FEP WORKING PAPERS
FEP WORKING PAPERS

Research Work in Progress

n. 262, Jan. 2008

Segmentação de Mercado e modelos mistura de regressão para variáveis normais

Ana Oliveira-Brochado*
Francisco Vitorino Martins**

* EDGE, CESUR, DECIVIL-IST, Universidade Técnica de Lisboa
**EDGE, Faculdade de Economia da Universidade do Porto

U.PORTO
FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO

SEGMENTAÇÃO DE MERCADO E MODELOS MISTURA DE REGRESSÃO PARA VARIÁVEIS NORMAIS

Ana Oliveira-Brochado*; Francisco Vitorino Martins**

*Centre for Urban and Regional Systems, DECivil-IST, Technical University of Lisbon

Av. Rovisco Pais, 1049-001 Lisbon, Portugal, abrochado@civil.ist.utl.pt

** Faculty of Economics, University of Porto - Rua Dr. Roberto Frias, 4200-464 Porto, Portugal,

vmartins@fep.up.pt

RESUMO: Este trabalho propõe-se apresentar uma revisão dos modelos mistura de regressão para variáveis normais, frequentes em estudos de pesquisa de mercados. De facto, exemplos de aplicação destes modelos continuam a acumular na literatura do marketing, dadas as suas vantagens relativas. Adicionalmente, estes modelos são facilmente implementados, devido à sua incorporação em vários programas comerciais de pesquisa de mercados. Pretende-se apresentar as raízes históricas que motivaram o desenvolvimento dos modelos mistura de regressão (regressions, *clusterwise* regression e modelos mistura finitos) e rever a formulação do modelo base, bem como as suas extensões principais aos cenários de dados em painel e estudos de análise conjunta.

Palavras-chave: segmentação de mercado, modelos mistura de regressão, dados normais.

ABSTRACT: The purpose of this work is to provide an overview of what is perhaps the most common analysis context in market research – that of regression models for normally distributed data. In fact, examples of applications of these models continue to accumulate in the marketing literature, given their relative advantages. Moreover, these models are ease implemented due to its incorporation in many commercial packages of marketing research. We aim at presenting the background for the development of mixture regression models (switching regressions, *clusterwise* regression and finite mixture models) and review the formulation of the basic model and its main extensions in the context of panel data analysis and conjoint studies.

Keywords: market segmentation; mixture regression models; normal data.

1 INTRODUÇÃO

Exemplos de aplicações de modelos mistura de regressão para variáveis normais na segmentação de mercado têm acumulado na literatura do marketing (Andrews *et al.*, 2002; DeSarbo e Cron, 1988; DeSarbo *et al.*, 1992; Helsen *et al.*, 1993; Ramaswamy, *et al.*, 1993; Vriens *et al.*, 1996; Wedel e DeSarbo, 1994,1995; DeSarbo *et al.*, 2001; Andrews e Currim, 2003; Jedidi *et al.,* 1996; Bowman *et al.*, 2004). O presente artigo pretende (*i*) identificar os principais contributos para que inspiraram o desenvolvimento dos modelos mistura de regressão para variáveis normais (ponto 2); (*ii*) apresentar o modelo mistura de regressão desenvolvido para variáveis dependentes normais (DeSarbo e Cron, 1988; Wedel e DeSarbo, 1995) e as estratégias utilizadas na estimação dos seus parâmetros (ponto 3); (*iii*) sintetizar as extensões ao modelo base para um contexto multivaridado propostas por Ramaswamy *et al.* (1993) e por DeSarbo *et al.* (1992) (ponto 4).

2 CONTRIBUTOS PARA O DESENVOLVIMENTO DOS MODELOS MISTURA DE REGRESSÃO

De seguida, com base na revisão da literatura efectuada, propõe-se uma síntese dos principais contributos para o desenvolvimento destes modelos mistura de regressão para variáveis normais e apresentam-se as principais formalizações divulgadas na literatura do marketing. Um trabalho central no contexto dos modelos mistura de regressão é o artigo de DeSarbo e Cron (1988), que desenvolveu um modelo mistura de regressão para uma variável dependente normal. Após uma revisão da literatura da análise classificatória e da análise de regressão, identificam-se três grandes contributos para o desenvolvimento do modelo proposto por DeSarbo e Cron (1988), que, em termos latos, coincidem com os principais factores que inspiraram o desenvolvimento dos modelos mistura de regressão. Refiram-se:

a) As *switching regressions,* analisadas nos trabalhos de Quandt (1972), Hosmer (1974), Quandt e Ramsey (1978); o modelo de DeSarbo e Cron (1988) é uma extensão das *switching regressions* para mais de dois regimes;
b) A *clusterwise regression* (Späth, 1979, 1981, 1982, 1985); os modelos mistura de regressão efectuam a extensão do conceito de *clusterwise regression* para um contexto estocástico permitindo a possibilidade não só de partições mutuamente exclusivas, como também de grupos com pertença probabilística;

c) Os modelos mistura finitos, ou '*unconditional mixture approaches'* (Wolfe, 1965, 1967, 1970; Day, 1969, Ganesalingam e McLachlan, 1981; McLachlan, 1982; Sclove, 1977; Symons, 1981; Scott e Symons, 1971; Marriott, 1975; Hartigan, 1975; Basford e McLachlan, 1985).

2.1 Switching Regression Models

Wedel e DeSarbo (1994) consideram que o desenvolvimento dos modelos mistura de regressão começou com o trabalho de Quandt (1972), ao introduzir os '*switching regression models*'; estes modelos foram estendidos posteriormente por Hosmer (1974) e por Quandt e Ramsey (1978), que propuseram novos desenvolvimentos no processo de estimação. O problema base endereçado aos ´*switching regimes*' na análise de regressão consiste, segundo Quandt (1972), na estimação de duas equações de regressão para a mesma amostra, caso se conclua que as observações foram geradas por dois conjuntos de parâmetros, sem se conhecer à partida nem a constituição, nem a dimensão, de cada um dos dois grupos. Este problema pode ser formulado da seguinte forma:

a) Teste da hipótese nula (H_0) de inexistência de alterações de regime (i.e. uma equação de regressão é suficiente), contra a hipótese alternativa (H_1) de que as observações são geradas por duas (ou possivelmente mais) regressões distintas;
b) Estimação dos parâmetros que caracterizam dois (ou mais) regimes, no caso da rejeição de (H_0).

Dadas N observações para a variável dependente y_n, $y = 1,...,N$ e para as P variáveis independentes, a hipótese nula (H_0) pode ser formulada através do modelo de regressão linear múltipla (1).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (1)$$

em que:

$\mathbf{y} = (y_n)$ – vector de dimensão $(Nx1)$ das observações da variável dependente

$\mathbf{X} = ((x_{np}))$ – matriz de dimensão (NxP) das N observações $(n = 1,...,N)$ para as P variáveis independentes $(p = 1,...,P)$

$\boldsymbol{\beta} = (\beta_p)$ - vector de dimensão $(Px1)$ dos coeficientes da regressão a serem estimados

$\boldsymbol{\varepsilon} = (\varepsilon_n)$ - vector de dimensão $(Nx1)$ de termos aleatórios não observados distribuídos como $N(\mathbf{0}, \sigma^2\mathbf{I})$

A hipótese alternativa (*H₁*) é expressa considerando que é possível reorganizar as linhas de $\mathbf{y}$ e nas colunas de $\mathbf{X}$, de tal forma em que estas sejam divididas em dois grupos (2), dando origem às formulações (3) e (4).

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \tag{2}$$

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \tag{3}$$

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2 \tag{4}$$

$\boldsymbol{\varepsilon}_1$ e $\boldsymbol{\varepsilon}_2$ distribuem-se, respectivamente, como $N(\mathbf{0}, \sigma_1^2\mathbf{I})$ e $N(\mathbf{0}, \sigma_2^2\mathbf{I})$. Uma formulação mais simples para a hipótese alternativa, que motivou o conhecido teste de Chow (1960), considera que as observações correspondentes a (3) e (4) são identificadas à priori, correspondendo a uma divisão da amostra em dois grupos. Configurações mais complexas são desenvolvidas com base no pressuposto de desconhecimento a partir de qual dos dois regimes (1 ou 2), cada uma das observações é gerada.

Quandt (1972) assume que há uma probabilidade desconhecida λ da natureza escolher o regime 1 e uma probabilidade $(1-\lambda)$ de escolher o regime 2 na geração das observações. Pressupondo que os termos aleatórios nos dois regimes são *i.i.d.* e seguem uma distribuição normal, a densidade condicional do valor da observação *n,* da variável dependente ***y*** (y_n), condicional nos valores das P variáveis independentes $\left(x_{n1},...,x_{np}\right)$ é dada pela expressão (5):

$$f\left(y_n \middle| x_{n1},...,x_{np}\right)=\frac{\lambda}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{1}{2\sigma_1^2}\left(y_n-\sum_{p=1}^{P}\beta_{1p}x_{np}\right)^2\right] \\ +\frac{1-\lambda}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{1}{2\sigma_2^2}\left(y_n-\sum_{p=1}^{P}\beta_{2p}x_{np}\right)^2\right] \tag{5}$$

em que $\boldsymbol{\beta}_1=\left(\beta_{1p}\right)$ e $\boldsymbol{\beta}_2=\left(\beta_{2p}\right)$ são os vectores dos coeficientes de regressão no regime 1 e no regime 2, respectivamente.

A função de verosimilhança logaritmizada (6) é obtida pelo cálculo do somatório do logaritmo de (5) para todas as *N* observações, $\log L=\sum_{n=1}^{N}\log f\left(y_n \middle| x_{n1},...,x_{np}\right)$:

$$\log L=\sum_{n=1}^{N}\log\left\{\begin{array}{l}\frac{\lambda}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{1}{2\sigma_1^2}\left(y_n-\sum_{p=1}^{P}\beta_{1p}x_{np}\right)^2\right] \\ +\frac{1-\lambda}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{1}{2\sigma_2^2}\left(y_n-\sum_{p=1}^{P}\beta_{2p}x_{np}\right)^2\right]\end{array}\right\} \tag{6}$$

Quandt (1972) propõe a maximização da função (6) em relação a β_{1p}, β_{2p}, $\left(p=1,...,P\right)$, $\sigma_1,\sigma_2\geq 0$ e $0\leq\lambda\leq 1$, com base no gradiente conjugado de Powell (Powell, 1964) '*Powell's conjugate gradient algorithm*'.

Goldfeld e Quandt (1973) assumem que existe uma variável não observável z_n $\left(n=1,...,N\right)$, que pode ser usada para classificar as observações nos dois regimes: assume-se que y_n foi gerada pelo regime 1 ou 2, consoante $z_n\leq z_0$ ou $z_n>z_0$ (o nível de corte z_0 é desconhecido. De seguida apresenta-se o processo de estimação dos parâmetros de regressão e de z_0.

Seja $\mathbf{D}$ uma matriz diagonal de dimensão (NxN), com elementos $d(z_n)$, $n=1,...,N$ na diagonal, em que:

$$d(z_n)=\begin{cases}0, \text{ sse } z_n \leq z_0\\ 1, \text{ caso contrário}\end{cases}$$

Deste modo, supõe-se que as observações são geradas por (7)

$$\mathbf{y}=(\mathbf{I}-\mathbf{D})\mathbf{X}\boldsymbol{\beta}_1+\mathbf{D}\mathbf{X}\boldsymbol{\beta}_2+\mathbf{W} \tag{7}$$

em que $\mathbf{W}$ é o vector de termos aleatórios não observáveis e heteroescedásticos $\mathbf{W}=(\mathbf{I}-\mathbf{D})\boldsymbol{\varepsilon}_1+\mathbf{D}\boldsymbol{\varepsilon}_2$; $\mathbf{W}$, $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ e os elementos de $\mathbf{D}$ devem ser estimados. Trata-se de um problema de optimização combinatória, computacionalmente admissível substituindo as funções $d(z_n)$ em $\mathbf{D}$ pela aproximação contínua (8) (Goldfeld e Quandt, 1973):

$$d(z_n)=\int_{-\infty}^{z_n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{\xi-z_0}{\sigma}\right)^2\right]d\xi \tag{8}$$

Note-se que esta formulação introduz dois novos parâmetros, z_0 (estimativa para o *cut-off*) e σ (medida da discriminação entre os dois regimes).

Seja $\boldsymbol{\Omega}$ a matriz de variâncias e covariâncias de $\mathbf{W}$ (9):

$$\boldsymbol{\Omega}=(\mathbf{I}-\mathbf{D})^2\sigma_1^2+\mathbf{D}^2\sigma_2^2 \tag{9}$$

A função de verosimilhança logaritmizada é assim dada por (10):

$$\begin{aligned}\log L = \text{constante} - \frac{1}{2}\log|\boldsymbol{\Omega}| \\ -\frac{1}{2}\left\{\mathbf{y}-\left[(\mathbf{I}-\mathbf{D})\mathbf{X}\boldsymbol{\beta}_1-\mathbf{D}\mathbf{X}\boldsymbol{\beta}_2\right]'\boldsymbol{\Omega}^{-1}\left[\mathbf{Y}-(\mathbf{I}-\mathbf{D})\mathbf{X}\boldsymbol{\beta}_1-\mathbf{D}\mathbf{X}\boldsymbol{\beta}_2\right]\right\}\end{aligned} \tag{10}$$

Quandt (1972) aplica o modelo e algoritmo proposto num estudo de simulação (dados experimentais) e em dados relativos à procura e oferta no mercado imobiliário. Uma das limitações deste método é não permitir a identificação das observações que pertencem a cada regime (Quandt, 1972).

Quandt (1972) propõe ainda extensões ao modelo base, entre as quais:

a) Generalização da *switching regression* para mais de dois regimes; se se assumir que o número de regimes é *S*, com probabilidades de serem seleccionados pela natureza $\lambda_1, \lambda_2, ..., \lambda_S$, $\sum_{s=1}^{S} \lambda_s = 1$, e que a densidade condicional da variável y_n dado o valor das P variáveis explicativas para cada regime é $f_s\left(y_n \middle| x_{n1}, ..., x_{nP}\right)$, então a função densidade condicional correspondendo a (5) será dada por (11):

$$f\left(y_n \middle| x_{n1}, ..., x_{nP}\right) = \sum_{s=1}^{S} \lambda_s f_s\left(y_n \middle| x_{n1}, ..., x_{nP}\right) \tag{11}$$

a partir do qual poderá ser derivada uma função de verosimilhança equivalente a (6).

b) Combinação do método proposto por Goldfeld e Quandt (1973) e do método proposto por Quandt (1972). Se se assumir que a probabilidade da natureza seleccionar um dado regime depende de uma variável não observável z, então definindo-se $d\left(z_n\right)$ de forma equivalente a (8), a função densidade condicional (5) seria definida por (12):

$$\begin{aligned} f\left(y_n \middle| x_{n1}, ..., x_{nP}\right) = & \frac{d\left(z_n\right)}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2\sigma_1^2}\left(y_n - \sum_{p=1}^{P} \beta_{1p} x_{np}\right)^2\right] \\ & + \frac{1 - d\left(z_n\right)}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2\sigma_2^2}\left(y_n - \sum_{p=1}^{P} \beta_{2k} x_{np}\right)^2\right] \end{aligned} \tag{12}$$

c) Necessidade de definição de metodologias que permitam a identificação da existência de um ou dois regimes e o teste dessa hipótese; claro que numa abordagem imediata, numa formulação com apenas dois grupos, se o valor de λ estiver muito próximo de 0 ou de 1, é provável que uma regressão seja suficiente para o ajustamento dos dados.

A primeira extensão proposta, i.e., a generalização da '*switching regression*' a mais de dois regimes foi efectuada mais de uma década depois por DeSarbo e Cron (1988). "*In fact, this* *expression generalizes the Quandt (1972), Hosmer (1974) e Quandt e Ramsey (1978) stochastic regression models to more than two 'regimes'*" (DeSarbo e Cron 1988: 255).

Em relação ao segundo desenvolvimento proposto, interpretando-o em termos latos como um apelo ao desenvolvimento de novos processos de estimação, podem então ser referidos os contributos de: Goldfeld e Quandt (1973, 1976) e Cosslet e Lee (1985) que propuseram os '*hidden Markov switching regression models',* nos quais as pertenças das observações num dado regime são modeladas por um processo de Markov; Quandt e Ramsey (1978) que propõem uma abordagem de estimação com base no método dos momentos; o trabalho de DeSarbo e Cron (1988) em que a estimação dos parâmetros desconhecidos é efectuada com recurso ao algoritmo EM - *Expectation-Maximization* de Dempster *et al.* (1977).

No entanto, em relação à terceira extensão proposta (alínea c)), que refere a necessidade de metodologias para a identificação do número de regimes presente nos dados, ainda não surgiram soluções satisfatórias que acompanhassem os desenvolvimentos ocorridos ao nível dos modelos e processos de estimação. A terceira extensão proposta por Quandt (1972) pode ser reinterpretada, deste modo, como a problemática da selecção do número de componentes em modelos mistura de regressão, se se considerar a extensão do modelo base a mais de dois regimes (desenvolvimento proposto na alínea a))

Hamilton (1989, 1990, 1991) e Engel e Hamilton (1990) estenderam os *switching regression models* para modelos de séries temporais; estes modelos descrevem movimentos discretos em parâmetros autoregressivos, em que as mudanças são modelizadas através de '*hidden discrete-time series models'*. Enquanto que inicialmente a estimação estava limitada a pequenos sistemas, dada a complexidade computacional envolvida na maximização da função de verosimilhança (Hamilton, 1989), Hamilton (1990) propôs um algoritmo EM (Dempster, *et al.* 1977), que aliviou este problema. Num terceiro artigo, Hamilton (1991) demonstrou as vantagens da abordagem Quase-Bayesiana sobre a abordagem de maximização da verosimilhança para a estimação dos parâmetros. O modelo foi aplicado na análise de taxas de câmbio (Engel e Hamilton, 1990; Hamiltom, 1990).

Titterington *et al.* (1985) apresentam alguns exemplos da aplicação *das switching regressions* na economia.

A Tabela 1. apresenta uma síntese dos artigos sobre *switching regression models* revistos.

Tabela 1. Primeiros contributos em modelos mistura de regressão

REFERÊNCIA	APLICAÇÃO	MÉTODO DE ESTIMAÇÃO
Quandt (1972)	Mercado imobiliário; construção de habitações	MV, NR
Hosmer (1974)	-	-
Quandt e Ramsey (1978)	Previsão de salários	MD
Goldfeld e Quandt (1973,1976)	Mercado imobiliário; construção de habitações	MV, NR
Cosslett e Lee (1985)	Estabilidade de Carteis	MV, NR
Hamilton (1989)	Crescimento do PIB	MV, NR
Hamilton (1990)	-	MV, EM
Hamilton (1991)	Taxas de juro	QB, EM
Engel e Hamilton (1990)	Taxas de juro	MV, EM
Legenda: MV – máxima verosimilhança; NR – algoritmo *Newton-Raphson*; EM – algoritmo EM *Expectation Maximization*; MD - distância mínima; QB – *quasi-Bayes*		

2.2 Clusterwise Regression

O modelo mistura de regressão proposto por DeSarbo e Cron (1988) propõe-se generalizar a *clusterwise regression* inicialmente desenvolvida por Späth (1979, 1981, 1982, 1985), para um contexto de classificação difusa. "*The primary goal of this research is to extend the concept of clusterwise regression to a stochastic context alowing for the possibility of fuzzy clusters, as well as mutually exclusive partitions*" (DeSarbo e Cron 1988: 224). A *clusterwise regression* é um dos primeiros métodos de agrupamento não hierárquicos motivados pelo estudo da relação entre uma variável dependente e um conjunto de variáveis independentes, enquadrando-se, por isso, nas abordagens preditivas de segmentação. Os autores propõem solucionar o problema combinatório de determinação de uma partição $P_1,...,P_S$ (não sobreposta) das *N* observações e a estimação dos *S* vectores de parâmetros $(\beta_{s1},...,\beta_{sP})$ associados ao modelo (13):

$$\mathbf{y}_s = \mathbf{x}_s\boldsymbol{\beta}_s + \boldsymbol{\varepsilon}_s,\ \boldsymbol{\varepsilon}_s \approx N\left(0,\sigma^2\mathbf{I}\right),\ \mathrm{s}=1,...,\mathrm{S} \tag{13}$$

Tal é efectuado minimizando a função $\sum_{s=1}^{S} E(C_s)'$, em que $E(C_s)$ é dado por:

$$E(C_s) = \left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right)\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right)'. \tag{14}$$

Späth (1979, 1981, 1982, 1985) desenvolveu um algoritmo de troca '*exchange algorithm*' para obter uma partição de S grupos e o correspondente conjunto de parâmetros[1]. O algoritmo desenvolvido é apresentado na Tabela 2.

É interessante notar que este método, desenvolvido no contexto da literatura estatística, foi exportado para a problemática da segmentação de mercado e inspirou vários autores na literatura do Marketing. Por exemplo, DeSarbo et. al (1989) e Wedel e Kistemaker (1989) estendem a *clusterwise regression* para um cenário de múltiplas observações por indivíduo.

[1] De acordo com Späth (1982), o resultado final depende da partição inicial e da dimensão mínima imposta para os grupos (necessariamente superior ao número de variáveis independentes). Para evitar os problemas com soluções óptimas locais, Späth recomenda o ensaio, para S grupos, de diferentes iniciações.

Tabela 2. ***Clusterwise linear regression***

INICIAÇÃO

Derivação de uma partição inicial dos dados em *S* segmentos. Fixar $n := n_0$ ($n_0 \in N$).

ITERAÇÃO

Colocar $n := n+1$ e $n := 1$ se $n > N$. Seja $n \in P_s$. Examinar se há grupos P_t $t \neq s$ em que:

$$E\left(P_t \cup \{n\}\right) + E\left(P_s - \{n\}\right) < E\left(P_t\right) + E\left(P_s\right) \quad (17)$$

se sim, seja t' um índice associado ao grupo que permite a máxima redução na função objectivo; Neste caso, redefinir:

$$P_s := P_s - \{n\},\ \mathrm{P}_{t'} = P_{t'} \cup \{n\}$$

se não, regressar ao passo 1.

CRITÉRIO DE PARAGEM

Repetir a iteração até que a função objectivo não possa ser reduzida, i.e., até que *n* tenha sido aumentado *N* vezes sem que (17) tenha diminuído.

N_s número de observações no segmento *S*

y_s vector de dimensão $\left(N_s \times 1\right)$ das preferências dos indivíduos no segmento S

X_s matriz de dimensão $\left(N_s \times P\right)$

β_s vector de dimensão $\left(P \times 1\right)$ dos pesos de preferências no segmento *S*

Fonte: Adaptado a partir de Späth (1979), pág. 368

2.3 Modelos Mistura Finitos

O desenvolvimento dos modelos mistura surge no século XIX, com os trabalhos de Pearson (1984) e de Newcomb (1886). Nos modelos mistura finitos ('*finite mixture models*') assume-se que as observações amostrais são extraídas de dois ou mais grupos, que foram misturados em proporções desconhecidas. O seu objectivo é "desmisturar" a amostra, i.e., identificar os segmentos implícitos (componentes da mistura), e estimar os parâmetros da função densidade (usada na descrição das probabilidades de ocorrência dos valores para uma dada variável) intra-segmento. Ao contrário das abordagens de segmentação, tradicionais, que fornecem heurísticas para a construção dos segmentos com base em valores amostrais, as distribuições mistura são uma abordagem de segmentação baseada num modelo. Deste modo, permitem a estimação e testes de hipóteses no quadro da teoria estatística tradicional. A abordagem de modelos mistura aplicados à segmentação apresenta uma classe de algoritmos de agrupamento extremamente flexíveis desenhados para a resolução de muitos problemas de Marketing. Os modelos mistura são modelos estatísticos que envolvem uma forma específica da função distribuição das observações em cada população subjacente (que é especificada). A função distribuição é usada para descrever as probabilidades de ocorrência dos valores observados das variáveis em análise.

De seguida propõe-se a apresentação do modelo geral, a descrição dos algoritmos de estimação dos seus parâmetros, uma discussão das limitações desta abordagem, e o papel destes modelos no desenvolvimento dos modelos mistura de regressão. Apresenta-se ainda uma síntese das principais aplicações dos modelos mistura finitos no contexto da segmentação de mercado. A exposição que se segue recorre essencialmente ao texto de Wedel e Kamakura (2000), ao artigo de Dillon e Kumar (1994) e aos trabalhos de Everitt e Hand (1981), McLachlan e Basford (1988) e Titterington *et al.* (1985).

2.3.1 Modelo

O modelo geral pressupõe que os objectos (consumidores), para os quais se dispõe do valor das variáveis $\mathbf{y}_n = (y_{nk})$, $n = 1,...,N$, $k = 1,...,K$, são extraídos de uma população constituída por uma mistura de *S* segmentos, em proporções λ_s $\mathrm{s} = 1,...,\mathrm{S}$:

$$\sum_{s=1}^{S} \lambda_s = 1 \qquad \lambda_s \geq 0,\ s = 1,...,S\ . \tag{15}$$

Supondo que a observação y_{nk} pertence ao segmento *s*, a função distribuição condicional para o vector y_n é definida por $f_s(\mathbf{y}_n | \boldsymbol{\theta}_s)$, em que $\boldsymbol{\theta}_s$ representa o vector de parâmetros desconhecidos necessários para a sua caracterização. Como a pertença aos segmentos das observações é desconhecida, a função densidade mistura (‘*mixture density*´) de y_n é dada por:

$$f\left(\mathbf{y}_n | \Phi\right) = \sum_{s=1}^{S} \lambda_s f_s\left(\mathbf{y}_n | \boldsymbol{\theta}_s\right) \tag{16}$$

em que $\Phi = \left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)$.

A função $f_s\left(\mathbf{y}_n | \boldsymbol{\theta}_s\right)$ pode assumir várias formas, no contexto das distribuições discretas ou contínuas (Wedel e Kamakura 2000). A sua caracterização é efectuada, para cada segmento *s*, através da média, μ_{ks}, e de um parâmetro de dispersão, σ^2. No contexto das distribuições contínuas, a distribuição normal multivariada é a mais utilizada (Wedel e Kamakura, 2000).

2.3.2 Estimação por máxima verosimilhança

O propósito da análise dos modelos mistura finitos é a estimação do vector de parâmetros $\Phi = \left(\boldsymbol{\lambda}, \boldsymbol{\theta}\right)$. Neste contexto, o método da máxima verosimilhança consagrou-se como o procedimento geral de estimação (Dillon e Kumar, 1994). A obtenção de estimadores consistentes e assimptoticamente normalmente distribuídos (Dillon e Kumar, 1994) é apontada como a sua principal propriedade estatística. Para as distribuições mistura, a função de verosimilhança de Φ assume a seguinte forma geral:

$$L = \prod_{n=1}^{N} f\left(\mathbf{y}_n | \Phi\right) \tag{17}$$

Os estimadores para os parâmetros são obtidos maximizando a função de verosimilhança (17) em relação ao vector de parâmetros Φ, sujeita às restrições definidas em (15). Neste contexto, são calculadas as derivadas parciais, a partir da função de verosimilhança logaritmizada aumentada (18):

$$\log \mathrm{L} = \sum_{n=1}^{N} \log\left[\sum_{s=1}^{S} \lambda_s f_s\left(\mathbf{y}_n | \boldsymbol{\theta}_s\right)\right] - w\left[\sum_{s=1}^{S} \lambda_s - 1\right] \tag{18}$$

em que w representa o multiplicador de Lagrange. Daqui resultam as seguintes equações de estacionariedade:

$$\frac{\partial \log L}{\partial \lambda_s} = \sum_{n=1}^{N} \frac{f_s(\mathbf{y}_n)}{f(\mathbf{y}_n)} - w = 0 \qquad (19)$$

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}_s} = \sum_{n=1}^{N} \frac{\partial f_s(\mathbf{y}_n)/\partial \boldsymbol{\theta}_s}{f(\mathbf{y}_n)} = 0 \qquad (20)$$

Titterington *et al.* (1985) fornecem uma excelente descrição do processo de estimação pelo método da máxima verosimilhança em modelos mistura finitos.

2.3.3 Algoritmos para a estimação dos parâmetros envolvidos no modelo

No contexto dos modelos mistura são utilizadas duas abordagens gerais na obtenção das soluções de máxima verosimilhança (Wedel e Kamakura, 2000; Dillon e Kumar, 1994). A primeira aplica métodos de optimização numérica às equações (19) e (20); as rotinas (de optimização) utilizadas - *Newton-Raphson*, *Quasi-Newton*, *Simplex*, *Fisher´s Scoring* - inspiram-se em métodos baseados no gradiente. Uma introdução a estes métodos pode ser encontrada nos trabalhos de Dennis e Schanabel (1983), Gill *et al.* (1989) e Everitt (1987). A abordagem mais usada (Wedel e Kamakura, 2000) recai na utilização do algoritmo EM - *Expectation-Maximization* (Dempster *et al.*, 1977). O algoritmo EM deriva a sua designação dos dois passos do algoritmo. Na fase E '*expectation step*' (Esperança), a partir das estimativas disponíveis para todos os parâmetros do modelo, Φ, são obtidos novos valores para as probabilidades de afectação dos consumidores aos segmentos, p_{ns}, $n = 1,...,N$, $s = 1,...,S$ (usando o Teorema de Bayes). No passo M '*maximization step*' (Maximização), são obtidas novas estimativas para Φ com base nessas estimativas temporárias de p_{ns}. Estes dois passos são repetidos iterativamente enquanto for possível uma melhoria na função de verosimilhança (Tabela 3).

2.3.4 Comparações de performance relativa

Em termos de eficiência, não é claro qual dos dois métodos (algoritmo EM ou optimização numérica) é preferível (Wedel e Kamakura, 2000). No entanto, algumas conclusões gerais sobre a sua *performance* relativa (no contexto da obtenção de estimativas de máxima verosimilhança para os parâmetros das componentes dos modelos mistura) podem ser apontadas. Comparado com os algoritmos de optimização (*Newton-Raphson*, *Quasi-Newton, Simplex*, *Fisher´s Scoring*), a convergência do algoritmo EM é considerada baixa, (medida pelo número de iterações necessárias). Um problema potencial associado à aplicação das duas abordagens relaciona-se com a propriedade não desejável dos algoritmos terminarem num óptimo local, tendo sido propostos vários procedimentos para reduzir este risco (Wedel e Kamakura, 2000). Leisch (2004) acrescenta mais duas limitações do algoritmo EM, a referir: instabilidades numéricas se uma componente possui poucas observações numa dada iteração e problemas na estimação dos parâmetros (no caso das distribuições mistura normais a função de verosimilhança tende a crescer sem limites se $\sigma^2 \to 0$).

Não obstante, a simplicidade computacional do algoritmo EM (com apenas dois tipos de iterações) é apontada como a sua principal fonte de popularidade.

Para a resolução das equações de verosimilhança é necessária a especificação de um conjunto de parâmetros iniciais. Note-se que a escolha de más configurações pode afectar de forma adversa a *performance* do algoritmo, conduzindo a problemas de não convergência.

Uma análise do quadro geral de segmentação baseado nos modelos mistura permite concluir que esta abordagem estatística possui várias vantagens em relação aos métodos de classificação baseados em heurísticas; no entanto, vários assuntos, como a existência de óptimos locais, ou os testes para a determinação do número de segmentos, necessitam ainda de uma resolução satisfatória. Revisões da aplicação desta abordagem a problemas de segmentação de mercado podem ser encontradas no artigo de Dillon e Kumar (1994) e no texto de Wedel e Kamakura (2000).

2.3.5 Modelos mistura finitos versus modelos mistura de regressão

Nos modelos mistura finitos não existem variáveis exógenas para explicar as médias e as variâncias de cada componente da distribuição mistura finita. Por exemplo, numa mistura de distribuições normais, a média e a variância em cada segmento latente (ou componente da mistura) são estimadas directamente. Nestas aplicações o principal objectivo é descritivo, i.e., formar grupos homogéneos de indivíduos/ consumidores com base em várias características observadas. Em contraste, os modelos mistura permitem não só a classificação probabilística das observações em segmentos latentes como também a estimação de modelos de regressão explicando as médias e as variâncias de cada variável dependente dentro de cada segmento.

Tabela 3. Algoritmo EM

(1) O procedimento é iniciado $(h := 1)$ com a fixação do número de segmentos, *S*, e com a geração de uma partição inicial $p_{ns}^{(1)}$ (aleatória ou baseada na utilização de um procedimento de análise de agrupamento convencional).

(2) Dado $p_{ns}^{(h)}$, obter as estimativas de λ_s e de $\boldsymbol{\theta}_s$ usando os estimadores de máxima verosimilhança:

$$\lambda_s = \frac{\sum_{n=1}^{N} p_{ns}}{N} \tag{21}$$

$$\theta_s = \frac{\sum_{n=1}^{N} y_n p_{ns}}{N\lambda_s} \tag{22}$$

(3) Teste de convergência: parar se a alteração em $\log L$ (16) da iteração (*h-1*) para a iteração (*h*) for residual.

(4) Aumentar o índice da iteração $h := h+1$ e calcular novas estimativas para $\mathrm{p}_{\mathrm{ns}}^{(\mathrm{h}+1)}$ usando a equação.

$$p_{ns} = \frac{\lambda_s f_s\left(\mathbf{y}_n | \boldsymbol{\theta}_s\right)}{\sum_{s=1}^{S} \lambda_s f_s\left(\mathbf{y}_n | \boldsymbol{\theta}_s\right)} \tag{23}$$

(5) Repetir os passos 2 a 4.

Fonte: Adaptado a partir de Wedel e Kamakura (2000), pág. 85

Os modelos mistura de regressão relacionam uma variável dependente, como as frequências de compra ou as preferências por marcas, com variáveis explicativas, como variáveis do *marketing mix,* dentro de cada classe latente.

3 Modelos Mistura de Regressão para Variáveis Normais

3.1 Modelo

Nesta sessão é apresentado o modelo mistura de regressão para variáveis dependentes normais de DeSarbo e Cron (1988), que é um caso especial do modelo GLIMMIX - *Generalized Linear Mixture Model* (Wedel e DeSarbo, 1995).

DeSarbo e Cron (1988) generalizam pela primeira vez os '*stochastic switching regression models*' (Quandt, 1972; Hosmer, 1974; Quandt e Ramsey, 1978) para mais de dois regimes. O artigo apresenta uma nova metodologia para derivação de grupos e a estimação simultânea das correspondentes funções regressão intra-grupo. A partir do conjunto de N observações para a variável dependente y e as P variáveis explicativas $\left(x_1,...,x_P\right)$, $p=1,...,P$, são utilizadas distribuições mistura condicionais, num quadro de máxima verosimilhança, com a utilização do algoritmo EM (Dempster *et al.,* 1977) para a estimação dos parâmetros.

Assume-se que as variáveis aleatórias y_n, $n=1,...,N$ são extraídas de uma população que é uma mistura de um número finito de segmentos (populações normais) S, em proporções $\lambda_1,...,\lambda_S$, $s=1,...,S$; no entanto, não é conhecida *à priori* a população da qual a observação y_n foi extraída. As probabilidades λ_s estão sujeitas às restrições (24):

$$\sum_{s=1}^{S}\lambda_s=1,\ \lambda_s\geq 0,\ s=1,...,S \tag{24}$$

A distribuição de y_n, dado que y_n é extraído do segmento S é dada por $f_s\left(y_n\middle|\sigma_s^2,\beta_{sp}\right)$;

$$f_s\left(y_n\middle|\sigma_s^2,\beta_{sp}\right)=\left(2\pi\sigma_s^2\right)^{-1/2}\exp\left[-\frac{\left(y_n-\mathbf{X}_n^{'}\boldsymbol{\beta}_s\right)}{2\sigma_s^2}\right] \tag{25}$$

As variáveis y_n são independentes[2], sendo condicionais ao segmento S. Deste modo, a função densidade de probabilidade de y_n pode ser expressa como uma mistura de densidades normais univariadas[3]:

$$\begin{aligned} f\left(y_n \middle| \Phi\right) &= \sum_{s=1}^{S} \lambda_s f_s\left(\mathbf{y}_n \middle| \boldsymbol{\theta}_s\right) = \sum_{s=1}^{S} \lambda_s f_s\left(y_n \middle| \sigma_s^2, \beta_{sp}\right) \\ &= \sum_{s=1}^{S} \lambda_s \left(2\pi\sigma_s^2\right)^{-1/2} \exp\left[-\frac{\left(y_n - \mathbf{X}_n^{'}\boldsymbol{\beta}_s\right)}{2\sigma_s^2}\right] \end{aligned} \tag{26}$$

em que os vectores de parâmetros são $\Phi = \left(\lambda_s, \boldsymbol{\theta}_s\right)$, $\boldsymbol{\theta}_s = \left(\beta_{sp}, \sigma_s^2\right)$.

Quando se aplicam modelos mistura de regressão, o objectivo é prever as médias das observações em cada segmento usando um conjunto de variáveis explicativas. Para esse objectivo, é especificado um previsor linear $\mu_{ns} = \sum_{p=1}^{P} x_{np}\beta_{sp}$, que é produzido

por P $\left(p = 1,...,P\right)$ variáveis independentes e um vector de parâmetros $\boldsymbol{\beta}_s = \left(\beta_{sp}\right)$ para o

segmento s.

As estimativas para os parâmetros $\Phi = \left(\lambda_s, \beta_{sp}, \sigma_s^2\right)$ são obtidas maximizando a função de verosimilhança L (27), ou o logaritmo da função de verosimilhança $\log L$ (28) em relação a $\boldsymbol{\phi}$, sujeita às restrições (24); no entanto só é possível a obtenção de estimadores consistentes se for imposta a condição $\sigma_s^2 > 0$ (DeSarbo e Cron, 1988).

---

[2] Na presença de K medidas repetidas $\left(k = 1,...,K\right)$ para cada consumidor n $\left(n = 1,...,N\right)$, os y_{nk} não se podem assumir como independentes, sendo apropriada uma distribuição da família exponencial multivariada, como a distribuição normal multivariada.

[3] No caso de K medidas repetidas por indivíduo obtém-se $f\left(y_{nk} \middle| \Phi\right) = \sum_{s=1}^{S} \lambda_s \prod_{k=1}^{K} f_s\left(\mathbf{y}_{nk} \middle| \boldsymbol{\theta}_s\right)$.

$$L = \prod_{n=1}^{N} f\left(y_n | \Phi\right) = \prod_{n=1}^{N} \sum_{s=1}^{S} \lambda_s f_s\left(y_n | \boldsymbol{\theta}_s\right)$$

$$L = \prod_{n=1}^{N} \left\{ \sum_{s=1}^{S} \lambda_s \left(2\pi\sigma_s^2\right)^{-1/2} \exp\left[-\frac{\left(y_n - \mathbf{X}_n^t \boldsymbol{\beta}_s\right)^2}{2\sigma_s^2} \right] \right\} \tag{27}$$

ou

$$\log L = \sum_{n=1}^{N} \log f\left(y_n | \Phi\right) = \sum_{n=1}^{N} \log \sum_{s=1}^{S} \lambda_s f_s\left(y_n | \boldsymbol{\theta}_s\right)$$

$$\log L = \sum_{n=1}^{N} \log \left\{ \sum_{s=1}^{S} \lambda_s \left(2\pi\sigma_s^2\right)^{-1/2} \exp\left[-\frac{\left(y_n - \mathbf{X}_n' \boldsymbol{\beta}_s\right)^2}{2\sigma_s^2} \right] \right\}. \tag{28}$$

3.2 Algoritmo de Estimação

A função (28) pode ser maximizada usando o algoritmo EM – *Expectation Maximization* (Dempster *et al.,* 1977). O nome do algoritmo resulta dos seus dois passos, *Expectation* e *Maximization*. Para se exemplificar o algoritmo EM, é necessária a introdução de dados não observados z_{ns}, indicando se a observação y_n pertence ou não ao segmento *s*. Assim, $z_{ns} = 1$ se *n* é extraído do segmento *s* e $z_{ns} = 0$, caso contrário. Assume-se que z_{ns} são *i.i.d.* multinomiais (29):

$$f\left(\mathbf{z}_n | \lambda\right) = \prod_{n=1}^{N} \lambda_s^{z_{ns}} \tag{29}$$

em que $\mathbf{z}_n = \left(z_{n1}, ..., z_{nS}\right)'$.

Assume-se que os y_n, dado $\mathbf{z}_n$, são condicionalmente independentes e possuem a densidade (30):

$$f\left(y_n | \mathbf{z}_n, \Phi\right) = \prod_{s=1}^{S} f_s\left(y_n \middle| \boldsymbol{\beta}_s, \sigma_s^2\right)^{z_{ns}}. \tag{30}$$

Sendo os valores z_{ns} considerados '*missing data*', o logaritmo da função de verosimilhança para os dados completos '*log-likelihood function for the complete data*' $\mathbf{X} = ((x_{np}))$ e $\mathbf{Z} = ((z_{ns}))$, $n = 1,...,N$, $p = 1,...,P$, $s = 1,...,S$, pode ser formada a partir das equações (31) e (32).

$$\mathrm{L}_c = \prod_{n=1}^{N}\prod_{s=1}^{S} f_s\left(y_n \middle| \boldsymbol{\beta}_s, \sigma_s^2\right)^{z_{ns}} + \prod_{n=1}^{N}\prod_{s=1}^{S} \lambda_s^{z_{ns}} \tag{31}$$

$$\log L_c = \sum_{n=1}^{N}\sum_{s=1}^{S} z_{ns} \log f_s\left(y_n \middle| \boldsymbol{\beta}_s, \sigma_s^2\right) + \sum_{n=1}^{N}\sum_{s=1}^{S} z_{ns} \log \lambda_s \tag{32}$$

A função (32) é maximizada com recurso ao algoritmo EM (Dempster *et al.*, 1977). Depois das estimativas para os parâmetros Φ terem sido obtidas, as probabilidades posteriores de pertença p_{ns} são calculadas no passo E. No passo M o valor esperado de $\log L_c$ (32) é maximizado em relação a Φ, permitindo a obtenção de novas estimativas para os parâmetros.

No passo E o logaritmo da função de verosimilhança é substituído pelo seu valor esperado, calculado com base nas estimativas provisórias de Φ. No passo M, o valor esperado de $\log L_c$ é maximizado em relação a Φ, permitindo a obtenção de novas estimativas para os parâmetros. Os passos E e M são alternados até que nenhuma melhoria na função de verosimilhança seja possível. Descrevem-se, de seguida, os passos E e M de forma mais detalhada.

Passo E

No passo E, é calculado o valor esperado de $\ln L_c$ em relação à distribuição condicional dos dados não observados $\mathbf{Z}$, dados os valores de $\mathbf{y}$ e as estimativas provisórias para Φ, $(\boldsymbol{\lambda}^*, \boldsymbol{\beta}^*, \sigma^{2*})$. Pode-se facilmente verificar que $E\left[\log L_c\left(\Phi \middle| \mathbf{y}, \mathbf{Z}\right)\right]$ é obtida pela substituição em $\log L_c$ de z_{ns} pelos seus valores esperados $E\left(z_{ns} \middle| \mathbf{y}, \Phi\right)$; estes valores esperados podem ser obtidos com base em e com base na regra de Bayes (33),

$$E\left(z_{ns}\left|y,\Phi\right.\right)=\frac{\lambda_s f_s\left(y_n\left|\boldsymbol{\beta}_s,\sigma_s^2\right.\right)}{\sum_{s=1}^{S}\lambda_s f_s\left(y_n\left|\boldsymbol{\beta}_s,\sigma_s^2\right.\right)}=p_{ns}\,. \tag{33}$$

Deste modo, os dados não observados $\mathbf{Z}$ em $\log L_c$ são substituídos pelos seus valores actuais p_{ns} (passo E),

$$E\left[\log L_c\left(\boldsymbol{\phi}\left|\mathbf{y},\mathbf{Z}\right.\right)\right]=\sum_{n=1}^{N}\sum_{s=1}^{S}p_{ns}\log f_s\left(y_n\left|\boldsymbol{\beta}_s,\sigma_s^2\right.\right)+\sum_{n=1}^{N}\sum_{s=1}^{S}p_{ns}^2\log\lambda_s \tag{34}$$

Passo M

Para maximizar o valor esperado de $\log L_c$ em relação a Φ sujeito às restrições (24), é formada uma função aumentada (35):

$$\Psi=\sum_{n=1}^{N}\sum_{s=1}^{S}p_{ns}^{*}\log f_s\left(y_n\left|\beta_{sp}^{*},\sigma_s^{2*}\right.\right)+\sum_{n=1}^{N}\sum_{s=1}^{S}p_{ns}^2\log\lambda_s^{*}-w\left(\sum_{s=1}^{S}\lambda_s^{*}-1\right) \tag{35}$$

em que w representa o multiplicador de Lagrange.

As equações de estacionariedade são obtidas igualando as derivadas parciais de primeira ordem a zero:

$$\frac{\partial\Psi}{\partial\lambda_s}=\sum_{n=1}^{N}\frac{p_{ns}}{\lambda_s}-w=0 \tag{36}$$

$$\frac{\partial\Psi}{\partial\beta_s}=\sum_{n=1}^{N}p_{ns}\frac{\partial\log f_s\left(y_n\left|\boldsymbol{\beta}_s,\sigma_s^2\right.\right)}{\partial\boldsymbol{\beta}_s}=0 \tag{37}$$

$$\frac{\partial\Psi}{\partial\sigma_s^2}=\sum_{n=1}^{N}p_{ns}\frac{\partial\log f_s\left(y_n\left|\boldsymbol{\beta}_s,\sigma_s^2\right.\right)}{\partial\sigma_s^2}=0 \tag{38}$$

Para estimar λ_s, é necessário estimar préviamente w; tal é obtido pela multiplicação de ambos os lados da equação (36) por λ_s e soma de ambos os lados para S:

$$\sum_{n=1}^{N}\sum_{s=1}^{S} p_{ns}^{*} - w\sum_{s=1}^{S}\lambda_s = 0$$

$$\hat{w} = N \qquad (39)$$

De seguida, substituindo (39) em (36) e multiplicando ambos os lados da equação por λ_s obtém-se:

$$\sum_{n=1}^{N} p_{ns} - N\lambda_s = 0$$

$$\hat{\lambda}_s = \frac{\sum_{n=1}^{N}\hat{p}_{ns}}{N} \qquad (40)$$

Note-se que maximizar $E\left[\ln L_c\left(\Phi\middle|y,z\right)\right]$ em ordem a $\boldsymbol{\beta}_s$ e σ_s^2 é equivalente a maximizar de forma independente cada uma das S expressões (41):

$$\log L^s = \sum_{n=1}^{N} p_{ns} \log f_s\left(y_n\middle|\beta_s,\sigma_s^2\right) \qquad (41)$$

Deste modo conclui-se que a optimização de (41) é equivalente ao problema de maximização de um modelo linear generalizado para os dados completos, com a particularidade de cada observação y_n contribuir para a função com um peso conhecido, p_{ns}, obtido no passo E. As equações de estacionariedade são obtidas igualando as derivadas parciais de (41) de primeira ordem a zero.

As estimativas para $\boldsymbol{\beta}_s$ e σ_s^2 são obtidas através do *Fisher Scoring Method* (que é equivalente ao procedimento de *Newton-Raphson* para funções de *link* canónico) (McCullagh e Nelder, 1989, Wedel e DeSarbo, 1995).

Sumariando, o algoritmo EM proposto para o ajustamento do modelo consiste nos seguintes passos (Wedel e DeSarbo, 1995), descrito na Tabela 4.

Tabela 4. Algorimo EM

1. No primeiro passo da iteração, $h := 0$, iniciar o procedimento fixando o número de classes, S, e gerando uma partição inicial $p_{ns}^{(0)}$.
2. Dados $p_{ns}^{(0)}$, estimativas de máxima verosimilhança para $\boldsymbol{\beta}_s$ podem ser obtidas através do método de mínimos quadrados ponderados. Obtenção de estimativas para λ_s (Passo M).
3. Teste de convergência: parar se $\left|\log L\left(\Psi^{(h+1)}|\mathbf{y}\right) - \log L\left(\Psi^{(h)}|\mathbf{y}\right)\right|$ for suficientemente pequeno.
4. Calcular novas estimativas para as probabilidades posteriores, $p_{ns}^{(h+1)}$, de acordo com a equação (42)[4]

$$p_{ns} = \frac{\lambda_s f_{nk|s}\left(y_{nk}\left|\beta_s, \sigma_s^2\right.\right)}{\sum_{s=1}^{S} \lambda_s f_{nk|s}\left(y_{nk}\left|\beta_s, \sigma_s^2\right.\right)} \tag{42}$$

5. Repetir os passos 2 a 4.

Fonte: Adaptado a partir de Wedel e DeSarbo (1995), pág. 28, 29

4 GENERALIZAÇÕES DO MODELO BASE

O modelo proposto por DeSarbo e Cron (1988), e incluído como um caso especial do modelo mistura linear generalizado por Wedel e DeSarbo (1995), foi também estendido por Ramaswamy *et al.* (1993) para um cenário de dados em painel, por DeSarbo *et al.* (1992) para dados de análise conjunta e por Jedidi *et al.* (1996) para modelos de equações estruturais. De seguida apresentam-se as formulações de DeSarbo *et al.* (1992) e Ramaswamy *et al.* (1993), que permitem a realização simultânea de regressões normais multivariadas e a segmentação

[4] No caso de medidas repetidas a probabilidade de pertença do consumidor *n* ao segmento *s* é dada por

$$p_{ns} = \frac{\lambda_s \prod_{k=1}^{K} f_s\left(y_{nk}\left|\beta_s, \sigma_s^2\right.\right)}{\sum_{s=1}^{S} \lambda_s \prod_{k=1}^{K} f_s\left(y_{nk}\left|\beta_s, \sigma_s^2\right.\right)}.$$

de mercado. Não é descrita a formulação de Jedidi *et al.* (1996), dado que no âmbito deste trabalho apenas se pretende estudar a problemática da selecção do número adequado de segmentos de mercado em modelos mistura de regressão uniequacionais para variáveis normais.

4.1.1 Modelos Mistura de Regressão para Análise Conjunta Métrica

Nesta sessão apresenta-se o modelo para análise conjunta métrica proposto por DeSarbo *et al.* (1992). Este modelo, designado por *Latent Class Metric Conjoint Analysis*, permite a estimação simultânea de um modelo de análise conjunta e a segmentação de mercado e generaliza o modelo proposto por DeSarbo e Cron (1988) para um contexto multivariadado. A formulação proposta introduz flexibilidade na modelização de situações em que os membros de um segmento de mercado particular exibem diferentes perfis de preferências. Trata-se de num modelo particularmente interessante dado que um estudo de simulação comparando nove modelos de análise conjunta métrica revelou que o modelo de classes latentes apresenta o melhor desempenho em termos de recuperação de parâmetros, qualidade do ajustamento e capacidade preditiva (Vriens *et al.*, 1996).

Considere-se, a seguinte notação:

$n = 1,...,N$ consumidores;

$j = 1,...,J$ perfis conjuntos (‘*conjoint profiles’)*;

$p = 1,...,P$ variáveis independentes na análise conjunta (codificadas como variáveis *dummy*);

$s = 1,...,S$ segmentos;

$y_{nj} =$ rating de resposta métrico ao perfil *j* pelo consumidor *n*;

$X_{jp} =$ valor da variável independente *p* no perfil *j* (por simplificação, não é efectuada a distinção entre factor e nível),

$\mathbf{X} = \left(\left(X_{jp}\right)\right)$;

$\mathbf{X}_j =$ vector linha de dimensão $\left(1xP\right)$ de variáveis independentes para o perfil *j*;

$\beta_{ps} =$ coeficiente de valor parcial (‘*conjoint part-worth*’) estimado para a variável *p* no segmento *s*;

$\boldsymbol{\beta}_s =$ vector linha de dimensão $(1xP)$ de valores parciais para o segmento *s*,

$\boldsymbol{\beta} = ((\beta_{ps}))$;

$\sum_s =$ matriz de covariâncias de dimensão (JxJ) para o segmento *s*,

$\sum' = (\Sigma_1, \Sigma_2, ..., \Sigma_S)$.

Assume-se que o vector linha $\mathbf{y}_n$ de dimensão $(1xJ)$ possui uma função densidade de probabilidade que pode ser modelizada como uma mistura finita das seguintes distribuições condicionais (43),

$$f(\mathbf{y}_n; \boldsymbol{\lambda}, \boldsymbol{\beta}, \Sigma) = \sum_{s=1}^{S} \lambda_s f_s(\mathbf{y}_n | \boldsymbol{\beta}_s, \Sigma_s) \tag{43}$$

em que $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_{S-1})$ são $S-1$ proporções mistura independentes da mistura finita sujeitas às restrições (24).

Cada f_s em (43) é definido a partir de uma distribuição multivariada condicional (44).

$$f_s(\mathbf{y}_n | \boldsymbol{\beta}_s, \Sigma_s) = (2\pi)^{-J/2} |\Sigma_s|^{-1/2} \exp\left[-1/2(\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta}_s^{'})\Sigma_s^{-1}(\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta}_s^{'})'\right] \tag{44}$$

Dado uma amostra de *N* consumidores independentes, é possível formar a expressão da função de verosimilhança (45):

$$L = \prod_{n=1}^{N} \left[\sum_{s=1}^{S} \lambda_s (2\pi)^{-J/2} |\Sigma_s| \exp\left[-1/2(\mathbf{y}_n - \mathbf{X}_j\boldsymbol{\beta}_s^{'})\Sigma_s^{-1}(\mathbf{y}_n - \mathbf{X}_j\boldsymbol{\beta}_s^{'})'\right]\right] \tag{45}$$

ou

$$\log L = \sum_{n=1}^{N} \log\left[\sum_{s=1}^{S} \lambda_s f_s(\mathbf{y}_n | \boldsymbol{\beta}_s, \Sigma_s) \right] \tag{46}$$

Depois das estimativas de $\boldsymbol{\lambda}$, Σ e $\boldsymbol{\beta}$ terem sido obtidas a partir de qualquer iteração do procedimento de máxima verosimilhança, é possível afectar cada consumidor *n* a cada classe latente ou segmento de mercado *s* através da probabilidade à posterior estimada, aplicando a regra de Bayes, resultando num agrupamento probabilístico (47),

$$p_{ns} = \frac{\lambda_s f_s\left(\mathbf{y}_n \middle| \boldsymbol{\beta}_s, \Sigma_s\right)}{\sum_{s=1}^{S} \lambda_s f_s\left(\mathbf{y}_n \middle| \boldsymbol{\beta}_s, \Sigma_s\right)}. \tag{47}$$

O algoritmo EM *Expectation-Maximization* (Dempster *et al.*, 1977) é utilizado, tal como no modelo descrito na secção anterior, para a estimação dos parâmetros.

4.1.2 Modelo Mistura de Regressão para Dados em Painel

Ramaswamy *et al.* (1993) generalizam a abordagem proposta por DeSarbo e Cron (1988), para dados em painel.

Considere-se a seguinte notação adicional:

$n = 1,...,N$ unidades seccionais

$t = 1,...,T$ períodos de tempo

$p = 1,...,P$ variáveis explicativas

$X_{npt} =$ valor da variável explicativa *p* no período *t* para a unidade seccional *n*

$\mathbf{X}_n = \left(\left(X_{npt}\right)\right)$ matriz dos valores das variáveis explicativas p no período *t*, para cada unidade seccional *n*

$y_{nt} =$ valor da variável dependente métrica para a unidade seccional *n* no período *t*

$\mathbf{y}_n = \left(y_{nt}\right)$ vector de observações da variável dependente *y* para o consumidor *n*

$\Delta_s =$ matriz de variâncias e covariâncias para o painel (segmento) *s*

$\Delta = \left(\Delta_1,...,\Delta_S\right)'$

$\boldsymbol{\beta}_s = \left(\beta_{sp}\right)$ vector de parâmetros para o painel (segmento) *s*

$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_s\right)'$

Como forma de capturar a heterogeneidade seccional potencial não observada, Ramaswamy *et al.* (1993) propõem a estimação simultânea de uma estrutura de painéis múltiplos e dos coeficientes de regressão para cada painel. Deste modo, supõem a existência de *S* painéis latentes, de tal forma que a relação estrutural dentro de cada painel é descrita por um vector de parâmetros específico desse painel.

Considere-se que o vector $\mathbf{y}_n$ se distribui de acordo com uma mistura de *S* densidades normais multivariadas (48):

$$f\left(\mathbf{y}_n \middle| \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\Delta}\right) = \sum_{s=1}^{S} \lambda_s f_s\left(\mathbf{y}_n \middle| \boldsymbol{\beta}_s, \boldsymbol{\Delta}_s\right) \quad (48)$$

em que:

$$f_s\left(\mathbf{y}_n \middle| \boldsymbol{\beta}_s, \boldsymbol{\Delta}_s\right) = \left(2\pi\right)^{-T/2} \left|\boldsymbol{\Delta}_s\right|^{-T/2} \exp\left[-1/2\left(\mathbf{y}_n - \mathbf{X}_n \boldsymbol{\beta}_s\right)' \Delta_s^{-1} \left(\mathbf{y}_n - \mathbf{X}_n \boldsymbol{\beta}_s\right)\right] \quad (49)$$

Da especificação condicional em (49) conclui-se que se uma unidade seccional pertence ao painel *s*, então a estrutura das relações entre as variáveis explicativas e a variável dependente é representada pelo vector de parâmetros $\boldsymbol{\beta}_s$ e a matriz de variâncias e covariâncias Δ_s. Tal equivale à definição de um modelo linear generalizado para cada painel.

A partir de uma amostra de *N* unidades seccionais, é possível definir a função de verosimilhança (50) e a função de verosimilhança logaritmizada (51):

$$L = \prod_{n=1}^{N} \left\{ \sum_{s=1}^{S} \lambda_s \left(2\pi\right)^{-T/2} \left|\Delta_s\right|^{-1/2} \exp\left[-1/2\left(\mathbf{y}_n - \mathbf{X}_n \boldsymbol{\beta}_s\right)' \boldsymbol{\Delta}_s^{-1} \left(\mathbf{y}_n - \mathbf{X}_n \boldsymbol{\beta}_s\right)\right] \right\} \quad (50)$$

ou

$$\log L = \sum_{n=1}^{N} \log\left[\sum_{s=1}^{S} \lambda_s f\left(\mathbf{y}_n \middle| \mathbf{X}_n, \boldsymbol{\beta}_s, \boldsymbol{\Delta}_s\right) \right] \quad (51)$$

Dados os valores de **y**, **X** e o valor de *S*, pretende-se estimar os seguintes parâmetros: $(S-1)$ proporções mistura, o vector de coeficientes $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_S)'$ e a matriz de variâncias e covariâncias $\boldsymbol{\Delta}=(\boldsymbol{\Delta}_1,...,\boldsymbol{\Delta}_S)'$ para cada segmento. Para o efeito é maximizada a função (50) ou (51), sujeita às condições especificadas em (24); adicionalmente, $\boldsymbol{\Delta}_s$ deverá ser uma matriz definida positiva, simétrica, de forma a se obterem estimativas consistentes dos parâmetros.

A estimação do modelo mistura de regressão com dados em painel para variáveis dependentes normais é efectuada através da implementação do algoritmo EM (Dempster *et al.,* 1977).

As probabilidades posteriores de pertença aos painéis (para cada unidade seccional) podem ser calculadas usando a regra de Bayes (condicional nas estimativas para os parâmetros obtidas) como:

$$p_{ns}=\frac{\lambda_s f\left(\mathbf{y}_n|\boldsymbol{\beta}_s,\boldsymbol{\Delta}_s\right)}{\sum_{s=1}^{S}\lambda_s f\left(\mathbf{y}_n|\boldsymbol{\beta}_s,\boldsymbol{\Delta}_s\right)} \tag{52}$$

Considera-se o caso de uma única observação para cada unidade seccional (i.e., a abordagem proposta por DeSarbo e Cron 1988) como um caso especial do método proposto (Ramaswamy *et al.,* 1993).

5 SÍNTESE CONCLUSIVA

Na segmentação de mercado, os modelos de natureza estatística são de grande utilidade, na medida em que a classificação probabilística dos consumidores e a estimação de regressões intra-segmento é efectuada de forma simultânea, permitindo a identificação de segmentos homogéneos na forma como respondem às variáveis do *marketing mix*. O modelo geral, proposto por Wedel e DeSarbo (1995) e designado por GLIMMIX, acomoda uma variedade de variáveis dependentes, usadas para descrever fenómenos diversos como as preferências ou a satisfação (distribuição normal), a frequência de compra (distribuição Poisson) ou as escolhas de marcas alternativas (distribuição binomial). É possível a identificação de um grande número de publicações divulgadas na literatura do marketing com aplicações destes modelos a problemas de segmentação de mercado. Neste trabalho foi apresentada a formulação para o modelo mistura de regressão proposto por DeSarbo e Cron (1988) que é um caso geral do modelo de Wedel e DeSarbo (1995), assim como as duas extensões para o contexto multivariado, desenvolvidas por Ramaswamy *et al.* (1993) e por DeSarbo *et al.* (1992). Este tipo de modelos são de utilização generalizada em estudos de segmentação de mercado, em grande parte dada a sua disponibilidade em programas comerciais, como, por exemplo, o *GLIMMIX* (primeiro programa), o *Latent Gold* (líder actual) e o NLogit.

REFERÊNCIAS

Andrews, R. e Currim, I. S. (2003). "Retention of Latent Segments in Regression-Based Marketing Models", *International Journal of Research in Marketing*, 20 (4): 315-321.

Andrews, R. L., Ansari, A. e Currim, I. S. (2002). "Hierarchical Bayes vs. Finite Mixture Conjoint Analysis Models: a Comparison of Fit, Prediction and Partworth Recovery", *Journal of Marketing Research*, 39 (1): 87-98.

Basford, K. E. e McLachlan, G. J. (1985). "Likelihood Estimation for Normal Mixture Models", *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 34 (3): 282-289.

Bowman, D., Heilman, C. e Seetharaman, PB. (2004). "Determinants of Product-Use Compliance Behavior", *Journal of Marketing Research*, 41 (3): 324-338.

Chow, G. (1960). "Tests of the Equality Between Two Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28 (3): 561-605.

Cosslett, Stephen R. e Lee, Lung-Fei (1985). "Serial Correlation in Latent Discrete Variable Models", *Journal of Econometrics*, 27 (1): 79-97.

Day, N. E. (1969). "Estimating the Components of a Mixture of Two Normal Distributions", *Biometrika*, 56 (3): 463-474.

Dempster, A. P., Laird, N. M. e Rubin, D.B. (1977). “Maximum Likelihood from Incomplete Data via EM-Algorithm”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39 (1): 1-38.

Dennis, J. E. e Schanabel, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Engelwood Cliffs, NJ, Prentice Hall.

DeSarbo, W. S. e Cron, William L. (1988). “A Maximum Likelihood Methodology for Clusterwise Linear Regression”, *Journal of Classification*, 5 (1): 249-282.

DeSarbo, W. S., Jedidi, K. e Sinha, I. (2001). “Customer Value Analysis in a Heterogeneous Market”, *Strategic Management Journal*, 22 (9): 845-857.

DeSarbo, W. S., Oliver, R. L. e Ramaswamy, A. (1989). “A Simulated Annealing Methodology for Clusterwise Linear Regression”, *Psychometrika*, 54 (4): 707-736.

DeSarbo, W.S., Wedel, M., Vriens, M. e Ramaswamy, V. (1992). “Latent Class Metric Conjoint Analysis”, *Marketing Letters*, 3 (3): 273-288.

Dillon, W. R. e A. Kumar (1994). “Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview”, in R. P. Bagozzi (Eds.), *Advanced Methods of Marketing Research*, Blackwell, pp. 295-351.

Donoho, D. L. (1988). “One-Sided Inference about Functionals of a Density”, *Annals of Statistics*, 16 (4): 1390-1420.

Dunn, J. C. (1974). “A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters”, *Journal of Cybernetics*, 3 (3): 32-57.

Edwards, A. L. (1970). *The Measurement of Personality Traits by Scales and Inventories*, New York: Holt, Rinehart e Winston.

Efron, B. (1979). “Bootstrap Methods: Another Look at the Jackknife”. *Annals of Statistics*, 7 (1): 1-26.

Efron, B. e Tibshirani, R. (1993). *An Introduction to the Bootstrap*, London: Chapman e Hall.

Engel, Charles e Hamilton, James D. (1990). “Long Swings in the Dollar: Are they in the Data or Markets Know It?”, *The American Economic Review*, 80 (4): 689-713.

Engel, Charles e Hamilton, James D. (1990). “Long Swings in the Dollar: Are they in the Data or Markets Know It?”, *The American Economic Review*, 80 (4): 689-713.

Everitt, B. S. (1987). *An Introduction to Optimization Methods and their Application in Statistics*, London, New York, Chapman e Hall.

Everitt, B. S. (2001). *Cluster Analysis*, Hodder Arnold Publication.

Everitt, B. S., e Hand, D. J. (1981). *Finite Mixture Distributions*, London: Chapman and Hall.

Everitt, B. S., e Hand, D. J. (1981). *Finite Mixture Distributions*, London: Chapman and Hall.

Fisher, R. A. (1936). “Uncertain inference”, *Proceedings of the American Academy of Arts and Sciences*, 71:. 245-258.

Fraley, C. e Raftery, A. E. (1998). “How many clusters? Which clustering method? Answers via model-based cluster analysis”, *The Computer Journal,* 41 (8): 578-588.

Frank, R. E., Massy, W. F., e Wind, Y. (1972). *Market Segmentation*, Prentice-Hall.

Ganesalingam, S. e McLachlan, G. J. (1981). “Some Efficiency Results of the Estimation of the Mixing Proportions in a Mixture of two Normal Distributions”, *Biometrics*, 37 (1): 23-33.

Gill, P. E., Murray, W. e Wright, M. H. (1989). *Practical Optimization*, Academic Press.

Goldfeld, S. M. e Quandt, R. E. (1973). “A Markov Model for Switching Regressions”, *Journal of Econometrics*, 1 (1): 3-16.

Goldfeld, S. M. e Quandt, R. E. (1976). *Studies in Nonlinear estimation*, Cambridge, MA: Ballinger.

Hamilton, J. D. (1989). “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle”, *Econometrica*, 57 (2): 357-384.

Hamilton, J. D. (1990). “Analysis of Time Series Subject to Changes in Regime”, *Journal of Econometrics*, 45 (1): 39-70.

Hamilton, James D. (1991). “A Quasi-Bayesian Estimating Parameters for Mixtures of Normal Distributions”, *Journal of Business & Statistics*, 9 (1): 27-39.

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley e Sons.

Helsen, K., Jedidi, K. e DeSarbo, W. S. (1993). “A New Approach to Country Segmentation Utilizing Multinational Diffusion Patters”, *Journal of Marketing*, 57 (4): 60-71.

Hosmer, D. W. (1974). “Maximum Likelihood Estimates of the Parameters of a Mixture of two Regression Lines”, *Communications in Statistics*, 3 (10): 995-1006.

Jedidi, K, Ramaswamy, V., DeSarbo, W.S. e Wedel, M. (1996). “The Disaggregate Estimation of Simultaneous Equation Models: an Application to the Price-Quality Relationship”, *Journal of Structural Equation Modelling*, 3 (3): 266-289.

Marriott, F. H. C. (1975). “Separating Mixtures of Normal Distributions”, *Biometrics*, 31 (3): 767-769.

McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models*. Chapman e Hall.

McLachlan, G. J. (1982). “The Classification and Mixture Maximum Likelihood Approaches to Cluster Analysis”, in P. R. Krishnaiah e L. N. Kanal (Eds.), *Handbook of Statistics*, Vol. 2, Amsterdam: North-Holland, pp. 199-208.

McLachlan, G. J. e Basford, K. E. (1988). *Mixture Models. Inference and Applications to Clustering*, Marcel Dekker.

Newcomb, S. A (1886). “Generalized Theory of the Combination of Observations so as to Obtain The Best Result”, *American Journal of Mathematics*, 8: 343-366.

Pearson, K. (1894). “Contributions to the Mathematical Theory of Evolution”, *Philosophical Transactions of the Royal Society of London: Series A*, 185: 71-110.

Powell, M. J. D. (1964). “An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives”, *Computer Journal*, 7 (2): 155-162.

Quandt, R. E. e Ramsey, J. B. (1978). “Estimating Mixtures of Normal Distributions and Switching Regressions”, *Journal of the American Statistical Association*, 73 (364): 730-738.

Quandt, Richard E. (1972). “A New Approach to Switching Regressions”, *Journal of the American Statistical Association*, 67 (338): 306- 310.

Ramaswamy, V., Anderson, E. W. e DeSarbo, W. S. (1993). “A Disaggregate Negative Binomial Regression Procedure for Count Data Analysis”, *Management Science*, 40 (3): 405-417.

Sclove, S. L. (1987). “Application of Model-Selection Criteria to Some Problems in Multivariate Analysis”, *Psychometrika*, 52 (3): 333-343.

Scott, A. J. e Symons, M. J. (1971). “Clustering Methods Based on Likelihood Ratio Criteria”, *Biometrics*, 27 (2): 287-298.

Späth, H. (1979). “Algorithm 39: Clusterwise Linear Regression”, *Computing*, 22 (4): 367-373.

Spath, H. (1980). *Cluster Analysis and Algorithms*, Ellis Horwood.

Spath, H. (1981). “Correction to Algorithm 39: Clusterwise Linear Regression”, *Computing*, 26 (3): 275.

Späth, H. (1982). “Algorithm 48: A Fast Algorithm for Clusterwise Linear Regression”, *Computing*, 29 (2): 175-181.

Späth, H. (1985). *Cluster Dissection and Analysis*, Wiley: New York.

Symons, M. J. (1981). “Clustering Criteria and Multivariate Normal Mixtures”. *Biometrics*, 37 (1): 35-43.

Titterington, D. M., Smith, A. F. M. e Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley.

Vriens, M., Wedel, M. e Wilms, T. (1996). “Metric Conjoint Segmentation Methods: A Monte Carlo Comparison”, *Journal of Marketing Research*, 33 (1): 73-85.

Wedel, M. e DeSarbo, W. S. (1994). “A Review of Recent Developments in Latent Class Regression Models”, in Bagozzi, Richard P. (Eds.), *Advanced Methods of Marketing Research*, Blackwell, pp. 352-388.

Wedel, M. e DeSarbo, W. S. (1995). “A Mixture Likelihood Approach for Generalized Linear Models”, *Journal of Classification*, 12 (1): 21-55.

Wedel, M. e Kamakura, W.A. (2000). *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers

Wedel, M. e Kistemaker, C. (1989). “Consumer Benefit Segmentation Using Clusterwise Linear Regression”, *International Journal of Research in Marketing*, 6 (1): 45-59.

Wolfe, J. H. (1965). “A Computer Program for the Computation of Maximum Likelihood Analysis of Types”, Research Memo. SRM 65-12. San Diego: U.S. Naval Personnel Research Activity.

Wolfe, J. H. (1967). “NORMIX: Computational Methods for Estimating the Parameters of Multivariate Normal Mixtures of Distributions”, Research Memo. SRM 68-2. San Diego: U.S. Naval Personnel Research Activity.

Wolfe, J. H. (1970). “Pattern Clustering by Multivariate Mixture Analysis”, Multivariate Behavioral Research”, 5: 329-350.

Wolfe, J. H. (1971). “A Monte Carlo Study of Sampling Distributions of the Likelihood Ratio for Mixtures of Multinormal Distributions”, Technical Bulletin STB 72-2. San Diego: U.S. Naval Personnel and Training Research Laboratory.

Recent FEP Working Papers

Nº 261	Ana Oliveira-Brochado and Francisco Vitorino Martins, *"Aspectos Metodológicos da Segmentação de Mercado: Base de Segmentação e Métodos de Classificação"*, January 2008
Nº 260	João Correia-da-Silva, *"Agreeing to disagree in a countable space of equiprobable states"*, January 2008
Nº 259	Rui Cunha Marques and Ana Oliveira-Brochado, *"Comparing Airport regulation in Europe: Is there need for a European Regulator?"*, December 2007
Nº 258	Ana Oliveira-Brochado and Rui Cunha Marques, *"Comparing alternative instruments to measure service quality in higher education"*, December 2007
Nº 257	Sara C. Santos Cruz and Aurora A.C. Teixeira, *"A new look into the evolution of clusters literature. A bibliometric exercise"*, December 2007
Nº 256	Aurora A.C. Teixeira, *"Entrepreneurial potential in Business and Engineering courses … why worry now?"*, December 2007
Nº 255	Alexandre Almeida and Aurora A.C. Teixeira, *"Does Patenting negatively impact on R&D investment?An international panel data assessment"*, December 2007
Nº 254	Argentino Pessoa, *"Innovation and Economic Growth: What is the actual importance of R&D?"*, November 2007
Nº 253	Gabriel Leite Mota, *"Why Should Happiness Have a Role in Welfare Economics? Happiness versus Orthodoxy and Capabilities"*, November 2007
Nº 252	Manuel Mota Freitas Martins, *"Terá a política monetária do Banco Central Europeu sido adequada para Portugal (1999-2007)?"*, November 2007
Nº 251	Argentino Pessoa, *"FDI and Host Country Productivity: A Review"*, October 2007
Nº 250	Jorge M. S. Valente, *"Beam search heuristics for the single machine scheduling problem with linear earliness and quadratic tardiness costs"*, October 2007
Nº 249	T. Andrade, G. Faria, V. Leite, F. Verona, M. Viegas, O.Afonso and P.B. Vasconcelos, *"Numerical solution of linear models in economics: The SP-DG model revisited"*, October 2007
Nº 248	Mário Alexandre P. M. Silva, *"Aghion And Howitt's Basic Schumpeterian Model Of Growth Through Creative Destruction: A Geometric Interpretation"*, October 2007
Nº 247	Octávio Figueiredo, Paulo Guimarães and Douglas Woodward, *"Localization Economies and Establishment Scale: A Dartboard Approach"*, September 2007
Nº 246	Dalila B. M. M. Fontes, Luís Camões and Fernando A. C. C. Fontes, *"Real Options using Markov Chains: an application to Production Capacity Decisions"*, July 2007
Nº 245	Fernando A. C. C. Fontes and Dalila B. M. M. Fontes, *"Optimal investment timing using Markov jump price processes"*, July 2007
Nº 244	Rui Henrique Alves and Óscar Afonso, *"Fiscal Federalism in the European Union: How Far Are We?"*, July 2007
Nº 243	Dalila B. M. M. Fontes, *"Computational results for Constrained Minimum Spanning Trees in Flow Networks"*, June 2007
Nº 242	Álvaro Aguiar and Inês Drumond, *"Business Cycle and Bank Capital: Monetary Policy Transmission under the Basel Accords"*, June 2007
Nº 241	Sandra T. Silva, Jorge M. S. Valente and Aurora A. C. Teixeira, *"An evolutionary model of industry dynamics and firms' institutional behavior with job search, bargaining and matching"*, April 2007
Nº 240	António Miguel Martins and Ana Paula Serra, *"Market Impact of International Sporting and Cultural Events"*, April 2007
Nº 239	Patrícia Teixeira Lopes and Lúcia Lima Rodrigues, *"Accounting for financial instruments: A comparison of European companies' practices with IAS 32 and IAS 39"*, March 2007
Nº 238	Jorge M. S. Valente, *"An exact approach for single machine scheduling with quadratic earliness and tardiness penalties"*, February 2007
Nº 237	Álvaro Aguiar and Ana Paula Ribeiro, *"Monetary Policy and the Political Support*

	for a Labor Market Reform", February 2007
Nº 236	Jorge M. S. Valente and Rui A. F. S. Alves, *"Heuristics for the single machine scheduling problem with quadratic earliness and tardiness penalties"*, February 2007
Nº 235	Manuela Magalhães and Ana Paula Africano, *"A Panel Analysis of the FDI Impact on International Trade"*, January 2007
Nº 234	Jorge M. S. Valente, *"Heuristics for the single machine scheduling problem with early and quadratic tardy penalties"*, December 2006
Nº 233	Pedro Cosme Vieira and Aurora A. C. Teixeira, *"Are Finance, Management, and Marketing Autonomous Fields of Scientific Research? An Analysis Based on Journal Citations"*, December 2006
Nº 232	Ester Gomes da Silva and Aurora A. C. Teixeira, *"Surveying structural change: seminal contributions and a bibliometric account"*, November 2006
Nº 231	Carlos Alves and Cristina Barbot, *"Do low cost carriers have different corporate governance models?"*, November 2006
Nº 230	Ana Paula Delgado and Isabel Maria Godinho, *"Long term evolution of the size distribution of Portuguese cities"*, September 2006
Nº 229	Sandra Tavares Silva and Aurora A. C. Teixeira, *"On the divergence of evolutionary research paths in the past fifty years: a comprehensive bibliometric account"*, September 2006
Nº 228	Argentino Pessoa, *"Public-Private Sector Partnerships in Developing Countries: Prospects and Drawbacks"*, September 2006
Nº 227	Sandra Tavares Silva and Aurora A. C. Teixeira, *"An evolutionary model of firms' institutional behavior focusing on labor decisions"*, August 2006
Nº 226	Aurora A. C. Teixeira and Natércia Fortuna, *"Human capital, trade and long-run productivity. Testing the technological absorption hypothesis for the Portuguese economy, 1960-2001"*, August 2006
Nº 225	Catarina Monteiro and Aurora A. C. Teixeira, *"Local sustainable mobility management. Are Portuguese municipalities aware?"*, August 2006
Nº 224	Filipe J. Sousa and Luís M. de Castro, *"Of the significance of business relationships"*, July 2006
Nº 223	Pedro Cosme da Costa Vieira, *"Nuclear high-radioactive residues: a new economic solution based on the emergence of a global competitive market"*, July 2006
Nº 222	Paulo Santos, Aurora A. C. Teixeira and Ana Oliveira-Brochado, *"The 'de-territorialisation of closeness' - a typology of international successful R&D projects involving cultural and geographic proximity"*, July 2006
Nº 221	Manuel M. F. Martins, *"Dilemas macroeconómicos e política monetária: o caso da Zona Euro"*, July 2006

Editor: Sandra Silva (sandras@fep.up.pt)
Download available at:
http://www.fep.up.pt/investigacao/workingpapers/workingpapers.htm
also in http://ideas.repec.org/PaperSeries.html

www.fep.up.pt

Faculdade de Economia da Universidade do Porto
Rua Dr. Roberto Frias, 4200-464 Porto | Tel. 225 571 100
Tel. 225571100 | www.fep.up.pt