEFINITION OF URBAN CORE *CONCELHOS* (UCCs)

The delimitation of FURs implies the identification of the respective urban cores which (according to Hall (2)) consist of cities containing at least 20 000 jobs and all their contiguous communities containing at least 12.35 jobs/ha (A

minimum population of 60 000 inhabitants for the core and its hinterland is also required). Unfortunately data about jobs density are not available for Portugal and therefore other analyses based on working population/ *concelho* data recently obtained and not yet published have been developed.

The rank size graph of residential population (A) working in industry and Services (including Commerce) is presented in figure 1 and the following conclusions can be drawn up :

- the first 5 *concelhos* (fig.1) are separated from the others by a major discontinuity expressing the outstanding level of the urban concentration for Lisboa, Porto and some of their dormitory areas (Group 1)
- a second and less important discontinuity can be detected for the 24 th (fig. 1) *concelho* and the corresponding non-agricultural working population is equal to 17 700 inhabitants. It should be remarked that this value is rather close to the limit of 20 000 jobs suggested by Hall and adopted by IIASA (Group 2)
- the *concelhos* of group 1 & 2 are located within a narrow strip of land near the coast (fig. 3) and therefore the study of the inner territory implies the selection of a few other units. With this purpose, the next 11 *concelhos* (fig. 1) are also considered.

Concelho is a portuguese administrative local division (in average, smaller than an english county). The *concelhos* are grouped in districts.

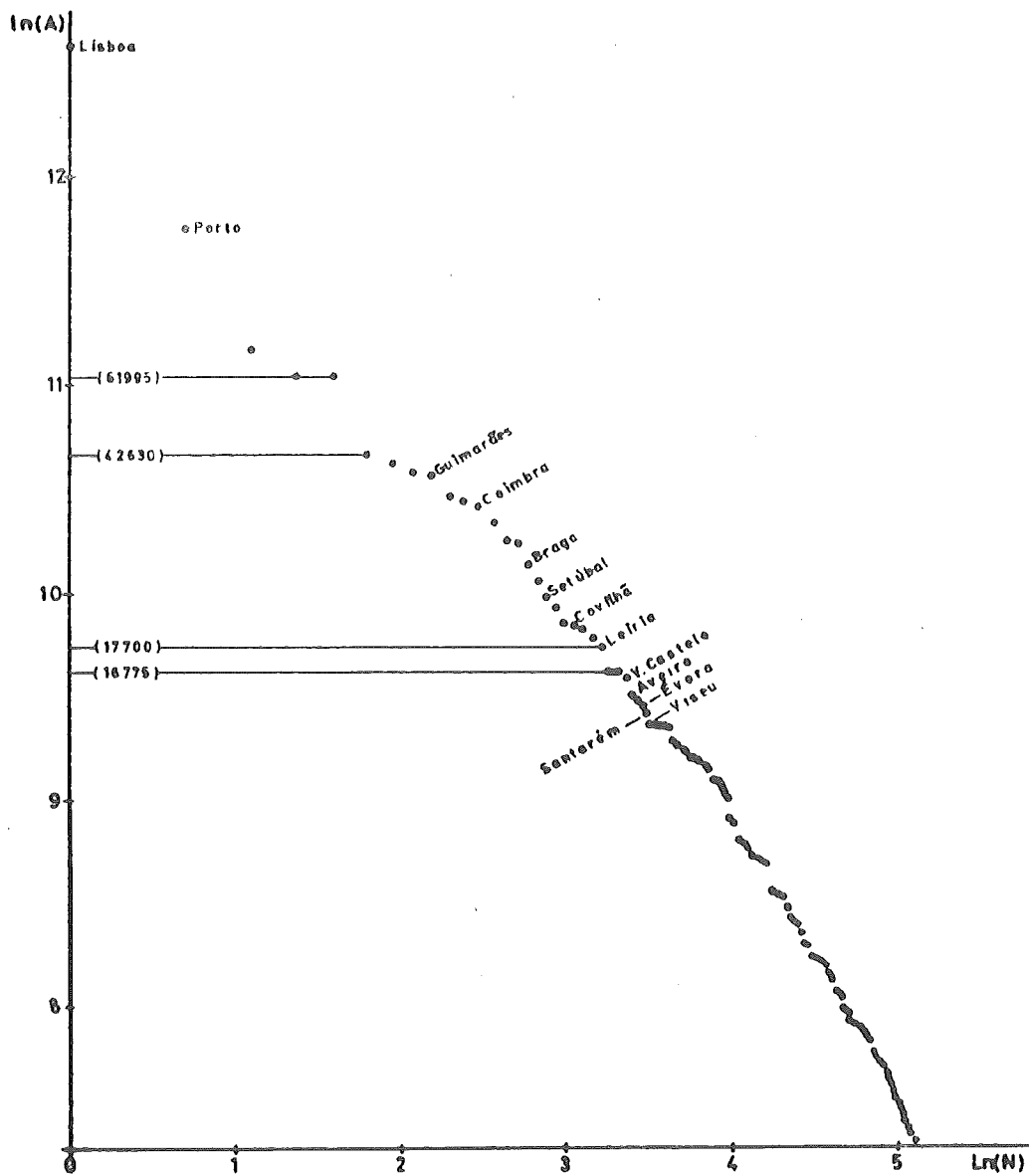


Fig. 1 - Rank Size Analysis for non-agricultural active population

Therefore, the most urbanized *concelhos* are those included in groups 1, 2 and 3 if the non-agricultural working population is used as an indicator of the urban phenomena. The urban cores are then selected as the major urban centres of these *concelhos* after excluding those which have a dormitory function or which are an industrial extension of major centres. The accepted *concelhos* (named *urban core concelhos*) are indicated in fig 2.

3. DEFINITION OF THE FUNCTIONAL URBAN REGIONS AS A SET OF INTERCONNECTED *CONCELHOS* USING A TAXONOMIC APPROACH

Updated flows of capital, labour services, passengers or goods between portuguese *concelhos* are unknown and therefore indirect indicators of inter - *concelhos* relationships have to be defined. The expected travel time (by public transport modes) between pairs of *concelhos* is adopted with this purpose because the supply of these services is directly related to its demand and because the other links are correlated with this indicator (master connection).

In a previous work, (7), the expected travel time by bus was adopted and a summary of the procedure and results is presented in paragraph 3.1. the analysis is now extended to cover possibility of travelling by train (paragraph 3.2).

3.1 Travelling by Bus

The expected time (A_{ij}) that has to be spent to go by bus from *concelho* i to any other j is defined as

$$A_{ij} = T_{ij} + P W_{ij} + (1-P) W_{ij}^* \quad (1)$$

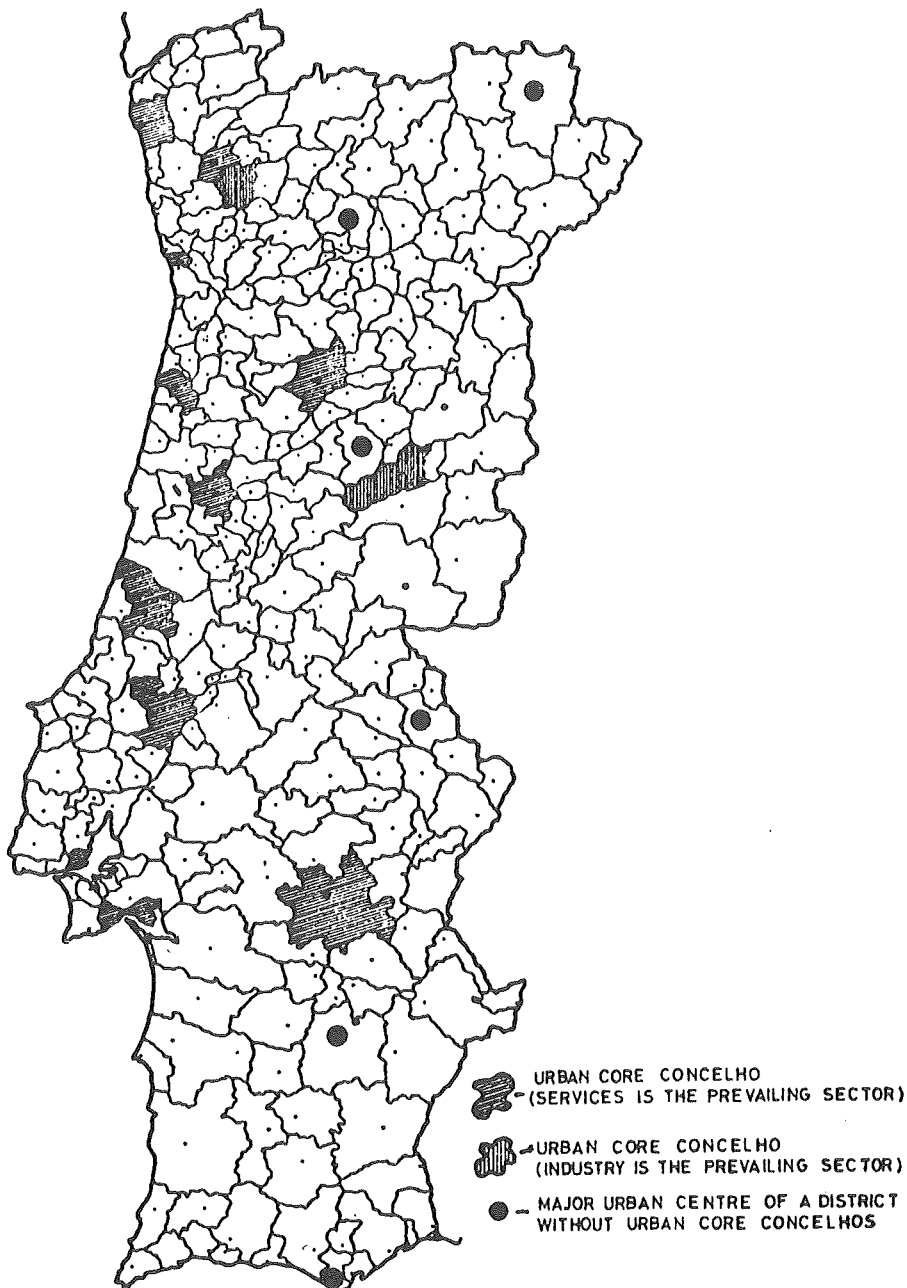


Fig. 2 - Urban Core Concelhos

where

T_{ij} - travel time between i and j

P - probability that the trip can be made at any time during the day. P is made equal to 0.75.

W_{ij} - waiting time assuming that the existing departure times are convenient for (i,j) trip. W_{ij} is made equal to 10 minutes.

W^*_{ij} - waiting time assuming that the trip has to be made at a random instant uniformly distributed from 5 a.m. until 1 a.m.. Thus:

$$W^*_{ij} = \frac{20 \text{ hours}}{2(S + 1)}$$

if there are S departures equally spaced during that period.

A_{ij} - matrix is computed from the collected time-tables of all existing inter-
- *concelhos* bus lines and suburban trains.

Now the delimitation of FURs implies the allocation of *concelhos* to each UCC in order to maximize the inter-FURs independence (or equivalently to maximize the within FURs interdependence).

With this purpose, a method of Numerical Taxonomy based on the concept of minimum spanning tree and developed at CESUR (3) is adopted because previous studies (3) have shown that the minimum spanning tree criterion is usually more convenient for taxonomic problems of regional analysis than the usual criteria based on the sum of squared deviations. Given a set of N individuals defined on a K - dimension space this method

determines the minimum number of clusters (*optimal solution*) for which the distance between any individual and the complementary set of individuals of the same cluster is lower than or equal to a specified value, L . The distance between one individual, P , and a set of individuals, C , is defined as the minimum distance between P and any individual contained by C .

This method is initially applied to the given set of individuals and L is usually made equal to $(1+\alpha) d_{\min}$ where d_{\min} is the minimum distance between any pair of individuals ($\alpha > 0$).

This method can be also applied to the generated clusters (some of them can be formed by a single individual) providing that the distance between a pair of clusters is defined. This distance is made equal to the distance between any pair of individuals contained by these two clusters because the increment introduced on the size of their minimum spanning tree by virtue of their linkage is equal to that distance. Therefore, M successive generations of clusters can be obtained using $L=(1+\alpha) d_{\min}$ where d_{\min} is the minimum distance between the obtained clusters in the previous generation. (Obviously, a set of clusters generation is obtained for each α). It can be easily shown that $M < (N-1)$ and the production of the clusters by successive generations is an interesting feature of this approach for sensitivity studies.

Obviously the space defined through A_{ij} is a non-euclidian one and the generated clusters by this method with $\alpha = 0.25$ are presented in figures 3, 4, 5, 6. The computing time spent to obtain a complete set of generations was about 3 minutes using an IBM 360-44.

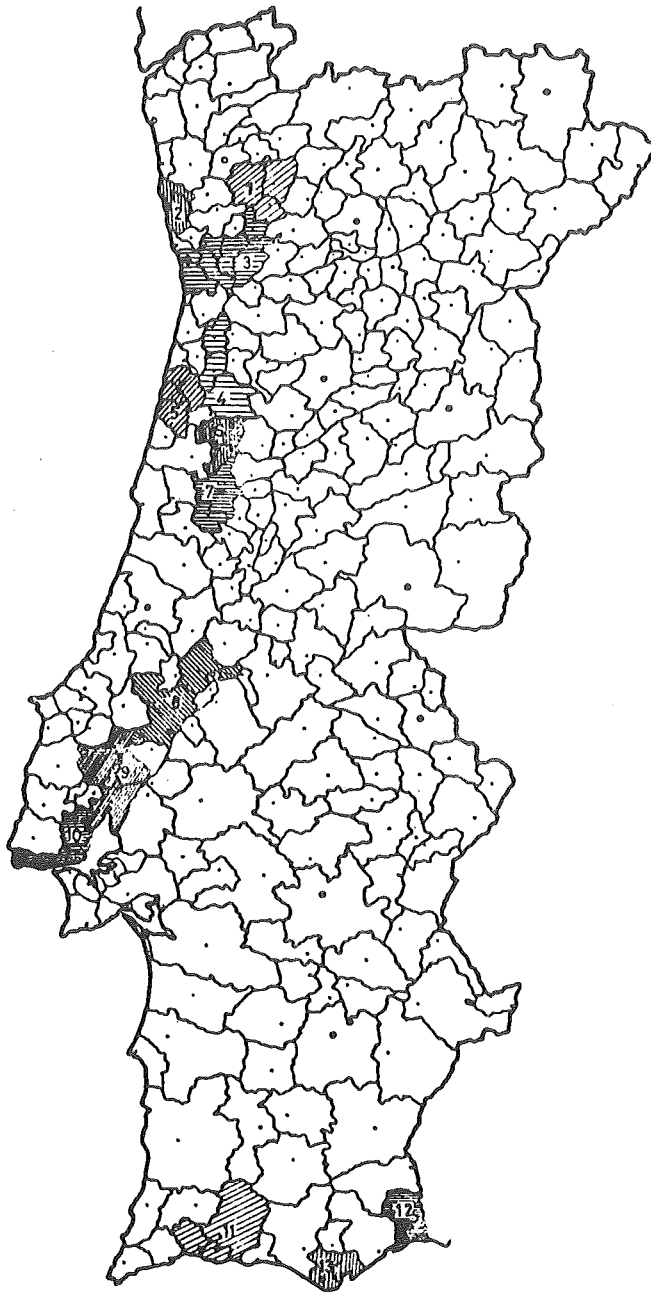


Fig. 3 - 3rd Generation Clusters

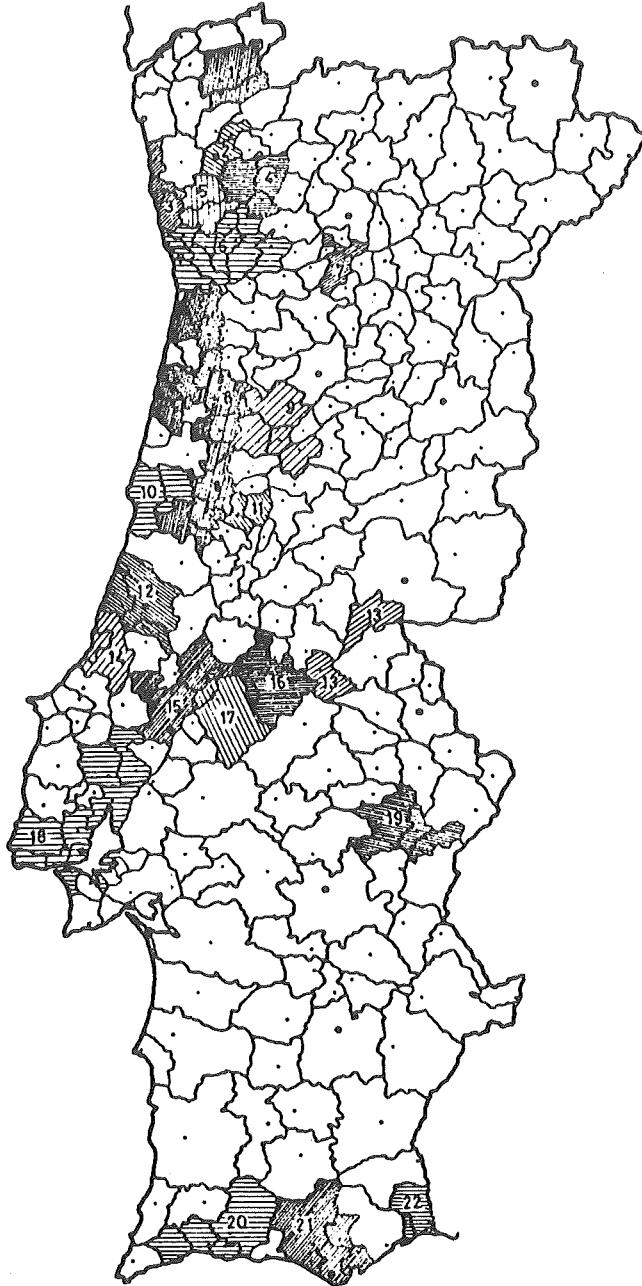


Fig. 4 - 4th Generation Clusters

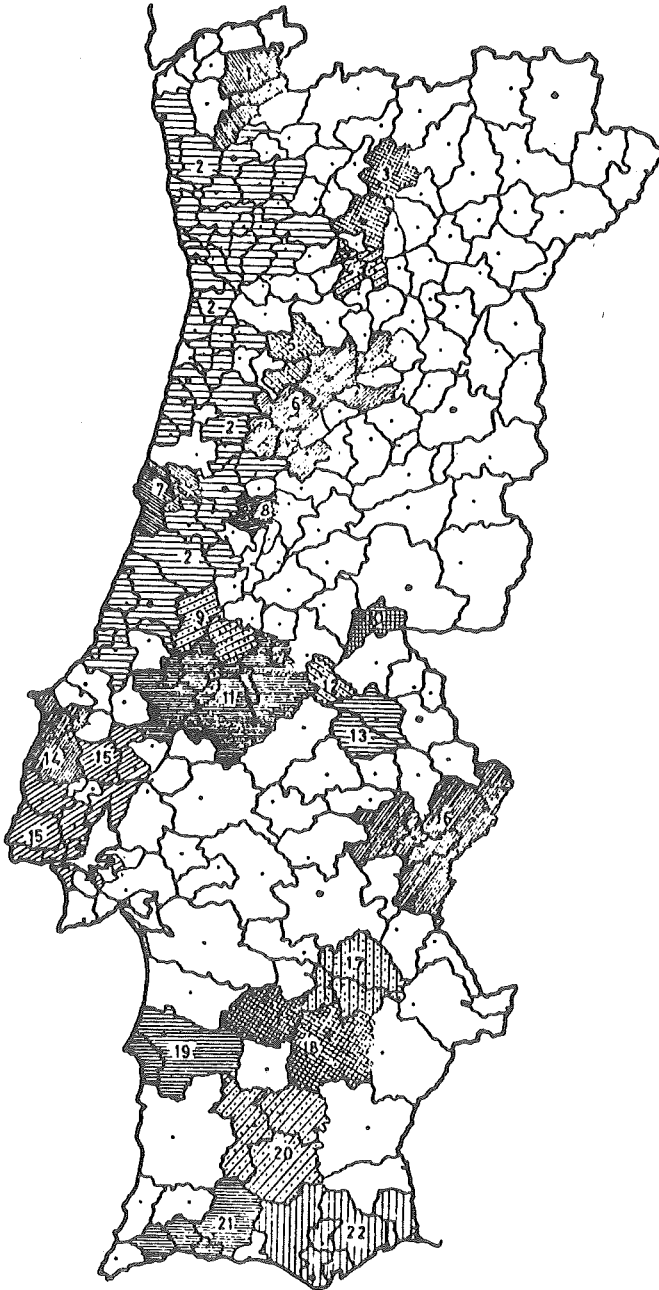


Fig. 5 - 5th Generation Clusters

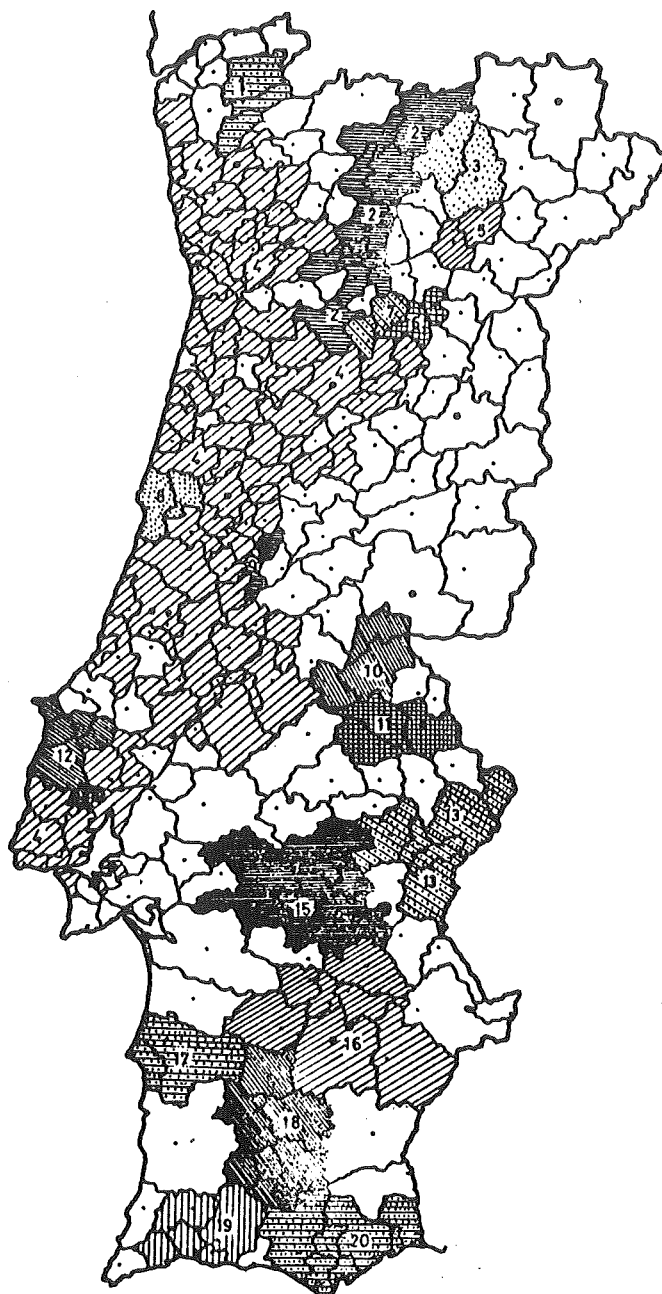


Fig. 6 - 6th-Generation Clusters

The successive generations show how strong is a link between two *concelhos* because the stronger the connection between a pair of individuals, the lower the generations where they enter the same cluster. Obviously, the obtained clusters in any generations are not necessarily FURs because they can also be fractions or sets of functional urban regions. The convenient generation to be used as a guide for the delimitation of FURs is selected according to the number and location of UCCs.

A first proposal of functional urban regions is presented in fig. 7 using the clusters of 4th and 5th generations.

The FURs of Viana do Castelo, Porto, Braga, Guimarães, Aveiro-Coimbra and Leiria are associated in a single 5th order cluster. Porto is more closely connected with the eastern neighbour *concelhos* than in the north coast or the southern district and the functional urban regions of Porto, Braga and Guimarães produce a continuous compact and inter-related urban region. Aveiro is strongly linked to Coimbra (they belong to the same cluster from the 4th generation) and therefore a single FUR is defined.

The connections of Coimbra with the south are not so evident and the southern limit of this FUR is the border between Soure and Pombal (4th generation) or Alcobaça and Caldas (5th generation).

Lisbon FUR is not linked to the south and the 4th generation FURs of the UCC detected through the 3rd group of *concelhos* (fig.2) contain only one *concelho* (excepting Santa-rém and Aveiro).

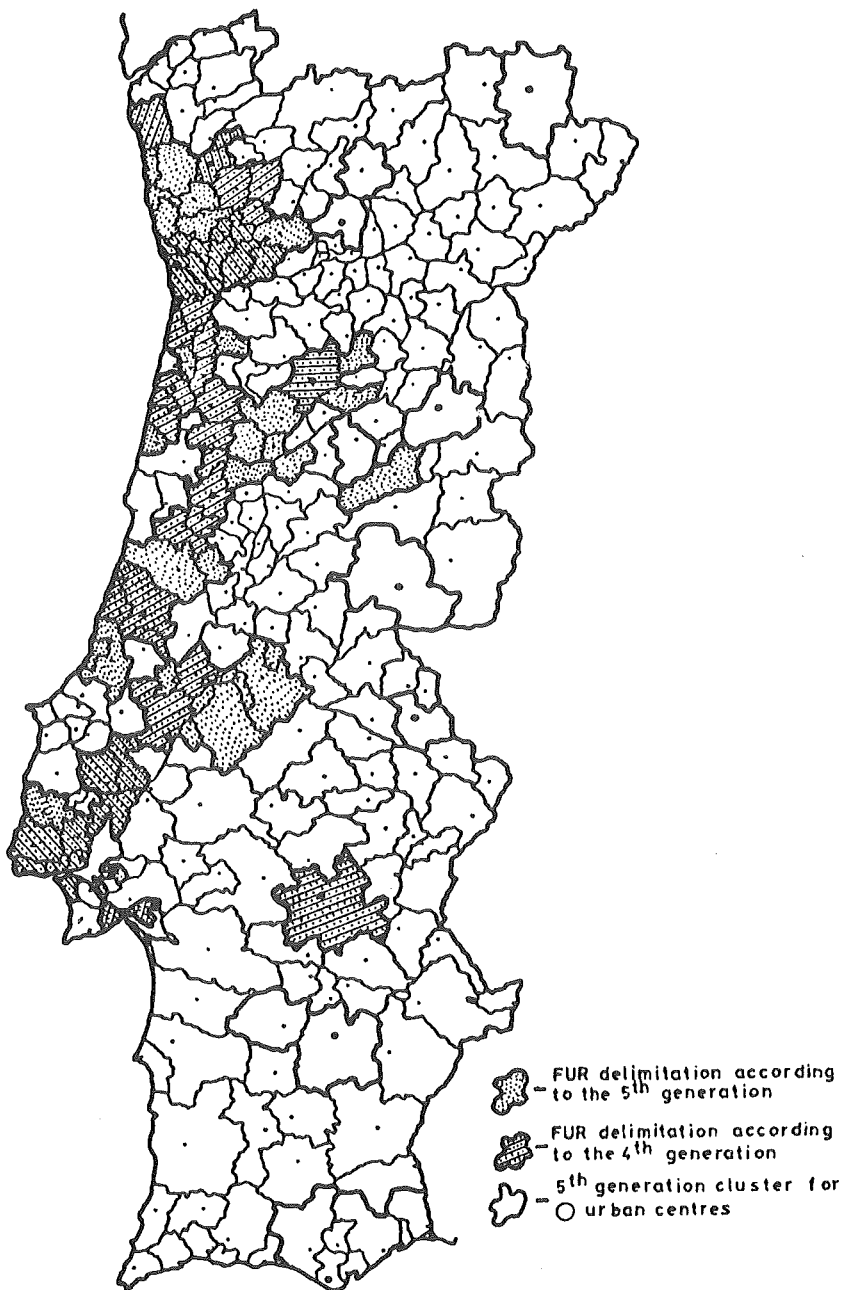


Fig. 7 - A first delimitation of major FUR's

The interaction between *concelhos* belonging to inner lands of the portuguese territory is much weaker and therefore they are clusterized at later generations (fig.6) becoming then evident the isolation of some units and the extension of rural areas.

The 5th generation clusters containing the major urban centres of districts without UCCs are presented in fig. 7 and only one (Faro) has a significant number of *concelhos* (Algarve eastern region).

All these results show that the 4th & 5th generation clusters are adequate to the FURs delimitation. Fourth level results may be preferred for the analysis of more interconnected urbanized regions (FURs of Viana do Castelo, Braga, Guimarães, Porto, Aveiro-Coimbra and Leiria) and the 5th generation clusters are particularly useful to define more isolated FURs (Viseu, Covilhã Santarém, Lisboa, Setúbal and Évora) and to identify eventual associations of functional urban regions (Viana do Castelo-Braga-Guimarães-Porto-Aveiro-Coimbra-Leiria).

Obviously, other distance indicators and tolerance levels (α) may be used in future studies to explore further potentialities of this methodology.

Finally, the delimitation of the functional urban cores using the *concelho* as spatial unit is done by including the *concelho* associated to each urban core *concelho* at the 4th generation if the travel time by bus (or suburban train) between their capital is lower than 60 minutes (fig.8).

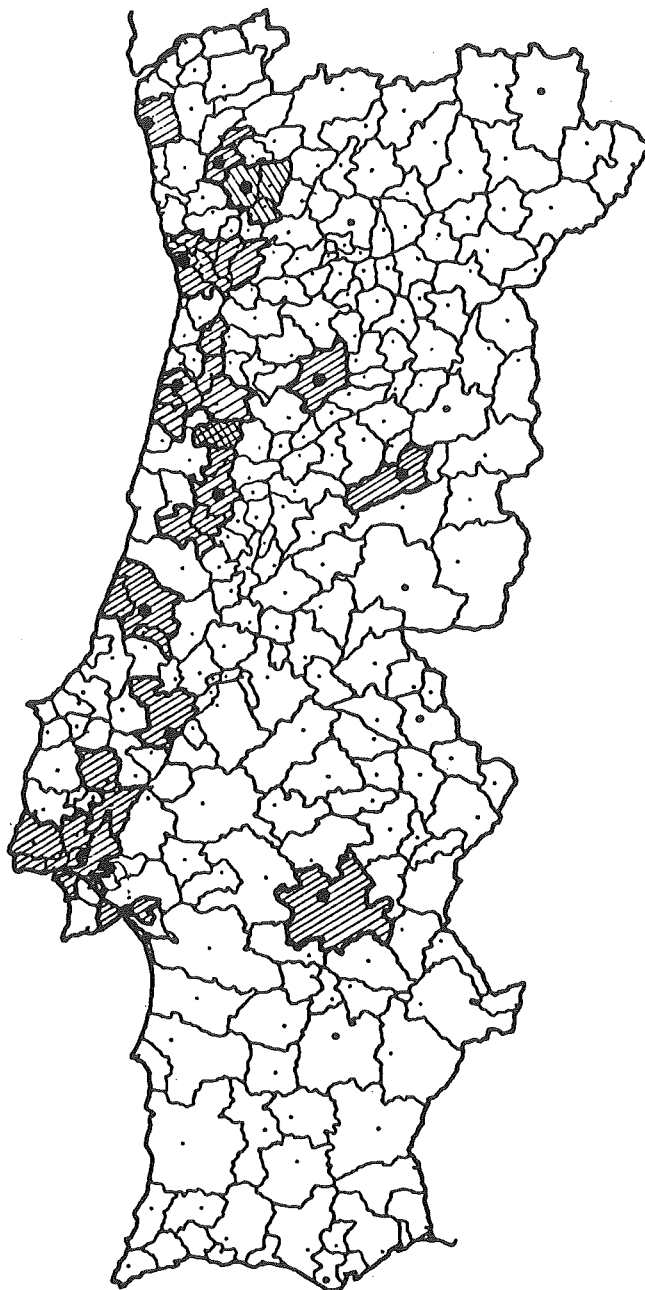


Fig. 8 - The Functional Urban Cores

3.2 TRAVELLING BY TRAIN

Figure 9 shows the Portuguese railway network and it can be observed that there are numerous *concelho* capitals that are not served by this network. Furthermore, there are also cases of railway stations that have the same name as the *concelho* capital when they are actually more than 10 km apart. Thus, the *concelhos* were classified according to the distance by road between their capitals and the nearest railway station (table 1).

| <i>concelho</i> classification | Distance between the <i>concelho</i> capital and the nearest railway station (Km) |
|-----------------------------------|---|
| 1st class | ≤ 3 |
| 2nd class | between 4 and 6 |
| 3rd class | between 7 and 9 |
| 4th class | > 9 |

For each pair of *concelhos* an expected time (E_{ij}) spent to go by train from a *concelho* i to any other j was defined as

$$E_{ij} = T_{ij} + P W_{ij} + (1-P) W_{ij}^* + Q_{ij}$$

where

T_{ij} - travel time between the railway station that serves the *concelho* i and the station serving j ;

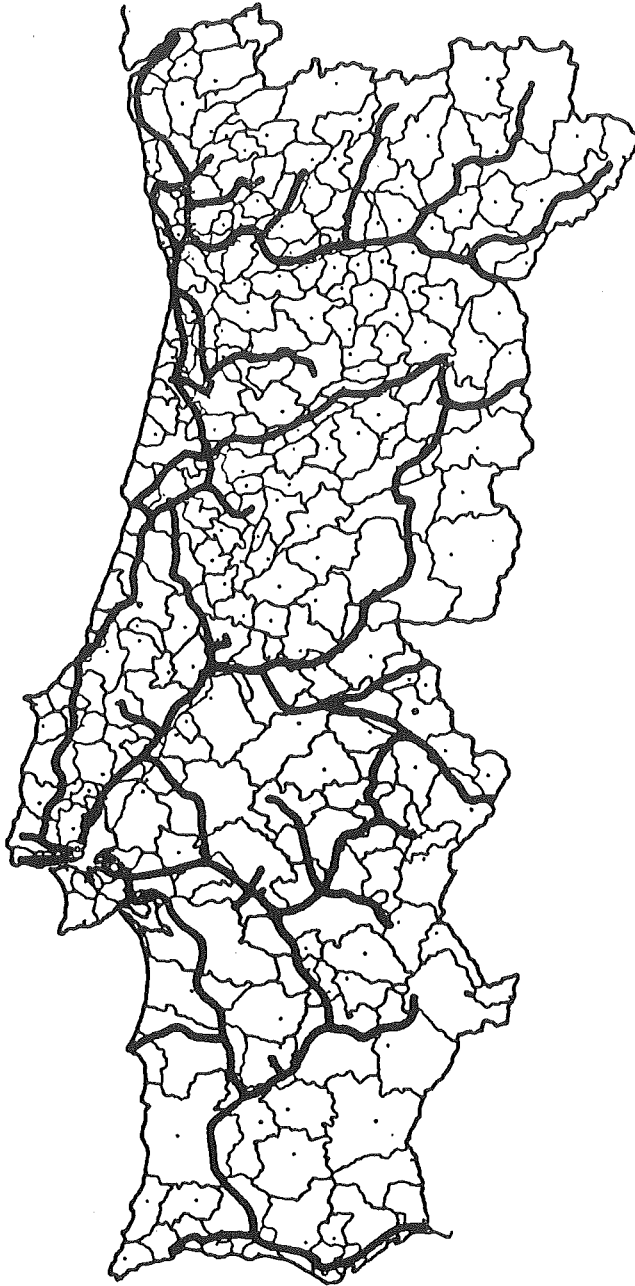


Fig. 9 - Portuguese Railway Network

Q_{ij} - additional time required to go from the *concelho* capital i to the nearest railway station and to go from the station serving *concelho* j to its capital. This additional time is computed according to the *concelhos* classification presented above;

and all the remaining symbols have the same meaning as in expression (1).

The taxonomic method described in the previous paragraph was then applied to the distance matrix E_{ij} . The generated clusters (3rd and 4th generation) are presented in figure 12.

Again, the FURs of Viana do Castelo, Porto, Braga, Aveiro, Guimarães and Coimbra are associated in a single 4th order cluster. However, Porto is now more closely connected with the southern neighbour *concelhos* and, as before, the FURs of Porto, Braga and Guimarães produce a continuous compact and inter-related urban region.

Again, Aveiro is strongly linked to Coimbra and therefore a single FUR is defined. Coimbra appears to be more closely connected to the northern neighbour *concelhos* (and now, to the eastern ones as well) than to the south.

The main difference, however, occurs for the FURs of Lisbon and Santarem which appear now more strongly linked and extending their influence farther east.

The inner lands *concelhos* are only clusterized at later generations and this result serves to confirm the isolation of these extensive rural areas and to illustrate the heterogeneous features of Portugal.

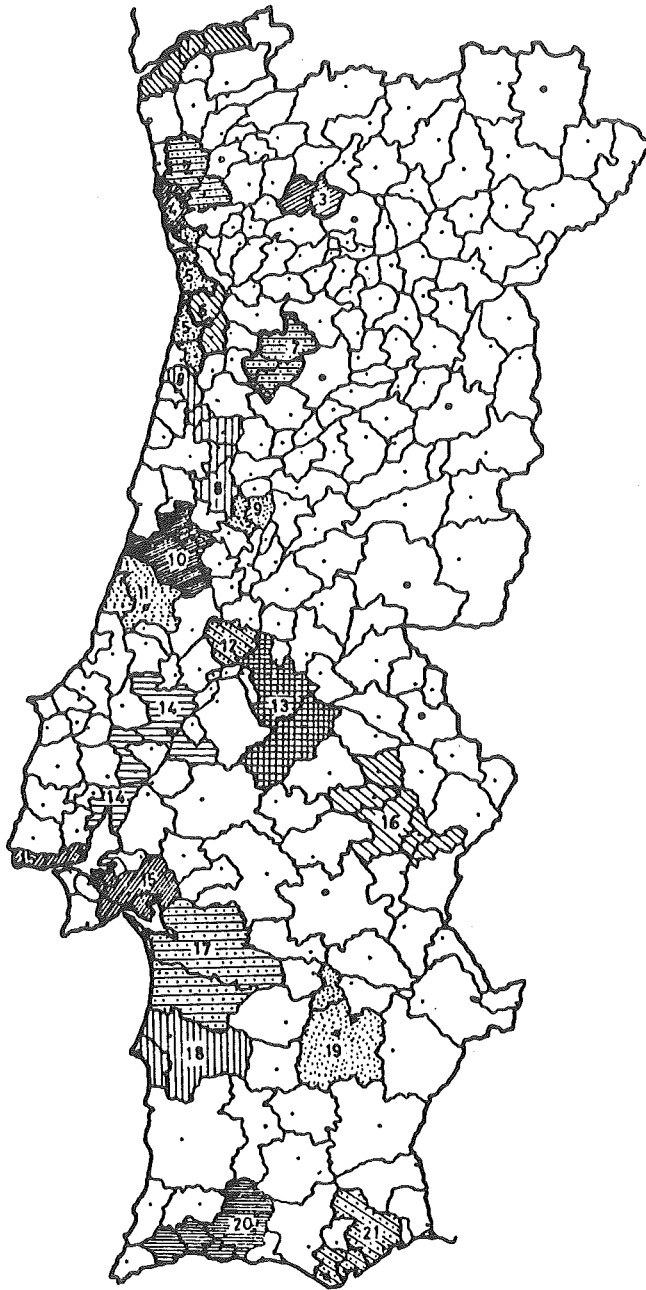


Fig. 10 - 3rd Generation Clusters (Railway Data)

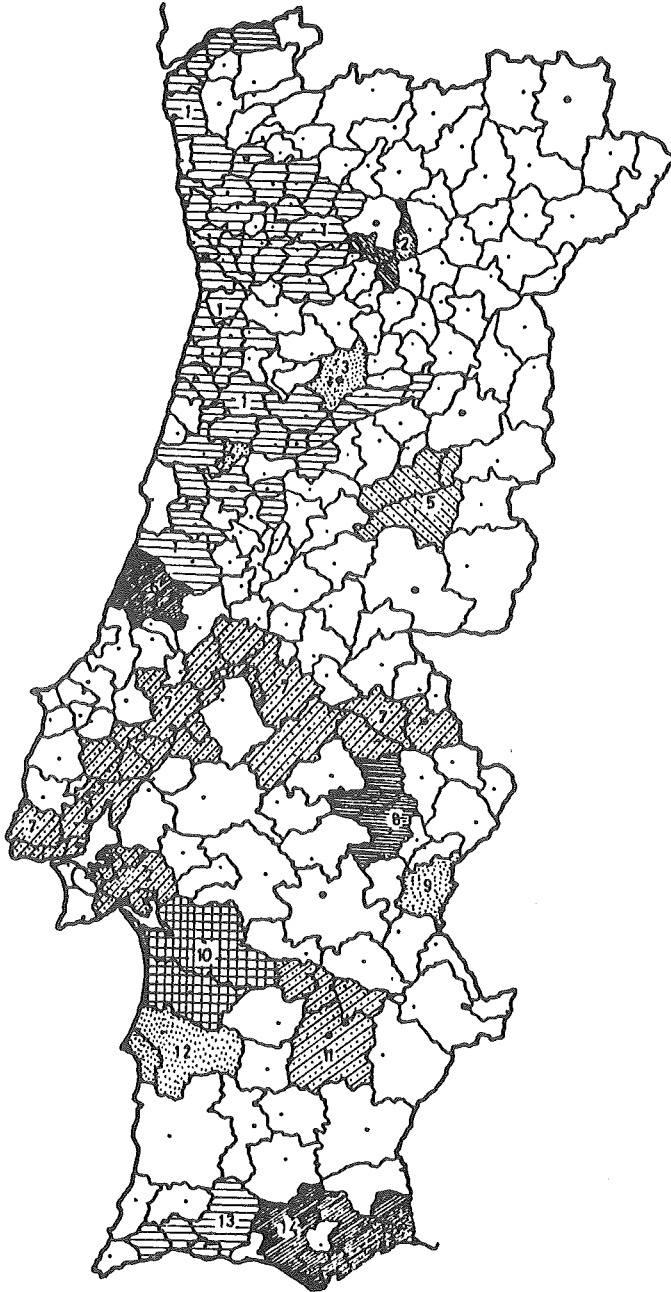


Fig. 11 - 4th Generation Clusters (Railway Data)

4. CONCLUSIONS

1. The lack of information on the spatial distribution of jobs for Portugal implies the adoption of alternative approaches to define the urban cores. With this purpose, a well succeeded application of an indirect indicator (non-agricultural working population) was carried out using recent unpublished statistical data.
2. Due to the non-availability of commuting data, indirect indicators of the inter-*concelhos* connections were defined : The expected waiting plus travel time by public transport modes between each pair of *concelhos*. The FURs were obtained using a taxonomic method which produced interesting generations of *concelhos* clusters reflecting their preferential connections.
3. The results obtained with the two indicators are similar for low order generations. For higher order generations, the significant differences can be explained by the high dependence of the results presented in paragraph 3.2 on the rigid structure and lay out of the railway network. Furthermore, the analysis just presented is based on the assumption that the demand for transportation (for which there is no data available) is related to the provision of transportation services (for which the data were collected). However, there is some evidence that this is not a valid assumption for some railway lines. This is the case for instance of international trains that represent a very significant fraction of total circulation on some tracks and on which most of their capacity is allocated to international traffic. This could explain the eastward extensions of the Coimbra and Santarém-Lisboa FURs observed in figure 12.

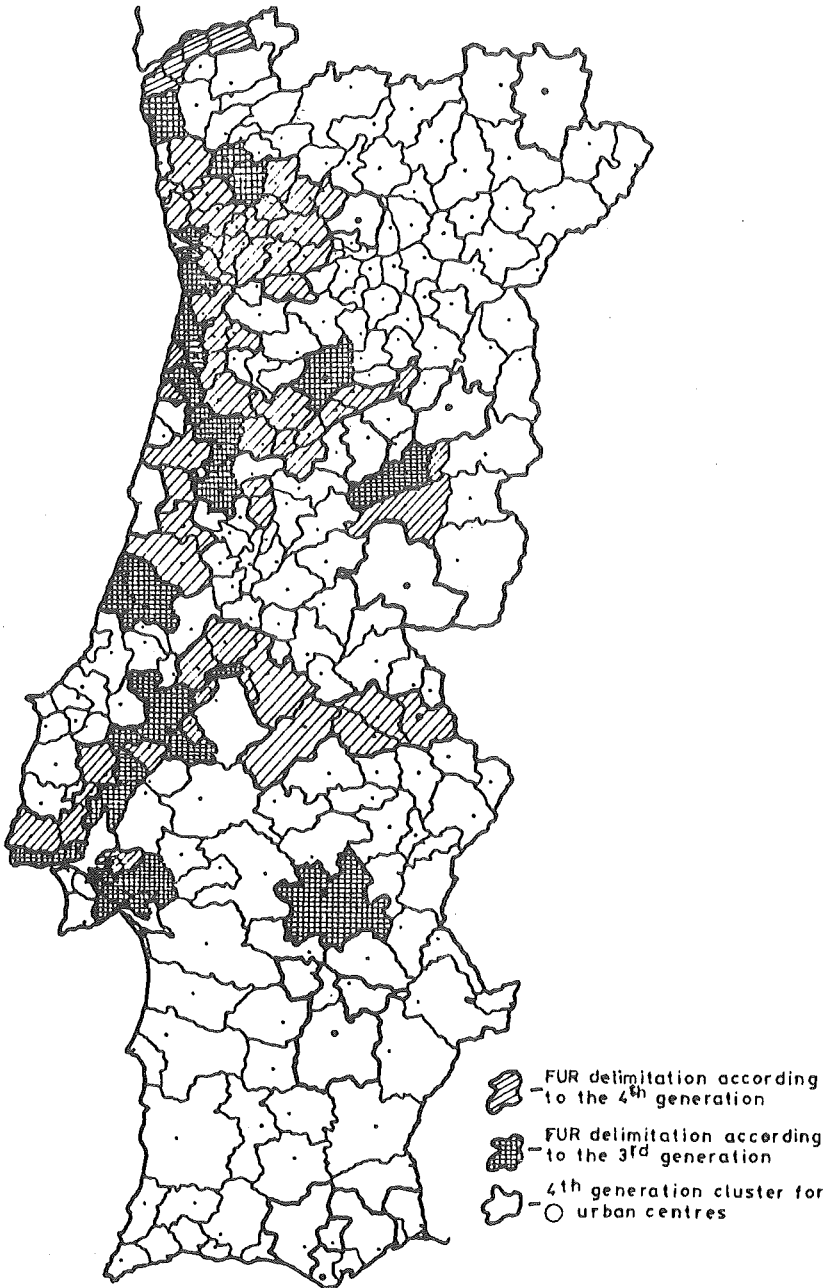


Fig. 12 - First delimitation of major FUR's

4. These results also illustrate the unbalanced pattern of the portuguese human settlement system. The connections are much stronger along the coast than along the coast-interior axis and this can be related to the inexistence of important urban centres in the innerlands. The isolation of the extensive rural areas of the interior is also well demonstrated and it can be concluded that these areas present a high degree of closure in terms of place of residence and workplace.

This analysis can also be useful for identifying the urban centres of the innerlands that present higher growth potentialities for their strategic location and more easily improvable accessibility.

5. The results just presented are hardly comparable with the FURs delineation proposed by Hall and Hay, 3, shown in figure 13.

These authors used a standardized framework previously applied to several northern european countries and justified on the grounds of a desirable uniformity of criteria so that international comparisons can be made. However, it may be questionable whether such a methodology is appropriate for a country like Portugal. Furthermore, Hall and Hay refer that they were constrained by data availability and recognize that very little evidence can be produced to support some of the results obtained except the data they have employed.

The application of a standardized procedure to a non-standard situation can produce sui generis results as

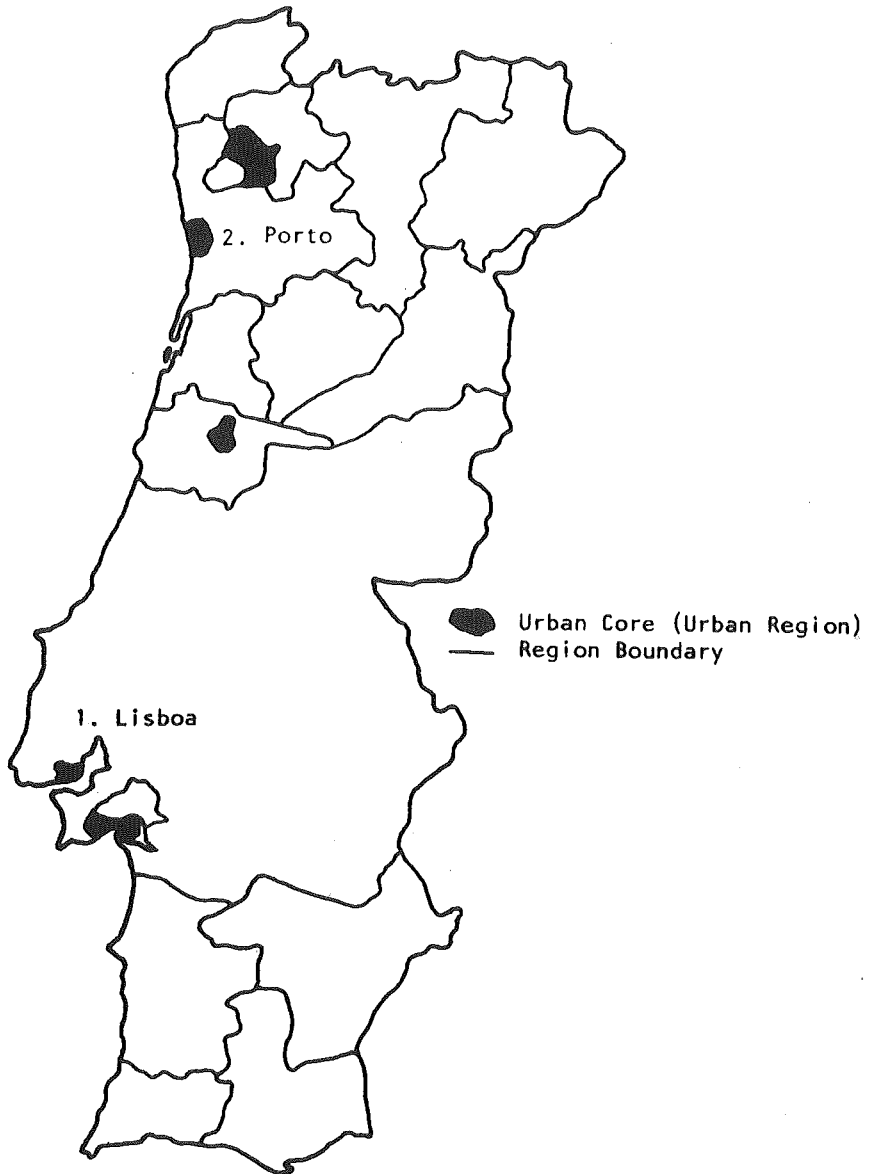


Fig. 13 - FUR's delimitation according to Hall & Hay

those obtained by Hall and Hay for Portugal. This serves to illustrate the argument that methodologies of analysis appropriate for developed countries might have to be adapted when it comes to analyze developing countries like Portugal. The authors believe that the procedure now presented may be a valid contribution in these situations.

5. REFERENCES

1. BERRY, B., GROWTH CENTRES IN THE AMERICAN URBAN SYSTEM, I, Ballinger Publishing Co, Cambridge, Mass (1978).
2. HAY, D. G., P. HALL and C. YOUNG, URBAN REGIONALIZATION OF GREAT BRITAIN, University of Reading, International Urban Systems, Working Paper 1 (1976).
3. HAY, D. G. AND P. HALL, URBAN REGIONALIZATION OF PORTUGAL University of Reading International Urban Systems, Working Paper 6.
4. HALTIER, S. and M. ENACHE, FUNCTIONAL URBAN REGIONS IN ROMANIA, paper presented at the Conference on the Analysis of Human Settlement Systems, Luxembourg, Austria (1978).
5. SERRIL, K., FUNCTIONAL URBAN REGIONS AND CENTRAL PLACE REGIONS IN THE FEDERAL REPUBLIC OF GERMANY AND SWITZERLAND, IIASA, Research Memorandum RM-77-17 (1977).
6. TAVARES, L. V. and J. A. A. FERREIRA, REGIONAL NUMERICAL TAXONOMY : PERSPECTIVES AND RESULTS, CESUR, Lisboa (1977).
7. TAVARES, L. V. and J. A. A. FERREIRA, A CLUSTERING APPROACH TO THE DEFINITION OF FUNCTIONAL URBAN REGIONS IN PORTUGAL, CESUR, Lisboa (1978).

NOTA DO EDITOR

É do conhecimento geral estar eminente a conclusão do longo processo de adesão de Portugal à CEE. Ora, no âmbito dos programas que agora se abrem a Portugal surge com especial importância o FEDER que visa apoiar o desenvolvimento regional.

O aproveitamento das potencialidades deste programa implica actuações rápidas com horizonte de apenas alguns meses, tendo por objectivo a definição das unidades regionais, o quadro das políticas de desenvolvimento e os respectivos programas a cinco anos.

Deste modo, considerou-se particularmente oportuna a publicação de um estudo sintético sobre as unidades regionais do nosso país, já apresentado internacionalmente mas pouco divulgado entre nós.

those obtained by Hall and Hay for Portugal. This serves to illustrate the argument that methodologies of analysis appropriate for developed countries might have to be adapted when it comes to analyze developing countries like Portugal. The authors believe that the procedure now presented may be a valid contribution in these situations.

5. REFERENCES

1. BERRY, B., GROWTH CENTRES IN THE AMERICAN URBAN SYSTEM, I, Ballinger Publishing Co, Cambridge, Mass (1978).
2. HAY, D. G., P. HALL and C. YOUNG, URBAN REGIONALIZATION OF GREAT BRITAIN, University of Reading, International Urban Systems, Working Paper 1 (1976).
3. HAY, D. G. AND P. HALL, URBAN REGIONALIZATION OF PORTUGAL University of Reading International Urban Systems, Working Paper 6.
4. HALTIER, S. and M. ENACHE, FUNCTIONAL URBAN REGIONS IN ROMANIA, paper presented at the Conference on the Analysis of Human Settlement Systems, Luxembourg, Austria (1978).
5. SERRIL, K., FUNCTIONAL URBAN REGIONS AND CENTRAL PLACE REGIONS IN THE FEDERAL REPUBLIC OF GERMANY AND SWITZERLAND, IIASA, Research Memorandum RM-77-17 (1977).
6. TAVARES, L. V. and J. A. A. FERREIRA, REGIONAL NUMERICAL TAXONOMY : PERSPECTIVES AND RESULTS, CESUR, Lisboa (1977).
7. TAVARES, L. V. and J. A. A. FERREIRA, A CLUSTERING APPROACH TO THE DEFINITION OF FUNCTIONAL URBAN REGIONS IN PORTUGAL, CESUR, Lisboa (1978).

NOTA DO EDITOR

É do conhecimento geral estar eminente a conclusão do longo processo de adesão de Portugal à CEE. Ora, no âmbito dos programas que agora se abrem a Portugal surge com especial importância o FEDER que visa apoiar o desenvolvimento regional.

O aproveitamento das potencialidades deste programa implica actuações rápidas com horizonte de apenas alguns meses, tendo por objectivo a definição das unidades regionais, o quadro das políticas de desenvolvimento e os respectivos programas a cinco anos.

Deste modo, considerou-se particularmente oportuna a publicação de um estudo sintético sobre as unidades regionais do nosso país, já apresentado internacionalmente mas pouco divulgado entre nós.

IDENTIFICAÇÃO DE MODELOS ARIMA(p,d,q) ATRAVÉS DA
FUNÇÃO DE AUTOCORRECÇÃO EXTENDIDA DE TIAO & TSAY

Reinaldo Castro Souza, Grupo de Sistemas, DEE
PUC, Rio de Janeiro, Brasil

É apresentado o conceito de regressão AR(p) iterativa, a partir do qual podem ser obtidos estimadores consistentes da parte autoregressiva de modelos ARMA pelo método dos mínimos quadrados ordinários. A função de autocorrecção extendida a partir da regressão AR(p) iterativa é então obtida e seu estimador facilmente calculado.

Tal função permite a identificação directa dos parâmetros "p" e "q" dos modelos ARMA (p,q), apresentando-se como alternativa ao procedimento tradicional de reconhecimento de padrões das funções de autocorrecção e autocorrelação parcial proposto por Box & Jenkins.

INTERPRETAÇÃO EMPÍRICA DO DUAL DE UM PROGRAMA INTEIRO MISTO

MOSS MADDEN, Dept. of Civic Design, Univ. of Liverpool, UK

PAULO V. D. CORREIA, Secção de Urbanização e Sistemas, Dep.
de Eng. Civil, Instituto Superior Técnico, Lisboa, Portugal

O artigo apresenta as técnicas de programação matemática utilizadas na formulação do dual de um programa inteiro misto e sua interpretação.

Os autores desenvolvem a abordagem proposta por Balas (1970), e através da formulação de um problema simples sobre aquisição de solo por uma Autarquia Local, ilustrado por uma aplicação numérica, revelam como opera o sistema generalizado de preços-sombra e variáveis de subsídio/penalidade.

A TAXOMIC DELIMITATION OF FUNCTIONAL URBAN REGIONS IN A HETEROGENEOUS COUNTRY

L.VALADARES TAVARES, J.A.ANTUNES FERREIRA and
R. CARVALHO OLIVEIRA

CESUR, Instituto Superior Técnico, Lisboa, Portugal

A identificação da interdependência entre centros urbanos e zonas territoriais envolventes é uma questão crucial nos processos de planeamento urbano e definição de políticas neste domínio.

Embora não exista um critério único para definir regiões funcionais urbanas, um processo standard baseado em informações de emprego e migrações pendulares foi aplicado no processo de regionalização de vários países europeus. No entanto, em Portugal não estão disponíveis dados fiáveis sobre estas variáveis, tendo-se por isso recorrido a um indicador indirecto das relações interconcelhias, definido para o efeito: o tempo esperado de viagem em transportes públicos.

Um método de taxonomia numérica foi então aplicado de modo a gerar "clusters" que pudessem ser usados para a definição de áreas funcionais urbanas.

IDENTIFICATION OF ARIMA(p,d,q) MODELS THROUGH THE
EXTENDED AUTOCORRECTION FUNCTION OF TIAO & TSAY

Reinaldo Castro Souza, Grupo de Sistemas, DEE
PUC, Rio de Janeiro, Brasil

The concept of iterative AR(p) regression is introduced, from which consistent estimates of the autoregressive part of the ARMA models can be obtained through the standard least-squares method. The extended auto-correction function is then obtained from this iterative AR(p) regression, and its estimate easily evaluated.

Such a function allows the direct identification of the "p" and "q" parameters in the ARMA (p,q) models, thus being an alternative to the traditional procedure of recognition of patterns of the partial autocorrection and autocorrelation functions proposed by Box & Jenkins.

INTERPRETAÇÃO EMPÍRICA DO DUAL DE UM PROGRAMA INTEIRO MISTO

MOSS MADDEN, Dept. of Civic Design, Univ. of Liverpool, UK

PAULO V. D. CORREIA, Dept. of Civil Engineering, Instituto Superior Técnico, Lisboa, Portugal

The paper presents the mathematical programming techniques used to set up the dual of a mixed-integer programme and its interpretation.

The authors develop the approach proposed by Balas (1970), and through the formulation of a simple problem about land purchasing by a municipal authority, illustrated by an empirical example, provide an insight to the generalised system of shadow-prices and subsidy/penalty variables.

A TAXONOMIC DELIMITATION OF FUNCTIONAL URBAN REGIONS IN A HETEROGENEOUS COUNTRY

L.VALADARES TAVARES, J.A.ANTUNES FERREIRA and
R. CARVALHO OLIVEIRA

CESUR, Instituto Superior Técnico, Lisboa, Portugal

The identification of the interdependence between the urban centres and the surrounding hinterlands is a crucial issue for the processes of urban planning and policy making.

Though a unique criterion to define functional urban regions does not exist, a standardized procedure based on employment and commuting data has been applied in the regionalization of several european countries. However, reliable data on employment and commuting flows are not available in Portugal and indirect indicators of the interdistricts relationships had to be defined : the expected travel time by public transport modes.

A numerical taxonomy method was then applied to generate clusters that can be used to define functional urban regions.

Índice dos artigos do Volume 4

| | |
|--|-----|
| B.C. Vasconcelos | |
| Utilidade em Modelos de Previsão | 6 |
| | |
| W.H. Swann | |
| The Management of O.R. Projects | 25 |
| | |
| A.S.P. Aor & L.F.A. Monteiro Gomes | |
| Análise Probabilística de um sistema de estacionamento | 46 |
| | |
| J.K.Lenstra & A.H.G. Rinnooy Kan | |
| Computational Complexity of Discrete Opti- mization Problems | 66 |
| | |
| R. Castro Souza | |
| Identificação de Modelos ARIMA(p,d,q) através da Função de Autocorreção Extendida de Tiao & Tsay | 111 |
| | |
| M. Madden & P.V.D. Correia | |
| Interpretação Empírica do Dual de um pro- grama inteiro misto | 131 |
| | |
| L. Valadares Tavares, J.A. Antunes Ferreira & R. Carvalho Oliveira | |
| A Taxonomic Delimitation of Functional Urban Regions in a Heterogeneous Country | 149 |

