



INSTITUTO NACIONAL DE ESTATÍSTICA

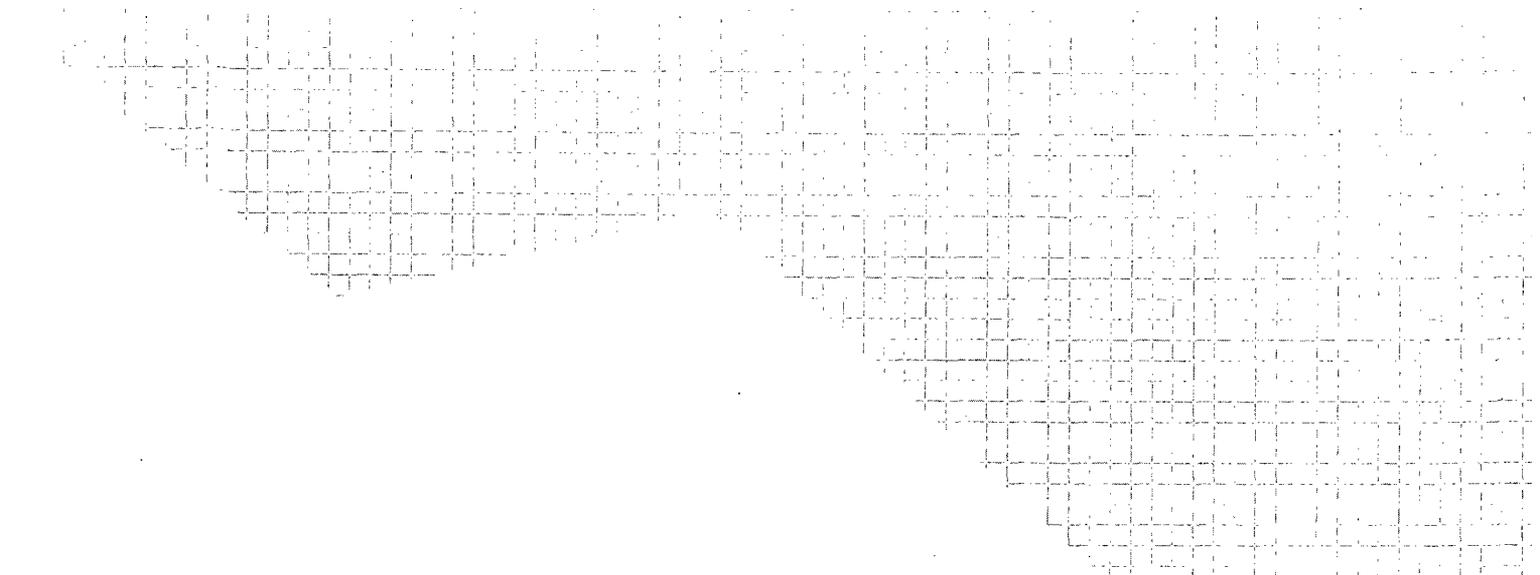
PORTUGAL

REVISTA DE ESTATÍSTICA

STATISTICAL REVIEW



VOLUME I
1º QUADRIMESTRE 2001



REVISTA DE ESTADÍSTICA

STATISTICAL REVIEW

VOLUME I

1º QUADRIMESTRE DE 2001

ÍNDICE

INDEX

- ARTIGOS

<i>ARTICLES</i>	5
DETERMINANTES DA COMPETITIVIDADE EMPRESARIAL NAS ÁREAS METROPOLITANAS DE LISBOA E PORTO <i>DETERMINANTS OF COMPANY COMPETITIVENESS IN THE METROPOLITAN AREAS OF LISBON AND OPORTO</i>	
Por/By: <i>Manuela Magalhães Hill</i>	7
M G ∞ QUEUE HEAVY-TRAFFIC SITUATION BUSY PERIOD LENGTH DISTRIBUTION (POWER AND PARETO SERVICE DISTRIBUTIONS) <i>DISTRIBUIÇÃO DO COMPRIMENTO DO PERÍODO DE OCUPAÇÃO DA FILA DE ESPERA M G ∞ EM SITUAÇÃO DE "HEAVY-TRAFFIC(DISTRIBUIÇÕES DE SERVIÇO POTÊNCIA E PARETO)</i>	
Por/By: <i>Manuel Alberto M. Ferreira</i>	27
REDUÇÃO DE DIMENSIONALIDADE ATRAVÉS DUMA ANÁLISE EM COMPONENTES PRINCIPAIS: UM CRITÉRIO PARA O NÚMERO DE COMPONENTES A RETER <i>REDUCING DIMENSIONALITY BY MEANS OF A PRINCIPAL COMPONENT ANALYSIS: A CRITERION TO CHOOSE HOW MANY PRINCIPAL COMPONENTS SHOULD BE RETAINED</i>	
Por/By: <i>Jorge F. C. L. Cadima</i>	37
MÉTODO DE TAGUCHI: CONTROLO DE QUALIDADE COM RECURSO NÃO-CONVENCIONAL AO DESENHO ESTATÍSTICO DE EXPERIÊNCIAS (DOE) <i>TAGUCHI'S METHOD: QUALITY CONTROL WITH A NON-CONVENTIONAL USE OF DESIGN OF EXPERIMENTS (DOE)</i>	
Por/By: <i>Paula M. Melo e Castro</i>	51
ÍNDICES DE CUSTO DE VIDA: METODOLOGIAS DE CÁLCULO <i>COST OF LIVING INDEXES: METHODOLOGIES</i>	
Por/By: <i>Paulo Parente e Teresa Bago d' Uva</i>	73
ÍNDICES DE CUSTO DE VIDA: UM ESTUDO SOBRE A REALIDADE PORTUGUESA <i>COST OF LIVING INDEXES: THE PORTUGUESE CASE THROUGH A MAGNIFYING GLASS</i>	
Por/By: <i>Teresa Bago d' Uva e Paulo Parente</i>	97
NUPCIALIDADE E CONJUNTURA ECONÓMICA. À PROCURA DE UM MODELO ECONOMÉTRICO PARA A TAXA BRUTA DE NUPCIALIDADE <i>NUPTIALITY AND ECONOMIC CONJUNCTURE. IN SEARCH OF AN ECONOMETRIC MODEL FOR THE GROSS MARRIAGE RATE</i>	
Por/By: <i>Paulo Parente</i>	139

EQUAÇÕES SALARIAIS E RENDIBILIDADE DA EDUCAÇÃO EM PORTUGAL
A CONDITIONAL QUANTILE REGRESSION APPROACH TO RETURNS TO EDUCATION IN
PORTUGAL

Por/By: Daniel Mota. 179

- INFORMAÇÕES
INFORMATIONS 215

ACTIVIDADES E PROJECTOS IMPORTANTES DO SISTEMA ESTATÍSTICO
NACIONAL

IMPORTANT ACTIVITIES AND PROJECTS OF THE NATIONAL STATISTICAL
SYSTEM 217

ACÇÕES DESENVOLVIDAS PELO INE NO ÂMBITO DA COOPERAÇÃO

ACTIONS ACHIEVED BY NSI IN THE SCOPE OF COOPERATION 231

CONGRESSOS, SEMINÁRIOS, COLÓQUIOS E CONFERÊNCIAS

CONGRESS, SEMINARS AND CONFERENCES 235

FUNDAMENTO, OBJECTO E ÂMBITO DA REVISTA

FOUNDATION, SUBJECT MATTER AND SCOPE OF THE REVIEW 249

NORMAS DE APRESENTAÇÃO DE ORIGINAIS PARA A REVISTA

RULES FOR SUBMITTING ORIGINALS TO THE REVIEW 251

ARTIGOS

VOLUME I

1º QUADRIMESTRE DE 2001

Determinantes da Competitividade Empresarial nas Áreas Metropolitanas de Lisboa e Porto

Autora:
Manuela Magalhães Hill

VOLUME I

1º QUADRIMESTRE DE 2001

DETERMINANTES DA COMPETITIVIDADE EMPRESARIAL NAS ÁREAS METROPOLITANAS DE LISBOA E PORTO¹

DETERMINANTS OF COMPANY COMPETITIVENESS IN THE METROPOLITAN AREAS OF LISBON AND OPORTO

Autora: Manuela Magalhães Hill

- Professora Associada do Instituto Superior de Ciências do Trabalho e da
Empresa

RESUMO:

- O artigo analisa factores associados com competitividade na indústria transformadora portuguesa obtidos a partir de um questionário aplicado a uma amostra representativa de empresas das áreas metropolitanas de Lisboa e Porto. Através da regressão logística, foram encontrados cinco factores de competitividade comuns às duas áreas, os quais conseguiram predizer correctamente se uma empresa pertencia ao grupo mais competitivo ou ao grupo menos competitivo em pouco mais de 70% das empresas. Encontraram-se ainda factores “regionais” associados com competitividade, ou seja, factores importantes para uma área mas não para a outra. Na globalidade, a competitividade surgiu associada com a inovação, com a eficiência na produção, com o sistema remuneratório e custos salariais, com o uso e gestão dos recursos humanos e com a utilização de serviços externos.

PALAVRAS-CHAVE:

- *Competitividade, indústria transformadora nas áreas metropolitanas de Lisboa e Porto, regressão logística.*

ABSTRACT:

- Factors associated with competitiveness in the Portuguese manufacturing industry were investigated using questionnaire data from representative samples of companies in the Metropolitan Areas of Lisboa and Porto. Logistic regression revealed five factors common to competitiveness in the two areas. These correctly predicted company status (Low competitive/High competitive) for a little over 70% of all companies. In addition, “regional” factors were found related to competitiveness in one area but not the other. Overall, competitiveness was

¹ O estudo aqui desenvolvido tem como ponto de partida a pesquisa realizada no DINÂMIA – Centro de Estudos sobre a Mudança Socioeconómica, do ISCTE, no âmbito de um projecto mais abrangente sobre *Competitividade e exclusão social nas áreas metropolitanas de Lisboa e Porto*, subsidiado pela FCT, programa PRAXIS, e levado a cabo por três instituições, DINÂMIA e CIES do ISCTE e Instituto de Sociologia da Faculdade de Letras da Universidade do Porto.

associated with Innovation and production efficiency, payment systems and salary costs, use and management of human resources and the use of external services.

KEY-WORDS:

- *Competitiveness, manufacturing industry in the Lisbon and Oporto metropolitan areas, logistic regression.*

1. INTRODUÇÃO

Portugal é frequentemente apelidado de «país intermédio» por se encontrar num estado de transição entre uma economia predominantemente agrícola e uma outra onde a indústria transformadora desempenha um papel importante. O ser membro da União Europeia e a evolução para uma globalização da economia faz com que, para ser eficaz no seu papel, a indústria transformadora portuguesa tenha de ser competitiva tanto no mercado interno como no mercado externo. Nos países mais industrializados, a produtividade e a competitividade dependem de vários factores tais como a capacidade para inovar (Ghodhal, Bartlett & Moran, 2000), maior investimento em novas tecnologias, maior investimento em I&D (Kald & Nilsson, 2000) e boa gestão dos recursos humanos (West and Patterson, 1999). Mas não é claro o modo como, num país intermédio como Portugal, estes factores estão associados com competitividade. Foi esta a razão do estudo cujos resultados aqui apresentamos. Este estudo foi desenvolvido sobre a indústria transformadora nas áreas metropolitanas de Lisboa (AML) e Porto (AMP) e os principais objectivos são:

- (1) Determinar a natureza dos principais factores associados com competitividade nas duas áreas metropolitanas em conjunto (factores gerais)
- (2) Determinar se há factores regionais, ou seja, se determinantes da competitividade numa área e não na outra, através de análises separadas.

2. MÉTODO

A amostra

O estudo teve como população alvo as empresas da indústria transformadora portuguesa com mais de 10 trabalhadores localizadas nas áreas metropolitanas de Lisboa e Porto e constantes da base BELEM (Base de estabelecimentos e empresas) de 1998 do INE. A amostra seleccionada é constituída por 678 empresas e representativa em termos de dimensão da empresa, sector de actividade (CAE a 2 dígitos) e área metropolitana.

Os dados

Os dados foram recolhidos em 1999 por meio de dois questionários: um deles continha 85 questões cobrindo seis áreas: caracterização da empresa (incluindo informação sobre o tipo de tecnologias utilizadas), actividades de I&D, gestão dos recursos humanos, organização da empresa e conteúdo do trabalho, e relações profissionais; o outro continha 15 questões dirigidas especificamente ao empresário ou gestor da empresa.

O indicador de competitividade

A literatura (ver, p. ex., Lança, 2000), aponta para que a competitividade das empresas possa ser medida em termos do ganho de quotas de mercado, do crescimento da empresa ou da sua rendibilidade. Por ser muito difícil dispor de informação sobre a quota de mercado das empresas e outras variáveis deste tipo², o indicador de competitividade aqui definido é o «*cash flow* em 1998 por trabalhador». Em rigor, este é mais um indicador de *performance* da empresa do que de competitividade. Mas, se consideramos que uma empresa que apresente um elevado rácio *cash flow*/dimensão pode ser tida como competitiva, caso contrário, a sua quota de mercado tende a baixar e, por consequência, a sua *performance*, parece-nos assim razoável utilizar *cash flow*/dimensão da empresa como um indicador indirecto de competitividade.

O rácio *cash flow*/dimensão é uma variável métrica mas não apresenta distribuição normal (é positivamente assimétrica e muito leptocúrtica). Deste modo, as empresas foram dicotomizadas em dois grupos (competitividade Alta e competitividade Baixa). Como esta variável vai ser utilizada como critério principal na análise estatística, usou-se a partição pelo método dos “grupos extremos”, ou seja, retiraram-se da amostra 6% de empresas situadas em torno da mediana da distribuição. Tem-se assim que um dos grupos (grupo de competitividade Baixa) contém empresas com valores do indicador métrico inferiores ao percentil 47 e o outro (grupo de competitividade Alta) contém empresas com valores superiores ao percentil 53. Este método é mais vantajoso que a simples partição pela mediana pois faz com que os grupos obtidos difiram mais um do outro em termos da medida considerada na dicotomização. A representatividade da amostra dos “grupos extremos” foi investigada.

Os testes de bondade de ajustamento mostram que a amostra reduzida é adequadamente representativa em termos de sector de actividade tanto para as duas áreas metropolitanas em conjunto ($\chi^2 = 20,25$, $P = 0,50$) como para a AML ($\chi^2 = 15,88$, $P = 0,80$) ou para a AMP ($\chi^2 = 13,59$, $P = 0,90$). Em termos do número de pessoas ao serviço tem-se uma boa correlação entre as amostras reduzidas e a amostra inquirida [$r = 0,864$ ($P < 0,0001$) para as duas áreas metropolitanas em conjunto, $r = 0,845$ ($P < 0,0001$) para a AML e $r = 0,881$, ($P < 0,0001$) para a AMP].

Os outros indicadores utilizados

Quatro novos indicadores foram construídos por forma a reduzir o número de variáveis usadas nas análises estatísticas – indicadores de *inovação*, de *intensidade das tecnologias de informação*, de *mudança organizacional* ocorrida nos últimos cinco anos e de *mudança na gestão dos recursos humanos* introduzida nos últimos dois anos. Cada um destes indicadores é a primeira componente resultante da aplicação de uma análise de componentes principais não linear a variáveis medidas em escalas ordinais dicotómicas. A percentagem de variância que cada um dos factores

² Entenda-se por quota de mercado o rácio entre o crescimento das vendas da empresa e o crescimento do respectivo mercado. São duas as razões porque é quase impossível obter informação sobre a quota de mercado das empresas. Primeiro, a maior parte das empresas, se é que dispõem desta informação, é relutante em fornecê-la, e segundo, as empresas que exportam (e 53% das empresas da amostra vendem para exportação), fazem-no para mercados diversificados em diferentes países e as bases de dados disponíveis não contém informação sobre este tipo de quota de mercado por empresa.

explica, a fiabilidade interna (*Cronbach's alpha*) e o número de variáveis que entram em cada factor são:

- *Inovação* (32,8%: $\alpha = 0,764$; 11 variáveis combinadas)
- *Intensidade das tecnologias de informação* (27,2%: $\alpha = 0,757$; 12 variáveis combinadas)
- *Mudança organizacional* (40,7%: $\alpha = 0,632$; 5 variáveis combinadas)
- *Mudança na gestão de recursos humanos* (31,3%: $\alpha = 0,612$; 7 variáveis combinadas)

Para além destes quatro indicadores, três outros conjuntos de variáveis foram sujeitos a análises de componentes principais separadas. O primeiro diz respeito a serviços adquiridos a terceiros, o segundo a características que a empresa considera serem a sua vantagem competitiva e o terceiro à estrutura organizacional da empresa. Cada conjunto de variáveis deu lugar a 4 componentes principais que vão ser usadas como variáveis explicativas nas análises estatísticas que a seguir se apresentam.

Análises estatísticas dos dados

Para encontrar os principais factores associados com competitividade, tanto para as duas áreas metropolitanas em conjunto (objectivo 1) como para cada uma das áreas em separado (objectivo 2), é preciso definir o que é um factor «principal» associado com competitividade. Para evitar o uso de uma opinião subjectiva consideram-se factores «principais» aqueles que contribuem significativamente para discriminar as empresas mais competitivas das menos competitivas. Como os dados não satisfaziam os pressupostos da análise discriminante, e como algumas das variáveis independentes eram não métricas, a discriminação foi feita usando a regressão logística binária (método forward *LR*). Aplicaram-se três regressões, uma para as duas áreas metropolitanas em conjunto e as outras duas para cada uma das áreas em separado.

Os quatro indicadores acima referidos foram usados nas regressões logísticas juntamente com um grupo de variáveis adicionais (incluindo as componentes principais acima referidas) para uma primeira selecção de variáveis explicativas. Este procedimento tornou-se necessário por causa do elevado número de potenciais variáveis independentes em relação ao número de casos. Esta pré-selecção de variáveis foi feita seleccionando pelo menos uma variável adicional relativa a cada uma das partes principais dos questionários: *Caracterização da empresa*, *Investigação e Desenvolvimento (I&D)*, *Recursos humanos*, *organização da empresa e conteúdo do trabalho*, *Relações profissionais e Inquérito aos empresários*.

O critério usado para seleccionar as variáveis foi o seguinte:

Variáveis não métricas: a variável tinha de ter uma associação significativa com a competitividade ao nível $P < 0,005$.

Variáveis métricas: a variável tinha de ter valores médios para os grupos de Competitividade Alta e Competitividade Baixa que diferissem significativamente ao nível $P < 0,005$.

Usou-se um nível de significância relativamente severo para manter um nível adequado da taxa de erro *family-wise* nas comparações múltiplas efectuadas no processo de pré-selecção. O teste de bondade do ajustamento de Hosmer-Lemeshow foi aplicado ao modelo obtido na última iteração e para a primeira regressão (aplicada

aos dados das duas áreas metropolitanas em conjunto) fez-se uma validação cruzada usando o método Holdout. Os modelos finais obtidos foram usados com um ponto de corte de 0,5 para classificar as empresas e a bondade da classificação foi avaliada usando o Kappa de Cohen.

3. RESULTADOS

3.1. A ANÁLISE PARA AS DUAS ÁREAS METROPOLITANAS AGREGADAS

Para além dos quatro indicadores – *intensidade da tecnologia de informação, inovação, mudança na gestão dos recursos humanos e inovação organizacional*, entraram na análise mais 34 variáveis, seis das quais são combinações ponderadas de variáveis obtidas por aplicação da análise de componentes principais ou factorial.

O modelo de regressão logística mostrou que o ajustamento era bom ($\chi^2 = 8,567$; g.l. = 8; $P = 0,380$). O modelo seleccionou as cinco variáveis seguintes: *intensidade de tecnologias de informação, inovação, utilização da técnica MRP no planeamento das necessidades de produção, percentagem dos custos salariais em relação aos custos totais de exploração, utilização dos serviços de uma Universidade para além do recrutamento*.

A classificação dos casos

As cinco variáveis em conjunto conseguiram classificar correctamente 72,8% das empresas (74,2% de predição correcta para as empresas do grupo de competitividade Baixa e 71,5% para as empresas do grupo de competitividade Alta). O valor de Kappa para a classificação global foi 0,648, ($P < 0,0001$) e o coeficiente de validade preditiva do modelo de regressão foi $\phi = 0,457$ ($P < 0,0001$).

Para a validação cruzada, a amostra global foi dividida aleatoriamente em duas subamostras A e B. A subamostra A foi então utilizada para deduzir o modelo de regressão logística a usar para predizer a que grupo pertencem as empresas nas duas subamostras. Este método apresentou um bom ajustamento aos dados ($\chi^2 = 9,140$; g.l. = 8; $P = 0,331$). Obteve-se uma predição correcta de 75,0% para a subamostra A, e de 72,1% para a subamostra B.

O modelo deduzido para as duas áreas metropolitanas agregadas apresentou uma predição correcta para 71,3% das empresas na AML (70,2% de predição correcta para as empresas do grupo competitividade Baixa e 72,1% para as empresas do grupo competitividade Alta) e uma predição correcta para 74,6% das empresas da AMP (77,5% das empresas do grupo competitividade Baixa e 70,5% das empresas do grupo competitividade Alta).

Deste modo, podemos concluir que o modelo de regressão é tecnicamente adequado pois ajustou-se bem aos dados, produziu uma boa validação cruzada e mostrou validade preditiva razoável. Foi particularmente interessante verificar que a validade preditiva era aproximadamente a mesma para as duas áreas metropolitanas (ϕ

= 0,420; $P < 0,0001$ na AML e $\phi = 0,480$; $P < 0,0001$ na AMP). A adequacidade técnica do modelo suporta a conclusão de que as cinco variáveis incluídas representam factores importantes de competitividade.

O perfil das empresas mais e menos competitivas

Examinemos agora mais em pormenor os cinco factores encontrados com o objectivo de desenvolver uma panorâmica mais detalhada de como é que as empresas mais competitivas diferem das menos competitivas em termos destes factores. O Quadro 1 apresenta os valores médios para os indicadores intensidade de tecnologias de informação e inovação, e para a variável percentagem dos custos salariais em relação aos custos totais de exploração. O mesmo quadro apresenta a percentagem de empresas que utilizam a técnica de planeamento das necessidades de produção MRP e a percentagem que costuma recorrer a outros serviços de uma universidade para além do recrutamento de jovens licenciados.

Os resultados apresentam um padrão idêntico para as duas áreas metropolitanas e os valores absolutos são muito semelhantes nas duas áreas. O grupo mais competitivo é em média mais inovador do que o menos competitivo, possui mais tecnologia de informação, tende a usar mais a técnica MRP no planeamento das necessidades de produção e tende a usar mais os serviços das universidades para outros fins que o recrutamento de quadros. Por outro lado, o rácio custos salariais/custos totais médio no grupo mais competitivo é significativamente inferior ao do menos competitivo,.

Quadro 1 - Valores médios e percentagens dos factores associados com competitividade

FACTOR DE COMPETITIVIDADE	LISBOA		PORTO		AS DUAS ÁREAS	
	Competitividade		Competitividade		Competitividade	
	Baixa	Alta	Baixa	Alta	Baixa	Alta
Intensidade das tecnologias de informação (valor médio)	-0,09	0,43	-0,23	0,32	-0,17	0,39
Inovação (valor médio)	4,75	5,30	4,82	5,27	4,79	5,29
Utilização da técnica de planeamento e gestão de produção MRP (percentagem das empresas)	24,2 %	39,9 %	17,7 %	38,6 %	20,6 %	39,4 %
Percentagem dos custos salariais em relação aos custos totais de exploração (percentagem das empresas)	33,0 %	24,1 %	38,0 %	20,2 %	35,7 %	22,5 %
A empresa costuma recorrer a outros serviços da universidade (percentagem das empresas)	11,1 %	20,3 %	7,3 %	21,6 %	9,0 %	20,8 %

Tentando ir um pouco mais além, constata-se que (ver Quadro 2), considerando as duas áreas em conjunto, é mais provável que sejam as empresas mais competitivas a possuírem as seguintes tecnologias de informação específicas: Computadores de médio e grande porte, componentes robotizados, equipamento CAD, equipamento CAD/CAM, rede de comunicação entre postos de trabalho, controlo automatizado de secções fabris e gestão assistida por computador. O Quadro 2 mostra ainda que, em geral, o mesmo resultado se verifica *qualquer que seja o escalão de dimensão da empresa*.

Quadro 2 - **Percentagem de empresas com vários tipos de tecnologia de informação por grupo de competitividade**

(análise por dimensão da empresa)

TIPO DE EQUIPAMENTO	DIMENSÃO DA EMPRESA (Nº de trabalhadores)							
	< 50		(50 – 100)		(101 – 200)		> 200	
	Competitividade		Competitividade		Competitividade		Competitividade	
	Baixa	Alta	Baixa	Alta	Baixa	Alta	Baixa	Alta
Computador de médio/grande porte	16,7	38,2	39,6	54,3	43,3	68,2	65,8	76,7
Rede de comunicação entre postos de trabalho	15,7	40,8	47,2	67,4	56,7	56,8	68,4	83,3
CAD	2,9	6,6	20,8	15,2	23,3	31,8	39,5	51,7
CAD/CAM	1,0	0,0	3,8	8,7	0,0	22,7	21,1	28,3
Gestão da produção assistida por computador	4,9	17,1	22,6	34,8	26,7	43,2	50,0	63,3
Controlo automatizado de secções fabris	1,0	7,9	1,9	21,7	10,0	27,3	23,7	41,7
Componentes robotizados	4,9	5,3	9,4	19,6	6,7	11,4	21,1	43,3

Nota: Para cada escalão de dimensão de empresa e para cada tipo de equipamento, a percentagem de empresas que possuem equipamento é maior para as empresas mais competitivas do que para as menos competitivas, à excepção das duas situações que se encontram a bold.

O indicador de inovação conjuga a inovação no produto com a inovação no processo produtivo. Em termos de inovação no produto, o Quadro 3 mostra que nos cinco anos que antecederam o estudo, a percentagem de empresas mais competitivas que introduziu novos produtos, melhorou os produtos existentes (mais em qualidade do que no desenho) e/ou introduziu alterações nas matérias primas do produto, foi significativamente maior que a percentagem de empresas menos competitivas. No mesmo período, foram também mais empresas mais competitivas a alterarem o processo produtivo.

A relação entre inovação no produto e competitividade parece depender, até certo ponto, do contexto em que se processa a inovação. Se a empresa introduziu produtos novos nos últimos cinco anos, a probabilidade condicionada de ser mais competitiva é 0,615, a qual aumenta para 0,763 se a empresa tiver um nível elevado de tecnologia de informação (acima da mediana) e usar a técnica MRP no planeamento das necessidades de produção. Analogamente, se a empresa apenas melhorou os produtos, a probabilidade condicionada de estar no grupo mais competitivo é 0,567, probabilidade esta que aumenta para 0,726 se a empresa tiver um elevado nível de tecnologia de informação (acima da mediana) e paralelamente usar a técnica MRP no planeamento das necessidades de produção.

Quadro 3 - Percentagem de empresas que fizeram vários tipos de inovação

TIPO DA INOVAÇÃO	AML competitividade		AMP competitividade		AML & AMP competitividade	
	Baixa	Alta	Baixa	Alta	Baixa	Alta
A empresa introduziu produtos completamente novos nos últimos cinco anos.	27,3	34,8	18,5	36,4	22,4	35,4 **
A empresa introduziu produtos melhorados nos últimos cinco anos	55,6	68,1	57,3	80,7	56,5	73,0 *
Mudança nas matérias primas do produto nos últimos cinco anos	19,2	24,6	25,0	26,1	22,4	25,2
Mudança no processo de fabrico do produto nos últimos cinco anos	28,3	37,7	19,4	39,8	23,3	38,5 **
Mudança na qualidade do produto nos últimos cinco anos	38,4	55,1	41,9	58,0	40,4	56,2 *
Mudança no design do produto nos últimos cinco anos	31,3	39,1	33,1	40,9	32,3	39,8
Reestruturações no processo produtivo nos últimos dois anos	32,3	47,1	35,5	39,8	34,4	44,2
Criação de novos produtos /mudança no(s) produto(s) nos últimos dois anos	28,3	44,2	27,4	34,1	27,8	40,3 *

Nota: Os números a bold e itálico indicam que a percentagem de empresas que faz inovação é significativamente maior no grupo de competitividade Alta que no grupo de competitividade Baixa.

(* = $P < 0,05$ and ** = $P < 0,01$).

Os serviços universitários mais solicitados pelas empresas competitivas são consultoria e investigação aplicada, os quais são normalmente requeridos de maneira *ad hoc* para apoio à resolução de problemas, em particular, problemas ambientais. Existe contudo uma influência da dimensão da empresa, 36,7% das empresas com mais de 200 trabalhadores usam estes serviços em comparação com apenas 15,1% de empresas com menos de 200 trabalhadores. Os valores correspondentes para as empresas menos competitivas são 13,2% e 8,1%.

3.2. A ANÁLISE PARA A ÁREA METROPOLITANA DE LISBOA

Para além dos quatro indicadores – *intensidade da tecnologia de informação, inovação, mudança na gestão dos recursos humanos e inovação organizacional*, entraram na análise outras 20 variáveis, cinco das quais são combinações ponderadas de variáveis obtidas por aplicação da análise de componentes principais ou análise factorial.

O modelo de regressão logística final mostrou existirem cinco factores principais de competitividade para as empresas da AML – *inovação organizacional, aquisição a terceiros de serviços financeiros/formação* (componente principal), *percentagem de custos salariais em relação aos custos totais de exploração, aumentos salariais individualizados considerando o mérito/desempenho e importância do preço do produto como vantagem concorrencial* (componente principal). O teste de bondade de ajustamento de Hosmer-Lemeshow mostra existir um bom ajustamento ($\chi^2 = 6,480$; g.l. = 8; $P = 0,594$).

A classificação dos casos

Os cinco factores predizem correctamente, 71,4% das empresas da AML (70,0% das empresas pertencentes ao grupo de competitividade Baixa e 72,6% das empresas pertencentes ao grupo de competitividade Alta). O valor de Kappa para a classificação global foi 0,415 ($P < 0,001$) e o coeficiente de validade preditiva do modelo foi $\phi = 0,417$ ($P < 0,0001$).

O perfil das empresas mais e menos competitivas

O Quadro 4 diz-nos que em média é significativamente mais provável que as empresas mais competitivas pratiquem aumentos salariais individualizados baseados no mérito /desempenho e apresentem baixos rácios custos salariais/custos totais de exploração. As empresas do grupo mais competitivo também efectuaram significativamente mais alterações na organização nos últimos dois anos. Considerando as variáveis que integram o indicador de mudança organizacional, constata-se que quer as maiores empresas quer as mais pequenas deste grupo tendem a criar novos procedimentos de planeamento, controlo e gestão da produção. Para as empresas com menos de 200 trabalhadores é significativamente mais provável que, nos últimos dois anos, tenham sido as mais competitivas a descentralizar a tomada de decisão, e/ou melhorar a circulação de informação, e/ou criar grupos de trabalho.

Quadro 4 - Factores de competitividade para a área metropolitana de Lisboa

FACTOR DE COMPETITIVIDADE	COMPETITIVIDADE	
	Baixa	Alta
Aumentos salariais individualizados considerando o mérito/desempenho (% de empresas que fazem)	42,4 %	60,9 %
Percentagem dos custos salariais em relação aos custos totais de exploração (valor médio)	33,0 %	24,1 %
Indicador: Inovação na organização (valor médio)	- 0,146	0,495
Aquisição a terceiros de vários serviços (valor médio da Componente Principal)	- 0,069	0,415
Importância do preço do produto como vantagem de concorrência (valor médio da Componente Principal)	0,281	- 0,201

Para as maiores empresas (mais de 200 trabalhadores) é significativamente mais provável serem as mais competitivas a criar novos departamentos ou secções.

As empresas mais competitivas adquirem significativamente mais serviços a terceiros. Este resultado é, contudo, enganoso porque o modo como as empresas adquirem esses serviços está a ser influenciado pela dimensão da empresa. Nas empresas com menos de 200 trabalhadores é mais provável que as mais competitivas recorram a terceiros para serviços de consultoria técnica, formação profissional ou consultoria financeira/auditoria, e é menos provável que o façam no caso de serviços de contabilidade. Nas empresas com mais de 200 trabalhadores, *não* há associação significativa entre competitividade e o recurso a serviços externos. A maioria das maiores empresas, quer sejam mais ou menos competitivas, utilizam terceiros para consultoria financeira/auditoria, consultoria técnica e formação profissional mas apenas 10% o faz no que respeita a serviços de contabilidade.

O Quadro 4 mostra ainda que o grupo mais competitivo apresenta um valor médio negativo e o menos competitivo um valor médio positivo na componente principal importância do preço do produto como vantagem concorrencial. Acontece porém que a maioria das empresas tanto mais como menos competitivas considera que o preço dos seus produtos é uma vantagem competitiva importante ou muito importante que possuem, sendo que as empresas mais competitivas atribuem menos importância ao preço do produto que as menos competitivas.

3.3. A ANÁLISE PARA A ÁREA METROPOLTANA DO PORTO

Para além dos quatro indicadores, entraram na análise outras 13 variáveis, 2 das quais são combinações ponderadas de variáveis obtidas por aplicação da análise de componentes principais.

O modelo de regressão logística apresentou seis factores principais de competitividade para as empresas da AMP – *mudança na gestão de recursos humanos, manutenção de contactos/relações de cooperação com centros de I&D, utilização da técnica MRP no planeamento das necessidades de produção, percentagem de custos salariais em relação aos custos totais de exploração, aumento salarial geral de acordo com as convenções colectivas aplicáveis e aumentos salariais individualizados considerando o contributo para os objectivos da empresa*. O teste de bondade de ajustamento de Hosmer-Lemeshow mostra que o ajustamento é adequado ($\chi^2 = 13,227$; g.l. = 8; $P = 0,104$).

A classificação dos casos

Os seis factores predizem correctamente 72,9% das empresas da AMP em termos de competitividade Alta ou Baixa (72,5% para o grupo competitividade Alta e 73,3% para o grupo competitividade Baixa). O valor de Kappa é 0,452 ($P < 0,001$) e o coeficiente de validade preditiva é $\phi = 0,454$ ($P < 0,0001$).

O perfil das empresas mais competitivas e menos competitivas

As empresas mais competitivas mantêm significativamente mais ligações com centros de I&D do que as menos competitivas, em particular, as que têm mais de 200 trabalhadores (66,7% mantiveram ligações nos últimos dois anos, sendo que destas 47,6% usaram os centros para resolver problemas de produção). A competitividade também surge relacionada com o planeamento de produção: uma percentagem significativamente maior de empresas mais competitivas usam a técnica MRP no planeamento das necessidades de produção.

O rácio médio custos salariais/custos totais é significativamente mais baixo no grupo mais competitivo, sendo *mais* provável que este grupo pratique aumentos salariais individualizados considerando o contributo para os objectivos da empresa, e *menos* provável que pratique aumentos salariais gerais de acordo com as convenções colectivas aplicáveis. De facto, esta prática de aumentos salariais está negativamente associada com competitividade na AMP.

Quadro 5 - Factores de competitividade para a área metropolitana do Porto

FACTOR DE COMPETITIVIDADE	COMPETITIVIDADE	
	BAIXA	ALTA
A empresa mantém contactos/relações de cooperação com centros de I&D. (percentagem das empresas)	8,9 %	26,1 %
Utilização da técnica MRP no planeamento das necessidades de produção. (percentagem das empresas)	17,7 %	38,6 %
Percentagem de custos salariais em relação aos custos totais de exploração. (valor médio)	38,0 %	20,2 %
Aumento salarial geral de acordo com as convenções colectivas aplicáveis. (percentagem das empresas)	68,5 %	45,5 %
Aumentos salariais individualizados considerando o contributo para os objectivos da empresa. (percentagem das empresas)	21,8 %	43,2 %
Indicador de Inovação em Recursos Humanos. (valor médio)	4,820	5,208

Foram as empresas mais competitivas que introduziram mais mudanças na gestão dos recursos humanos. O Quadro 6 apresenta 7 variáveis usadas na construção do indicador. As empresas mais competitivas só diferem significativamente das menos competitivas em três tipos de mudança na gestão dos recursos humanos – selecção e recrutamento de pessoal, formas de divulgação da informação e condições de higiene e segurança no trabalho.

Quadro 6 - Percentagem de empresas que nos últimos dois anos praticou alteração na gestão dos recursos humanos

TIPO DE INOVAÇÃO (Alterações na gestão de recursos humanos nos últimos dois anos)	TODAS AS EMPRESAS		EMPRESAS ATÉ 200 TRABALHADORES		EMPRESAS COM MAIS DE 200 TRABALHADORES	
	Competitividade		Competitividade		Competitividade	
	Baixa	Alta	Baixa	Alta	Baixa	Alta
Seleção e recrutamento de pessoal	17,7 %	29,5 % *	16,0 %	23,9 %	27,8 %	47,6 %
Política de remunerações	18,5 %	21,6 %	17,9 %	19,4 %	22,2 %	28,6 %
Redução do efectivo	27,4 %	23,9 %	24,5 %	14,9 %	44,4 %	52,4 %
Política de promoções	4,8 %	8,0 %	4,7 %	6,0 %	5,6 %	14,3 %
Política de formação profissional	16,9 %	23,9 %	10,4 %	16,4 %	55,6 %	47,6 %
Formas de divulgação da informação	6,5 %	25,0 % ***	5,7 %	17,9 % **	11,1 %	47,6 % **
Condições de higiene e segurança no trabalho	29,0 %	45,5 % **	24,5 %	38,8 % *	55,6 %	66,7 %

Nota: Números a bold e itálico indicam que a diferença entre as empresas mais competitivas e menos competitivas é significativa

* = P < 0,05

** = P < 0,01

*** = P < 0,0001

4. DISCUSSÃO

Os resultados mostram que os factores principais associados com competitividade podem ser agrupados em quatro áreas: *inovação e eficiência na produção, salários, recursos humanos e uso de serviços externos*. Analisemos cada um deles em separado.

Inovação e eficiência na produção

É do conhecimento geral que a competitividade na indústria transformadora dos países mais industrializados depende em grande parte da inovação (Ghoshal, Bartlett e Moran, 2000). Mas, para que uma empresa se mantenha ou torne competitiva não é suficiente criar novos produtos. Para conseguir vantagem competitiva através da inovação, a criação ou melhoria de novos produtos deve atingir o mercado na quantidade certa, na qualidade certa e em tempo certo. Para isso, importa que o planeamento da produção e o processo produtivo sejam eficientes e, como Kald e Nilsson (2000) afirmam, a eficiência da produção nos países tecnologicamente mais avançados tem aumentado grandemente com investimento em tecnologia da informação baseada em automação. Nas duas áreas metropolitanas passa-se, em certa medida, algo de semelhante, ou seja, as empresas mais competitivas são mais inovadoras em termos do desenvolvimento de produtos novos ou produtos melhorados (em especial, melhoria na qualidade), usam mais tecnologias de informação do tipo automação ou controlo automático da produção e métodos eficientes de planeamento das necessidades de produção. É preciso contudo não exagerar nestas semelhanças; a inovação no produto das empresas mais competitivas foi mais em termos de melhoria nos produtos do que produtos novos. Como positivo, observa-se a existência de associação entre competitividade e o uso de tecnologia da informação do tipo automação ou controlo automático da produção tanto nas pequenas ou médias empresas como nas grandes empresas. Embora a percentagem de pequenas ou médias empresas competitivas a utilizar este tipo de tecnologia da produção seja apreciavelmente mais pequena que a percentagem de grandes empresas na mesma situação, a evidência conduz a um razoável reconhecimento das vantagens resultantes do investimento em novas tecnologias da produção. Do ponto de vista nacional, isto é encorajador pois a dimensão média das empresas portuguesas pertencentes ao sector da indústria transformadora é inferior às suas congéneres nos países mais industrializados.

Salários

São dois os aspectos dos salários associados com competitividade – custos salariais como percentagem dos custos totais e tipo de aumento salarial praticado. Em ambas as áreas, o rácio médio custos salariais/custos totais de exploração nas empresas mais competitivas é significativamente mais baixo do que nas menos competitivas. No entanto, esse rácio é apreciavelmente mais pequeno na AMP que na AML. Na AML o rácio para as empresas mais competitivas é, em média, 73% do das empresas menos competitivas enquanto que, na AMP, o valor correspondente é 53%. Tem-se assim que, em cada uma das áreas, e em especial na AMP, as empresas mais competitivas têm em média uma vantagem sobre as menos competitivas em termos do peso dos seus custos salariais nos seus custos totais. Afirma-se com

frequência (veja-se p. ex., Rodrigues, 1991) que uma das vantagens competitivas da indústria transformadora portuguesa são os baixos custos da mão-de-obra. Esta conclusão baseia-se na comparação de Portugal com outros países e, muito embora isso possa explicar, pelo menos em parte, a competitividade das empresas portuguesas que operam em mercados de exportação, não consegue explicar o porquê de, *em Portugal*, as empresas competitivas apresentarem baixos rácios custos salariais/custos totais. Infelizmente, o nosso estudo não conseguiu mostrar isto. É possível que o maior investimento feito pelas empresas competitivas em tecnologias de informação associadas com produção tenha feito aumentar os custos totais e conseqüentemente baixar o rácio. Também é possível que as empresas mais competitivas (que são em média as maiores) paguem salários mais baixos. Os dados que dispomos não nos permitem saber se qualquer um destes pontos se verifica pelo que é necessário prosseguir a investigação para tentar perceber quais as determinantes dos rácios custos salariais/custos totais de exploração.

Passando para o segundo aspecto dos salários acima referido, observa-se que, em ambas as áreas metropolitanas, a competitividade está relacionada com um sistema remuneratório baseado num determinado tipo de critério comportamental. Na AML, o critério é o mérito/desempenho enquanto que na AMP é a contribuição para os objectivos da empresa. Em contraste, aumentos gerais de acordo com as convenções colectivas aplicáveis a grandes grupos de trabalhadores apresentam-se negativamente relacionados com competitividade em ambas as áreas, embora apenas significativamente na AMP. Estes resultados são compreensíveis em termos da motivação do trabalhador e da contribuição que isto possa ter para a competitividade da empresa. Não há dúvida que para uma empresa se tornar ou manter competitiva precisa de ter uma mão-de-obra motivada e a motivação depende de um conjunto de factores, inclusivé factores de natureza financeira. Do nosso ponto de vista, aumentos salariais praticados com base num dos critérios comportamentais acima referidos estão relacionados com competitividade porque são entendidos pelos trabalhadores como um reconhecimento do seu valor individual para a empresa, e tal reconhecimento é uma fonte poderosa de motivação (Herzberg, Mausner e Snyderman, 1959). Por outro lado, aumentos gerais (em vez de individuais) de acordo com as convenções colectivas aplicáveis tendem a não ser entendidos como um reconhecimento pela empresa pelo valor individual de cada trabalhador. Tais aumentos podem ter como efeito reduzir ou prevenir a insatisfação do trabalhador mas é pouco provável que tenham um efeito motivacional positivo necessário a uma empresa competitiva

Recursos humanos

Na AML, em particular nas pequenas e médias empresas (<200 trabalhadores), a competitividade está associada com a inovação efectuada na utilização dos recursos humanos como parte do processo de produção, em termos de descentralização na tomada de decisão, melhoria da circulação de informação e criação de grupos/equipas de trabalho. Não é de surpreender que estas mudanças estejam relacionadas com competitividade pois representam tipos de enriquecimento do trabalho que West e Patterson (1999), num estudo longitudinal em larga escala efectuado no Reino Unido, encontraram estarem associados tanto com a satisfação do trabalhador como com a performance económica da empresa. Na AMP, observaram-se mudanças menos radicais na gestão dos recursos humanos associadas com competitividade, em particular, melhoria no processo de selecção e recrutamento, na circulação de informação e nas condições de higiene e segurança no trabalho.

A aquisição de serviços a terceiros

É prática usual nos vários países industrializados a ligação universidade-empresa para efeitos que não o recrutamento de jovens licenciados. Os benefícios para a produtividade da empresa conseguidos pela chamada «ligação de conhecimento» foram recentemente apresentados por Lissenburgh (1999). A situação é semelhante em Portugal pois, em ambas as áreas metropolitanas, a competitividade surge relacionada com o uso das universidades para efeitos de consultoria ou investigação aplicada para resolução de problemas. Cerca de 1 em 5 empresas mais competitivas utilizam tais serviços embora mais de 1 em 3 empresas com mais de 200 trabalhadores o fazem.

As duas áreas metropolitanas diferem em termos dos tipos de serviços usados por agências externas. No Porto, mas não em Lisboa, os serviços externos usados pelas empresas mais competitivas estão relacionados com I&D e produção, em particular no caso das maiores empresas. A maioria destas (67%) mantiveram ligações com centros de I&D nos últimos anos e quase metade (47%) usaram os centros para assistência na resolução de problemas de produção. Em Lisboa, apenas uma minoria (41%) de grandes empresas mantiveram contactos com centros de I&D nos últimos anos e só 28% usaram os centros para assistência na resolução de problemas de produção. Contudo, as duas áreas não diferem em termos da percentagem de empresas mais competitivas que desenvolvem a sua própria I&D (36,4% na AMP e 36,2% na AML). O facto de em cada área pouco mais de 1 em 3 empresas desenvolverem actividades de I&D é, até certo ponto, encorajador, embora, em termos absolutos, isso não seja uma proporção elevada e talvez explique porque é que, como afirmámos antes, é mais provável que a inovação no produto nas duas áreas tenha sido em termos de melhoria de produtos existentes e não em termos de desenvolvimento de novos produtos. A criação de produtos novos requer um maior uso de actividades de I&D – ponto suportado pela constatação de que 49,7% das empresas que desenvolveram actividades de I&D nos últimos anos introduziram recentemente produtos completamente novos (em comparação com 21,1% das empresas que não desenvolveram actividades de I&D). A AMP mantém mais ligações com centros de I&D que a AML, o que sugere que I&D seja talvez mais importante naquela área, sendo possível que o uso de centros de I&D seja encarado como um método mais efectivo em termos de custos de investimento em I&D:

Em Lisboa, é significativamente mais provável que as pequenas e médias empresas mais competitivas adquiram a terceiros serviços de consultoria financeira/auditoria e de formação profissional mas *não* serviços de contabilidade. A maioria das empresas com mais de 200 trabalhadores, quer sejam mais ou menos competitivas, adquirem a terceiros aqueles serviços e serviços técnicos mas não serviços de contabilidade. As razões deste facto não são bem claras, embora *talvez* possam estar relacionadas com o tipo de gestor da empresa. Muito embora não se encontrem diferenças significativas em termos de idade ou antiguidade na empresa entre os gestores das empresas mais e menos competitivas com menos de 200 trabalhadores, o facto é que o nível de instrução é significativamente mais elevado nas empresas mais competitivas: 76,6% destes gestores tinham como habilitações literárias o ensino médio/superior (ou mais) enquanto que com estas habilitações há apenas 47,1% de gestores nas empresas menos competitivas. Talvez os gestores com maior nível de instrução considerem ser conveniente para a sua empresa disporem de um sistema de contabilidade interno para que possam fazer um acompanhamento sistemático da posição financeira da empresa, e que a compra ocasional de outros serviços especializados tais como formação profissional, consultoria

financeira/auditoria é um investimento capaz de ser compensado por um aumento de competitividade. Em alternativa, é possível que as pequenas ou médias empresas mais competitivas tenham, como política, admitir gestores com níveis de habilitação mais elevados. Não temos evidência para suportar nem um nem outro destes pontos de vista, pelo que sugerimos que deverão ser melhor explorados numa futura investigação.

REFERÊNCIAS BIBLIOGRÁFICAS

- GHOSHAL, S., BARTLETT, C. & MORAN, P. (2000) Um novo Manifesto para a gestão. *Revista Portuguesa de Gestão*, Inverno, 15, 1, Inverno, 18-29.
- HERZBERG, F. MAUSNER, B. & SNYDERMAN, B. (1959) *The Motivation to Work*. John Wiley & Sons, New York.
- KALD, M. e NILSSON, F. (2000) Performance Measurement at Nordic Companies, *European Management Journal*, vol. 18, nº 1, Fevereiro, 113-127.
- LISSENBURGH, S. (1999) Knowledge kinks, *New Economy*, vol. 6, Março, 28-32.
- PFEFFER, J. (1998) "Six Dangerous Myths about pay", *Harvard Business Review*, Maio-Junho, 109-119.
- RAPPAPORT, A. (1999) "New Thinking on how to Link Executive Pay with Performance", *Harvard Business Review*, Março-Abril, 91-101.
- RODRIGUES, M. J. (1991) *Competitividade e Recursos Humanos*, Publicações Dom Quixote, Lisboa.
- SALAVISA LANÇA, I. & HILL, A.B. (1999) Patterns of technological innovation in intermediate countries – An alternative index and some conclusions. Artigo apresentado na Conferência EAEPE, Praga, Republica Checa.
- SALAVISA LANÇA, I. S. (2000) *A Indústria Portuguesa. Especialização Internacional e Competitividade*. Lisboa, Celta.
- SIMON, H. (1991) Organizations and markets. *Journal of Economic Perspectives*, vol. 5, Primavera, 25-44.
- WEST, M. e PATTERSON, M. (1999) The workforce and productivity, *New Economy*, vol. 6, Março, 22-27.

VOLUME I

1º QUADRIMESTRE DE 2001

**M | G | ∞ Queue Heavy-Traffic Situation Busy
Period Length Distribution (Power and Pareto
Service Distributions)**

Autor:
Manuel Alberto M. Ferreira

VOLUME I

1º QUADRIMESTRE DE 2001

M|G| ∞ QUEUE HEAVY-TRAFFIC SITUATION BUSY PERIOD LENGTH DISTRIBUTION (POWER AND PARETO SERVICE DISTRIBUTIONS)

DISTRIBUIÇÃO DO COMPRIMENTO DO PERÍODO DE OCUPAÇÃO DA FILA DE ESPERA M|G| ∞ EM SITUAÇÃO DE "HEAVY-TRAFFIC" (DISTRIBUIÇÕES DE SERVIÇO POTÊNCIA E PARETO)

Autor: Manuel Alberto M. Ferreira
- Professor Associado - Departamento de Métodos Quantitativos - I.S.C.T.E.

ABSTRACT:

- (Ferreira e Ramalhoto, 1994) showed that for service distributions such that $\lim_{\alpha \rightarrow \infty} G(t) = 0$ ($G(\cdot)$ is the d.f. and α is the mean, for α and ρ (traffic intensity) great enough the M|G| ∞ queue busy period length distribution is approximately exponential.

In this work we show that for service distributions for which this condition is not necessarily true, it is possible to have in some situations an approximately exponential behaviour for the M|G| ∞ queue busy period length.

KEY-WORDS:

- M|G| ∞ , busy period, power distribution, Pareto distribution.

RESUMO:

- (Ferreira e Ramalhoto, 1994) mostraram que para distribuições de serviço tais que $\lim_{\alpha \rightarrow \infty} G(t) = 0$ ($G(\cdot)$ é a função de distribuição e α a média), para α e ρ (intensidade de tráfego) suficientemente grandes a distribuição do comprimento do período de ocupação da fila de espera M|G| ∞ é aproximadamente exponencial.

Neste trabalho mostramos que para distribuições de serviço para as quais aquela condição não se verifica necessariamente, é possível em certas situações ter um comportamento aproximadamente exponencial para o comprimento do período de ocupação da fila de espera M|G| ∞ .

PALAVRAS-CHAVE:

- M|G| ∞ , período de ocupação, distribuição potência, distribuição de Pareto.

VOLUME I

1º QUADRIMESTRE DE 2001

1. INTRODUCTION

A $M|G|\infty$ queue busy period is thought here as being a period that begins when a customer arrives at the system being it empty, ends when a customer leaves the system letting it empty and in it there is, always, at least one customer in the system.

So be a $M|G|\infty$ queue where λ is the arrival Poisson process rate, $G(\cdot)$ is the service d.f., $\alpha = \int_0^\infty [1-G(t)]dt$ is the mean service time, $\rho = \lambda\alpha$ is the traffic intensity and each customer, when it arrives, finds always an available server.

B is the busy period length and $\bar{B}(s)$ is its Laplace-Stieltjes transform.

(Takács, 1962) showed that

$$\bar{B}(s) = 1 + \lambda^{-1} \left(s - \frac{1}{\int_0^\infty e^{-st-\lambda \int_0^t [1-G(v)]dv} dt} \right) \quad (1.1)$$

The inversion of (1.1) is very hard to carry out. The exception is the service distributions collection (Ferreira, 1998)

$$G(t) = 1 - \frac{(1-e^{-\rho})(\lambda+\beta)}{\lambda e^{-\rho}(e^{(\lambda+\beta)t}-1)+\lambda}, \quad t \geq 0, \quad -\lambda \leq \beta \leq \frac{\lambda}{e^\rho - 1} \quad (1.2)$$

for which B is exponentially distributed at rate $e^{-\rho}(\lambda+\beta)$, with an atom at the origin whose value is $\frac{e^{-\rho}(\lambda+\beta)-\beta}{\lambda}$.

In (Ferreira e Ramalhoto, 1994) we showed that for service distributions such that $\lim_{\alpha \rightarrow \infty} G(t) = 0$ (as it happens, for instance, with the exponential), for α and ρ great enough B is approximately exponentially distributed. It is an important result because it gives emphasis to a B distribution insensibility situation related with that service distributions class in heavy-traffic conditions.

In this work we will study service distributions cases for which we have not necessarily $\lim_{\alpha \rightarrow \infty} G(t) = 0$, as it happens with power and Pareto ones, in order to identify situations for which B has a similar behaviour.

To achieve that goal we will use the busy period assimetry coefficient

$$\beta_1 = \frac{(E[B^3] - 3E[B]E[B^2] + 2E^3[B])^2}{(E[B^2] - E^2[B])^3} \quad (1.3)$$

and the Kurtosis one

$$\beta_2 = \frac{E[B^4] - 4E[B]E[B^3] + 6E^2[B]E[B^2] - 3E^4[B]}{(E[B^2] - E^2[B])^2} \quad (1.4)$$

that in the exponential distribution case have the values 4 and 9 respectively (see, for instance, (Kendall and Stuart, 1979) and (Murteira, 1979)).

To compute $E[B^n]$ we have, according to (Ferreira e Ramalhoto, 1994),

$$E[B^n] = (-1)^{n+1} \left\{ \frac{e^\rho}{\lambda} n C^{n-1} - e^\rho \sum_{p=1}^{n-1} \binom{n}{p} E[B^{n-p}] C^p \right\}, \quad n = 1, 2, \dots \quad (1.5)$$

where

$$C^n = \int_0^\infty (-t)^n e^{-\lambda \int_0^t [1-G(v)] dv} \lambda (1-G(t)) dt, \quad n = 0, 1, 2, \dots \quad (1.6)$$

and calculated the integrals numerically with the aid of a computer.

2. POWER SERVICE DISTRIBUTION

If the service distribution is a power function with parameter C , $C > 0$

- $G(t) = \begin{cases} t^C, & 0 \leq t < 1 \\ 1, & t \geq 1 \end{cases}$
- $\alpha = \frac{C}{C+1}$ and $0 < \alpha < 1$.

$$\text{So } \lim_{C \rightarrow \infty} G(t) = \begin{cases} 0, & 0 \leq t < 1 \\ 1, & t \geq 1 \end{cases} \text{ and } \lim_{C \rightarrow \infty} \alpha = 1.$$

Being $B(t)$ the d.f. of B , from (1.1) we have, see (Stadje, 1985),

$$B(t) = 1 - \lambda^{-1} \sum_{n=1}^{\infty} \left[\frac{e^{-\lambda \int_0^t [1-G(v)] dv} \lambda (1-G(t))}{1 - e^{-\rho}} \right]^{*n} (1 - e^{-\rho})^n \quad (2.1)$$

where $*$ is the convolution operator.

For the service distribution with which we are working, since ρ and C are great enough,

$$\frac{e^{-\lambda \int_0^t [1-G(v)] dv} \lambda (1-G(t))}{1 - e^{-\rho}} \cong \lambda e^{-\lambda t}, \quad 0 \leq t \leq 1.$$

Computing the Laplace-Stieltjes transform of (2.1) with this approximation we have

$$\int_0^1 e^{-st} \lambda e^{-\lambda t} dt = \lambda \int_0^1 e^{-(s+\lambda)t} dt = \lambda \left[-\frac{e^{-(s+\lambda)t}}{s+\lambda} \right]_0^1 = \lambda \frac{1 - e^{-(s+\lambda)}}{s+\lambda}.$$

But, for λ great enough, we have approximately $\frac{\lambda}{\lambda+s}$. So, from (2.1) taking

the Laplace-Stieltjes transform of $\frac{e^{-\lambda \int_0^t [1-G(v)] dv} \lambda (1-G(t))}{1 - e^{-\rho}}$ as being $\frac{\lambda}{\lambda+s}$ we

conclude that $B(t) \cong 1 - e^{-\lambda e^{-\rho} t}$, $t \geq 0$ that is:

- In a $M|G|\infty$ queue where the service distribution is a power function, for α near 1, since α and ρ are great enough B is approximately exponential with mean $\frac{e^{\rho}}{\lambda}$.

For this system we computed the values of β_1 and β_2 given for (1.3) and (1.4), respectively, for $\alpha = .25, .5$ and $.8$ making, in each case, ρ take values from $.5$ till 100.

We got the following results:

ρ	$\alpha = .25$		$\alpha = .5$		$\alpha = .8$	
	β_1	β_2	β_1	β_2	β_1	β_2
.5	3.0181197	9.5577742	1.5035507	5.9040102	3.8933428	9.3287992
1	4.4211164	9.1402097	2.7111584	7.4994861	3.9854257	9.0702715
1.5	5.3090021	10.433228	3.3711526	8.2784408	3.9749455	8.9969919
2	5.8206150	11.140255	3.7332541	8.6924656	3.9751952	8.9815770
2.5	6.0803833	11.489308	3.9322871	8.9173048	3.9809445	8.9828631
3	6.1786958	11.619970	4.0388433	9.0369125	3.9871351	8.9877124
6	5.7006232	11.020248	4.0969263	9.1024430	3.9996462	8.9996459
7	5.5034253	10.774653	4.0765395	9.0804332	3.9999342	8.9999341
8	5.3382992	10.570298	4.0596336	9.0623268	3.9999992	8.9999992
9	5.2037070	10.404722	4.0467687	9.0486486	4.0000086	9.0000086
10	5.0944599	10.271061	4.0372385	9.0385796	4.0000068	9.0000068
15	4.7702550	9.8790537	4.0152698	9.0156261	4.0000005	9.0000005
20	4.6102773	9.6888601	4.0082556	9.0083980	4.0000000	9.0000000
50	4.3045903	9.3338081	4.0012425	9.0012513	4.0000000	9.0000000
100	4.1715617	9.1842790	4.0003047	9.0003057	4.0000000	9.0000000

The analysis of the results shows a strong trend of β_1 and β_2 , to 4 and 9, respectively, after $\rho = 10$. This trend is faster the greatest is the value of α .

3. PARETO SERVICE DISTRIBUTION

In this section we will see, mainly, examples. So let us consider a Pareto distribution such that

$$1 - G(t) = \begin{cases} 1, & t < k \\ \left(\frac{k}{t}\right)^3, & t \geq k \end{cases}, \quad k > 0 \quad (3.1).$$

Then, $\alpha = \frac{3}{2}k$ (see, for instance, (Murteira, 1979)).

The values calculated for β_1 and β_2 given for (1.3) and (1.4), respectively with $\lambda = 1$ and, so, $\rho = \alpha$ were

$\alpha = \rho$	β_1	β_2
.5	1028.5443	1373.4466
1	1474.7159	1969.0197
10	38.879220	54.896896
20	4.0048588	9.0049233
50	4.0000000	9.0000000
100	4.0000000	9.0000000

and show a strong trend from β_1 and β_2 to 4 and 9, respectively, after $\rho = 20$. This is natural because, in this case, the convergence of α to infinite makes k to have the same behaviour and, so, after (3.1) we have $\lim_{\alpha \rightarrow \infty} G(t) = 0$

But, considering now a Pareto distribution, such that

$$1 - G(t) = \begin{cases} 1, & t < .4 \\ \left(\frac{.4}{t}\right)^\theta, & t \geq .4 \end{cases}, \quad \theta > 1 \quad (3.2),$$

$\alpha = \frac{.4\theta}{\theta - 1}$ (see, still, (Murteira, 1979)) and the values got for β_1 and β_2 in the same conditions of the previous case are

$\alpha = \rho$	β_1	β_2
.5	10.993704	16.675733
1	6.8553306	12.010791
10	4.5112470	9.5724605
20	4.4832270	9.5397410
50	4.4669879	9.5208253
100	4.4616718	9.5146406

and do not go against the hypothesis of the existence of a trend from β_1 and β_2 to 4 and 9, respectively, although much slower than in the previous case. But, now, the convergence of α to infinite implies do convergence of θ to 1. So

$$\lim_{\alpha \rightarrow \infty} G(t) = \begin{cases} 0, & t < .4 \\ 1 - \frac{.4}{t}, & t \geq .4 \end{cases} \text{ and we can not guarantee at all that for } \alpha \text{ great enough}$$

$$1 - G(t) \cong 1.$$

4. CONCLUSIONS

In the case of service distributions for which is possible to get conditions for which $G(t) \cong 0$, for these conditions and with the possible consideration of another ones it is possible to guarantee that B is approximately exponentially distributed.

So, if $G(t) \cong 0$ for α great enough it is sufficient to consider ρ great enough.

But, for instance, if the service distribution is a power function, as we have seen we can not guarantee such conditions. However, for α near 1 and λ and ρ great enough it is possible to guarantee that B is approximately exponentially distributed.

But in the case of Pareto distribution where we have not $G(t) \cong 0$ for α great enough, although we can not give identical guarantees to those of the power function service, the results got for β_1 and β_2 are not against that for ρ great enough B is approximately exponentially distributed.

REFERENCES

- FERREIRA, M.A.M., (1998) "Aplicação da Equação de Ricatti ao Estudo do Período de Ocupação do Sistema $M|G|\infty$ ", Revista de Estatística, Volume 1, I.N.E..
- FERREIRA, M.A.M. e RAMALHOTO M.F., (1994) "Estudo dos parâmetros básicos do período de ocupação da fila de espera $M|G|\infty$ ", A Estatística e o Futuro e o Futuro da Estatística, Actas do I Congresso Anual da S.P.E., Edições Salamandra, Lisboa.
- KENDALL and STUART, (1979) "The Advanced Theory of Statistics". Distributions Theory, London. Charles Griffin and Co., Ltd., 4th Edition.
- MURTEIRA, B., (1979) "Probabilidades e Estatística", Vol. 1, Editora McGraw-Hill de Portugal, Lda., Lisboa.
- STADJE, W., (1985) "The busy period of the queueing system $M|G|\infty$ ", J.A.P., 22, pp.697-704.
- TAKÁCS, L., (1962) "An introduction to queueing theory". Oxford University Press, New York.

Redução de dimensionalidade através duma Análise em Componentes Principais: Um critério para o número de Componentes Principais a reter

Autor:
Jorge F. C. L. Cadima

VOLUME I

1º QUADRIMESTRE DE 2001

REDUÇÃO DE DIMENSIONALIDADE ATRAVÉS DUMA ANÁLISE EM COMPONENTES PRINCIPAIS: UM CRITÉRIO PARA O NÚMERO DE COMPONENTES PRINCIPAIS A RETER

REDUCING DIMENSIONALITY BY MEANS OF A PRINCIPAL COMPONENT ANALYSIS: A CRITERION TO CHOOSE HOW MANY PRINCIPAL COMPONENTS SHOULD BE RETAINED

Autor: Jorge F. C. L. Cadima
- Professor Auxiliar - Departamento de Matemática - Instituto Superior de Agronomia

RESUMO:

- A escolha do número de factores a reter, no contexto duma redução de dimensionalidade através duma Análise em Componentes Principais, tem sido objecto de numerosos critérios, em que geralmente são necessárias opções de natureza subjectiva por parte do utilizador, e/ou hipóteses distribucionais subjacentes. Nesta comunicação apresenta-se um novo critério, que evita os referidos problemas. O critério baseia-se em propriedades geométricas, associadas à estrutura do cone de matrizes semi-definidas positivas, que sugerem uma dimensionalidade mínima admissível. A estrutura do cone, resultante da introdução do habitual produto interno matricial, veda determinadas zonas do cone à presença de matrizes de características reduzidas.

PALAVRAS-CHAVE:

- *Componentes Principais, análise factorial, redução de dimensionalidade, cone de matrizes semi-definidas positivas.*

ABSTRACT:

- There are numerous criteria to choose how many Principal Components should be retained when reducing dimensionality via a Principal Component Analysis. Subjective decisions by the user, and/or distributional hypotheses are usually necessary when applying these criteria. This communication proposes a new criterion which avoids these problems. The criterion is based on geometric properties of the cone of positive semi-definite matrices, that suggest a smallest admissible dimensionality. The structure of the cone, which is associated with the use of the usual matrix inner product, precludes the presence of low-rank matrices in certain areas of the cone.

KEY-WORDS:

- *Principal components, factorial analysis, dimensionality reduction, cone of positive semi-definite matrices.*

VOLUME I

1º QUADRIMESTRE DE 2001

1. INTRODUÇÃO

Na análise de conjuntos de dados multivariados, a possibilidade de reduzir a dimensionalidade dos dados sem grande perda de informação desempenha um papel crucial. Caso seja admissível proceder a uma aproximação em espaços bi- ou tri-dimensionais, tornar-se-á possível uma visualização gráfica aproximada do conjunto de dados. Mesmo quando não seja possível uma redução de dimensionalidade tão acentuada, as simplificações associadas à consideração de menos dimensões têm provado ser úteis para a compreensão, exploração e modelação de dados.

A redução de dimensionalidade é geralmente efectuada por meio de técnicas de análise factorial³, e no caso concreto de se analisar uma matriz X de observações de p variáveis sobre n indivíduos, através duma Análise em Componentes Principais (ACP). Numa ACP – contexto que será aqui analisado em pormenor, apesar de muito do que se segue ser também válido para análises factoriais em geral – identificam-se sucessivas combinações lineares das p variáveis originais que sejam de variância máxima, sujeitas às restrições de não-correlacionamento com anteriores combinações lineares. Estas combinações lineares designam-se Componentes Principais (CPs) ou factores. Os coeficientes que definem cada combinação linear são dados pelos elementos de cada vector próprio da matriz Σ de variâncias-covariâncias definida pelos dados (ou pela matriz de correlações, caso se opte por trabalhar com dados normalizados). O valor próprio associado a cada vector próprio indica a variância da CP correspondente. O quociente desse valor próprio sobre a soma dos valores próprios (isto é, sobre o traço da matriz Σ) designa-se a proporção da inércia da nuvem de pontos definida pela matriz dos dados associada à CP em questão. Esta interpretação é extensível a conjuntos de CPs. Veja-se (Jolliffe, 1986) para uma discussão mais pormenorizada sobre a Análise em Componentes Principais.

A questão que se pretende aqui abordar prende-se com a determinação duma dimensionalidade mínima admissível para a representação dos dados originais, isto é, a identificação do número k de Componentes Principais (factores) a reter de forma a se poder considerar aceitável uma aproximação k -dimensional dos dados.

Inúmeras regras e critérios têm sido propostos com esta finalidade. Recenseamentos desses critérios, no contexto de ACP, encontram-se em (Jolliffe, 1986, Capítulo 6), e mais recentemente em (Richman, Angel & Gong, 1992). A ideia central subjacente a tais critérios é a de que factores associados a valores próprios próximos de zero podem ser desprezados sem grande perda de informação. Os critérios propostos podem agrupar-se em duas categorias fundamentais: os critérios ‘descritivos’, que não consideram qualquer modelo ou distribuição subjacente, e os critérios ‘inferenciais’, nos quais as hipóteses distribucionais desempenham um papel central. De entre a primeira categoria, refiram-se, a título de exemplo, o critério de

³ A expressão “análise factorial” é aqui usada no sentido mais lato, predominante na literatura de origem francesa, indicando de forma genérica técnicas baseadas em análises espectrais ou em valores singulares. Não deve ser confundida com a técnica mais específica, correntemente designada por *Factor Analysis* na literatura anglo-saxónica. Para uma discussão das relações entre esta última técnica e a Análise em Componentes Principais, veja-se (Jolliffe, 1986, Capítulo 7).

ignorar CPs após se ter atingido uma determinada proporção cumulativa de inércia explicada (provavelmente o critério mais frequentemente utilizado); o critério de Kaiser, de excluir CPs associados a valores próprios inferiores ao valor próprio médio (ou a variante de Jolliffe, que baixa esse limiar para 70% do valor próprio médio); e critérios baseados na análise visual de gráficos de valores próprios, ou logaritmos de valores próprios, contra as respectivas ordens (gráficos *scree* ou *LEV*), onde é frequente encontrar um padrão de rápido decréscimo inicial seguido de uma estabilização dos valores próprios em torno de valores próximos de zero, que dá aos gráficos o aspecto duma recta quase horizontal, procedendo-se à eliminação de CPs associadas a essa região do gráfico. De entre a segunda categoria, cite-se o exemplo dum teste assintótico à igualdade dos q menores valores próprios duma matriz de variâncias-covariâncias, admitindo a multinormalidade dos dados subjacentes (Jolliffe, 1986, pg.44).

No entanto, nenhum critério recolhe consenso generalizado. Em relação aos critérios inferenciais, são evidentes as restrições impostas pelas hipóteses distribucionais, ou modelos exigíveis. E a maior generalidade dos critérios determinísticos envolve geralmente apreciações de natureza subjectiva (qual o limiar a partir do qual se considera admissível uma dada proporção da inércia total explicada; qual o limiar para se considerar um determinado valor próprio como sendo desprezável; qual o limiar a partir do qual um gráfico *scree* ou *LEV* se considera aproximadamente horizontal; etc.) que podem conduzir a opções diferentes por diferentes utilizadores. Alguns critérios são ainda de aplicação algo trabalhosa.

O critério aqui proposto tem as vantagens de não exigir hipóteses distribucionais, de ser de aplicação simples e de ser um critério objectivo, no sentido de conduzir a uma dimensionalidade mínima admissível com base em considerações de natureza geométrica que não envolvem opções subjectivas por parte do utilizador.

Na Secção 2 apresentam-se os resultados teóricos que sustentam o critério. Este último é apresentado na Secção 3. Finalmente, a Secção 4 discute o critério.

2. ALGUMAS PROPRIEDADES DO CONE DE MATRIZES SEMI-DEFINIDAS POSITIVAS

Admitir a redução de dimensionalidade numa ACP dum conjunto de dados $n \times p$ equivale a admitir que os dados podem ser projectados sobre um subespaço k -dimensional sem perda substancial de informação. Do ponto de vista da matriz de variâncias-covariâncias (ou de correlações) associada aos dados, essa hipótese equivale a considerar que a matriz é “aproximadamente de característica k ”, conceito que será formalizado seguidamente.

As matrizes de variâncias-covariâncias (ou de correlações) são sempre matrizes semi-definidas positivas, ou seja, matrizes simétricas $\mathbf{A}_{p \times p}$ cujas formas quadráticas são sempre não-negativas, i.e., tais que $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0, \forall \mathbf{x} \in \mathcal{R}^p / \{\mathbf{0}\}$. Conversamente,

qualquer matriz semi-definida positiva (s.d.p.) é sempre uma matriz de variâncias-covariâncias para uma infinidade de matrizes de dados (qualquer que seja o número n de indivíduos observados), pelo que os dois conceitos se confundem. Relembremos agora o conceito de cone.

Definições.

- 1) Um subconjunto C dum espaço linear designa-se um **cone** se fôr fechado para combinações lineares não-negativas: $\forall \mathbf{x}, \mathbf{y} \in C$ e $\forall \alpha, \beta \in \mathcal{R}_0^+$, $\alpha \mathbf{x} + \beta \mathbf{y} \in C$.
- 2) Num cone C , dado qualquer elemento V do cone, designa-se por **raio** associado a V ao conjunto de todos os múltiplos escalares de V , com escalares não-negativos.

É fácil de verificar que o conjunto das matrizes semi-definidas positivas de dimensão $p \times p$ forma um cone no espaço linear das matrizes quadradas $p \times p$, uma vez que se \mathbf{V} e \mathbf{W} forem matrizes s.d.p. e $\alpha, \beta \geq 0$, então $\mathbf{x}'(\alpha \mathbf{V} + \beta \mathbf{W})\mathbf{x} = \alpha \mathbf{x}'\mathbf{V}\mathbf{x} + \beta \mathbf{x}'\mathbf{W}\mathbf{x} \geq 0$, $\forall \mathbf{x} \neq \mathbf{0}$. Designemos por D_p o cone das matrizes s.d.p. de dimensão p , cone que é discutido em mais pormenor por (Hill & Waters, 1987).

A definição dum produto interno no espaço linear de matrizes $p \times p$ – e por conseguinte no cone D_p – permite definir o conceito de ângulo entre duas matrizes (ou seja, entre os raios definidos por duas matrizes). Como se verá seguidamente, este conceito dará uma estrutura interessante ao cone, que sugere, de forma natural, uma solução para o problema sob consideração.

Definição. Seja $M_{n \times p}$ o espaço linear das matrizes $n \times p$, munido do habitual produto interno matricial, definido por $\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}'\mathbf{B})$, $\forall \mathbf{A}, \mathbf{B} \in M_{n \times p}$, onde tr designa o traço. Para $\mathbf{A}, \mathbf{B} \neq \mathbf{0}$, defina-se o **ângulo** entre as matrizes \mathbf{A} e \mathbf{B} como sendo o arco cujo coseno é dado por:

$$\cos(\mathbf{A}, \mathbf{B}) = \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\|\mathbf{A}\| \cdot \|\mathbf{B}\|}$$

onde $\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}'\mathbf{A})}$ é a norma induzida pelo produto interno.

Na estrutura do cone D_p de matrizes s.d.p., um papel central é desempenhado pelo raio definido pela matriz identidade $p \times p$, \mathbf{I}_p . Os resultados seguintes, baseados no trabalho de (Tarazaga, 1990) e desenvolvidos em (Cadima, 1997) sintetizam a estrutura que nos irá interessar.

Teorema. Considere-se o cone D_p de matrizes semi-definidas positivas $p \times p$. Considere-se a matriz identidade $\mathbf{I}_p \in D_p$, e qualquer outra matriz $\mathbf{V} \in D_p$, cujos valores próprios sejam os elementos do vector $\lambda \in (\mathfrak{R}_0^+)^p$. Então:

$$\text{i) } \cos(\mathbf{V}, \mathbf{I}_p) = \frac{\|\lambda\|_1}{\sqrt{p} \cdot \|\lambda\|_2} = \frac{1}{\sqrt{1 + p^2 \sigma_\pi^2}} \geq \frac{1}{\sqrt{p}}, \text{ onde } \sigma_\pi^2 = \frac{1}{p} \sum_{i=1}^p (\pi_i - \bar{\pi})^2 \text{ é}$$

a variância dos p valores próprios relativos da matriz \mathbf{V} , isto é, dos $\pi_i = \lambda_i / \sum_j \lambda_j$.

$$\text{ii) } \cos(\mathbf{V}, \mathbf{I}_p) = \frac{1}{\sqrt{p}} \text{ se e só se } \mathbf{V} \text{ fôr uma matriz de característica } 1.$$

iii) Se \mathbf{V} é uma matriz de característica k , então $\cos(\mathbf{V}, \mathbf{I}_p) \leq \sqrt{\frac{k}{p}}$, verificando-se a igualdade se e só se os k valores próprios não-nulos de \mathbf{V} forem todos iguais.

Demonstração. Nesta demonstração serão usadas as conhecidas desigualdades relacionando as normas ℓ_1 e ℓ_2 dum qualquer vector $\mathbf{x} \in \mathfrak{R}^n$: $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n} \|\mathbf{x}\|_2$ (Horn & Johnson, 1985, pg. 279).

i) Seja $\mathbf{V} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$ a decomposição espectral de \mathbf{V} , onde $\mathbf{\Lambda}$ indica a matriz diagonal cuja diagonal é constituída pelos elementos do vector λ (admita-se, sem perda de generalidade, que estão ordenados por ordem decrescente) e \mathbf{P} é a matriz ortogonal cujas colunas são os vectores próprios correspondentes. Pela definição de produto interno entre duas matrizes, e pela circularidade do traço do produto matricial, tem-se:

$$\langle \mathbf{V}, \mathbf{I}_p \rangle = \text{tr}(\mathbf{V}) = \sum_{i=1}^p \lambda_i = \|\lambda\|_1 \quad (\text{já que todos os valores próprios são não-negativos});$$

$$\|\mathbf{V}\| = \sqrt{\text{tr}(\mathbf{V}^2)} = \sqrt{\text{tr}(\mathbf{P}\mathbf{\Lambda}^2\mathbf{P}')} = \sqrt{\text{tr}(\mathbf{\Lambda}^2)} = \sqrt{\sum_{i=1}^p \lambda_i^2} = \|\lambda\|_2;$$

$$\text{e } \|\mathbf{I}_p\| = \sqrt{\text{tr}(\mathbf{I}_p)} = \sqrt{p}. \text{ Logo, fica provada a primeira igualdade. Ora,}$$

$$\sigma_\pi^2 = \frac{1}{p} \sum_{i=1}^p \pi_i^2 - \frac{1}{p^2}, \quad \text{uma vez que } \bar{\pi} = \frac{1}{p} \sum_{i=1}^p \pi_i = \frac{1}{p}.$$

$$\text{Logo, } \frac{\|\lambda\|_1}{\sqrt{p} \cdot \|\lambda\|_2} = \frac{1}{\sqrt{p \cdot \sum_{i=1}^p \pi_i^2}} = \frac{1}{\sqrt{1 + p^2 \sigma_\pi^2}}.$$

A desigualdade final resulta directamente de $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$ para qualquer vector \mathbf{x} .

$$\text{ii) Tendo em conta, como foi visto, que } \cos(\mathbf{V}, \mathbf{I}_p) = \frac{1}{\sqrt{p \cdot \sum_{i=1}^p \pi_i^2}}, \text{ tem-se}$$

$$\cos(\mathbf{V}, \mathbf{I}_p) = 1/\sqrt{p} \Leftrightarrow \sum_{i=1}^p \pi_i^2 = 1 \Leftrightarrow \|\pi\|_2^2 = \|\pi\|_1, \text{ onde } \pi \text{ designa o}$$

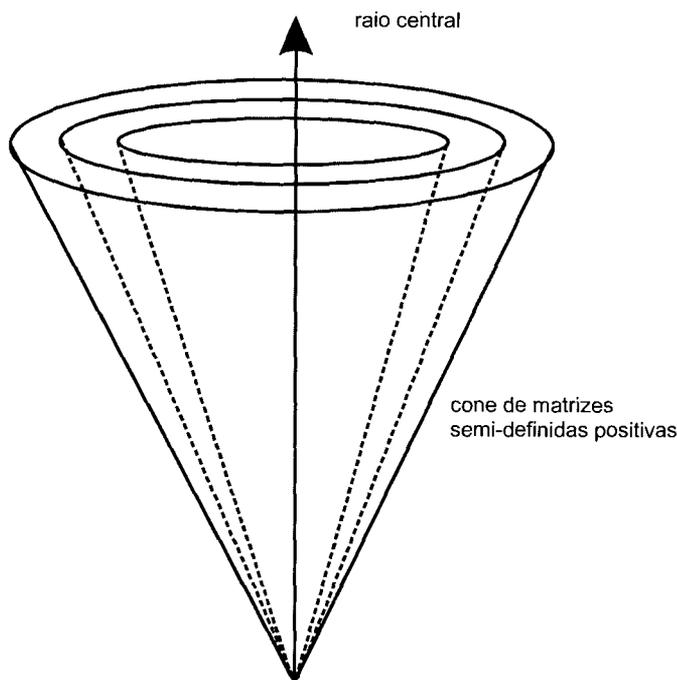
vector p -dimensional dos valores próprios relativos, π_i , já que $\|\pi\|_1 = 1$. Nesse caso, π tem de ser colinear com um dos eixos coordenados, ou seja ter $p-1$ coordenadas nulas, e a restante coordenada com valor ± 1 . Como a matriz V é semi-definida positiva, e dada a ordenação admitida para os valores próprios, isso significaria $\pi_1 = 1$ e $\pi_i = 0, \forall i > 1$. Assim, V tem de ser uma matriz de característica 1.

iii) Já vimos que $\cos(V, I_p) = \frac{1}{\sqrt{p \cdot \sum_{i=1}^p \pi_i^2}}$ para qualquer matriz s.d.p. V .

No caso de V ser de característica k , os últimos $p-k$ valores próprios de V serão nulos, pelo que $\sum_{i=1}^p \pi_i^2 = \sum_{i=1}^k \pi_i^2$. Designando por π o vector k -dimensional cujos elementos são os k valores próprios relativos não-nulos de V , $\{\pi_i\}_{i=1}^k$, vemos que $\cos(V, I_p) = \frac{1}{\sqrt{p \cdot \|\pi\|_2}} \leq \sqrt{\frac{k}{p}}$, já que $\|\pi\|_2 \geq \frac{1}{\sqrt{k}} \|\pi\|_1$ e $\|\pi\|_1 = 1$. A igualdade verifica-se se e só se $\|\pi\|_2 = \frac{1}{\sqrt{k}} \|\pi\|_1$, situação que equivale a ter todos os elementos do vector π iguais, isto é, todos os valores próprios não-nulos de V iguais.

(c.q.d.)

O Teorema mostra que o cone de matrizes $p \times p$ semi-definidas positivas D_p tem uma estrutura estratificada, com sucessivos sub-cones encaixados, conforme ilustrado na Figura.



No centro do cone encontra-se o raio central, definido pela matriz identidade \mathbf{I}_p . Na região que rodeia o raio central, e que é constituída por matrizes cujos raios associados definem ângulos com o raio central menores que $\arccos\sqrt{\frac{p-1}{p}}$, apenas poderemos encontrar matrizes de característica plena (característica p). Esse núcleo é rodeado por uma fronteira constituída pelas matrizes $p \times p$ cujo ângulo com o raio central é precisamente $\arccos\sqrt{\frac{p-1}{p}}$. Essas matrizes podem ser de característica plena, ou de característica $p-1$, mas neste último caso apenas se forem múltiplos escalares de matrizes de projecção ortogonal sobre algum subespaço de dimensão $p-1$. Segue-se um estrato constituído por matrizes cujo ângulo θ com o raio central se encontra no intervalo $\arccos\sqrt{\frac{p-1}{p}} < \theta < \arccos\sqrt{\frac{p-2}{p}}$. Nesse estrato encontram-se apenas matrizes de característica p ou $p-1$. Na fronteira desse estrato constituída pelas matrizes cujo ângulo com o raio central é precisamente $\arccos\sqrt{\frac{p-2}{p}}$ encontramos, além de matrizes de característica p ou $p-1$, também matrizes de característica $p-2$, mas apenas se forem múltiplos escalares de matrizes de projecção sobre subespaços de \mathbb{R}^p de dimensão $p-2$. Segue-se um novo estrato, composto por matrizes cujo ângulo θ com o raio central satisfaz $\arccos\sqrt{\frac{p-2}{p}} < \theta < \arccos\sqrt{\frac{p-3}{p}}$, onde residem matrizes de características p , $p-1$ e $p-2$, terminando numa fronteira definida por $\theta = \arccos\sqrt{\frac{p-3}{p}}$, onde surgem também múltiplos escalares de matrizes de projecção sobre subespaços de dimensão $p-3$. Sucessivos estratos, definidos de forma análoga à medida que aumenta o ângulo θ , irão passar a incluir matrizes de características cada vez mais baixas. A região exterior do cone é constituída pelas matrizes que definem raios que formam um ângulo $\arccos\sqrt{\frac{1}{p}}$ com o raio central. Essas matrizes são precisamente as matrizes $p \times p$ de característica 1, isto é, matrizes da forma $\lambda \mathbf{x}\mathbf{x}'$ para algum vector $\mathbf{x} \in \mathbb{R}^p$ e $\lambda \in \mathbb{R}^+$.

Embora matrizes de características baixas não se possam encontrar perto do núcleo central do cone, que rodeia o eixo central, podem-se encontrar matrizes de características elevadas, ou até de característica plena, em todo o cone (excepto se formam o ângulo máximo com o raio central, admissível exclusivamente para matrizes de característica 1). Tal facto não é motivo de surpresa, pois matrizes de característica plena podem ter valores próprios arbitrariamente próximos de zero, comportando-se como matrizes de característica reduzida do ponto de vista da variância dos seus valores próprios relativos que, como vimos no Teorema, determina em última análise o posicionamento angular de cada matriz no cone das matrizes semi-definidas positivas. E é precisamente este facto que sugere um critério para a escolha da redução de dimensionalidade admissível para um qualquer conjunto de dados.

3. O CRITÉRIO

Começemos por introduzir o conceito de pseudo-característica duma matriz semi-definida positiva.

Definição. Seja V uma matriz $p \times p$ semi-definida positiva. A **pseudo-característica** de V é o menor inteiro k^* tal que $\cos(V, I_p) \leq \sqrt{\frac{k^*}{p}}$.

Trabalhando com as expressões anteriormente obtidas para $\cos(V, I_p)$ é fácil de verificar que a pseudo-característica duma matriz s.d.p. é dada por:

$$k^* = \left\lceil \frac{1}{\sum_{i=1}^p \pi_i^2} \right\rceil = \left\lceil \frac{\text{tr}^2(V)}{\text{tr}(V^2)} \right\rceil$$

(onde $\lceil x \rceil$ designa o menor inteiro maior ou igual a x), já que $\sum_{i=1}^p \pi_i^2 = \frac{\sum_{i=1}^p \lambda_i^2}{(\sum_{j=1}^p \lambda_j)^2}$.

Tendo em conta a estrutura do cone D_p atrás discutida, a pseudo-característica duma matriz s.d.p. V corresponde à menor característica das matrizes s.d.p. que se encontram no mesmo estrato do cone que a matriz V . Esse posicionamento de V sugere que *pode fazer sentido reduzir a dimensionalidade de V até à sua pseudo-característica, mas não ulteriormente.*

A título ilustrativo, apresentam-se na Tabela as dimensões e pseudo-características de algumas matrizes de covariâncias e correlações referidas na literatura estatística. Na Tabela indicam-se ainda as percentagens cumulativas de inércia associadas a uma redução para a dimensionalidade da pseudo-característica.

Dados	Tipo de matriz	No. variáveis	pseudo-característica	% inércia cumulativa
<i>Iris verginica</i> ⁴	covariância	4	2	90,26
<i>Iris verginica</i>	correlação	4	3	96,79
<i>Iris setosa</i>	covariância	4	2	88,41
<i>Iris setosa</i>	correlação	4	3	93,71
<i>Iris versicolor</i>	covariância	4	2	89,67
<i>Iris versicolor</i>	correlação	4	2	86,82
<i>Alimentos Lebart</i> ⁵	covariância	7	2	96,46
<i>Alimentos Lebart</i>	correlação	7	3	97,06
<i>Pitprop</i> ⁶	correlação	13	6	86,9
<i>Adelges</i> ⁷	correlação	19	2	85,5
<i>Orheim</i> ⁸	correlação	9	4	85,75
<i>Lagostins Somers</i> ⁹	covariância	13	2	90,00
<i>Lagostins Somers</i>	correlação	13	3	83,35
<i>Explorações agrícolas</i> ¹⁰	correlação	62	15	78,80

⁴ Os dados relativos aos lírios devem-se a Fisher e encontram-se em numerosas fontes, e.g., (Krzanowski, 1988, pg. 46).

⁵ Este conjunto de dados é discutido em (Lebart *et al*, 1982).

⁶ Dados discutidos em (Jeffers, 1967).

⁷ Dados discutidos em (Jeffers, 1967).

⁸ Dados apresentados por Orheim e discutidos em (McCabe, 1984).

⁹ Dados discutidos em (Cadima & Jolliffe, 2001).

¹⁰ Dados referidos em (Cadima & Jolliffe, 2001).

4. A DISCUSSÃO

A escolha da dimensionalidade a reter, por meio da pseudo-característica da matriz de variâncias-covariâncias Σ (ou de correlações) dos dados sob análise, apresenta as seguintes vantagens:

- i) é um critério objectivo, no sentido de não estar sujeito a apreciações diferenciadas por parte de utilizadores diferentes;
- ii) é de cálculo muito fácil, envolvendo apenas os traços de Σ e de Σ^2 ;
- iii) entra em linha de conta com todos os valores próprios de Σ (através da variância dos valores próprios a dividir pelo traço de Σ), e não apenas com alguns, isto é, utiliza a totalidade da informação disponível;
- iv) tem uma justificação geométrica, independentemente da natureza dos dados sob estudo;
- v) não exige qualquer hipótese distribucional ou modelo subjacente.

Algumas destas vantagens (em particular as últimas) poderão ser encaradas como *desvantagens* em situações particulares, como sejam quando se deseja estudar a adequabilidade dum determinado modelo, ou quando a natureza da aplicação aconselha outros critérios de escolha. Mas na generalidade das situações onde uma análise factorial se aplica como técnica exploratória, a inexistência de hipóteses prévias sobre a natureza dos dados é, sem dúvida, uma vantagem.

Como para qualquer outro critério, haverá aspectos ou situações particulares onde se poderá considerar que a pseudo-característica indica um número excessivamente elevado ou reduzido de factores (CPs) a reter. Importa sublinhar que a pseudo-característica deve ser vista como uma dimensionalidade *mínima* admissível, no sentido de indicar que a matriz Σ está numa região do cone D_p que está vedada a matrizes de características inferiores, mas que nada obsta a que se considerem aproximações de dimensionalidade superior. Por outro lado, o facto de cada estrato do cone ser delimitado por fronteiras onde se encontram matrizes de característica inferior (desde que tenham todos os valores próprios não-nulos iguais) criará situações onde poderia ser tentador considerar dimensionalidades inferiores às sugeridas pela pseudo-característica. Por exemplo, uma matriz de variâncias-covariâncias de dimensão 3 que tenha os dois primeiros valores próprios iguais, terá pseudo-característica 3 para valores arbitrariamente pequenos (mas não-nulos) do último valor próprio. No entanto, procurar-se ajustar o critério a fim de obviar a algumas situações deste tipo é, por um lado, um exercício condenado antecipadamente ao fracasso (haverá sempre matrizes arbitrariamente próximas de qualquer fronteira que se queira considerar) e por outro lado, destrói uma das vantagens do critério: a sua natureza objectiva.

Por tudo quanto ficou dito, considera-se que a pseudo-característica da matriz de variâncias-covariâncias (ou de correlações) dum conjunto de dados é uma boa escolha para a dimensionalidade mínima admissível na redução da dimensionalidade desses dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- CADIMA, J. “Algumas aplicações do Produto Interno Matricial Usual na Análise de Componentes Principais”, in *A Estatística a Decifrar o Mundo, Actas do IV Congresso Anual da Sociedade Portuguesa de Estatística*, Edições Salamandra, pgs. 273-280, 1997.
- CADIMA, J. & JOLLIFFE, I.T. “Variable selection and the interpretation of principal subspaces”, Aceite para publicação no *Journal of Agricultural, Biological and Environmental Statistics*, 2001.
- HILL, R.D. & WATERS, S.R. “On the cone of positive semi-definite matrices”, *Linear Algebra and its Applications*, 90, pgs. 81-88, 1987.
- HORN, R. & JOHNSON, C. “Matrix Analysis”, Cambridge University Press, 1986.
- JEFFERS, J.N.R. “Two case studies in the application of principal component analysis”, *Applied Statistics*, 16, pgs. 225-236, 1967.
- JOLLIFFE, I.T. “Principal Component Analysis”, Springer-Verlag, 1986.
- KRZANOWSKI, W.J., “Principles of Multivariate Analysis”, Oxford University Press, 1988.
- LEBART, L.; MORINEAU, A. & FÉNELON, J.-P., “Traitement des données statistiques”, Dunod, 1982.
- MCCABE, G.P., “Principal Variables”, *Technometrics*, 26(2), pgs. 137-144.
- RICHMAN, M.B.; ANGEL, J.R. & GONG, X. “Determination of Dimensionality in Eigenanalysis”, Proceedings of the 5th International Meeting on Statistical Climatology, Toronto, 1992.
- TARAZAGA, P. “Eigenvalue estimates for symmetric matrices”, *Linear Algebra and its Applications*, 135, pgs. 171-179, 1990.

VOLUME I

1º QUADRIMESTRE DE 2001

Método de Taguchi: Controlo de Qualidade com Recurso Não-Convencional ao Desenho Estatístico de Experiências (DOE)

Autora:
Paula M. Melo e Castro

VOLUME I

1º QUADRIMESTRE DE 2001

MÉTODO DE TAGUCHI: CONTROLO DE QUALIDADE COM RECURSO NÃO-CONVENCIONAL AO DESENHO ESTATÍSTICO DE EXPERIÊNCIAS (DOE)

TAGUCHI'S METHOD: QUALITY CONTROL WITH A NON-CONVENTIONAL USE OF DESIGN OF EXPERIMENTS (DOE)

Autora: Paula M. Melo e Castro

- Docente - Instituto Português de Administração de Marketing (IPAM) - Lisboa

RESUMO:

- O sucesso (ou mesmo a sobrevivência) de qualquer empresa, num mercado globalizado e altamente competitivo depende, inquestionavelmente, da qualidade do seu produto.

O método de Taguchi adquiriu adeptos entusiásticos, internacionalmente, por ter conseguido induzir melhorias consideráveis na qualidade do produto, com uma razão custo/benefício atractivamente baixa.

Este artigo expõe o método de Taguchi de um modo simplificado, apresentando as suas principais directrizes e alertando para algumas objecções que suscita, sob o ponto de vista teórico.

PALAVRAS-CHAVE:

- *Produto robusto, desenho de experiências, matriz intrínseca, matriz extrínseca, sinal-para-ruído (SN), factor-de-controlo, factor-de-sinal, parâmetro, tolerâncias.*

ABSTRACT:

- The success (or even simple survival) of a company undoubtedly depends on the quality of its own product.

Taguchi's method has enthusiastic supporters all over the world for producing huge quality improvements at an attractively high benefit/cost ratio.

This article presents, in a simplified way the guidelines of Taguchi's method and highlights some of the theoretical objections to it.

KEY-WORDS:

- *Robust product, design of experiments, inner array, outer array, signal-to-noise (SN), control-factor, signal-factor, parameter, tolerances.*

VOLUME I

1º QUADRIMESTRE DE 2001

1. INTRODUÇÃO

A importância que a qualidade do produto adquiriu, nos nossos dias, é inquestionável e as empresas estão cada vez mais cientes de que dela dependerá o seu sucesso e, mesmo, a sua sobrevivência.

A progressiva globalização do mercado tem conduzido quer a uma melhor informação do consumidor em relação a produtos concorrenciais, quer a novas (e crescentes) facilidades de acesso, possibilitando-lhe um leque de escolha mais alargado. Deste modo, o perfil do consumidor evolui no sentido de uma exigência cada vez maior, dificultando, ao produtor, a sua cativação.

Numa primeira análise, o consumidor procura o produto que lhe garanta a maior qualidade ao menor preço mas, frequentemente, sentir-se-á receptivo a um produto que evidencie melhor qualidade, ainda que a um preço ligeiramente superior (dentro da mesma gama¹¹). Este aspecto é particularmente interessante para o produtor ao qual não é viável reduzir o preço do seu produto, mas que pode alcançar uma diferenciação desejável, investindo na melhoria da qualidade do mesmo (obviamente, procurando minimizar os custos de tal investimento).

O método de Taguchi tem-se revelado economicamente interessante, além de bem sucedido, em Qualidade.

Após a abordagem do conceito de qualidade de Taguchi, seguem-se a apresentação (simplificada) do método de Taguchi e a referência a algumas das objecções teóricas que suscitou.

2. QUALIDADE SEGUNDO TAGUCHI

O Dr. Genichi Taguchi, engenheiro e estatístico japonês cuja obra, com vista à melhoria da qualidade, lhe proporcionou, em 1960, a atribuição do prémio Deming, é frequentemente referenciado, em Qualidade, a par de Ishikawa, de Juran e do próprio Deming¹².

¹¹ Não faz sentido comparar as qualidades de produtos de gamas diferentes como, por exemplo, artigos de luxo e de primeira necessidade.

¹² Sumidades, em Qualidade.

A qualidade de um produto pode ser encarada sob diversas perspectivas e, por isso, associada a diferentes acepções:

- Sob o ponto de vista do consumidor, adquire um carácter subjectivo, em qualquer das suas vertentes — física, sensorial, temporal¹³ —, dado que não obedece a qualquer escala de avaliação universal. No entanto, existe uma proporcionalidade directa entre o nível de qualidade que este atribui ao produto, a adequabilidade ao uso e o grau de satisfação das suas expectativas perante o mesmo.
- Sob o ponto de vista do produtor, a qualidade do produto integra dois parâmetros:
 1. *qualidade do projecto*, isto é, das características que se lhe pretende conferir;
 2. *qualidade da conformação*, que diz respeito ao grau de aproximação entre as características que o produto realmente adquiriu e as previamente pretendidas (especificações). É este aspecto da qualidade que, revelando-se problemático para o produtor, necessita de ser controlado, uma vez que a qualidade do projecto é arbitrariamente decidido.

Foi em relação à qualidade da conformação — o aspecto da qualidade que importa controlar — que Taguchi sugeriu uma abordagem completamente inovadora.

Tradicionalmente, a aceitação (ou a rejeição) de um produto dependia da inclusão (ou não) dos seus parâmetros nas respectivas especificações, pré-definidas, sob a forma de um intervalo de valores. Exemplificando: se as especificações, para uma determinada característica (Y) do produto, fossem estabelecidas como

$[LIE, LSE]$,

LIE - limite inferior das especificações,

LSE - limite superior das especificações,

o produto só seria considerado defeituoso e, como tal, rejeitado, se o valor que exibisse, para a referida característica, fosse inferior a LIE , ou superior a LSE . Por outras palavras, assumindo que a característica medida apresentou o valor a , isto é,

$$Y=a,$$

o produto seria considerado defeituoso se, e só se

$$a < LIE \quad \vee \quad a > LSE.$$

¹³ Exemplos das três vertentes da qualidade de um produto:

- Física - comprimento, peso, viscosidade,
- Sensorial - sabor, cor, aspecto,...
- Temporal - durabilidade, usabilidade, fiabilidade,...

Note-se que, de acordo com este critério, incluir-se-ão entre as unidades de produto conformes¹⁴, sem qualquer possibilidade de distinção, quer as que apresentem valores vizinhos das extremidades do intervalo de especificações, quer as que apresentem valores afastados destas.

Taguchi idealizou um conceito de qualidade distinto que, reflectindo considerações de ordem socio-económica, revolucionou os objectivos da produção: em vez de se procurar produzir em função de um intervalo de especificações, passou-se a pretender produzir para valores-alvo específicos e precisos.

A ideia da qualidade de um produto, segundo Taguchi, pode ser expressa como

*“a perda causada à Sociedade pelo produto,
desde o momento em que este é expedido para o cliente”.*

Repare-se que, a partir do momento em que, não tendo conseguido alcançar os seus valores-alvo, o produto é entregue ao consumidor, é susceptível de lhe causar toda uma série de inconvenientes (mais ou menos graves), aquando da sua utilização.

Entre outras eventuais consequências, contam-se a publicidade negativa e a reparação/substituição do produto, com os devidos custos para o produtor.

O conjunto de danos resultantes para o consumidor e para o produtor constituiriam a perda causada à Sociedade, segundo Taguchi¹⁵.

A vantagem da produção para um valor-alvo, defendida por Taguchi, é a desta resultar numa menor variabilidade do produto (“inimiga” da qualidade).

Compare-se os esboços gráficos das distribuições da produção, segundo um intervalo de especificações e segundo um valor-alvo:

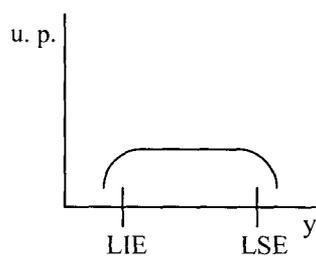


Figura 2.1 – Produção segundo intervalo de especificações

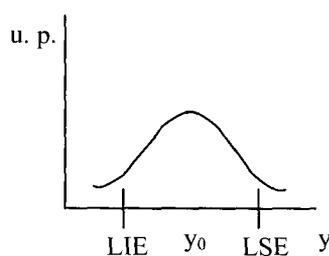


Figura 2.2 – Produção segundo um valor-alvo

y - valores da característica em medição
u. p. - unidades produzidas
LIE - limite inferior das especificações
LSE - limite superior das especificações
y₀ - valor-alvo (que se pretende conferir ao produto)

¹⁴ Unidades consideradas *não-defeituosas*, isto é, cujos valores pertencem ao intervalo de especificações das respectivas características em medição.

¹⁵ Kacker (1986) sugere o alargamento desta definição à fase produtiva, alegando que, ao ser fabricado, o produto é, também, susceptível de gerar perdas para a Sociedade, como, por exemplo, o desperdício de matérias-primas e de trabalho, bem como a resultante de eventual poluição ambiental associada.

3. MÉTODO DE TAGUCHI

O método de Taguchi pertence à categoria dos métodos que, em Qualidade, são classificados de “fora-de-linha”¹⁶, por serem de utilização exterior à linha de produção, em oposição aos métodos “em-linha”¹⁷ que, sendo os mais frequentemente utilizados, executam o controlo durante a fase produtiva. Mais concretamente, o método de Taguchi pode intervir desde a fase de desenho (concepção) do produto¹⁸, ou seja, antes do fabrico, propriamente dito.

3.1 LINHAS GERAIS

- Objectivo do método:

Minimizar a variabilidade do produto, identificando os meios de lhe conferir robustez, durante o processo de fabrico, perante as fontes de variação a que ele é submetido.

- Identificação da solução:

O conhecimento do processo sugere que determinados factores de produção influenciem os valores exibidos pelo produto, para determinada característica a controlar. Se a relação funcional entre os *input's* (níveis desses factores) e o valor do *output* (da característica a controlar) fosse conhecida, poder-se-ia esperar que os cálculos dos níveis em que os primeiros deveriam intervir, para se obter o valor-alvo do produto, fosse imediato. No entanto, identificadas essas variáveis¹⁹, revelar-se-ão, ainda, alguns factores externos, não controláveis durante a produção e que exercem uma acção perturbadora, ocasionando desvios no valor real do *output*, em relação ao esperado.

Denominar-se-á intrínsecas as variáveis controláveis do processo e extrínsecas as de perturbação (ruído).

A situação é, na realidade, ainda mais complicada, pelo facto de não se conhecer a relação funcional entre as variáveis intrínsecas, de *input*, e o *output*.

¹⁶ *Off-line*, em inglês.

¹⁷ *On-line*, em inglês. Exemplo: *cartas de controlo*.

¹⁸ Taguchi identifica 3 fases de desenvolvimento do produto:

1. desenho do sistema (engenharia)
2. desenho dos parâmetros (experimental)
3. desenho das tolerâncias (facultativo)

Na primeira fase, são definidos os parâmetros a considerar durante o desenvolvimento do protótipo. Na segunda fase, são especificados os valores-alvo que os parâmetros, identificados na fase anterior, deverão tomar. Na terceira fase (facultativa), procura-se reduzir mais a variabilidade, caso as experiências realizadas na fase anterior se revelem insuficientes.

¹⁹ Como o factor (de produção ou não), pode intervir em níveis distintos, passa a ser encarado como uma variável.

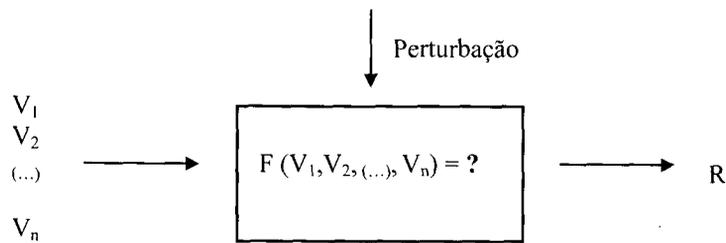


Figura 3.1 – Esquema representativo do sistema

V_i - Variável de produção intrínseca (controlável), de ordem i
 R - Resposta, ou valor obtido para a característica a estudar
 Perturbação - Acção combinada das variáveis (factores) extrínsecas

Para ultrapassar esta limitação, Taguchi sugere que se realize um conjunto de experiências estatisticamente planeadas, em que as variáveis intrínsecas intervenham em níveis diferentes. É nesta fase que o método recorre ao desenho estatístico de experiências (DOE²⁰), devendo, cada experiência-tipo,

- corresponder a uma combinação específica de níveis para as variáveis intrínsecas;
- ser replicada para diferentes condições das variáveis extrínsecas (também previamente planeadas).

As experiências simulam condições específicas de produção sob o efeito de fontes de variação e, analisados os resultados, identificar-se-á quais as variáveis intrínsecas que mais afectam a variabilidade, bem como quais os níveis em que elas deverão ser mantidas, durante a produção, de modo a que essa variação seja mínima.

- Realização de experiência de confirmação:

Identificada a solução proposta pelo método, é ainda necessário confirmar a melhoria (diminuição) da variabilidade, isto é, a obtenção de robustez, através da realização de nova experiência, desta vez com as variáveis intrínsecas a assumir os níveis propostos na etapa anterior.

- Redefinição de tolerâncias (etapa facultativa)

Eventualmente, a experiência de confirmação poderá contradizer a solução proposta ou, noutros casos, apresentar resultados manifestamente insuficientes. Para qualquer destas situações, existe ainda a possibilidade de se tentar melhorar a qualidade do produto através do desenho de tolerâncias para a característica em observação. Esta solução de recurso é condicionada pelos custos envolvidos²¹.

²⁰ DOE, do inglês *design of experiments*.

²¹ A substituição de matérias-primas ou equipamentos, por outros, análogos, mas de tolerâncias mais estreitas, envolve encargos adicionais, uma vez que estes serão mais onerosos.

Um dos aspectos mais atractivos do método de Taguchi é o facto dele sugerir que se realize um número de experiências consideravelmente inferior ao teoricamente previsto pelo DOE, do que resulta uma redução no custo experimental.

As experiências poderão, também, ser simuladas em computador, em vez de realizadas laboratorialmente, com a vantagem de se tornar desnecessário contabilizar quer o erro experimental, quer o efeito do ruído.

3.2 FERRAMENTAS DO MÉTODO

- Função-perda

A função-perda, proposta por Taguchi, é uma função de custo associada ao desvio à idealidade — valor-alvo - apresentado pelo produto. É proporcional ao quadrado do referido desvio, podendo ser expressa como

$$L(y) = k f(y-y_0)^2,$$

$L(y)$ - valor da função, em unidades monetárias

y - valor apresentado pelo produto, para a característica em observação

k - constante de proporcionalidade

y_0 - valor-alvo

e, graficamente, como

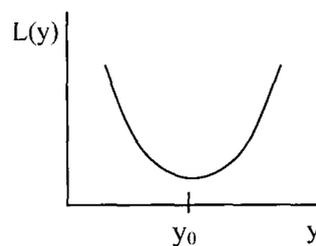


Gráfico 3.1 – Função-perda

- Quocientes SN (sinal-para-ruído²²)

Com base na função-perda, Taguchi construiu diversos quocientes SN — a que chamou de sinal-para-ruído, sugerindo que a maximização desses quocientes constitua a via de minimização da função-perda.

²² *Signal-to-noise*, em inglês.

Cada um dos vários tipos de SN foi concebido para solucionar um problema de natureza específica. Os que se apresenta a seguir, de acordo com três naturezas distintas de problemas, são os mais conhecidos:

A. O valor-alvo é o melhor

$$SN = 10 \log_{10}(y_m^2/s^2)$$

B. O menor valor é o melhor (quanto menor, melhor)

$$SN = -10 \log_{10}[(1/n) \sum y_i^2]$$

C. O maior valor é o melhor (quanto maior, melhor)

$$SN = -10 \log_{10}[(1/n) \sum (1/y_i^2)]$$

y_m - média dos resultados (resposta)

s^2 - variância dos resultados (resposta)

n - nº de experiências do conjunto realizado

Repare-se na reciprocidade entre os quocientes SN referentes aos casos em que “o menor valor é o melhor” e “o maior valor é o melhor”, que não é de admirar, visto que os quocientes foram todos construídos para serem maximizados, com o objectivo subjacente de minimização da função-perda.

- DOE

Taguchi baseou-se nos princípios do desenho estatístico de experiências factoriais e desenhou diversos conjuntos de experiências altamente fraccionadas²³, sob a forma matricial, a serem utilizados consoante o número de factores e o dos respectivos níveis envolvidos no estudo.

Exemplo: As combinações possíveis dos níveis de factores, caso se pretenda testar 7 factores, cada um dos quais a dois níveis, conduziram a 2^7 (=28) experiências distintas (desenho factorial completo).

Taguchi propõe, para a mesma situação, um quadro de apenas 8 destas experiências²⁴, criteriosamente escolhidas — desenho factorial fraccional — de modo a suprimir do estudo a maior parte da informação redundante²⁵ e diminuindo, conseqüentemente, o custo experimental. Para contabilizar o efeito de ruído, Taguchi propõe que se replique cada uma dessas experiências-tipo sob as condições de um desenho adequado aos factores de ruído (variáveis extrínsecas).

²³ Ver o anexo A.1., respeitante ao DOE.

²⁴ Ver o anexos A.2. e A.3., que exemplificam a situação exposta.

²⁵ Taguchi defende que a informação relevante se concentra num reduzido número de experiências (em relação ao total, do desenho factorial completo), pelo que a grande maioria das restantes produziria redundâncias de informação.

- Grafos lineares e tabelas de interacções

Da redução do conjunto de experiências, resulta, inevitavelmente, alguma redução na informação que é possível retirar do estudo. Uma das suas consequências mais indesejáveis é que os efeitos de alguns factores se apresentarão confundidos com os efeitos de algumas interacções²⁶ entre factores.

Para contornar este inconveniente, Taguchi sugere que a afectação dos factores às condições do desenho seja realizada com o auxílio dos grafos lineares ou das tabelas de interacções²⁷. Os dois auxiliares localizam as colunas do desenho que deverão representar interacções entre factores. Deve-se evitar afectar qualquer factor a qualquer das referidas colunas.

Além do exposto, Taguchi defende que a experiência e um conhecimento razoável do processo industrial permitem concluir quais das potenciais interacções poderão ser desprezadas. Advoga ainda que, na prática da realidade industrial, as interacções de ordem superior à segunda são inexistentes (não se manifestam).

- ANOVA

Após a realização do conjunto de experiências seleccionado, Taguchi propõe que se analise estatisticamente os resultados, com um emprego especial da ANOVA²⁸:

- A. Submeter ao método ANOVA os valores de SN, calculados sobre os valores da resposta (y), para identificar os factores que influenciem significativamente a variância — factores-de-controlo.
- B. Submeter ao método ANOVA as médias das respostas do conjunto de experiências, para identificar os factores que apresentem uma relação linear com a resposta — factores-sinal.

A intenção subjacente é a de empregar os factores-de-controlo nos respectivos níveis que minimizem a variabilidade e tirar proveito da relação linear entre o factor-sinal²⁹ e a resposta, empregando-o no nível que ajuste a resposta ao valor-alvo.

3.3 PRESSUPOSTOS

O método de Taguchi foi desenvolvido pressupondo a validade do modelo aditivo para os efeitos dos factores e admitindo que os desvios a este modelo são, na prática, mínimos.

²⁶ Ver o anexos A.4. e A.5., respeitantes a: interacções entre factores; confusão entre efeitos principais de factores e efeitos de interacções.

²⁷ Ver os anexos A.6. e A.7., respeitantes a grafos lineares e a tabelas de interacções.

²⁸ Ver o anexo A.8., referente à ANOVA.

²⁹ O factor-sinal escolhido será aquele que tiver mostrado afectar menos a variabilidade, na primeira análise ANOVA.

A assumpção da independência dos efeitos dos factores está, também, implícita, na aceitação da possibilidade de separação dos respectivos efeitos principais.

Taguchi considera que, na realidade das situações industriais, as interacções de ordem superior à segunda são insignificantes.

3.4 ARTICULAÇÃO DAS DIVERSAS ETAPAS DO MÉTODO

As diversas etapas do método de Taguchi pode ser resumidas de acordo com a seguinte sequência:

- I. Identificação dos factores do sistema
- II. Escolha dos níveis dos factores a experimentar
- III. Selecção do modelo de experiências a adoptar
- IV. Condução da experiência
- V. Apuramento dos factores-de-controlo e dos respectivos óptimos
- VI. Identificação do factor-sinal e respectivo nível a utilizar
- VII. Realização de experiência de confirmação
- VIII. Desenho de tolerâncias

Conhecidos a natureza do problema e o valor-alvo da característica à qual se pretende conferir robustez, a aplicação do método pode ser concretizada mas, se não se tiver, *a priori*, um bom conhecimento do processo (situação mais frequente), poderá ser necessário repetir esta sequência de procedimentos, para um conjunto modificado de factores e/ou respectivos níveis³⁰.

Apresenta-se, em seguida, um conjunto de orientações para a realização das etapas acima discriminadas:

- *Identificação dos factores do sistema:*

Deve envolver a colaboração de todo o pessoal envolvido no projecto, de preferência em sessões do tipo *brainstorming*³¹.

- *Escolha dos níveis dos factores, a utilizar na experiência:*

Deve reflectir as limitações de recursos, e considerações de ordem funcional ou de engenharia. Em relação ao número de níveis a utilizar, a decisão deverá ser parcimoniosa³², não descurando a eventual necessidade de nova utilização do método.

³⁰ É frequente concluir-se que deveriam ter sido incluídos novos factores (suspeita levantada pela experiência anteriormente efectuada) e/ou eliminados outros, conseguindo-se, assim, um conjunto de experiências mais eficaz.

³¹ O *brainstorming* é uma reunião do tipo participativo, estimulando a criatividade e a produção de ideias em períodos de tempo curtos.

³² É desaconselhável o uso de mais de 25% dos recursos na primeira utilização do método.

- *Seleção do modelo de experiências a adoptar:*

Escolher-se-á um desenho de Taguchi compatível com o número de factores intrínsecos e respectivos níveis, e outro que seja adequado ao número de factores extrínsecos (de ruído) e respectivos níveis, a intervir na experiência. O conjunto de experiências definido, no primeiro caso, para os factores intrínsecos deverá ser repetido segundo cada uma das condições definidas para o segundo caso (replicação).

O desenho escolhido para a matriz intrínseca deverá ter um número de experiências igual ou superior ao total de graus de liberdade do sistema³³.

Deve-se atender às limitações que se imponham sob o ponto de vista da exequibilidade.

Afectar-se-á judiciosamente os factores às colunas do desenho: não afectar factores a colunas representativas de interacções, segundo os grafos lineares ou as tabelas de interacções associados ao tipo de desenho escolhido.

- *Condução da experiência*

A experiência não deverá, se concretizada laboratorialmente, ser efectuada segundo a ordem apresentada pelo desenho: esta deverá ser substituída por uma completamente aleatória (salvaguardado o cumprimento de todas as experiências-tipo integrantes do conjunto), a fim de evitar o enviesamento dos resultados, por qualquer influência de carácter sistemático.

- *Apuramento dos factores-de-controlo e dos respectivos óptimos; Identificação do factor-sinal e respectivo nível a utilizar*

Aplicar a ANOVA sobre os quocientes SN, para a identificação dos factores-de-controlo (os que se apresentarem como significativos).

Aplicar a ANOVA sobre as médias das respostas, para identificação do factor-sinal (de entre os que se apresentaram significativos nesta fase, aquele que, em relação a SN, se revelou o menos significativo).

Após a análise ANOVA ter sido realizada sobre os quocientes SN e sobre as médias das respostas, dever-se-á escolher, para os primeiros, os níveis que produziram a menor variabilidade e, para o segundo, o nível que conduza a resposta ao seu valor-alvo. Os restantes factores poderão ser mantidos nos níveis habitualmente utilizados, ou nos que se afigurem mais económicos.

- *Realização de experiência de confirmação*

Realizar uma nova experiência, sob as condições apuradas como óptimas na fase anterior.

³³ Total de g. l. do sistema: $(\sum \text{g. l. factor}) + 1$;
g. l. factor = $(\text{n}^\circ \text{níveis factor}) - 1$.

Se os resultados forem considerados satisfatórios e não sugerirem que possam ser substancialmente melhorados a partir de um novo estudo, implementar-se-á a solução apurada ao processo de fabrico. No caso oposto, as alternativas serão ou uma aplicação melhorada do método, de acordo com as pistas reveladas pela anterior, ou o desenho de tolerâncias.

- *Desenho de tolerâncias*

O redimensionamento das tolerâncias só se justifica quando o desenho dos parâmetros tiver conduzido a resultados manifestamente insuficientes, quanto à redução da variabilidade do produto.

É nesta fase que se decide que factores deverão ser substituídos por análogos de tolerâncias mais estreitas e, por isso, de custos mais elevados.

O desenho das tolerâncias segue a lógica do desenho dos parâmetros, mas as matrizes intrínseca e extrínseca serão, naturalmente, diferentes:

- a) ambas conterão os mesmos factores, (apenas aqueles cujas tolerâncias se procura melhorar), mas em níveis diferentes; as respectivas dimensões e afectações de factores deverão ser realizadas em conformidade;
- b) a matriz intrínseca mantém as definições dos níveis dos factores empregues na fase de desenho dos parâmetros; exemplo: “1”= “baixo”; “2”= “elevado”;
- c) a matriz extrínseca utiliza novas definições dos níveis dos factores, relacionadas com os intervalos de tolerâncias usuais (de baixo custo); exemplo: “1”= “nominal”; “2”= “- $\alpha\%$ nominal”; “3”= “+ $\alpha\%$ nominal”.

Dos exemplos acima referidos, a combinação resultante seria, para o nível “1” da matriz intrínseca, o próprio valor “baixo” das tolerâncias, se a extrínseca indicasse “1”, uma redução desse valor em $\alpha\%$, se a indicação fosse “2”, ou o valor acrescido de $\alpha\%$, no caso de indicar “3”.

4. CONTROVÉRSIA GERADA PELO MÉTODO

Algumas das críticas mais frequentes, suscitadas pelo método de Taguchi, incidem sobre

- *Interações*

Aponta-se como a principal fraqueza do método a escassez de meios para a avaliação de potenciais interações entre os factores controláveis do sistema, bem como dos respectivos efeitos. Taguchi acredita que, na maioria das situações práticas, as interações entre dois factores ou não se fazem sentir ou o seu efeito se afigura manifestamente desprezável, podendo, eventualmente ser aproximado.

- *Quocientes SN*

Nalguns casos, a maximização de SN é de interpretação dúbia, como, por exemplo, no problema em que o valor-alvo é o melhor – tanto pode ser obtida pela minimização da variância amostral, como pela maximização da resposta média. Taguchi prefere esse uso simultâneo no quociente, por permitir a optimização a dois passos (primeiro em relação à variância, depois, em relação à resposta média).

- *Médias marginais*

O uso de gráficos de médias marginais, ignorando eventuais interacções, para identificar os níveis dos factores-de-controlo que maximizam SN (minimizando a variabilidade) é criticado, por não garantir, por si, a identificação do óptimo. Também se aponta como ineficaz o uso de grafos lineares e de tabelas de interacções como auxiliares na afectação dos factores às colunas do desenho.

- *ANOVA*

O modo como a ANOVA é utilizada no método é apontado como conducente a resultados espúria.

- *Matrizes intrínseca e extrínseca*

O uso combinado das duas matrizes é criticado com o argumento de que origina experiências de dimensão excessiva, quando comparado com o uso de uma única matriz que integre todos os factores (intrínsecos e extrínsecos).

- *Variância*

Também surgem objecções quanto à generalização do método aos casos em que haja replicação e em que a variância não seja constante, defendendo-se a necessidade de transformar os dados.

5. CONCLUSÕES

Em relação ao assunto exposto, sobressaem as seguintes reflexões:

- É crucial, para a competitividade das empresas, a atenção aos sinais do mercado, com o fim da adopção atempada de uma estratégia compatível com a evolução deste.
- Tem-se assistido ao crescimento continuado, a nível internacional, do uso do método de Taguchi como impulsionador da melhoria da qualidade, e são muitos os casos de sucesso divulgados (constata-se, facilmente, por simples pesquisa na Internet). Assim, torna-se desejável o conhecimento do método.

- A simplicidade e o baixo custo de utilização do método tornam-no particularmente atractivo, face aos benefícios que poderá proporcionar.
- O método torna-se acessível a não-estatísticos, graças à utilização sistematizada e simplificada dos princípios teóricos subjacentes: não é necessário desenhar a experiência, basta a selecção de um desenho compatível com o número de variáveis e dos respectivos níveis que se pretende utilizar. No entanto, o desconhecimento das implicações envolvidas na utilização de um desenho fraccional, nomeadamente a confusão entre efeitos, pode revelar-se contraproducente.
- A controvérsia que envolve o método é de cariz exclusivamente teórico e a sua importância é, frequentemente, minimizada, face à ânsia de obtenção da melhoria por ele anunciada. Esta atitude está relacionada com a necessidade de obtenção rápida de resultados e com o uso do conhecimento empírico para efectuar simplificações que conduzam à exequibilidade. Sob esta óptica, apresenta-se preferível empregar o método que apresentou resultados positivos, ainda que não ideal sob o ponto de vista formal, dado que não é conhecido outro que se afigure igualmente atractivo e simultaneamente consensual, do ponto de vista teórico. A colaboração dos estatísticos no emprego do método de Taguchi e/ou no desenvolvimento de outro, melhorado, adquire, então, uma importância evidente.

REFERÊNCIA BIBLIOGRÁFICA

- CASTRO, P. M. (1999) "Método de Taguchi: Apresentação e caso prático com utilização do SAS", *Tese de Mestrado em Estatística e Gestão de Informação*, Instituto Superior de Estatística e Gestão de Informação. Universidade Nova de Lisboa

ANEXOS

A.1. DESENHO DE EXPERIÊNCIAS (DOE)

A técnica do DOE³⁴ foi inicialmente desenvolvida e apresentada por Fisher (anos 20, Inglaterra), em investigação agrícola, para testar a influência de factores como a chuva e a composição orgânica do solo no rendimento das colheitas. Com o apoio da análise de variância, estimar-se-ia a significância dos efeitos dos referidos factores.

Em DOE, faz-se variar conjuntamente os níveis dos factores e observa-se as correspondentes alterações na variável-resposta. Esta técnica de planeamento de experiências facilita o estudo do funcionamento dos sistemas, quando se desconhece as relações funcionais entre as variáveis de resposta (ou de saída) e as variáveis de entrada (factores) representativas do sistema: perante a impossibilidade de efectuar a modelização teórica, analisa-se empiricamente o efeito que os factores produzem na resposta, ao assumirem determinados níveis. A experiência em que se estuda o efeito da variação simultânea de factores múltiplos é denominada factorial. O desenho (plano da experiência) é elaborado na forma matricial: a cada coluna, corresponde um factor e, a cada linha (ponto do desenho), uma combinação única de níveis dos factores, que caracteriza uma determinada experiência-tipo.

A diferença entre a experiência factorial completa e a experiência factorial fraccional é que, enquanto a primeira é constituída sobre o espaço completo das combinações possíveis³⁵, a segunda é constituída sobre um subespaço do caso anterior, estrategicamente definido³⁶. A experiência factorial fraccional resulta de uma redução efectuada com base nos princípios das matrizes ortogonais, segundo técnicas de projecção.

A.2. EXPERIÊNCIA DE TAGUCHI (EXEMPLO)

matriz intrínseca: L8 (8 ensaios)
matriz extrínseca: L4 (4 ensaios)

Ensaio	Réplicas	Observações	y_m	SN
1	(1), (2), (3), (4)	$y_{11}, y_{12}, y_{13}, y_{14}$	y_{m1}	SN_1
2	(1), (2), (3), (4)	$y_{21}, y_{22}, y_{23}, y_{24}$	y_{m2}	SN_2
(...)	(...)	(...)	(...)	(...)
8	(1), (2), (3), (4)	$y_{81}, y_{82}, y_{83}, y_{84}$	y_{m8}	SN_8
Total de Observações		32 (= 8 x 4)		

(y_m : resposta média; SN: quociente sinal-para-ruído)

³⁴ Do inglês "design of experiments"

³⁵ Dos níveis escolhidos para os factores que nela intervêm

³⁶ De modo a que a omissão na informação decorrente da redução no tamanho da experiência se confine à de menor relevância

A.3. O DESENHO L8 DE TAGUCHI

L8	Nº Coluna						
Nº Ensaio	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

(2⁷)

2 níveis, 7 factores

A.4. INTERACÇÕES ENTRE FACTORES

Cada factor exerce um efeito individual, sobre a variável-resposta, designado efeito principal. No entanto, essa influência sobre a variável-resposta pode assumir perfis distintos, em função do nível de outro factor presentes no estudo. Quando tal acontece, está-se na presença de interacção entre dois factores³⁷.

A.5. INTERACÇÕES ENTRE FACTORES

Decorre da construção de um desenho factorial fraccional (a partir da redução do respectivo desenho factorial completo) que não será possível estimar os efeitos de todas as interacções concebíveis, para os factores em estudo, pois elas estarão confundidas entre si, ou, eventualmente, com efeitos principais de factores.

Cada desenho fraccional terá a sua estrutura específica de confusão entre efeitos — em conformidade com as próprias regras de fraccionamento do desenho factorial —, que pode ser descrita através de um conjunto de regras.

A resolução de um desenho fraccional define os tipos de confusões entre efeitos que lhe poderão estar associados, tais como se exemplifica:

- Resolução III – Os efeitos principais não são confundidos entre si, mas pelo menos um deles está confundido com uma interacção (ou mais) entre dois factores.

³⁷ Se representarmos, graficamente, o comportamento da variável-resposta em função de um desses factores, obteremos linhas distintas para diferentes níveis de outro factor. A existência de uma interacção entre ambos é identificada com o não-paralelismo nas referidas linhas.

- Resolução IV – Os efeitos principais não estão confundidos entre si nem com interações entre dois factores, mas algumas das últimas poderão estar confundidas entre si.
- Resolução V – Nem os efeitos principais, nem as interações entre dois factores se confundem entre si, mas existe confusão entre interações de dois factores e interações de três factores.

Para se prever se dois tipos de efeitos podem apresentar-se confundidos, basta somar as ordens desses efeitos e comparar o resultado com o grau de resolução associado ao desenho: não haverá confusão se o valor calculado for inferior a este³⁸.

A.6. GRAFOS LINEARES (EXEMPLOS)

Estão construídos grafos lineares para as diversas matrizes ortogonais de Taguchi, podendo haver mais do que um tipo para cada desenho.

Exemplo para o desenho de Taguchi L8:

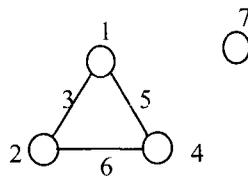


Figura A.6.1 – Grafo linear L8 (1)

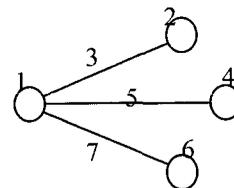


Figura A.6.2 – Grafo linear L8 (2)

A.7. TABELAS DE INTERAÇÕES ENTRE DOIS FACTORES (EXEMPLO)

À semelhança do que acontece com os grafos lineares, existem tabelas de interações entre duas colunas, construídas para diferentes tipos de desenhos,

³⁸ Exemplo: o desenho de resolução III confunde efeitos principais de factores com efeitos de interações entre dois factores ($1 + 2 = 3$), mas não confunde efeitos principais de dois factores entre si ($1 + 1 < 3$).

O exemplo que se apresenta em seguida, refere-se à matriz ortogonal L16:

Interacções entre duas Colunas

		Nº Coluna														
Nº Coluna		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	1	3	2	5	4	7	6	9	8	11	10	13	12	15	14	
3	2	1	6	7	4	5	10	11	8	9	14	15	12	13		
4	3	7	6	5	4	11	10	9	8	15	14	13	12			
5	4	1	2	3	12	13	14	15	8	9	10	11				
6	5	3	2	13	12	15	14	9	8	11	10					
7	6	1	14	15	12	13	10	11	8	9						
8	7	15	14	13	12	11	10	9	8							
9	8	1	2	3	4	5	6	7								
10	9	3	2	5	4	7	6									
11	10	1	6	7	4	5										
12	11	7	6	5	4											
13	12	1	2	3												
14	13	3	2													
15	14	1														
16	15															

Figura A.7.1 – Tabela de interacções para o desenho L16

A.8. RECURSO À ANOVA

Resumidamente, a ANOVA parte da decomposição da variabilidade total dos dados nas suas componentes, do seguinte modo:

$$SS_{\text{Total}} = SS_{\text{Factor}} + SS_{\text{Erro}} \quad (\text{SS} = \text{soma de quadrados})$$

Com base nesta partição, são calculados os quadrados médios do factor e do erro (MS_{Factor} e MS_{Erro} ³⁹, respectivamente), para testar a hipótese (H_0) de que as médias dos níveis do factor são iguais, ou seja, que os factores não afectam a resposta.

A aceitação ou a rejeição de H_0 é feita por comparação de $F_0 = MS_{\text{Factor}}/MS_{\text{Erro}}$, com F_c (valor crítico, retirado da tabela da distribuição F)

$$\text{Se } F_0 > (F_c = F_{\alpha, n_F, n_E}),$$

α – significância,

n_F – graus de liberdade de MS_{Factor} ,

n_E – graus de liberdade de MS_{Erro} ,

a hipótese nula é rejeitada, e admite-se que o factor seja significativo.

³⁹ O cálculo dos valores de MS efectua-se dividindo cada SS pelo respectivo número de graus de liberdade.

Dado que o modelo de base, considerado neste tipo de análise de variância, é o de efeitos fixos, as conclusões só são válidas para os níveis do factor em estudo, pelo que cada factor terá de ser analisado separadamente⁴⁰.

A.8.1. UTILIZAÇÃO DE TAGUCHI

Para um nível de significância de $\alpha\%$, dizer que um factor é significativo, equivale a afirmar que a probabilidade de a resposta se manter inalterada, quando este muda de nível, é igual ou inferior a $\alpha\%$.

Os factores serão considerados significativos se os respectivos valores de F_0 (calculados) forem superiores ao valor crítico F_c (tabelado). Taguchi propõe duas utilizações sequenciais da ANOVA:

- Em relação à estatística SN,
determinar os factores que a influenciam significativamente e colocá-los nos níveis que maximizam SN
- Em relação à média da resposta,
determinados os factores que a influenciam significativamente, seleccionar aquele que menor influência tiver apresentado sobre SN e usá-lo como factor-sinal, para ajustar a resposta ao seu valor-alvo.

⁴⁰ O teste F utilizado no método ANOVA pressupõe que os erros sejam i.i.d. e tenham distribuição normal. Assim, seria de se esperar que, num panorama diferente, os resultados não fossem fiáveis. No entanto, admite-se que o teste F seja relativamente insensível a desvios em relação aos pressupostos referidos, e que, portanto, apresente resultados válidos, mesmo que estes existam.

Índices de Custo de Vida: Metodologias de Cálculo

Autores:
Paulo Parente
Teresa Bago d' Uva

VOLUME I

1º QUADRIMESTRE DE 2001

ÍNDICES DE CUSTO DE VIDA: METODOLOGIAS DE CÁLCULO

COST OF LIVING INDEXES: METHODOLOGIES

Autores: Paulo Parente

- Técnico Superior de Estatística no Gabinete de Estudos e Conjuntura do Instituto Nacional de Estatística – Núcleo de Apoio e Desenvolvimento

Teresa Bago d' Uva

- Técnica Superior de Estatística no Gabinete de Estudos e Conjuntura do Instituto Nacional de Estatística – Núcleo de Apoio e Desenvolvimento

RESUMO:

- A comparação de custos de vida em diferentes áreas geográficas pode ser feita através do cálculo de índices de custo de vida. Têm vindo a ser desenvolvidas diversas metodologias de cálculo destes índices, no âmbito das comparações internacionais, não existindo consenso na escolha do método mais adequado.

Neste artigo, descrevemos algumas das metodologias mais utilizadas, bem como as suas propriedades mais relevantes, com vista à sua aplicação a regiões dentro do mesmo país. A aplicação ao caso português é apresentada em Bago d'Uva e Parente (2001).

PALAVRAS-CHAVE:

- *Índices de custo de vida, paridades de poder de compra.*

ABSTRACT:

- The comparison of the cost of living in different geographical areas can be done using cost of living indexes. Several methodologies have been developed in order to make international comparisons, none of which is generally accepted to be the most adequate.

In this article, not only do we describe some of most known methodologies, but also their most important properties. Our goal is to apply these methods to compare geographical areas within the same country. The empirical application to the Portuguese case is presented in Bago d' Uva and Parente (2001).

KEY-WORDS:

- *Cost of living indexes, purchasing power parities.*

VOLUME I

1º QUADRIMESTRE DE 2001

1. INTRODUÇÃO

Neste estudo, pretendemos apresentar diversas metodologias que permitem a comparação entre os custos de vida em diferentes áreas geográficas.

Esta comparação pode ser feita utilizando os chamados índices de custo de vida. O verdadeiro índice de custo de vida entre duas regiões é igual ao rácio dos custos de aquisição dos cabazes que permitem obter o mesmo nível de utilidade⁴¹ u^* :

$$ICDV = \frac{C(P_b, u^*)}{C(P_a, u^*)},$$

em que P_b e P_a são os preços de cada uma das regiões, $C(.)$ é a despesa mínima que permite atingir o nível de utilidade u^* . Por outras palavras, o índice de custo de vida dá o montante relativo em que se deve aumentar (ou diminuir) a despesa numa determinada região, em relação a outra, de modo a que o seu nível de utilidade seja idêntico.

Na secção 2, apresentamos algumas metodologias de cálculo dos índices de custo de vida entre duas regiões. Os métodos apropriados para comparações entre mais do que duas regiões são expostos na secção 3. Na secção 4, apresentamos algumas abordagens de avaliação de números índices. As propriedades dos índices estudados são referidas na secção 5. Por último, na secção 6, fazemos uma síntese das principais conclusões.

2. ÍNDICES BILATERAIS

Começemos por supor, para simplificar, a existência de apenas duas regiões. Vamos considerar a situação mais usual em que a informação sobre os preços está disponível a um nível mais desagregado do que a das despesas. Os dados disponíveis são os preços de bens e serviços que são classificados em M categorias e as despesas totais de cada uma dessas categorias.

Para cada categoria $i = 1, 2, \dots, M$, o índice de preços entre a região j e a região k é dado por:

⁴¹ Utilidade é a satisfação obtida com o consumo de um determinado cabaz de bens e serviços.

$$IP_i = \left(\frac{P_j}{P_k} \right)_i = \left[\prod_{r=1}^{A_i} \left(\frac{P_{rj}}{P_{rk}} \right)_i \right]^{1/A_i}$$

em que P_{rj} é o preço do item r na região j e A_i é o número de bens ou serviços nessa categoria.

Os índices de Laspeyres e Paasche que apresentamos de seguida permitem agregar os índices de preços de todas as categorias e obter uma estimativa do índice de custo de vida.

Índice de Laspeyres

$$L_{jk} = \frac{\sum_{i=1}^M IP_i \times w_{ik}}{\sum_{i=1}^M w_{ik}}$$

Índice de Paasche

$$S_{jk} = \frac{\sum_{i=1}^M w_{ij}}{\sum_{i=1}^M (1/IP_i) \times w_{ij}}$$

em que w_{ij} representa a despesa na categoria i na região j .

Geralmente, toma-se a média geométrica destas duas estimativas do índice de custo de vida que é designada por índice ideal de Fisher:

$$F_{jk} = \sqrt{S_{jk} \times L_{jk}} .$$

É consensual que este índice é o mais indicado para comparações bilaterais mas não é apropriado quando a análise envolve mais do que duas regiões dado que não verifica uma propriedade importante: a transitividade. Diz-se que um índice I é transitivo se o rácio dos índices de duas regiões em relação a uma mesma base, I_{AB} / I_{CB} , for igual ao índice entre aquelas duas regiões, I_{AC} .

3. MÉTODOS MULTILATERAIS

O cálculo dos índices multilaterais envolve dois passos: primeiro, obtêm-se índices para cada categoria de bens e serviços, chamados *basic headings*, que são, depois, agregados para níveis mais elevados.

3.1. MÉTODOS DE CÁLCULO DE ÍNDICES AO NÍVEL DOS BASIC HEADINGS

A análise de mais do que uma região traz quase sempre um problema no cálculo dos índices ao nível dos *basic headings*. A falta de informação sobre os preços de alguns bens ou serviços nalgumas regiões não permite utilizar as fórmulas apresentadas para o caso bilateral, em que se supõe que todos os preços são conhecidos. A utilização exclusiva dos preços dos produtos comuns a todas as regiões implicaria um grande desperdício de informação, pelo que a solução deste problema passa por procedimentos adequados que permitam utilizar o máximo de informação possível.

Os índices ao nível da categoria de bens e serviços devem ser transitivos e podem ser calculados através de dois métodos: o método *Country-Product-Dummy* (CPD) e o método de Elteto, Koves e Szulc (EKS)⁴². Estes métodos utilizam todos os preços disponíveis em pelo menos duas regiões.

O Método CPD

O método CPD envolve uma regressão para cada categoria de bens e serviços em que se incluem dois conjuntos de variáveis *dummy* – um para a região e outro para o produto. Assim, a regressão a estimar para a categoria $i = 1, 2, \dots, M$ é a seguinte:

$$\ln(P_{rj}^i) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{N-1} X_{N-1} + \gamma_1 Y_1 + \gamma_2 Y_2 + \dots + \gamma_{A_i} Y_{A_i} + u, \quad j = 1, 2, \dots, N, \\ r = 1, 2, \dots, A_i,$$

em que N é o número de regiões, A_i é o número de produtos existentes para cada nível de despesa e $\ln(P_{rj}^i)$ é o logaritmo natural do preço do bem r na região j . As variáveis X são as variáveis *dummy* correspondentes a $N - 1$ regiões, a região base, N , não é incluída. Cada produto é representado pela respectiva variável *dummy* Y . Os coeficientes β 's das variáveis *dummy* correspondentes a cada região são os logaritmos dos índices de preços (que designaremos por CPD_{ijN}) para a categoria de bens e serviços e os γ 's são os logaritmos naturais dos preços dos produtos em moeda do país base.

⁴² O primeiro autor a propor este índice foi Gini pelo que o índice também é conhecido por *Gini-EKS*.

Note-se que, se não existirem células vazias na matriz de preços, o índice de preços calculado pelo método *CPD* para cada categoria é a média geométrica das paridades dos respectivos produtos individuais.

O Método *EKS*

O método *EKS* permite que a transitividade seja estabelecida tendo em conta comparações directas e indirectas entre países. Segundo este método as paridades entre os países *j* e *k* devem ser dadas por uma média geométrica das paridades directas entre os países *j* e *k* e todos os índices indirectos que podem ser calculados com as restantes *N - 2* regiões.

$$EKS_{ijk}^{BH} = \left[IP_{ijk}^2 \prod_{l=1, l \neq j, k}^N \frac{IP_{ijl}}{IP_{ikl}} \right]^{\frac{1}{N}}$$

com

$$IP_{ijk} = \left(\frac{P_j}{P_k} \right)_i = \left[\prod_{r=1}^{A_i} \left(\frac{P_{rj}}{P_{rk}} \right)_i \right]^{1/A_i}$$

Estes IP_{ijk} são calculados apenas com produtos comuns entre aos países *j* e *k* ou com determinados produtos seleccionados.

3.2. MÉTODOS DE AGREGAÇÃO

Existem vários métodos de agregação dos índices de preços em índices de custo de vida entre zonas geográficas. Geralmente, este tipo de estudo compara vários países, tomando os índices de custo de vida o nome de paridades do poder de compra. Um dos raros exemplos da aplicação deste tipo de índices a comparações dentro do mesmo país é dado por Kokoski et al. (1994).

Ao contrário do caso bilateral, não existe consenso na escolha do método de cálculo dos índices multilaterais. Neste estudo, apresentamos os dois métodos mais utilizados, o método Geary-Khamis (*GK*) e o método *EKS*, e algumas suas generalizações.

O Método GK

O método *GK* é baseado na noção de que existem dois conjuntos de variáveis desconhecidas - as paridades de poder de compra e os preços internacionais de cada categoria - cujos valores podem ser calculados resolvendo um sistema de equações. A nomenclatura do *GK*, que manteremos neste estudo, é coerente com o seu objectivo de comparar países, devendo o leitor notar que, na sua aplicação à comparação de regiões dentro de um mesmo país, as paridades e os preços internacionais correspondem, respectivamente, aos índices de custo de vida regionais e aos preços nacionais.

O preço internacional de uma categoria, Π_i , é uma média, ponderada por quantidades, dos seus preços em todos os países, depois de convertidos à unidade monetária do país base através da respectiva paridade. A paridade de cada país j , que será designada por GK_j , $j = 1, 2, \dots, N$, é definida como o rácio da sua despesa total valorizada aos seus próprios preços e essa despesa total valorizada a preços internacionais. Mais especificamente, o sistema Geary-Khamis é constituído pelas seguintes equações:

$$\Pi_i = \sum_{j=1}^N \frac{P_{ij}}{GK_j} \left[\frac{q_{ij}}{\sum_{j=1}^N q_{ij}} \right], i = 1, \dots, M$$

$$GK_j = \frac{\sum_{i=1}^M P_{ij} q_{ij}}{\sum_{i=1}^M \Pi_i q_{ij}}, j = 1, \dots, N$$

Este sistema é constituído por $N + M$ equações e $N + M$ variáveis. No entanto, como se toma a paridade do país base igual a 1, $GK_N = 1$, uma das equações é redundante resultando, assim, um sistema de $N + M - 1$ equações de fácil resolução por ser linear em Π_i , $i = 1, \dots, M$, e $1/GK_j$, $j = 1, \dots, N$.

A principal vantagem do método *GK* é a produção de resultados consistentes em termos aditivos: para cada região, a soma das despesas de todas as categorias valorizadas aos respectivos preços internacionais é igual à despesa total convertida para a moeda da região base através da sua paridade.

Cuthbert (1999) criou o método *Generalised GK*, *GGK*, que, como o nome indica, é uma generalização do descrito acima, sendo também aditivo no sentido exposto.

O Método GGK

Seja b um vector de elementos constantes positivos β_j . Então, o método GGK correspondente a b define os valores Π_i e GGK_j como solução das equações:

$$\Pi_i = \sum_{j=1}^N \frac{P_{ij} \beta_j}{GGK_j} \left[\frac{q_{ij}}{\sum_{j=1}^N \beta_j q_{ij}} \right], i=1, \dots, M$$

$$GGK_j = \frac{\sum_{i=1}^M P_{ij} q_{ij}}{\sum_{i=1}^M \Pi_i q_{ij}}, j=1, \dots, N$$

Dentro da classe dos índices GGK existe um subgrupo com particular importância⁴³. Para $0 \leq \alpha \leq 1$, escolha-se β_j tal que satisfaz:

$$\beta_j = \left(\sum_{i=1}^M \Pi_i q_{ij} \right)^{-\alpha}.$$

Com $\alpha = 0$ temos o método GK e com $\alpha = 1$ temos o método proposto por Iklé (1972).

Na prática não é este sistema que é utilizado, uma vez que é necessário ter a informação sobre as quantidades físicas compradas e os preços das categorias. Utiliza-se um sistema modificado, em que se tem em conta as paridades por categoria calculadas pelo método CPD⁴⁴.

$$\bar{\Pi}_i = \sum_{j=1}^N \frac{CPD_{ijN} \beta_j}{\bar{GK}_j} \left[\frac{Q_{ij}}{\sum_{j=1}^N \beta_j Q_{ij}} \right], i=1, \dots, M$$

$$\bar{GK}_j = \frac{\sum_{i=1}^M CPD_{ijN} q_{ij}}{\sum_{i=1}^M \bar{\Pi}_i q_{ij}}, j=1, \dots, N$$

$$\beta_j = \left(\sum_{i=1}^m \bar{\Pi}_i Q_{ij} \right)^{-\alpha}$$

⁴³ Esta classe de índices foi introduzida por Cuthbert (2000).

⁴⁴ Note-se que poderíamos utilizar também o índice EKS para cada categoria de bem e serviço em alternativa ao CPD. Contudo, tradicionalmente, os métodos CPD e GK são utilizados em conjunto.

em que $Q_{ij} = P_{ij}q_{ij} / CPD_{ij}$. Para resolver este sistema, é necessário conhecer as paridades e as despesas de todas as categorias. Pode provar-se que $\bar{\Pi}_i = \Pi_i / P_m$ e que $\overline{GK}_j = GK_j$.

Os preços internacionais $\bar{\Pi}_i$ não são invariantes à base, porém, isso não constitui problema uma vez que não são utilizados isoladamente mas, sim, multiplicados pelas quantidades Q_{ij} . Note-se que o produto $\bar{\Pi}_i \times Q_{ij} = (\Pi_i / P_m) \times [(P_{ij}q_{ij}) / CPD_{ijn}]$ é invariante à base.

O Método EKS e algumas generalizações

O método EKS ao nível agregado consiste nos seguintes passos:

1. Se utilizarmos o método EKS no cálculo das paridades ao nível de categoria de produto elementar então é necessário calcular:

Índice de Laspeyres

$$L_{jk} = \frac{\sum_{i=1}^M EKS_{ijk}^{BH} \times w_{ik}}{\sum_{i=1}^M w_{ik}}$$

Índice de Paasche

$$S_{jk} = \frac{\sum_{i=1}^M w_{ij}}{\sum_{i=1}^M (1 / EKS_{ijk}^{BH}) \times w_{ij}}$$

em que w_{ij} representa a despesa no bem ou serviço i na região j .

Índice ideal de Fisher:

$$F_{jk} = \sqrt{S_{jk} \times L_{jk}}$$

Com F_{jn} o índice de Fisher da região j com a região n .

2. O índice de custo de vida entre duas regiões é dado por:

$$EKS_{jN} = \left[F_{jN}^2 \prod_{k=1, k \neq j, N}^N \frac{F_{jk}}{F_{Nk}} \right]^{\frac{1}{N}}$$

A grande desvantagem do método *EKS* face ao método *GK* é que não é consistente em termos aditivos.

Caves et al. (1982) propuseram substituir no ponto 1 o índice de Fisher pelo índice de Theil-Tornqvist.

$$T_{jk} = \prod_{i=1}^M (EKS_{ijk}^{BH})^{s_{ijk}} \text{ com } s_{ijk} = \frac{1}{2} \left(\frac{w_{ij}}{\sum_{i=1}^M w_{ij}} + \frac{w_{ik}}{\sum_{i=1}^M w_{ik}} \right).$$

O índice *EKS* revisto por Caves et al. (1982) tem o nome de *EKS* generalizado.

Este método pode ser criticado por dar o mesmo peso a todas as trocas indirectas (a média calculada no ponto 2 é não ponderada) o que é pouco consistente com a realidade, em que algumas destas trocas indirectas são mais importantes do que outras.

Seguindo este princípio Kokoski et al. (1994) utilizaram no cálculo de índices de custo de vida regionais a seguinte média geométrica ponderada:

$$KCM_{jN} = \prod_{l=1}^N \left(\frac{T_{kl}}{T_{Nl}} \right)^{\delta_l} \text{ com } \delta_l = \frac{\sum_{i=1}^M w_{il}}{\sum_{l=1}^N \sum_{i=1}^M w_{il}}.$$

Os ponderadores da média geométrica correspondem ao peso na despesa da região base em relação ao total nacional pelo que, neste índice, têm mais peso as paridades em relação às regiões mais ricas.

4. ABORDAGENS DE AVALIAÇÃO DE NÚMEROS ÍNDICES

Existem várias abordagens para avaliar a qualidade de um número índice. Entre estas, as principais são a abordagem axiomática e a abordagem económica⁴⁵.

⁴⁵ Para um estudo mais pormenorizado das diferentes abordagens dos números índices ver Diewert (1987)

4.1. ABORDAGEM AXIOMÁTICA

Suponhamos que temos os pares de preços e quantidades (p_{ij}, y_{ij}) para duas regiões tal que $i = 1, \dots, M$ e $j = 1, 2$. Seja $p_j = (p_{1j}, p_{2j}, \dots, p_{Mj})$ e $y_j = (y_{1j}, y_{2j}, \dots, y_{Mj})$. O primeiro passo para a determinação de um número índice consiste em resolver o problema de agregação. Nesta primeira etapa, pretende desenvolver-se uma metodologia para agregar a informação de preços e quantidades por forma a obter-se um único nível de preços e quantidades, ou seja, calcular P_j e Y_j tais que:

$$\sum_{i=1}^M p_{ij} y_{ij} = P_j Y_j.$$

O segundo passo no caso dos índices bilaterais consiste na criação de índices de preços IP e índices de quantidades IQ :

$$IP = \frac{P_1}{P_2} \text{ e } IQ = \frac{Y_1}{Y_2}.$$

Na abordagem axiomática, P_j e Y_j são vistos como funções dos vectores p_j e y_j , que são independentes, ou seja:

$$P_j = P(p_j, y_j) \text{ e } Y_j = Q(p_j, y_j)$$

Segundo esta abordagem, um bom número índice deve passar um conjunto de testes fundamentais. Nas comparações bilaterais, é indiferente aplicar os testes aos índices de preços IP ou aos índices de quantidades IQ uma vez estes, juntamente com a hipótese $P_1 = 1$, permitem calcular os valores agregados P_2 , Y_1 e Y_2 . Um conjunto de axiomas que um índice bilateral deve respeitar são os famosos critérios de selecção de Fisher (Murteira, 1993).

Quando trabalhamos com índices multilaterais, os testes são definidos em termos das *share functions* $\{S^1, S^2, \dots, S^N\}$, ou seja, das proporções do produto regional em relação ao total nacional, uma vez que, fazendo $Y_i = S^i$, é possível calcular $P^i = p^i y^i / Y_i$, para $i = 1, \dots, N$. Sejam P a matriz de preços e Y a matriz de quantidades de todas as regiões, S_k é função de P e Y , $S_k = S_k(P, Y)$. Ao conjunto de funções $\{S^1(P, Y), S^2(P, Y), \dots, S^N(P, Y)\}$ chamamos um sistema multilateral de *share functions*.

Se estamos a comparar apenas duas regiões, o índice bilateral de quantidades é dado por $Q(p^1, p^2, y^1, y^2) = S^2(p^1, p^2, y^1, y^2) / S^1(p^1, p^2, y^1, y^2)$.

Diewert (1996)⁴⁶ considerou os seguintes 12 axiomas a que um sistema multilateral deve obedecer:

1. *Teste de Partilha:*

As N funções $S^k(P, Y), k = 1, \dots, N$ são positivas, contínuas e tais que $\sum_{k=1}^N S^k(P, Y) = 1$ para todo o P e Y .

2. *Teste das Quantidades proporcionais:*

Suponhamos que $y^k = \beta_k y$, $\beta_k > 0$ e $\sum_{k=1}^N \beta_k = 1$, para $k = 1, \dots, N$. Segundo este teste, se para todas as quantidades a proporção relativamente à região k é β_k , é razoável pedir que $S^k(P, Y) = \beta_k$.

3. *Teste dos preços proporcionais:*

Suponhamos que $p^k = \alpha_k p$ para $k = 1, \dots, N$ com $\alpha_k > 0$ para algum $p \gg 0_M$. Então:

$$S^k(P, Y) = p^k y^k / \left(\sum_{j=1}^N p^j y^j \right), \text{ para } k = 1, \dots, N.$$

Se o vector de preços de cada região, p^k , é proporcional a um vector positivo comum p , então, este pode ser utilizado para determinar a proporção de produto consumida na região k .

4. *Invariância a alterações das unidades de medida:*

Seja $\delta_i > 0$ para $i = 1, \dots, M$ e seja $\hat{\delta}$ a matriz diagonal com elementos δ_i . Então:

$$S^k(\hat{\delta}P, \hat{\delta}^{-1}Y) = S^k(P, Y) \text{ para } k = 1, \dots, N.$$

Este teste requer que as *share functions* sejam invariantes a alterações das unidades de medida dos M bens.

⁴⁶ Ver, também, Diewert (1987).

5. *Invariância à ordenação dos bens:*

Seja Π a matriz $M \times M$ de permutação. Então:

$$S^k(\Pi P, \Pi Y) = S^k(P, Y), \text{ para } k = 1, \dots, N.$$

Segundo este teste, as *share functions* devem manter-se quando se altera a ordem dos países.

6. *Tratamento simétrico das regiões:*

Seja $S(P, Y)^T = [S^1(P, Y), \dots, S^N(P, Y)]$ o vector linha com as *share functions* e seja Π^* uma matriz permutação $N \times N$. Então:

$$S(P\Pi^*, Y\Pi^*)^T = S(P, Y)^T \Pi^*.$$

Este teste é conhecido por invariância à região/país base e implica que nenhuma região é tratada de forma desigual.

7. *Teste das unidades Monetárias:*

Seja $\alpha_k > 0$, para $k = 1, \dots, N$. Então:

$$S^k(\alpha_1 p^1, \dots, \alpha_N p^N, Y) = S^k(p^1, \dots, p^N, Y).$$

A utilização de diferentes unidades monetárias em cada região não afecta a percentagem do produto de cada região. Matematicamente, esta propriedade indica homogeneidade de grau zero das *share functions*.

8. *Teste de Homogeneidade das quantidades:*

Para $i = 1, \dots, N$, $\lambda_i > 0$, $i \neq j$, $j = 1, \dots, N$, tem-se:

$$\begin{aligned} S^i(P, y^1, \dots, y^{i-1}, \lambda_i y^i, y^{i+1}, \dots, y^N) / S^j(P, y^1, \dots, y^{i-1}, \lambda_i y^i, y^{i+1}, \dots, y^N) = \\ = \lambda_i S^i(P, Y) / S^j(P, Y) \end{aligned}$$

Segundo este axioma, a percentagem do produto do país i relativamente ao país j , S^i / S^j , é linearmente homogénea nos produtos do país i .

9. *Teste de Monotonicidade:*

$S^k(P, y^1, \dots, y^N)$ é crescente nas componentes do vector y^k para $k = 1, \dots, N$. Se qualquer componente do vector de quantidades do país k aumenta, a percentagem de produto do país k também aumenta.

10. *Teste de Partição das Regiões:*

Seja A um subconjunto do conjunto $\{1, \dots, N\}$ com pelo menos dois elementos. Suponhamos que, para $i \in A$: $p^i = \alpha_i p^a$, com $\alpha_i > 0$, $p^a \gg 0_M$; $y^i = \beta_i y^a$, com $\beta_i > 0$, $y^a \gg 0_M$ e $\sum_{i \in A} \beta_i = 1$. Seja B o subconjunto de $\{1, \dots, N\}$ que não pertence a A , e P^b e Y^b as matrizes de preços e quantidades das regiões pertencentes a B . Então:

(i) Para $i \in A, j \in A, S^i(P, Y) / S^j(P, Y) = \beta_i / \beta_j$

(ii) Para $i \in B, S^i(P, Y) = S^{i*}(p^a, P^b, y^a, Y^b)$, em que $S^{i*}(p^a, P^b, y^a, Y^b)$ é o sistema de *share functions* obtido se acrescentarmos os vectores p^a e y^a às matrizes P^b e Y^b .

Este teste significa que, se o vector de quantidades agregadas do bloco A , denominado por y^a , é distribuído proporcionalmente pelos seus membros e se os preços de cada membro de A também são proporcionais, as *share functions* da regiões do bloco reflectem a distribuição proporcional do produto e as das restantes regiões são iguais às que seriam calculadas se o bloco A fosse uma única região.

11. *Teste de agregação bilateral:*

Sejam A e B dois subconjuntos disjuntos de $\{1, \dots, N\}$. Suponhamos que, para $i \in A$, $p^i = \alpha_i p^a$, $y^i = \beta_i y^a$, $\alpha_i > 0$, $\beta_i > 0$, $\sum_{i \in A} \beta_i = 1$ e, para $j \in B$, $p^j = \gamma_j p^b$, $y^j = \delta_j y^b$, $\gamma_j > 0$, $\delta_j > 0$ e $\sum_{j \in B} \delta_j = 1$. Então:

$$\sum_{j \in B} S^j(P, Y) / \sum_{i \in A} S^i(P, Y) = Q_F(p^a, p^b, y^a, y^b),$$

em que Q_F é o índice ideal de Fisher definido como:

$$Q_F = \sqrt{\frac{\left(\frac{p^{a'} y^b}{p^{a'} y^a} \right)}{\left(\frac{p^{b'} y^b}{p^{b'} y^a} \right)}}$$

Segundo este teste, se o conjunto de regiões é dividido em dois blocos, A e B , dentro dos quais as quantidades e os preços são proporcionais, o melhor índice bilateral $Q(p^a, p^b, y^a, y^b)$ será obtido dividindo a soma das *share functions* para o bloco B pela soma para o bloco A . Diewert (1996) escolheu o índice ideal de Fisher Q_F pois é o índice que satisfaz o maior número de testes bilaterais, contudo, podemos substituir este índice por um alternativo.

3. Teste de aditividade:

O sistema multilateral deve ser aditivo. Um sistema aditivo tem a seguinte definição: Existem M funções positivas e diferenciáveis de $2NM$ variáveis $g_l(P, Y)$, $l = 1, \dots, M$ tais que:

$$S^k(P, Y) = \frac{\sum_{l=1}^M g_l(P, Y) y_l^k}{\sum_{l=1}^M \left[g_l(P, Y) \sum_{j=1}^M y_l^j \right]}$$

onde as funções $g_l(P, Y)$, $l = 1, \dots, M$, têm a seguinte propriedade:

$$g_l(p, \dots, p, Y) = p, \text{ para todo o } p \gg 0_M \text{ e } Y,$$

em que $p = [p_1, \dots, p_M]$ é um vector de preços comum a todos os países.

4.2. ABORDAGEM ECONÓMICA

Esta abordagem baseia-se na hipótese de que os agentes económicos existentes em todas as regiões têm um comportamento otimizador, ou seja, os consumidores têm como objectivo maximizar a sua utilidade e os produtores procuram maximizar o seu lucro. Antes de expor a abordagem económica é necessário introduzir o conceito de função agregadora.

Uma função agregadora $f(x)$ é uma função de M variáveis não negativas $x \geq 0_M$ que verifica as seguintes propriedades:

- (i) f é uma função positiva quando todos os argumentos são positivos ($f(x) > 0$ se $x \gg 0_M$);
- (ii) f é linearmente homogénea ($f(\lambda x) = \lambda f(x)$ para $\lambda \geq 0$, $x \geq 0_M$);
- (iii) f é côncava ($f[\lambda x^1 + (1-\lambda)x^2] \leq \lambda f(x^1) + (1-\lambda)f(x^2)$ para $0 \leq \lambda \leq 1$, $x^1 \geq 0_M$, $x^2 \geq 0_M$).

Vamos considerar a teoria do consumidor e admitir que a função utilidade, que designaremos por $f(y)$, é uma função agregadora comum a todas as regiões. Cada agregado familiar maximiza a sua utilidade sujeita a uma restrição orçamental. Agregando estes problemas para o total regional, verificamos que o vector das quantidades procuradas em cada região, y^k , é solução de:

$$\max_y \{f(y) : p^k \cdot y = p^k \cdot y^k\}, k=1, \dots, N$$

Considere-se a função custo total $C(u, p) = \min\{p \cdot x : f(x) \geq u\}$. A homogeneidade da função f implica $C(u, p) = uc(p)$ em que $c(p) = \min_z \{p \cdot z : f(z) \geq 1\}$ é a função custo unitário. Logo, devido à relação entre os problemas primal e dual, tem-se:

$$p^k \cdot y^k = c(p^k) f(y^k)$$

Consideremos em primeiro lugar índices bilaterais. Note-se que os índices de preços e os índices de quantidade têm que verificar o critério de reversão de factores, isto é, quando, através de uma mesma fórmula, se calcula um índice de quantidades e um índice de preços, o produto destes índices deve dar origem um índice de valor:

$$IP(p^k, p^j, y^k, y^j) IQ(p^k, p^j, y^k, y^j) = p^k \cdot y^k / p^j \cdot y^j,$$

logo, uma vez que os consumidores têm um comportamento otimizador:

$$IP(p^k, p^j, y^k, y^j) IQ(p^k, p^j, y^k, y^j) = c(p^k) f(y^k) / c(p^j) f(y^j).$$

Assim, uma função agregadora f é definida como exacta para o índice de quantidades $IQ(p^k, p^j, y^k, y^j)$ se, para cada $p^k \gg 0_M$ e $p^j \gg 0_M$, y^k é solução do problema $\max_y \{f(y) : p^k \cdot y = p^k \cdot y^k\}$, y^j é solução do problema $\max_y \{f(y) : p^j \cdot y = p^j \cdot y^j\}$ e:

$$IQ(p^k, p^j, y^k, y^j) = f(y^k) / f(y^j).$$

De modo semelhante, a função agregadora f com função custo unitária c é definida como exacta para o índice de preços $IP(p^k, p^j, y^k, y^j)$ se, para cada $p^k \gg 0_M$ e $p^j \gg 0_M$, y^k é solução do problema $\max_y \{f(y) : p^k \cdot y = p^k \cdot y^k\}$, y^j é solução do problema $\max_y \{f(y) : p^j \cdot y = p^j \cdot y^j\}$ e:

$$IP(p^k, p^j, y^k, y^j) = c(p^k) / c(p^j).$$

Uma função agregadora é “flexível” se é capaz de dar uma aproximação diferencial de segunda ordem de qualquer função duas vezes diferenciável e linearmente homogénea. Isto significa que a função que aproxima é igual à função que pretende aproximar num determinado ponto, verificando-se o mesmo para as primeiras e segundas derivadas, ou seja:

$$c(p^*) = c^*(p^*), \nabla_p c(p^*) = \nabla_p c^*(p^*), \nabla_{pp} c(p) = \nabla_{pp} c^*(p^*) \text{ para algum } p^*.$$

Diewert (1976) definiu um índice de preços (quantidades) como superlativo se a função custo unitário c (função utilidade f) que é exacta para ele é flexível. Em Diewert (1996) um sistema multilateral superlativo é definido de forma análoga. Sob as hipóteses optimizadoras dos agentes económicos, uma função agregadora é definida como exacta para uma *share function* se:

$$S^i(P, Y) / S^j(P, Y) = f(y^i) / f(y^j) \text{ para } 1 \leq i, j \leq N.$$

Assim, um sistema multilateral é superlativo se a função utilidade que é exacta para a *share function* é flexível⁴⁷.

Um índice superlativo tem em conta a possibilidade de substituição no consumo de bens induzida pela variação dos seus preços. Considerando que a utilidade se mantém, a composição do cabaz óptimo (de custo mínimo) vai ser diferente.

Collier (1998) criticou esta abordagem por ignorar a lei de Engel. Uma função utilidade é linearmente homogénea se e só se as preferências são homotéticas. Recorde-se que as preferências são homotéticas se os conjuntos de indiferença estão relacionados de uma forma proporcional ao longo de raios, ou seja, se o cabaz x é indiferente ao cabaz y então o cabaz αx é indiferente ao cabaz αy . Neste caso, a elasticidade da procura em relação ao rendimento é igual a um. No entanto, a lei de Engel diz que, à medida que o rendimento aumenta, diminui a sua proporção que é gasta em alimentação, ou seja, a elasticidade procura-rendimento neste grupo de produtos é inferior a um.

Podemos concluir que um sistema superlativo só resulta nos índices multilaterais correctos se as preferências forem homotéticas. No caso bilateral, Diewert (1976) mostrou que o índice de Fisher (que é um índice superlativo) é consistente com o axioma fraco das preferências reveladas. Este resultado implica que este índice dá a resposta correcta mesmo quando as preferências não são homotéticas.

⁴⁷ Para mais detalhes sobre a definição de sistemas multilaterais superlativos ver Diewert (1996).

5. PROPRIEDADES DOS ÍNDICES APRESENTADOS

Diewert (1996) estudou as propriedades de dez classes de índices multilaterais, dois dos quais são apresentados neste artigo, o *EKS* e o *GK*. Diewert (1996) demonstrou não só que o sistema multilateral *EKS* passa todos os testes excepto o teste 10 e o teste de aditividade 12, mas também que este sistema é superlativo. Em relação ao sistema multilateral *GK*, este autor provou que passa todos os testes, à excepção do 8, 9 e 11 (passaria o teste 11 se Q_F fosse substituído por outro índice apropriado). Adicionalmente, Diewert (1996) mostrou que a única função utilidade agregadora que é exacta para ele é a função utilidade linear (que corresponde ao caso de bens substitutos perfeitos). Como esta função não é flexível, este índice não é superlativo. Diewert (1996) concluiu que, do ponto de vista axiomático, não existe nenhum sistema dominante. Contudo, a abordagem económica sugere a utilização do sistema *EKS* uma vez que é superlativo.

O método *GK* origina o efeito Gerschenkron, ou seja, sobreavalia a despesa de uma região, se os seus preços relativos forem muito diferentes dos preços relativos internacionais. Este é um dos problemas do índice de Paasche que o *GK*, por ser deste tipo, reflecte.

Relembremos que o verdadeiro índice de custo de vida entre duas regiões é igual ao rácio dos custos de aquisição dos cabazes que permitem obter o mesmo nível de utilidade u^* . Pode ser representado por:

$$ICDV_j = \frac{\sum_{i=1}^m P_{ij} q_{ij}}{\sum_{i=1}^m \Pi_i \bar{q}_{ij}},$$

em que \bar{q}_{ij} são as quantidades óptimas associadas aos preços Π_i que permitem obter o mesmo nível de utilidade, u^* , que se obtém se o nível de preços é P_{ij} consumindo as quantidades óptimas q_{ij} . Um consumidor racional, dada a utilidade u^* , consome as quantidades q_{ij} quando os preços são P_{ij} e as quantidades \bar{q}_{ij} caso sejam Π_i . O cabaz óptimo no primeiro caso deixa de o ser se os preços forem diferentes, portanto, o seu custo é superior ao do novo cabaz óptimo:

$$\sum_{i=1}^m \Pi_i q_{ij} \geq \sum_{i=1}^m \Pi_i \bar{q}_{ij}.$$

Logo, tem-se que $ICDV_j \geq GK_j$, ou seja, o índice *GK* subestima sempre⁴⁸ o verdadeiro índice de custo de vida. O factor determinante desta diferença é a alteração

⁴⁸ Note-se que do lado do produtor existe um efeito substituição que tem o sentido contrário, contudo em geral o efeito substituição do consumidor é dominante (veja-se Hill, 2000)

do cabaz escolhido devida à alteração de preços, a deslocação da procura para bens cujos preços descem. O índice *GK* de uma região vai estar tanto mais distante do seu verdadeiro índice do custo de vida, quanto maior for a distância entre os seus preços relativos e os preços relativos internacionais. Como os preços internacionais estão mais próximos dos países mais ricos, os índices dos países mais pobres vão estar mais subestimados o que faz com que as suas quantidades sejam relativamente sobreavaliadas, quando valorizadas a preços internacionais.

Uma solução que tem vindo a ser apontada para o problema da aproximação excessiva dos preços internacionais aos dos países mais ricos (veja-se Cuthbert, 1999) é a utilização de quantidades fictícias para ponderar os preços de cada país. Estas quantidades permitem que o peso a dar a cada país seja o mesmo na medida em que têm o mesmo valor quando avaliadas a preços internacionais.

$$\sum_{i=1}^M \Pi_i q_{ij}^* = C, \text{ para qualquer } j.$$

O índice proposto por Iklé (1972) verifica esta propriedade uma vez que se considera $q_{ij}^* = \beta_j q_{ij}$ e $\beta_j = \left(\sum_{i=1}^M \Pi_i q_{ij} \right)^{-\alpha}$. Segundo alguns estudos empíricos (veja-se Hill, 2000) este índice é mais robusto ao efeito Gerschenkron que o método *GK*.

As propriedades dos índices *GGK* foram pouco estudadas na literatura, relativamente aos outros índices que apresentamos, contudo, é possível concluir imediatamente que dão origem a um sistema multilateral aditivo. Um índice multilateral é aditivo quando verifica a seguinte igualdade:

$$\sum_{l=1}^M g_l(P, Y) y_l^k = \frac{P^k \cdot y^k}{ICDV^k}.$$

A aditividade implica que é indiferente avaliar um cabaz numa determinada região a preços internacionais/nacionais ou avaliar a preços regionais e converter para os preços de uma determinada região através do índice de custo de vida. Esta propriedade é verificada pelos índices *GGK*. A partir desta propriedade e da análise de Diewert (1996) para o índice *GK* é fácil verificar que se um índice é aditivo então um sistema multilateral é aditivo.

6. CONCLUSÕES

Neste artigo pretendemos fazer uma síntese crítica das principais metodologias de cálculo dos índices de custo de vida.

É consensual que o Índice Ideal de Fisher é o mais indicado para comparar preços de dois países (regiões). No entanto, este não é adequado quando a análise envolve mais regiões e a escolha de um método, neste caso, é controversa.

Os índices multilaterais são calculados em dois passos. O primeiro passo envolve o cálculo de índices em cada categoria de bens e serviços, os chamados *basic headings*. Neste estudo, são apresentados dois métodos de cálculo de índices a este nível, o método *CPD* e o método *EKS*, ambos transitivos e invariantes à base escolhida. Para além destas propriedades desejáveis, estes métodos têm a vantagem de não exigir a disponibilidade de todos os preços em todas as regiões, no cálculo dos índices em cada categoria, evitando o desperdício de informação.

O segundo passo consiste na agregação dos índices obtidos na primeira fase ponderados pelas despesas das respectivas categorias. Os métodos mais comuns são o Geary-Khamis (*GK*) e o Elteto-Koves-Schultz (*EKS*), tendo sido desenvolvidas generalizações, respectivamente, o *KCM* e o *GGK*, como forma de dar resposta a algumas críticas.

Estudamos duas abordagens do problema de cálculo de índices de custo de vida multilaterais, a abordagem económica e a axiomática. Os métodos apresentados são avaliados à luz destas duas abordagens. Enquanto a abordagem económica dá suporte teórico ao método *EKS*, a abordagem axiomática não permite fazer uma escolha clara entre o método *GK* e o *EKS*.

As características inerentes aos índices dados pelas diversas metodologias podem levar a resultados diferentes. Por exemplo, a sensibilidade do índice *GK* ao efeito Gerschenkron faz com que este produza índices diferentes dos dados pela metodologia *EKS* que não é afectada por aquele problema.

Em Bago d'Uva e Parente (2001), é feito um exercício de aplicação das metodologias apresentadas à realidade portuguesa que permite analisar até que ponto os índices de custo de vida calculados são sensíveis ao método utilizado.

REFERÊNCIAS BIBLIOGRÁFICAS

BAGO D'UVA, T. e PARENTE, P. (2001), "Índices de Custo de Vida: Um estudo sobre a realidade Portuguesa", *Revista de Estatística*, 1º Quadrimestre de 2001, INE;

CAVES, D. W.; CHRISTENSEN, L. R. & DIEWERT, W. E. (1982), "Multilateral Comparisons of Output, Input and Productivity using Superlative Index Numbers", *The Economic Journal*, 92, March, pp. 73-86;

COLLIER JR., I.L. (1998), "Comment on W. Erwin Diewert's Axiomatic and Economic Approaches to International Comparisons", *Diskussionsbeiträge des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin*, nº 1998/3, March;

- CUTHBERT, J.R. (1999), "Categorisation of Additive Purchasing Power Parities", *Review of Income and Wealth*, Series 45, Number 2, June, pp. 235-249;
- CUTHBERT, J.R. (2000) "Theoretical and practical issues in purchasing power parities illustrated with reference to the 1993 Organization for Economic Co-operation and Development data", *Journal of the Royal Statistical Society*, Series A, 163, Part 3, pp. 421-444;
- DIEWERT, W.E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics*, Vol.4, No. 2, reeditado em Diewert, W.E., & Nakamura, A.O. (1993), *Essays in Index Number Theory*, Vol. 1, North-Holland;
- DIEWERT, W.E., (1987), "Index Numbers", *The New Palgrave: A Dictionary of Economics*, Vol.2, pp. 767-780, reeditado em Diewert, W.E., & Nakamura, A.O. (1993), *Essays in Index Number Theory*, Vol. 1, North-Holland;
- DIEWERT, W.E., (1996), "Axiomatic and Economic Approaches to International Comparisons", *Discussion paper n° 96-11*, Department of Economics, The University of British Columbia, Vancouver, Canada;
- HILL, J. R., (2000), "Measuring substitution bias in international comparisons based on additive purchasing power parity methods", *European Economic Review*, 44, pp. 145-162;
- IKLÉ, D. M. (1972), "A New Approach to the Index Number Problem", *The Quarterly Journal of Economics*, 86, pp.188-211;
- KOKOSKI, M.; CARDIFF, P. & MOULTON, B. (1994), "Interarea Price indices for Consumer Goods and Services: An Hedonic Approach Using CPI Data", *BLS Working Paper 256*, July, Office of Prices and Living Conditions, U. S. Department of Labor, Washington, D.C.;
- MURTEIRA, B.J.F. (1993) *Análise Exploratória de Dados, Estatística Descritiva*, McGraw-Hill.

VOLUME I

1º QUADRIMESTRE DE 2001

Índices de Custo de Vida: Um estudo sobre a realidade Portuguesa

Autores:
Teresa Bago d' Uva
Paulo Parente

VOLUME I

1º QUADRIMESTRE DE 2001

ÍNDICES DE CUSTO DE VIDA: UM ESTUDO SOBRE A REALIDADE PORTUGUESA

COST OF LIVING INDEXES: THE PORTUGUESE CASE THROUGH A MAGNIFYING GLASS

Autores: Teresa Bago d' Uva

- Técnica Superior de Estatística no Gabinete de Estudos e Conjuntura do Instituto Nacional de Estatística – Núcleo de Apoio e Desenvolvimento

Paulo Parente

- Técnico Superior de Estatística no Gabinete de Estudos e Conjuntura do Instituto Nacional de Estatística – Núcleo de Apoio e Desenvolvimento

RESUMO:

- O objectivo deste estudo é avaliar as diferenças de custo de vida em Portugal. Existem diversas metodologias que permitem a comparação de custos de vida entre países diferentes. Neste artigo, fazemos um exercício de aplicação de algumas dessas metodologias à comparação de custos de vida entre as regiões portuguesas. Consideramos as regiões NUTSII (Norte, Centro, Lisboa e Vale do Tejo, Alentejo, Algarve, Região Autónoma dos Açores e Região Autónoma da Madeira) nos anos 1994 a 1997.

Independentemente do método usado, Lisboa e Vale do Tejo é a região em que o custo de vida é mais elevado, ao longo do período considerado, seguida da R.A. da Madeira. No entanto, as diferenças observadas dentro de Portugal não são muito significativas.

PALAVRAS-CHAVE:

- *Comparações regionais, Índices de custo de vida, paridades de poder de compra.*

ABSTRACT:

- The aim of this study is to discover whether or not the cost of living differs significantly within Portugal. There are several methodologies that allow the comparison of costs of living among different countries. In this article, some of the most common methodologies are applied in order to compare the costs of living of Portuguese regions. We considered the regions NUTSII (Norte, Centro, Lisboa e Vale do Tejo, Alentejo, Algarve, Região Autónoma dos Açores and Região Autónoma da Madeira) in the years 1994 to 1997.

Independently of the method used, Lisboa e Vale do Tejo is the 'most expensive' region throughout the period considered, followed by Madeira. However, the differences observed between the Portuguese regions are not very significant.

KEY-WORDS:

- *Regional comparisons, Cost of living indexes, purchasing power parities.*

VOLUME I

1º QUADRIMESTRE DE 2001

1. INTRODUÇÃO

O objectivo deste estudo é investigar a existência de diferenças significativas no custo de vida dentro de Portugal. A comparação entre custos de vida de diferentes áreas geográficas pode ser feita utilizando os chamados índices de custo de vida.

Existem diversas metodologias de cálculo de índices de custo de vida entre países (ver Parente e Bago d'Uva, 2001, para um resumo dos métodos mais usuais e respectivas propriedades). As características inerentes aos índices dados pelas diversas metodologias podem levar a resultados diferentes. Por exemplo, a sensibilidade do índice Geary-Khamis ao efeito Gerschenkron faz com que este produza índices diferentes dos dados pela metodologia Elteto-Koves-Schultz que não é afectada por aquele problema.

Neste estudo, fazemos um exercício de aplicação de algumas dessas metodologias ao cálculo de índices de custo de vida entre as regiões portuguesas. Consideramos as regiões NUTSII (Norte, Centro, Lisboa e Vale do Tejo, Alentejo, Algarve, Região Autónoma dos Açores e Região Autónoma da Madeira) nos anos 1994 a 1997. A aplicação de várias metodologias permitirá a análise da sensibilidade dos índices ao método adoptado.

Na secção 2 é feita a descrição dos dados utilizados no estudo, cujos resultados são apresentados e analisados na secção 3. Por último, na secção 4, fazemos uma síntese das principais conclusões.

2. DADOS

Os preços utilizados para o cálculo dos índices de custo de vida dos anos 1994 a 1997 resultam dos preços recolhidos ao longo deste período com o objectivo de calcular os respectivos Índices de Preços do Consumidor (Base: 1991=100)⁴⁹. O Índice de Preços do Consumidor mensal (Base: 1991=100) de cada região utilizou as médias aritméticas dos preços recolhidos mensalmente em diversos estabelecimentos na respectiva região. Neste estudo os preços de cada artigo, para cada ano, são obtidos a partir das médias aritméticas daquelas médias mensais. O painel de preços é constituído por cerca de 580 artigos, oscilando este número ligeiramente ao longo do período considerado e de região para região, de acordo com as características inerentes a cada mercado.

Os dados referentes às despesas em cada uma das regiões são as estimativas obtidas através do Inquérito aos Orçamentos Familiares 1994/1995⁵⁰. A utilização da

⁴⁹ Ver Índice de Preços do Consumidor (Base: 1991=100) - Metodologia - Instituto Nacional de Estatística.

⁵⁰ Ver Inquérito aos Orçamentos Familiares 1994/1995 - Metodologia - Instituto Nacional de Estatística.

mesma estrutura de despesas para os quatro anos estudados, devida à não existência de informação para além da referida, tem implícita a hipótese de que esta não se alterou significativamente ao longo do período em causa.

Existe informação sobre despesas de 1280 categorias de bens ou serviços, número que excede o número de preços disponíveis. Os métodos de cálculo de índices de custo de vida pressupõem que existe informação sobre os preços de um cabaz de bens e serviços e sobre despesas de grupos desses bens e serviços, ou seja, que a informação disponível sobre as despesas é mais agregada do que a dos preços (Parente e Bago d' Uva, 2001). A diferença entre o nível de detalhe da informação existente obrigou a um estudo exaustivo com vista à organização de uma base de dados que permitisse a aplicação dos métodos de cálculo habituais. Esta adaptação passou pela agregação da informação sobre as despesas, obtendo-se um número de categorias inferior ao número de preços disponíveis.

Os métodos que utilizaremos para o cálculo dos índices de preços em cada categoria, *CPD* e *EKS*, não exigem o conhecimento de todos os preços em todas as regiões (Parente e Bago d' Uva, 2001). Requerem apenas que, para cada par de regiões, exista pelo menos um produto em comum. A constituição das categorias foi condicionada por este requisito pois existem algumas diferenças entre o conjunto de produtos cujo preço é recolhido nas várias regiões, sobretudo entre as regiões do Continente e as regiões autónomas. Nos casos em que as regiões autónomas não recolhem nenhum preço de uma categoria, há que agregá-la a outra em que isso não aconteça para que seja possível calcular o índice de preços.

O objectivo essencial deste estudo é o cálculo de índices de custo de vida entre as sete regiões, porém, foi também feita uma análise separada para as regiões do Continente e para as regiões autónomas. Houve, então, dois processos de constituição de categorias de produtos, um destinado à análise conjunta de todas as regiões do país e apenas das ilhas e outro destinado à análise das regiões do continente que resultaram, respectivamente, em 188 e 218 categorias.⁵¹

Quando um bem ou serviço existe apenas numa região, o seu preço não é considerado no âmbito da comparação de preços entre regiões porque não faz sentido a inclusão de um bem ou serviço que não exista em pelo menos duas regiões.

O quadro seguinte dá um exemplo de constituição de uma categoria, a do arroz. Os preços de arroz disponíveis são utilizados para o cálculo do índice de preços da categoria. Este índice vai ser ponderado, no cálculo do índice agregado, pela despesa total em arroz.

Despesas	Preços
ARROZ GRÃO EXTRA LONGO	ARROZ EXTRA LONGO BRANQUEADO
ARROZ GRÃO LONGO	ARROZ ESPECIAL MÉDIO BRANQUEADO
ARROZ GRÃO MÉDIO	
OUTROS TIPOS DE ARROZ	

⁵¹ Foi utilizada informação de preços de 577 produtos e 1187 categorias de despesa, as razões pelas quais não foi considerada toda a informação disponível são apresentadas mais à frente.

Alguns aspectos relacionados com os dados introduzem certas limitações no estudo que estamos a desenvolver.

O mais importante prende-se com o facto de as bases de dados utilizadas serem destinadas à comparação de preços ao longo do tempo para cada região e não entre regiões num dado momento do tempo. A preocupação que existe na recolha de dados para o cálculo do Índice de Preços do Consumidor é a de que os produtos observados se mantenham de um ano para o outro, em cada região. Devem considerar-se os produtos mais representativos de cada região, o que pode levar à impossibilidade de estabelecer uma equivalência entre todos eles. Ora é precisamente esta equivalência que seria desejável no nosso caso para evitar enviesamentos nos índices de preços dos *basic headings*. Tentámos colmatar esta limitação pela análise detalhada, para todas as regiões, de todos os produtos cujos preços são recolhidos e o estabelecimento de correspondências o mais razoáveis possível.

Outra condicionante da qualidade dos resultados que obteremos é a falta de informação sobre algumas categorias de despesas. Esta lacuna levou a não considerar essas categorias pois, apesar de existir informação suficiente para determinar o seu índice de preços, não existe a despesa para ponderá-lo.

O Índice de Preços do Consumidor considera os produtos mais representativos do consumo das famílias, em cada região, pelo que a informação sobre a despesa de produtos cujo preço não é recolhido não deveria ser considerada. De facto, alguma informação sobre despesas não foi utilizada por este motivo. No entanto, considerámos que não devíamos desprezar a informação sobre despesas de alguns produtos, a nosso ver, importantes, apesar de o seu preço não ser recolhido. Estas despesas, não podendo ponderar os índices de preços dos respectivos produtos, vão ponderar os índices de preços de produtos relacionados. Caso estes pertençam a várias categorias, a despesa é distribuída por todas elas nas mesmas proporções que as suas despesas originais. Este procedimento pareceu-nos mais adequado, nos casos em que foi utilizado, do que o abandono das respectivas despesas o que seria equivalente a distribuí-las proporcionalmente por todos as categorias e não só pelas que estão relacionadas com os produtos em causa⁵².

Por último, é necessário salientar que a categoria relativa às rendas de habitação não foi incluída no cálculo dos índices de custo de vida, apesar de ter um peso considerável nas despesas de consumo. A não inclusão desta categoria deveu-se à indisponibilidade da informação necessária ao cálculo dos respectivos índices de preços.⁵³

⁵² Note-se que este procedimento é utilizado correntemente no cálculo do Índice de Preços do Consumidor.

⁵³ Para os anos até 1996, a informação sobre as rendas de habitação existe apenas ao nível do país. A informação com o nível de detalhe necessário para o nosso estudo, ou seja, para cada região, só existe para o ano de 1997 e seguintes não tendo, ainda assim, sido considerada para o último ano do período que estudámos por pretendermos a comparabilidade dos índices ao longo desse período.

3. ÍNDICES DE CUSTO DE VIDA EM PORTUGAL ENTRE 1994 E 1997

Nesta secção, são apresentados índices de custo de vida para as sete regiões NUTSII (Norte, Centro, Lisboa e Vale do Tejo, Alentejo, Algarve, Região Autónoma dos Açores e Região Autónoma da Madeira), para os anos de 1994 a 1997. Estes índices foram calculados de acordo as metodologias habitualmente utilizadas com o objectivo de comparar custos de vida entre países. Os métodos mais comuns são o Geary-Khamis (*GK*) e Elteto-Koves-Schultz (*EKS*), dos quais derivam algumas generalizações, respectivamente, o *Generalised GK (GGK)* e o *Kokoski-Cardiff-Moulton (KCM)* (ver Parente e Bago d'Uva, 2001). Apresentamos os resultados obtidos através dos métodos *EKS*, *KCM* e o *GK* generalizado com parâmetro igual a 0 (*GK*), 0,5 e 1 (*Iklé*).

As diferenças existentes entre as regiões autónomas e Portugal Continental levaram-nos a pensar que a análise destes dois grupos em separado poderia resultar em relações mais correctas entre as regiões do que as obtidas considerando todo o país. A motivação para considerar apenas as regiões de Portugal Continental foi também a de testar a sensibilidade dos índices calculados à constituição das categorias de bens e serviços que, como referido na secção anterior, é diferente consoante se incluem as regiões autónomas ou não. Esta experiência foi feita, tendo-se comparado os índices de custo de vida entre regiões de Portugal Continental (e entre Açores e Madeira) com as relações dadas pelos índices entre todas as regiões. As relações estimadas entre regiões de Portugal Continental não sofrem alterações quando deixamos de considerar as regiões autónomas. Pelo contrário, a relação bilateral entre Açores e Madeira difere significativamente da que resulta da análise multilateral de todas as regiões. Contudo, optámos por não apresentar aqui os resultados da análise separada das regiões de Portugal Continental e das regiões autónomas uma vez que o objectivo primordial deste estudo é comparar o custo de vida em todas as regiões portuguesas.

No que diz respeito à agregação das categorias de bens e serviços, foram considerados três níveis. Calculámos os índices de custo de vida com todas as categorias e, a um nível mais desagregado, considerámos 10 grandes grupos cujos índices de preços resultam de agregar as categorias que as compõem ponderadas pelas respectivas despesas.⁵⁴ O grupo da alimentação é ainda dividido em 7 subgrupos.

3.1 ÍNDICES DE CUSTO DE VIDA

O Quadro 1 e as Figuras 1 e 2 apresentam os índices de custo de vida para as regiões de Portugal obtidos através de alguns métodos, tendo como base a região Norte.

⁵⁴ Considerámos, também, um nível de desagregação com 31 classes cujos resultados não são apresentados por esse detalhe não ser do âmbito deste artigo mas que serão disponibilizados aos eventuais interessados.

Quadro 1: Resultados dos vários índices para o período 1994 a 1997

Anos	Regiões	GK	GGK(0,5)	Iklé	EKS	KCM
1994	Norte	100,0	100,0	100,0	100,0	100,0
	Centro	98,9	99,0	99,1	99,5	99,4
	Lisboa e Vale do Tejo	105,3	105,4	105,5	105,4	105,3
	Alentejo	100,5	100,7	100,9	100,8	100,6
	Algarve	100,7	101,0	101,4	101,5	101,4
	R.A. Açores	97,8	98,6	99,9	101,4	101,4
	R.A. Madeira	102,4	102,8	103,6	104,5	104,5
1995	Norte	100,0	100,0	100,0	100,0	100,0
	Centro	98,8	98,8	98,9	99,3	99,3
	Lisboa e Vale do Tejo	105,3	105,5	105,7	105,5	105,4
	Alentejo	100,8	100,9	101,1	101,2	101,1
	Algarve	100,7	101,0	101,5	101,6	101,5
	R.A. Açores	99,4	100,1	101,3	102,7	102,7
	R.A. Madeira	103,0	103,3	104,0	104,9	105,0
1996	Norte	100,0	100,0	100,0	100,0	100,0
	Centro	98,4	98,5	98,5	99,0	98,9
	Lisboa e Vale do Tejo	104,6	104,7	104,9	104,7	104,6
	Alentejo	100,7	100,7	100,9	100,9	100,9
	Algarve	100,3	100,6	101,1	101,2	101,1
	R.A. Açores	98,8	99,6	100,9	102,1	102,2
	R.A. Madeira	102,5	102,9	103,6	104,3	104,4
1997	Norte	100,0	100,0	100,0	100,0	100,0
	Centro	98,0	98,1	98,1	98,7	98,6
	Lisboa e Vale do Tejo	104,2	104,4	104,6	104,4	104,2
	Alentejo	100,8	100,9	101,0	101,0	101,0
	Algarve	100,1	100,5	101,0	101,1	100,9
	R.A. Açores	97,5	98,4	99,8	101,2	101,4
	R.A. Madeira	101,8	102,2	103,0	103,7	103,8

Numa primeira análise, julgamos importante referir dois aspectos. O primeiro está relacionado com as diferenças entre os vários índices calculados que estão de acordo com o que seria de esperar, dada a sua diferente natureza (Parente e Bago d'Uva, 2001). O segundo é o facto de todos os métodos apontarem para níveis de custo de vida que não podem ser considerados muito díspares, a amplitude máxima para todos os métodos é cerca de 8% em 1994 e 6% nos restantes anos, relativamente ao custo de vida do Norte.

As regiões que suportam o custo de vida mais elevado ao longo de todo o período considerado são Lisboa e Vale do Tejo e a Madeira. Quanto às restantes regiões, apesar de verem a sua posição relativa variar com o método utilizado e de ano para ano, têm um custo de vida que não difere muito entre si. A região Centro situa-se ligeiramente abaixo do Norte que, por sua vez, se situa ligeiramente abaixo do Alentejo e do Algarve. Os Açores são a região cuja posição relativa estimada mais varia ao longo do tempo e com o método utilizado.

Exceptuando a R.A. dos Açores, as relações entre o custo de vida das várias regiões e o Norte não sofreram alterações significativas entre 1994 e 1997. Houve um ligeiro decréscimo nos índices do Centro, Lisboa e Vale do Tejo e Madeira o que corresponde a um afastamento do custo de vida do Centro do custo de vida no Norte e a uma aproximação do custo de vida das restantes ao custo nesta. A variação dos índices do Alentejo e do Algarve foi ainda menos significativa. O índice de custo de vida da R.A. dos Açores aumentou em 1995, tendo diminuído nos anos seguintes.

A representação gráfica dos resultados permite uma melhor compreensão da evolução ao longo do tempo, bem como a comparação entre métodos. A Figura 1 e a Figura 2 representam os resultados de dois métodos em simultâneo, o *EKS* e o *KCM* e o *GK* e o *Iklé*.

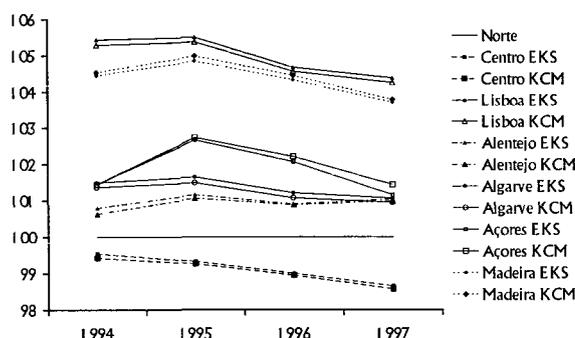


Figura 1: *EKS* e *KCM*

O método *EKS* e o método *KCM* dão origem a resultados quase coincidentes. O facto de no segundo ser dada uma maior importância às trocas indirectas com as regiões mais ricas não tem, no nosso caso, impacto nos resultados.

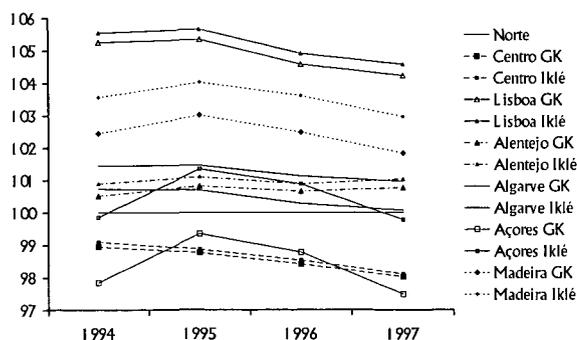


Figura 2: *GK* e *Iklé*

Os resultados dados pelo método *GGK* apresentam diferenças em relação aos baseados no *EKS* e são sensíveis ao parâmetro α . A análise da Figura 1 e da Figura 2 permite constatar que as evoluções dos índices fornecidos pelos vários métodos, para cada região, são paralelas. Para cada região, é o nível dos índices, e não a sua evolução ao longo do tempo, que varia com o método. A diferença entre os níveis dados pelos métodos não é muito significativa, à excepção das Regiões Autónomas e do Algarve. Analisemos os resultados destas regiões graficamente.

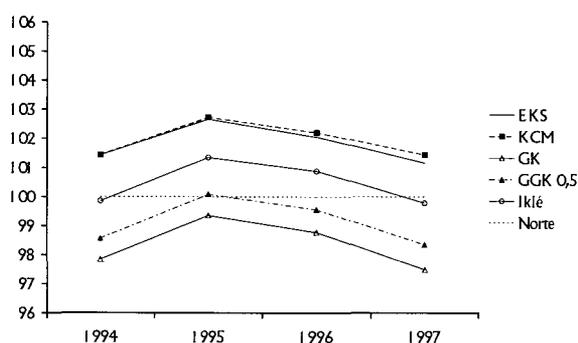


Figura 3: Índices de custo de vida nos Açores

Claramente, os índices *EKS* e *KCM* são semelhantes, situando os Açores sempre acima do Norte. A posição relativa dos Açores muda, se considerarmos um dos índices da classe *GGK*, notando-se, nitidamente, uma subida do índice em cada um dos anos à medida que o parâmetro α se aproxima de 1. Este comportamento pode estar relacionado com o efeito Gerschenkron que afecta os índices *GGK* e que se atenua com a aproximação de α à unidade (ver Parente e Bago d'Uva, 2001). Os índices *EKS* e *KCM*, por não se basearem em estruturas de preços “internacionais”, não sofrem este efeito e estimam um nível de preços para os Açores superior ao do Norte.

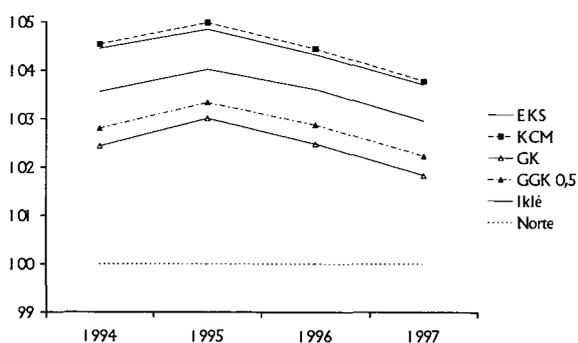


Figura 4: Índices de custo de vida na Madeira

O comportamento dos vários métodos no que diz respeito ao índice de custo de vida da Madeira em relação ao Norte é semelhante ao que foi analisado acima sendo,

no entanto, os níveis obtidos menos sensíveis ao método utilizado. Os índices *GGK* aproximam-se mais entre si e dos índices *EKS* e a posição desta R.A. em relação às outras regiões é a mesma, segundo todas as metodologias utilizadas.

Para as regiões de Portugal Continental, os índices estimados são pouco sensíveis ao método aplicado. A região do Algarve é a mais afectada mas, como a Figura 5 permite ver, os resultados são menos díspares do que no caso das regiões autónomas.

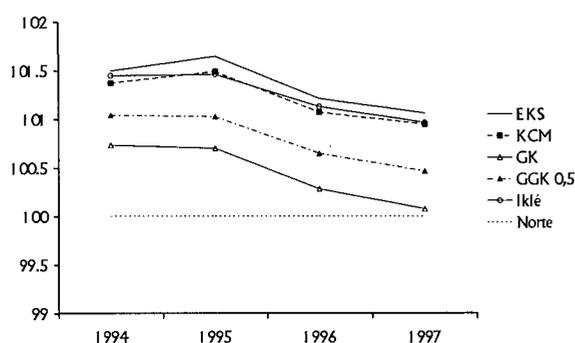


Figura 5: Índices de custo de vida no Algarve

3.2 ÍNDICES DE PREÇOS EM ALGUNS GRUPOS DE BENS E SERVIÇOS

A desagregação dos índices de custo de vida permite-nos identificar quais os grupos de bens e serviços que mais contribuem para as relações observadas entre os custos de vida das regiões. A análise dos próximos parágrafos baseia-se nos resultados desagregados por 10 grupos, bem como no peso de cada um destes na despesa total em cada uma das regiões que são apresentados nas secções seguintes. Para cada grupo, é apresentada, para cada região e para Portugal, a proporção da despesa na despesa total. Por ser nossa intenção analisar de que modo cada grupo contribui para o custo de vida, as proporções foram calculadas em relação à despesa utilizada na estrutura de ponderação do índice e não em relação à despesa total das famílias que inclui categorias de despesa não consideradas no cálculo do índice como a das rendas de habitação.

O facto de Lisboa e Vale do Tejo ter sido a região com o custo de vida mais elevado deve-se, em grande parte, ao custo da alimentação e bebidas não alcoólicas, o grupo mais importante dos dez considerados.⁵⁵ No segundo grupo mais importante, o dos transportes e comunicações, o índice de preços da região de Lisboa e Vale do Tejo não é o mais elevado sendo, contudo, superior a 100.

⁵⁵ O grupo da alimentação e bebidas não alcoólicas é o mais importante no sentido em que é aquele cuja despesa tem o maior peso na estrutura de despesa total de Portugal. Se analisarmos a estrutura de despesas de cada região, verificamos que este grupo é o mais importante em todas, à excepção do Centro, onde o grupo com maior peso é o dos transportes e comunicações

Os grupos dos serviços relacionados com a habitação e da saúde tiveram também uma contribuição relevante para o elevado custo de vida naquela região. Apesar de o custo da educação na região de Lisboa e Vale do Tejo ter sido bastante mais elevado do que nas restantes regiões, a contribuição deste grupo para o índice global não foi muito significativa.

A posição relativa da região da Madeira deve-se sobretudo ao elevado custo do mobiliário, acessórios para lar, equipamento doméstico e de manutenção corrente e dos hotéis, cafés e restaurantes. Quanto aos grupos mais importantes, os seus custos na Madeira não diferem significativamente dos da região Norte.

A evolução do índice de custo de vida da R.A. dos Açores no período 1994-1997 foi semelhante à que ocorreu no grupo da alimentação e bebidas não alcoólicas.

Da aplicação dos métodos *EKS* e a sua generalização *KCM* resultam índices muito semelhantes, como foi visto na secção anterior, pelo que dispensámos a apresentação do segundo na análise mais detalhada. Dentro da classe *GGK*, optámos por apresentar apenas os resultados dados pelo método *Iklé* por ser o menos afectado pelo efeito Gerschenkron (ver Parente e Bago d' Uva, 2001).

3.2.1 ALIMENTAÇÃO E BEBIDAS NÃO ALCOÓLICAS

Esta categoria de bens é muito relevante na determinação do custo de vida em Portugal uma vez que representa cerca de 25% das despesas dos consumidores. Das 10 grandes categorias consideradas, é a esta que corresponde a maior parte da despesa em todas as regiões, à excepção do Centro em que é superada apenas pela dos transportes e comunicações.

Quadro 2

Despesas em Alimentação e Bebidas não alcoólicas (% da despesa total)	
Norte	24,4%
Centro	23,2%
Lisboa e Vale do Tejo	25,1%
Alentejo	26,7%
Algarve	24,6%
Açores	30,3%
Madeira	26,8%
Portugal	24,7%

A região de Portugal em que as despesas em bens alimentares pesam mais no consumo total das famílias é a R.A. dos Açores, seguindo-se-lhe a Madeira e o Alentejo.

Quadro 3: Alimentação e bebidas não alcoólicas

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	100,7	100,7	101,4	101,3	100,3	100,3	101,0	100,9
Lisboa e Vale do Tejo	106,6	106,6	106,7	105,9	106,9	107,1	107,3	106,4
Alentejo	104,1	104,6	104,3	104,1	103,7	104,2	103,9	103,8
Algarve	106,1	107,2	107,7	106,6	106,3	107,5	108,1	107,1
R.A. Açores	96,2	97,8	97,5	96,4	93,1	95,3	95,5	93,7
R.A. Madeira	99,2	100,1	101,8	100,5	98,5	99,2	100,9	99,6

Os resultados dados pelos dois métodos são muito semelhantes, ambos indicam que é em Lisboa e Vale do Tejo e no Algarve que os preços dos bens alimentares são mais elevados, seguindo-se-lhe o Alentejo. Não existem diferenças muito significativas entre os índices de preços das regiões Norte, Centro e Madeira.

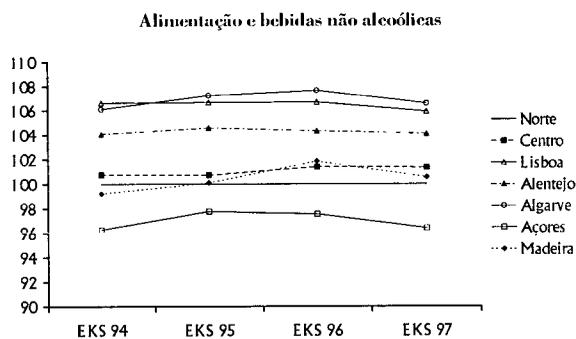


Figura 6

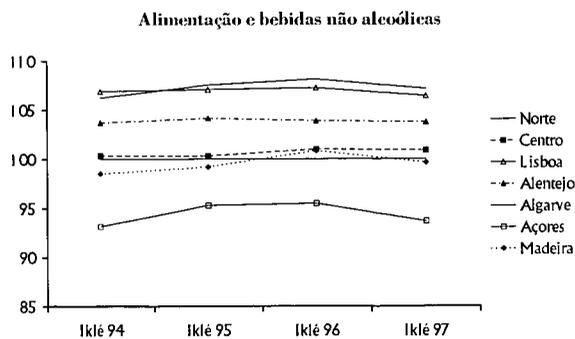


Figura 7

O grupo da Alimentação foi analisado com maior detalhe (nas secções seguintes apresentamos os índices dos seus subgrupos) o que permite ver de que forma cada subgrupo contribui para as posições relativas observadas.

Começamos pela região de Lisboa e Vale do Tejo onde o custo da alimentação é, em todos os anos, muito semelhante ao do Algarve e superior ao de todas as outras regiões, em particular, cerca de 7% superior ao do Norte. Isto deve-se a níveis de preços mais elevados nos subgrupos mais importantes da alimentação, sobretudo a carne e o peixe mas também os cereais e os legumes. No subgrupo que corresponde à maior parte da despesa em alimentação, o da carne, a região de Lisboa e Vale do Tejo tem o nível de preços mais elevado. O índice de preços da alimentação é mais baixo do que o daquele subgrupo, pois a disparidade observada nos subgrupos do peixe, cereais e legumes não é tão significativa e o nível de preços do Leite e derivados é semelhante em todas as regiões do Continente.

O elevado índice de preços da alimentação da região do Algarve é devido, sobretudo, aos índices da carne, do peixe e dos legumes. Os subgrupos dos legumes e dos cereais compensam-se na determinação da posição relativa entre o Algarve e Lisboa e Vale do Tejo.

Na região do Alentejo, o custo dos bens alimentares é cerca de 4% superior ao do Norte. A diferença entre a região base e o Alentejo resulta da que é observada nos subgrupos da carne e peixe. No entanto, estas diferenças, bem como a observada no grupo dos legumes, são menos significativas do que nas regiões que analisámos nos parágrafos anteriores. Por esta razão, e também porque os custos dos cereais no Alentejo e no Norte são semelhantes, o índice de preços da alimentação no Alentejo é inferior aos de Lisboa e Vale do Tejo e Algarve.

O custo da alimentação no Centro não difere significativamente do que se verifica no Norte, em nenhum dos anos considerados, o que está de acordo com o que acontece nos subgrupos do peixe, carne e leite. Esta semelhança não é observada no grupo dos cereais nem no dos legumes, porém, estes índices compensam-se.

A região autónoma dos Açores tem o índice de custo dos bens alimentares mais baixo do nosso país. O detalhe da análise que efectuámos indica-nos que isto resulta do reduzido custo do peixe e do leite e seus derivados, não sendo, contudo, a diferença na alimentação tão grande como a que acontece nestes subgrupos pelo facto de o custo dos legumes ser mais elevado nesta do que em qualquer outra região.

3.2.1.1 CEREAIS E PRODUTOS À BASE DE CEREAIS

Quadro 4

Despesa em cereais		
	% da despesa em alimentação	% da despesa total
Norte	18,3%	4,2%
Centro	19,8%	4,3%
Lisboa e Vale do Tejo	14,8%	3,5%
Alentejo	20,1%	5,0%
Algarve	15,5%	3,6%
Açores	20,9%	5,8%
Madeira	22,4%	5,7%
Portugal	17,3%	4,0%

Os cereais e produtos à base de cereais têm um peso importante na alimentação no nosso país, sobretudo nas regiões autónomas.

Quadro 5: Cereais e produtos à base de cereais

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	105,6	106,5	106,8	105,0	106,8	107,7	107,9	105,8
Lisboa e Vale do Tejo	107,6	107,5	106,1	104,7	110,0	110,1	108,6	107,0
Alentejo	99,3	100,5	101,0	99,0	99,9	101,1	101,6	99,3
Algarve	102,5	103,5	102,8	99,5	104,0	105,0	104,2	100,6
R.A. Açores	99,7	101,7	99,7	97,1	101,3	103,5	101,2	97,5
R.A. Madeira	99,7	102,6	105,3	101,4	100,1	103,2	105,9	101,8

Cereais e Produtos à base de cereais

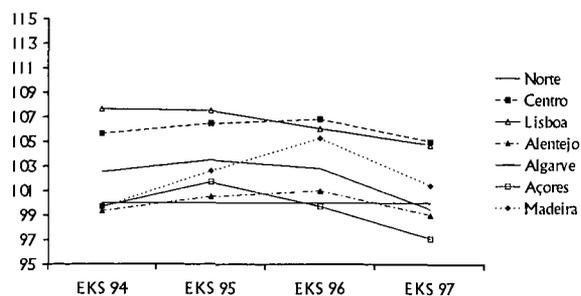


Figura 8

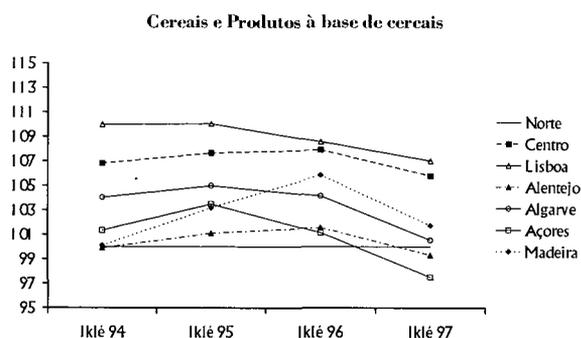


Figura 9

Os resultados dados pelos métodos *EKS* e *Iklé* têm algumas diferenças. O método *Iklé* estima diferenças mais significativas entre os custos das regiões.

3.2.1.2 CARNE

Quadro 6

Despesa em carne		
	% da despesa em alimentação	% da despesa total
Norte	29,2%	6,7%
Centro	26,6%	5,8%
Lisboa e Vale do Tejo	28,5%	6,7%
Alentejo	29%	6,0%
Algarve	25,8%	5,9%
Açores	22,7%	6,3%
Madeira	25,3%	6,5%
Portugal	28,0%	6,5%

Quadro 7: Carne

Região	<i>EKS</i>				<i>Iklé</i>			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	101,9	100,8	99,8	100,6	99,3	98,2	97,1	97,8
Lisboa e Vale do Tejo	114,7	114,4	114,0	114,1	113,0	112,9	112,7	112,8
Alentejo	111,9	111,7	110,4	110,7	108,9	108,5	106,9	107,4
Algarve	113,7	113,9	113,9	113,7	110,4	110,7	110,6	110,5
R.A. Açores	104,1	103,1	98,4	104,8	101,5	100,4	95,6	101,6
R.A. Madeira	106,8	107,1	104,9	104,0	103,3	104,0	100,9	99,8

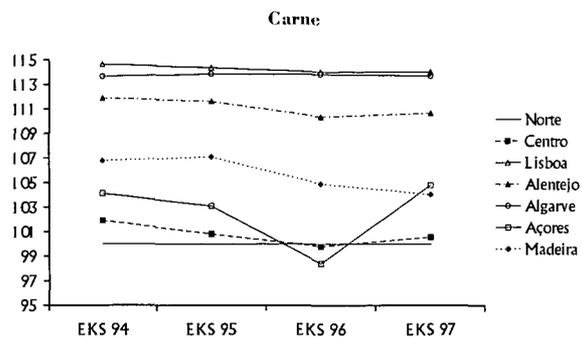


Figura 10

Verificamos que nas regiões de Lisboa e Vale do Tejo, Alentejo e Algarve os preços da carne são muito mais elevados que nos outros locais do país. O Norte é a região onde os preços da carne são mais baixos. O elevado custo da carne em Lisboa deverá ter como causa a elevada procura nesta região.

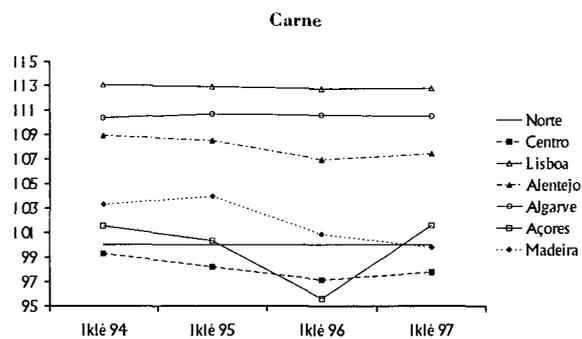


Figura 11

Notamos, ainda, que neste subgrupo os índices *EKS* e *Iklé* diferem cerca de três pontos percentuais.

3.2.1.3 PEIXE, CRUSTÁCEOS E MOLUSCOS

Quadro 8

Despesa em Peixe, crustáceos e moluscos		
	% da despesa em alimentação	% da despesa total
Norte	16,7%	3,8%
Centro	17,2%	3,8%
Lisboa e Vale do Tejo	18,8%	4,4%
Alentejo	15,1%	3,8%
Algarve	21,5%	4,9%
Açores	14,5%	4,0%
Madeira	8,6%	2,2%
Portugal	17,5%	4,1%

Quadro 9: Peixe

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	103,9	102,8	104,0	103,3	102,7	101,9	102,5	102,5
Lisboa e Vale do Tejo	109,5	107,7	107,3	106,0	110,5	109,3	108,4	107,2
Alentejo	107,7	106,4	105,1	102,8	109,1	108,1	106,6	104,4
Algarve	109,4	109,3	107,8	105,0	110,9	111,6	110,6	107,4
R.A. Açores	75,9	75,0	73,8	68,5	68,9	69,5	70,1	62,6
R.A. Madeira	93,8	90,0	88,5	92,1	93,3	87,0	84,0	90,3

Vemos no Quadro 9 que os preços são muito mais baixos na R.A. dos Açores. Na R.A. da Madeira, o custo do peixe também é menor que nas restantes regiões de Portugal Continental. Isto dever-se-á ao facto de a pesca ser uma actividade importante nestas regiões.

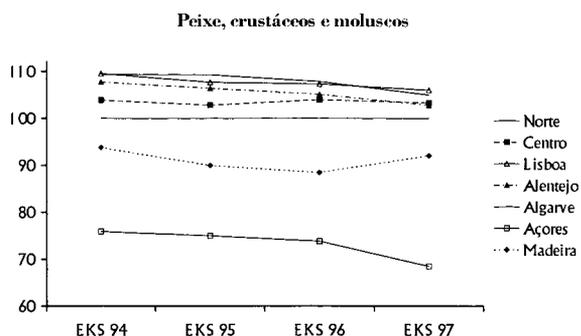


Figura 12

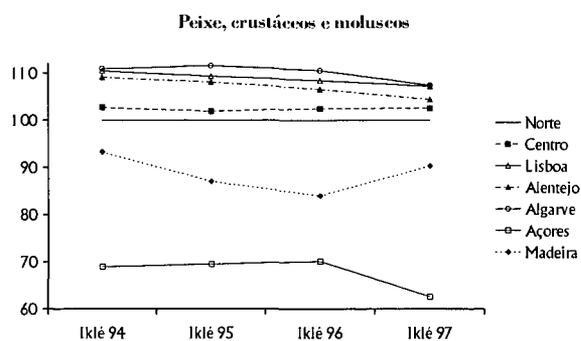


Figura 13

3.2.1.4 LEITE, PRODUTOS LÁCTEOS (EXCEPTO MANTEIGA) E OVOS

Quadro 10

Despesa em Leite, produtos lácteos (excepto manteiga) e ovos		
	% da despesa em alimentação	% da despesa total
Norte	12,7%	2,9%
Centro	13,8%	3,0%
Lisboa e Vale do Tejo	13,7%	3,2%
Alentejo	14,6%	3,6%
Algarve	13,0%	3,0%
Açores	16,7%	4,6%
Madeira	13,9%	3,5%
Portugal	13,4%	3,1%

Quadro 11: Leite, produtos lácteos (excepto manteiga) e ovos

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	97,7	97,7	98,1	99,0	97,6	97,5	98,0	99,3
Lisboa e Vale do Tejo	98,5	98,6	97,6	98,4	98,5	98,4	97,6	98,4
Alentejo	99,1	99,7	98,6	99,8	99,3	99,8	98,7	100,0
Algarve	98,4	99,4	98,9	99,7	98,4	99,3	98,8	99,8
R.A. Açores	76,2	80,1	80,5	81,4	74,5	78,8	79,4	80,5
R.A. Madeira	88,2	92,1	91,9	89,8	88,3	92,5	92,3	89,9

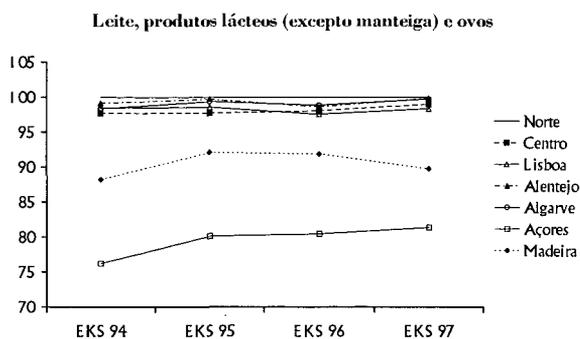


Figura 14

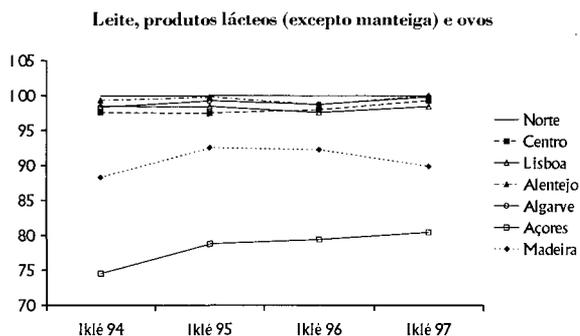


Figura 15

A R.A. dos Açores apresenta os preços mais baixos, o que é natural devido à sua grande produção de produtos lácteos. A região que apresenta um custo mais alto deste tipo de produtos é o Norte.

3.2.1.5 ÓLEOS E GORDURAS

Quadro 12

Despesa em Óleos e gorduras		
	% da despesa em alimentação	% da despesa total
Norte	6,2%	1,4%
Centro	7,0%	1,5%
Lisboa e Vale do Tejo	5,7%	1,4%
Alentejo	7,4%	1,9%
Algarve	6,8%	1,6%
Açores	5,8%	1,6%
Madeira	5,1%	1,3%
Portugal	6,2%	1,4%

A região que apresenta diferenças mais significativas é a R.A. da Madeira o que pode dever-se aos baixos custos da manteiga e do azeite. O custo nas restantes regiões foi semelhante nos anos de 1994 e 1995, tendo-se verificado uma diminuição do custo dos Açores relativamente ao Continente no ano seguinte. No ano de 1997, as diferenças entre as regiões foram mais significativas.

Quadro 13: Óleos e gorduras

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	98,3	100,4	99,6	101,3	98,1	100,2	99,4	101,4
Lisboa e Vale do Tejo	100,0	101,0	100,4	98,4	100,1	101,0	100,3	98,5
Alentejo	99,6	102,3	101,0	104,0	99,5	102,2	101,0	104,9
Algarve	98,1	101,6	101,3	103,3	98,1	101,5	101,2	103,8
R.A. Açores	100,3	100,9	96,0	102,5	98,2	98,7	94,8	101,0
R.A. Madeira	87,1	88,9	89,8	91,1	87,3	89,3	89,6	91,3

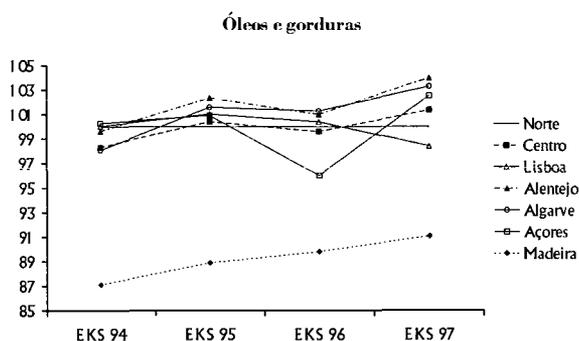


Figura 16

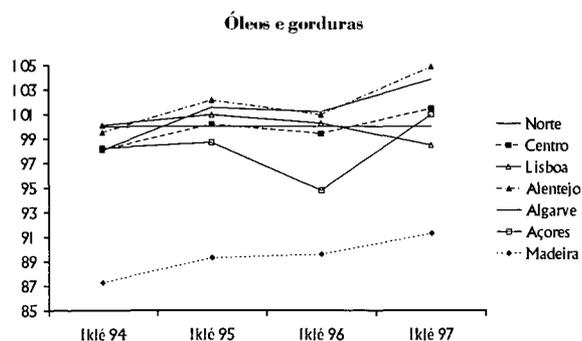


Figura 17

3.2.1.6 LEGUMES, LEGUMINOSAS, FÉCULAS, AMIDOS E FRUTAS

Quadro 14

Despesa em Legumes, leguminosas, féculas, amidos e frutas		
	% da despesa em alimentação	% da despesa total
Norte	14,1%	3,3%
Centro	12,3%	2,7%
Lisboa e Vale do Tejo	15,7%	3,7%
Alentejo	15,7%	3,9%
Algarve	14,9%	3,4%
Açores	14,9%	4,1%
Madeira	22,0%	5,6%
Portugal	14,7%	3,4%

Este subgrupo representa uma parte importante da despesa em alimentação, que é particularmente significativa na R.A. da Madeira.

Quadro 15: Legumes, leguminosas, féculas, amidos e frutas

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	92,4	93,3	97,1	96,2	92,1	93,1	98,1	96,1
Lisboa e Vale do Tejo	102,6	105,3	109,5	106,8	102,3	105,2	110,6	107,1
Alentejo	101,6	103,5	106,2	105,8	101,2	103,0	106,8	105,7
Algarve	107,3	110,6	116,7	115,6	106,1	109,5	116,5	114,8
R.A. Açores	121,2	128,5	143,8	131,6	117,3	127,3	144,6	129,9
R.A. Madeira	114,3	114,0	126,5	122,8	114,5	112,7	127,0	122,2

Legumes, leguminosas, féculas, amidos e frutas

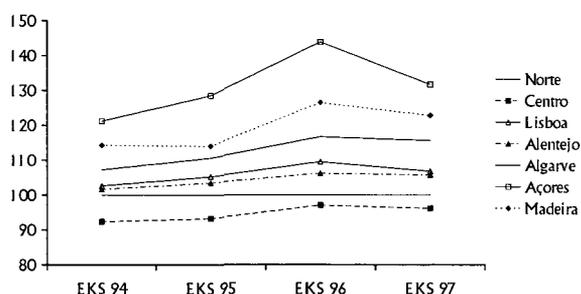


Figura 18

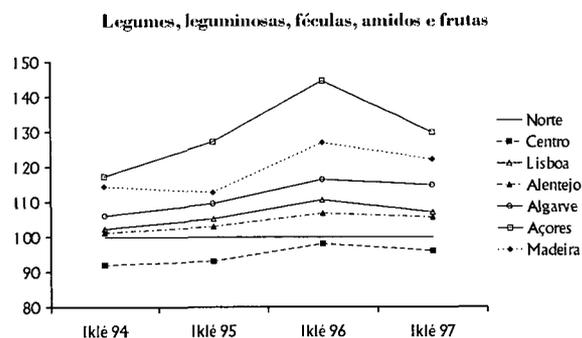


Figura 19

As regiões que apresentam custos mais elevados dos legumes, leguminosas, féculas, amidos e frutas são a R.A. Açores e a R.A. da Madeira. Este facto fica a dever-se aos elevados preços de alguns produtos, nomeadamente, o tomate, a tangerina e a uva de mesa.

3.2.1.7 AÇÚCAR, PRODUTOS AÇUCARADOS E GÉNEROS DE CONFEITARIA

Quadro 16

Despesa em Açúcar, produtos açucarados e géneros de confeitaria		
	% da despesa em alimentação	% da despesa total
Norte	2,9%	0,7%
Centro	3,3%	0,7%
Lisboa e Vale do Tejo	2,8%	0,6%
Alentejo	3,1%	0,8%
Algarve	2,4%	0,6%
Açores	4,5%	1,3%
Madeira	2,8%	0,7%
Portugal	2,9%	0,7%

Este último subgrupo da alimentação é o menos importante em termos de despesa em todas as regiões contribuindo pouco para o custo da alimentação e, ainda menos, para o custo de vida global.

Quadro 17: Açúcar, produtos açucarados e géneros de confeitaria

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	98,5	97,7	97,5	97,9	98,6	97,8	97,6	98,1
Lisboa e Vale do Tejo	97,8	96,1	96,7	97,6	97,4	95,7	96,1	97,1
Alentejo	98,7	98,6	99,7	100,4	98,6	98,6	99,7	100,4
Algarve	97,8	97,9	98,9	98,0	97,7	97,8	99,0	97,9
R.A. Açores	91,5	93,5	96,6	99,3	92,0	93,9	97,2	99,7
R.A. Madeira	65,2	67,9	68,1	67,6	61,8	64,8	64,7	64,5

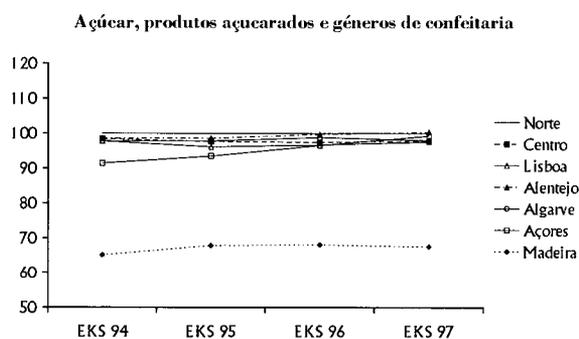


Figura 20

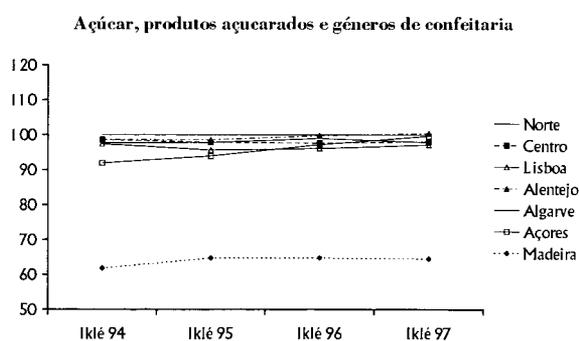


Figura 21

As diferenças entre os resultados dados pelos métodos *EKS* e *Iklé* são muito pequenas. Ambos indicam que o custo deste grupo quase não difere em Portugal Continental, tendo a região dos Açores vindo a aproximar-se até que, em 1997, o seu custo “igualou” o observado no Norte. As famílias residentes na R.A. da Madeira suportam um custo muito inferior ao suportado pelas famílias do resto do país.

3.2.2 BEBIDAS ALCOÓLICAS E TABACO

Quadro 18

Despesas em Bebidas alcoólicas e tabaco (% da despesa total)	
Norte	3,5%
Centro	2,4%
Lisboa e Vale do Tejo	3,0%
Alentejo	3,7%
Algarve	3,2%
Açores	4,9%
Madeira	3,0%
Portugal	3,2%

A região dos Açores é a que destina a maior percentagem de despesa ao consumo de bebidas alcoólicas e tabaco.

Quadro 19: Bebidas Alcoólicas e tabaco

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	99,2	99,8	101,0	103,2	98,6	99,3	100,6	103,0
Lisboa e Vale do Tejo	99,3	100,4	100,9	103,0	99,6	100,8	101,4	103,6
Alentejo	104,6	105,2	106,8	108,9	104,3	105,0	106,7	108,9
Algarve	102,1	103,6	105,7	107,6	102,3	103,9	105,9	108,0
R.A. Açores	98,1	95,6	94,1	96,4	97,5	95,0	93,5	95,9
R.A. Madeira	94,7	94,1	92,8	94,1	94,7	94,0	93,5	95,1

Os índices calculados mostram que nas Regiões Autónomas os preços destes bens são menores do que os das restantes regiões. Os produtos relacionados com o tabaco são os que contribuíram mais para a disparidade de preços entre a R.A. dos Açores e o Norte. O Alentejo e o Algarve são as regiões em que estes índices de preços apresentaram valores mais altos. A tendência de todos os índices das regiões de Portugal Continental foi crescente, indicando que o custo destes produtos aumentou mais nestas regiões do que no Norte. Os índices das ilhas evoluíram no sentido oposto de 1994 para 1996, tendência que se inverteu no ano seguinte.

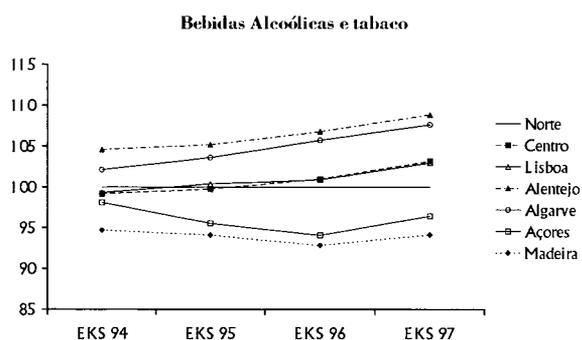


Figura 22

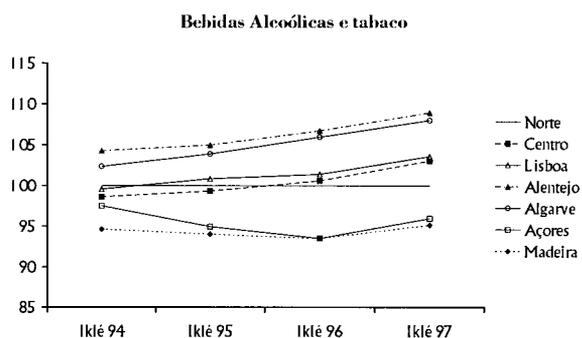


Figura 23

3.2.3 VESTUÁRIO E CALÇADO

Quadro 20

Despesas em Vestuário e Calçado (% da despesa total)	
Norte	8,4%
Centro	8,2%
Lisboa e Vale do Tejo	7,5%
Alentejo	9,5%
Algarve	8,3%
Açores	8,8%
Madeira	11,1%
Portugal	8,1%

O peso das despesas em Vestuário e Calçado nas despesas de cada região varia entre 7,5% (Lisboa e Vale do Tejo) e 11,1% (R.A. Madeira).

Quadro 21: Vestuário e calçado

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	92,5	92,1	89,8	88,8	91,8	91,4	89,5	88,6
Lisboa e Vale do Tejo	102,2	99,8	95,6	95,2	101,5	99,1	95,2	95,1
Alentejo	100,2	101,1	99,9	101,7	99,1	99,9	99,0	101,2
Algarve	99,3	98,5	97,1	98,5	98,4	97,7	96,7	98,3
R.A. Açores	104,3	106,0	104,6	102,6	103,0	104,7	103,5	102,0
R.A. Madeira	109,9	111,8	109,3	109,3	108,6	110,7	109,0	109,2

Verificamos que o custo do vestuário e calçado é mais elevado nos Açores e na Madeira do que nas outras regiões o que pode ser justificado pelos custos de transporte dos artigos provenientes de Portugal Continental. Na região Centro, os custos destes produtos são baixos relativamente ao Norte. Os índices de preços da maior parte das regiões não variam muito durante o período considerado neste estudo. Contudo, os preços diminuíram de forma significativa em relação ao Norte nas regiões de Lisboa e Vale do Tejo e do Centro.

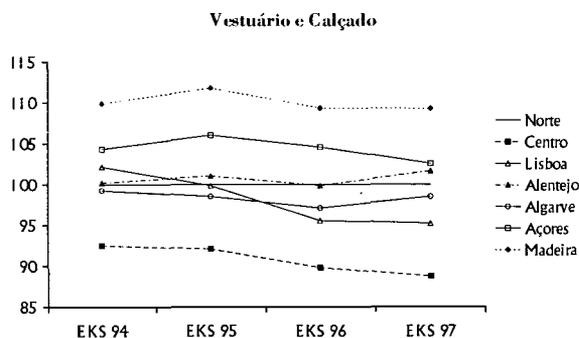


Figura 24

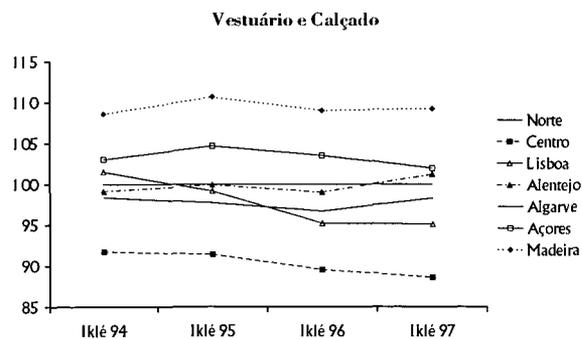


Figura 25

3.2.4 SERVIÇOS RELACIONADOS COM HABITAÇÃO (EXCEPTO RENDAS), ÁGUA, ELECTRICIDADE, GÁS E OUTROS COMBUSTÍVEIS

Quadro 22

Despesas em Serviços relacionados com habitação (excepto rendas), água e outros combustíveis (% da despesa total)	
Norte	4,9%
Centro	5,3%
Lisboa e Vale do Tejo	5,4%
Alentejo	6,7%
Algarve	5,2%
Açores	6,9%
Madeira	8,2%
Portugal	5,3%

O peso deste grupo nas despesas da Madeira é maior do que em qualquer outra região.

Quadro 23: Serviços relacionados com Habitação (excepto rendas), água, electricidade, gás e outros combustíveis

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	100,0	98,8	99,7	101,2	100,2	99,3	100,2	101,4
Lisboa e Vale do Tejo	110,5	109,9	108,9	108,5	111,0	110,7	110,0	109,6
Alentejo	97,2	96,5	95,7	96,6	97,8	97,4	96,7	97,5
Algarve	92,6	92,7	92,5	92,4	92,4	93,1	93,0	92,6
R.A. Açores	92,4	97,8	101,5	99,5	92,3	97,7	101,7	99,4
R.A. Madeira	90,7	93,6	93,9	93,1	89,9	92,9	93,4	92,7

A região que se destaca por ter o índice de preços mais elevado é Lisboa e Vale do Tejo. A Madeira e o Algarve são as regiões onde o custo deste grupo é mais baixo. Os custos destes serviços nas regiões de Portugal Continental tiveram uma evolução semelhante. Podemos ver na Figura 26 e na Figura 27 que os índices de preços tiveram uma tendência crescente nas ilhas nos anos 1994 a 1996, mais acentuada na R.A. dos Açores, que se inverteu em 1997.

Serviços relacionados com Habitação (excepto rendas),
água, electricidade, gás e outros combustíveis

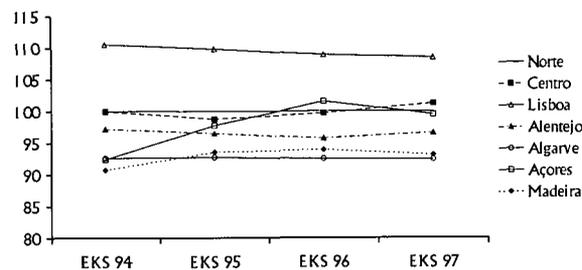


Figura 26

Serviços relacionados com Habitação (excepto rendas),
água, electricidade, gás e outros combustíveis

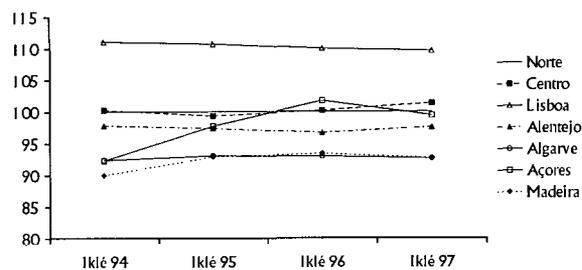


Figura 27

3.2.5 MOBILIÁRIO, ACESSÓRIOS PARA LAR, EQUIPAMENTO DOMÉSTICO E DE MANUTENÇÃO CORRENTE

Quadro 24

Despesas em Mobiliário, acessórios para lar, equipamento doméstico e de manutenção corrente (% da despesa total)	
Norte	9,9%
Centro	10,9%
Lisboa e Vale do Tejo	10,9%
Alentejo	8,6%
Algarve	10,5%
Açores	13,5%
Madeira	10,8%
Portugal	10,5%

O Alentejo é a região em que a proporção das despesas em mobiliário, acessórios para lar, equipamento doméstico e de manutenção corrente na despesa total é menor, em oposição à R.A. dos Açores.

Quadro 25: Mobiliário, acessórios para lar, equipamento doméstico e de manutenção corrente

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	99,1	98,5	97,8	97,2	98,8	98,1	97,4	96,7
Lisboa e Vale do Tejo	108,0	109,2	108,4	108,8	107,8	109,0	108,4	108,9
Alentejo	96,3	96,6	95,9	96,2	96,7	96,7	96,0	96,4
Algarve	96,8	95,4	94,6	95,0	95,4	93,8	93,2	93,6
R.A. Açores	110,7	112,8	110,5	111,0	110,4	112,8	110,5	110,9
R.A. Madeira	116,0	115,7	115,9	116,7	115,4	115,6	114,1	113,4

O custo destes bens e serviços é mais elevado nas regiões autónomas o que pode estar relacionado com os custos de transporte associados. Em relação a Portugal Continental, todas as regiões se situam abaixo do Norte, à excepção de Lisboa e Vale do Tejo onde o custo é muito superior. Neste grupo, os índices não sofreram variações significativas ao longo do tempo, tendo-se mantido a posição relativa das regiões, com excepção do Centro cujo índice diminuiu muito ligeiramente de 1994 para 1997.

Mobiliário, acessórios para lar, equipamento doméstico e de manutenção corrente

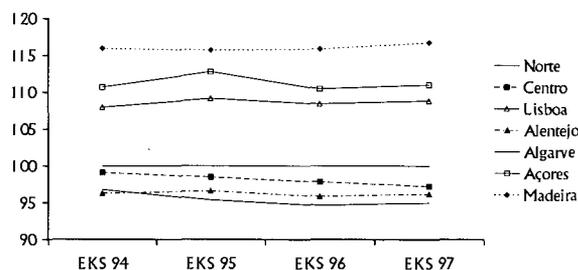


Figura 28

Mobiliário, acessórios para lar, equipamento doméstico e de manutenção corrente

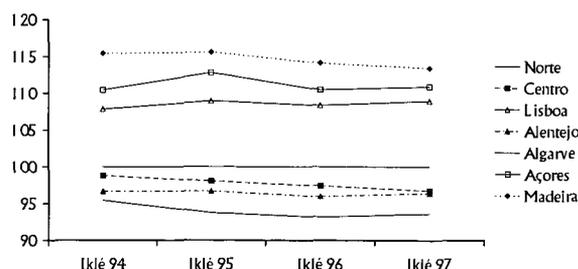


Figura 29

Quadro 26

Despesas em saúde (% da despesa total)	
Norte	5,9%
Centro	6,5%
Lisboa e Vale do Tejo	5,9%
Alentejo	5,3%
Algarve	7,3%
Açores	5,6%
Madeira	4,5%
Portugal	6,0%

As despesas em Saúde representam uma fatia importante das despesas totais, sobretudo nas regiões Centro e Algarve. Esta proporção é menor nas restantes regiões de Portugal Continental, que não se diferenciam muito entre si, e ainda menor na R.A. da Madeira.

Quadro 27: Saúde

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	93,9	94,7	93,8	95,4	89,0	90,3	89,1	91,0
Lisboa e Vale do Tejo	106,5	106,2	105,5	106,9	106,3	105,9	104,8	106,3
Alentejo	94,3	93,8	93,7	93,8	93,7	93,3	92,8	93,1
Algarve	97,4	96,3	94,9	96,8	96,1	95,1	93,3	95,3
R.A. Açores	99,4	100,2	98,7	99,5	97,9	98,6	96,9	97,8
R.A. Madeira	106,0	104,6	102,2	101,6	105,0	103,4	100,8	100,5

O custo dos bens e serviços relacionados com a saúde na região de Lisboa e Vale do Tejo é sempre superior ao das outras regiões. Na R.A. da Madeira, este custo foi semelhante ao de Lisboa no ano de 1994, tendo-se aproximado do da região Norte ao longo do período considerado. Os índices mais baixos são os das regiões Centro e Alentejo.

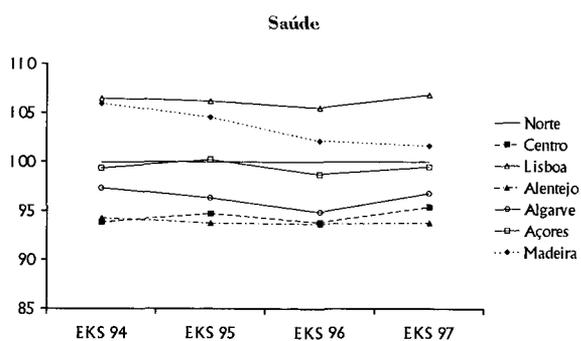


Figura 30

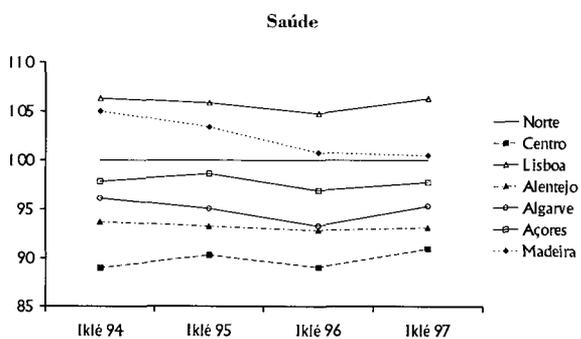


Figura 31

3.2.7 TRANSPORTES E COMUNICAÇÕES

Quadro 28

Despesas em Transportes e Comunicações (% da despesa total)	
Norte	22,7%
Centro	25,8%
Lisboa e Vale do Tejo	20,9%
Alentejo	21,8%
Algarve	20,7%
Açores	15,7%
Madeira	20,0%
Portugal	22,3%

Este grupo é constituído pela aquisição de veículos, utilização e manutenção de veículos próprios, serviços de transportes e comunicações. A proporção das despesas na despesa total é muito significativa porque as despesas em aquisição, utilização e

manutenção de veículos representam, para Portugal, cerca de 18% da despesa total, ou seja, cerca de 83% deste grupo.

A região em que as despesas em Transportes e Comunicações têm menor importância é a R.A. dos Açores o que está relacionado com o facto de o peso das suas despesas com aquisição, utilização e manutenção de veículos na sua despesa total ser de apenas 12%. A proporção das despesas em serviços de transportes observada nos Açores não difere significativamente da média nacional, cerca de 1%, o mesmo acontecendo no subgrupo das comunicações que tem um peso de cerca de 3% em todas as regiões.

Quadro 29: Transportes e comunicações

Região	EKS				Ikté			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Centro	102,89	101,51	101,22	100,51	103,26	101,74	101,37	100,62
Lisboa e Vale do Tejo	103,27	102,66	102,35	102,38	103,46	102,67	102,36	102,22
Alentejo	101,36	100,89	101,40	101,27	101,99	101,06	101,77	101,47
Algarve	102,44	102,76	101,08	100,35	103,25	103,12	101,51	100,46
R.A. Açores	102,72	103,10	103,84	103,60	102,71	102,87	103,37	102,94
R.A. Madeira	106,55	106,38	104,99	104,97	107,63	106,98	105,91	106,29

A Madeira é a região com o índice de custo de Transportes e Comunicações mais elevado o que se deve, sobretudo, ao custo dos transportes urbanos e das comunicações telefónicas.

Os índices de custo dos transportes e comunicações do Centro e do Algarve tiveram uma tendência decrescente, aproximando-se do nível do Norte. As variações dos índices para as outras regiões, durante o período considerado, não são significativas.

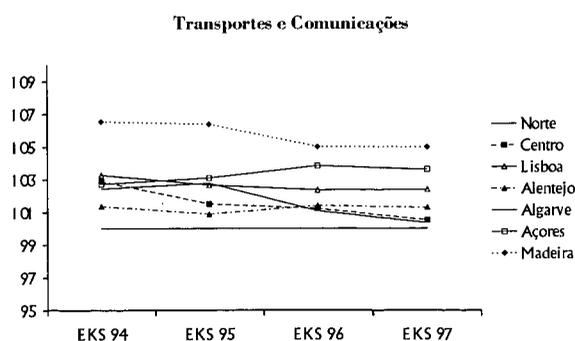


Figura 32

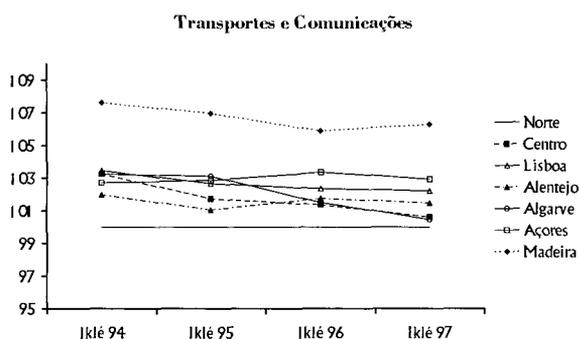


Figura 33

3.2.8 LAZER, RECREAÇÃO E CULTURA

Quadro 30

Despesas em Lazer, Recreação e Cultura (% da despesa total)	
Norte	3,7%
Centro	3,9%
Lisboa e Vale do Tejo	3,3%
Alentejo	3,2%
Algarve	3,4%
Açores	3,3%
Madeira	3,9%
Portugal	3,5%

Podemos ver no Quadro 30 que os portugueses reservam uma pequena percentagem das suas despesas para o Lazer, Recreação e Cultura.

Quadro 31: Lazer, Recreação e Cultura

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	99,0	99,2	97,5	97,5	99,3	99,0	97,2	97,2
Lisboa e Vale do Tejo	103,5	105,0	103,8	104,3	103,7	104,9	103,8	104,3
Alentejo	103,2	103,7	102,5	102,9	103,3	103,5	102,1	102,4
Algarve	102,8	104,3	104,0	104,0	102,1	103,4	103,0	102,9
R.A. Açores	116,9	121,2	116,0	114,9	114,8	119,0	113,6	111,9
R.A. Madeira	111,4	116,5	112,7	110,6	105,0	110,7	106,8	104,6

O custo em artigos relacionados com Lazer, Recreação e Cultura é claramente mais elevado nas ilhas. Neste grupo estão, provavelmente, incluídos produtos provenientes das regiões de Portugal Continental, pelo que os habitantes das ilhas têm que acarreter com o custo de transporte.

Os índices *EKS* e *Iklé* apresentam valores muito semelhantes para as regiões de Portugal Continental, contudo, para a R.A. dos Açores e R.A. da Madeira estes métodos produzem valores muito díspares. A diferença entre os índices *EKS* e *Iklé* na Madeira é cerca de 6 pontos percentuais e para nos Açores é de 2 pontos percentuais.

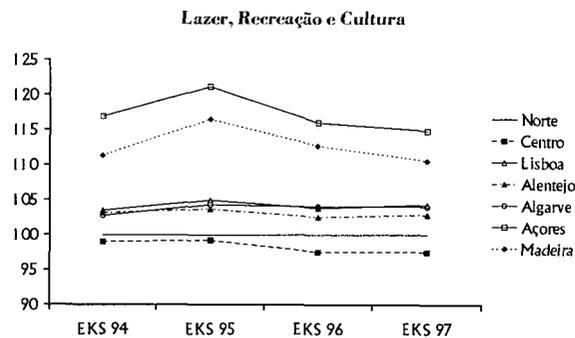


Figura 34

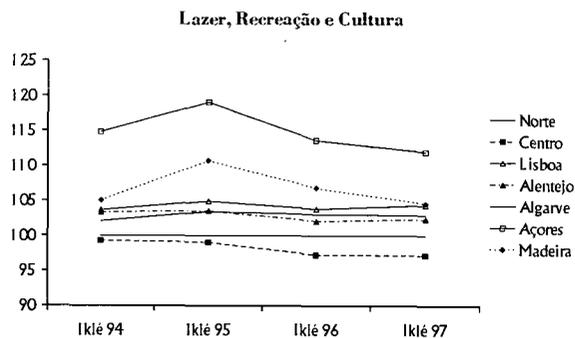


Figura 35

3.2.9 EDUCAÇÃO

As despesas em Educação são caracterizadas por valores relativamente baixos. Este resultado dever-se-á, talvez, ao facto de estas despesas serem constituídas principalmente por propinas do ensino público que, geralmente, correspondem a valores baixos. Note-se que os livros e material escolar não estão incluídos neste grupo, fazem parte da categoria do Lazer, Recreação e Cultura.

É necessário ter em conta a possibilidade de as despesas em educação suportadas pelas famílias de uma determinada região dizerem respeito a serviços

prestados noutras regiões. O procedimento correcto seria considerar, no cálculo dos índices, os preços observados nas regiões onde os serviços foram prestados, o que não pode ser feito uma vez que a informação acerca da despesa não discrimina a região em que esta teve lugar. Neste estudo, optámos por utilizar o procedimento que é correntemente aplicado no Índice de Preços do Consumidor, que consiste em ponderar o custo dos bens e serviços relacionados com educação de cada região com as despesas em educação dos consumidores da mesma região.

Quadro 32

Despesas em Educação (% da despesa total)	
Norte	1,6%
Centro	1,2%
Lisboa e Vale do Tejo	2,1%
Alentejo	0,9%
Algarve	1,1%
Açores	0,9%
Madeira	1,4%
Portugal	1,7%

Quadro 33: Educação

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	108,2	107,3	110,2	105,6	110,0	109,2	112,9	107,7
Lisboa e Vale do Tejo	127,5	125,3	123,1	116,1	130,8	128,7	127,0	119,3
Alentejo	106,8	106,3	107,9	104,2	109,4	108,7	109,8	105,3
Algarve	109,2	107,5	107,7	104,5	108,3	106,4	106,4	102,8
R.A. Açores	84,6	81,7	78,6	74,6	82,7	79,5	76,3	71,7
R.A. Madeira	88,2	86,4	83,4	78,6	88,3	86,5	83,3	77,8

Verificamos que o custo da educação é muito mais elevado em Lisboa e Vale do Tejo do que no resto do país o que pode ser devido à elevada procura do ensino privado naquela região. Os Açores e a Madeira são as regiões com os menores índices. Uma vez que, no período considerado, não existiam escolas públicas nas ilhas para certos graus de ensino, os Governos Regionais financiaram uma parte do ensino particular levando a uma diminuição do custo em educação.

A relação estimada do custo da educação nos Açores com o custo no resto do país pode exemplificar as consequências das limitações dos dados utilizados na secção 2. A possibilidade de restarem algumas correspondências entre produtos ou serviços

que não são equivalentes pode enviesar as relações entre os seus preços. Este grupo inclui a categoria serviços de educação no 3º ciclo do ensino básico particular. Para o cálculo dos índices de preços desta categoria os dados disponíveis são as propinas do ensino particular até ao 9º ano das regiões do Continente e da R.A. da Madeira e até ao 6º ano para a R.A. dos Açores. Uma vez que não existe uma equivalência entre o serviço cujo preço queremos comparar para todas regiões, foi feito um ajustamento às propinas dos Açores com base em dados adicionais por forma a obter uma aproximação dessa equivalência. No entanto, é possível que essa aproximação não seja a desejada e que o índice de preços dos Açores para esta categoria e, consequentemente, para o grupo da educação esteja enviesado.

Todas as regiões mostraram de 1994 para 1997 uma tendência para uma diminuição dos custos da educação relativamente ao Norte o que corresponde dizer que esta se aproximou das restantes regiões de Portugal Continental e se afastou das ilhas.

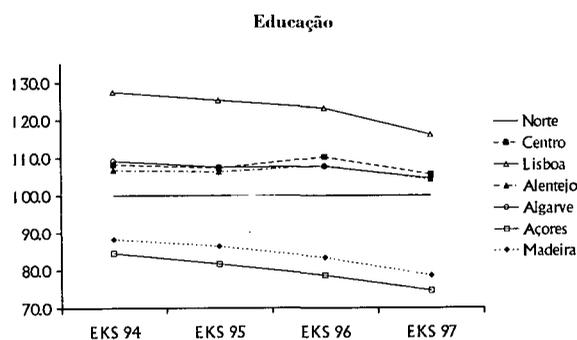


Figura 36

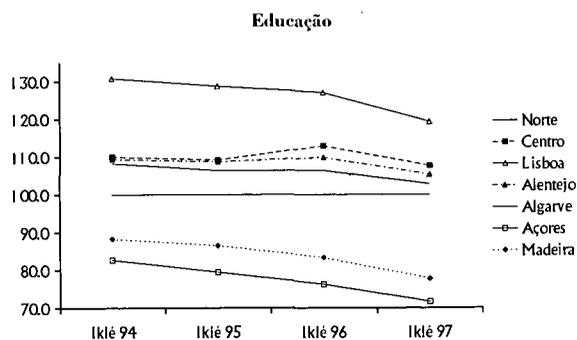


Figura 37

3.2.10 HOTÉIS, CAFÉS E RESTAURANTES

Quadro 34

Despesas em Hotéis, cafés e restaurantes (% da despesa total)	
Norte	9,7%
Centro	6,6%
Lisboa e Vale do Tejo	10,4%
Alentejo	8,2%
Algarve	9,9%
Açores	5,9%
Madeira	5,8%
Portugal	9,3%

As regiões que reservam uma maior fatia das suas despesas para os bens e serviços deste grupo são Lisboa e Vale do Tejo, Algarve e Norte. Neste grupo, os custos associados a cada região podem não reflectir correctamente os que foram suportados pelas famílias residentes. Na verdade, a maior parte das pessoas que procuram um Hotel numa determinada região será proveniente de outra região. Para cada região, temos um custo que corresponde aos preços lá praticados que é ponderado pela despesa das famílias residentes feita, também, noutras regiões. Neste caso adoptamos o mesmo procedimento que foi utilizado em relação à educação.

Quadro 35: Hotéis, cafés e restaurantes

Região	EKS				Iklé			
	1994	1995	1996	1997	1994	1995	1996	1997
Norte	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Centro	97,9	101,0	99,5	96,8	97,9	100,8	99,4	96,5
Lisboa e Vale do Tejo	103,8	106,5	104,5	101,9	102,6	105,6	103,7	100,8
Alentejo	100,4	103,2	103,0	102,6	100,2	103,1	102,8	102,1
Algarve	101,6	101,3	101,6	100,5	102,1	101,7	102,0	100,7
R.A. Açores	110,9	111,2	108,9	107,0	108,2	108,8	106,7	104,6
R.A. Madeira	122,6	117,6	111,1	108,6	122,3	117,1	111,0	108,2

Verificamos que neste grupo os preços são muito mais altos nas ilhas do que nas restantes regiões. Este facto deverá estar relacionado com a pouca oferta que existe nas nestas R.A. e com a grande procura provocada pelo turismo. Quanto à evolução no tempo, é de notar que neste grupo os índices de todas as regiões, com excepção do Centro, aproximaram-se do Norte.

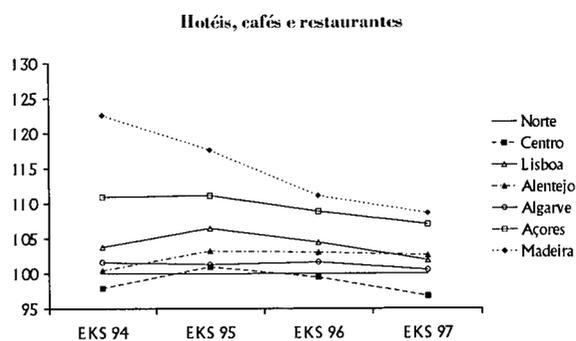


Figura 38

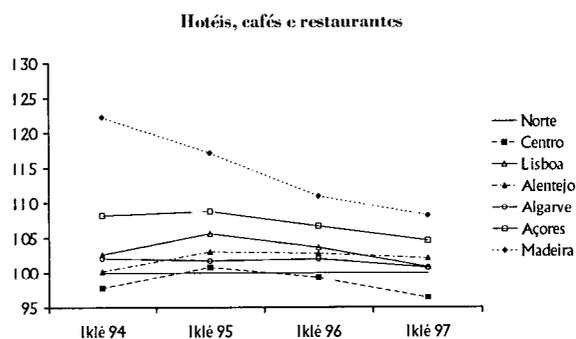


Figura 39

4. CONCLUSÕES

Com este estudo, pretendemos investigar até que ponto o custo de vida varia em Portugal. Recorremos às metodologias usualmente utilizadas para fazer comparações de custos de vida de países diferentes por forma a comparar os custos de vida de regiões portuguesas. As características inerentes aos índices dados pelas diversas metodologias podem levar a resultados diferentes. A aplicação de várias metodologias permitiu-nos comparar o seu comportamento no nosso caso e analisar a sensibilidade dos índices ao método adoptado.

Considerámos as regiões NUTSII (Norte, Centro, Lisboa e Vale do Tejo, Alentejo, Algarve, Açores e Madeira) no período 1994 a 1997 o que permitiu a análise da evolução das relações entre os custos de vida regionais. Os índices de custo de vida foram calculados a partir das bases de dados do Índice de Preços do Consumidor (Base 1991=100, para o período 1994-1997) e do Inquérito aos Orçamentos Familiares (1994/1995).

Os resultados obtidos através das várias metodologias utilizadas apresentam algumas diferenças, tal como era esperado. No entanto, elas estão de acordo quanto aos aspectos mais importantes. Por um lado, os níveis de custo de vida das regiões consideradas não podem ser considerados muito díspares, a amplitude máxima é 8% no ano de 1994 e 6% nos três anos seguintes. Por outro lado, todas indicam que a região que suporta o custo de vida mais elevado é a de Lisboa e Vale do Tejo, seguida da Madeira. Quanto às restantes regiões, apesar de verem a sua posição relativa variar com o método utilizado e de ano para ano, têm um custo de vida que não difere muito entre si. Exceptuando a R.A. dos Açores, as relações entre os custos de vida das regiões não sofreram alterações significativas entre 1994 e 1997. Na R.A. dos Açores, o custo de vida aumentou mais do que na região Norte no ano de 1995 e menos nos anos seguintes.

Interessou-nos, para além da comparação de custos de vida entre regiões, comparar os custos de alguns grupos mais específicos de bens e serviços. Esta análise mais detalhada permitiu, também, avaliar de que forma cada um destes grupos contribuiu para o custo de vida global. O grupo da alimentação e bebidas não alcoólicas foi o que teve a maior contribuição para a posição relativa de Lisboa e Vale do Tejo, sobretudo devido ao custo da carne, peixe, cereais e legumes. No caso da R.A. da Madeira, foram os grupos do mobiliário, acessórios para lar, equipamento doméstico e de manutenção corrente e dos hotéis, cafés e restaurantes que tiveram uma maior influência no custo de vida.

Na análise dos resultados obtidos, há que ter em conta certas limitações que dizem respeito aos dados utilizados. A única informação disponível sobre os preços nas várias regiões é a recolhida para o cálculo do Índice de Preços no Consumidor que pode não ser adequada ao objectivo deste estudo. Tentámos colmatar esta limitação pela investigação exhaustiva das bases de dados e, tanto quanto possível, a sua adaptação.

Procuraram dar-se algumas justificações para os resultados, contudo, a descoberta das suas verdadeiras causas carece de uma investigação mais detalhada das especificidades regionais.

Neste estudo, comparamos apenas os custos de vida das regiões portuguesas. Um trabalho de investigação, também, interessante seria a comparação dos poderes de compra de várias regiões europeias, tendo como base preços regionais, e não preços das capitais nacionais, que corresponde ao procedimento adoptado correntemente pelo EUROSTAT.

REFERÊNCIAS BIBLIOGRÁFICAS

PARENTE, P., BAGO D' UVA, T. (2001), "Índices de Custo de Vida: Metodologias de Cálculo",
Revista de Estatística, 1º Quadrimestre, INE;

ÍNDICE DE PREÇOS DO CONSUMIDOR (Base 1991=100) – *Metodologia - Instituto Nacional de Estatística;*

INQUÉRITO AOS ORÇAMENTOS FAMILIARES 1994/1995 - *Metodologia - Instituto Nacional de Estatística;*

Nupcialidade e Conjuntura Económica. À procura de um modelo econométrico para a Taxa Bruta de Nupcialidade

Autor:
Paulo Parente

VOLUME I

1º QUADRIMESTRE DE 2001

NUPCIALIDADE E CONJUNTURA ECONÓMICA. À PROCURA DE UM MODELO ECONOMÉTRICO PARA A TAXA BRUTA DE NUPCIALIDADE

NUPTIALITY AND ECONOMIC CONJUNCTURE. IN SEARCH OF AN ECONOMETRIC MODEL FOR THE GROSS MARRIAGE RATE

Autor: Paulo Parente⁵⁶

- Técnico Superior de Estatística no Gabinete de Estudos e Conjuntura do
Instituto Nacional de Estatística – Núcleo de Apoio e Desenvolvimento

RESUMO:

- Neste trabalho procuramos averiguar qual a relação entre o clima económico e o número de casamentos realizados em Portugal, durante o período 1980-1998. Com este propósito, é desenvolvido um modelo econométrico com dados mensais em que a variável dependente é a Taxa Bruta de Nupcialidade e os regressores são alguns indicadores de conjuntura económica. O modelo estudado permite concluir que a nupcialidade depende de uma forma positiva da conjuntura económica.

Paralelamente é desenvolvido um estudo sobre a estacionaridade das séries utilizadas. Os resultados obtidos sugerem que todas as séries estudadas têm uma raiz unitária simples. Notavelmente a série Taxa Bruta de Nupcialidade parece conter, também, algumas raízes unitárias sazonais, um resultado raramente encontrado na literatura.

PALAVRAS-CHAVE:

- *Nupcialidade, Conjuntura Económica, Estacionaridade, Raízes Unitárias Sazonais.*

ABSTRACT:

- In this article we try to discover the relationship between the economic climate and the number of marriages in Portugal during the period 1980-1998. With this purpose, an econometric model is developed using monthly data, in which the dependent variable is the Gross Marriage Rate and the regressors are some indicators of economic conjuncture. The studied model allows us to conclude that the nuptiality depends in a positive way of the economic conjuncture.

⁵⁶ Este artigo corresponde a uma actualização do relatório de estágio curricular da licenciatura em Matemática Aplicada à Economia e à Gestão do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa e elaborado pelo autor no ano lectivo 1997/1998. Este estágio foi realizado no Gabinete de Estudos - Área Demográfica e Social - do Instituto Nacional de Estatística sob a orientação do Professor Doutor João Santos Silva e da Dra. Maria José Carrilho a quem se agradece toda a ajuda dispensada. Agradece-se, ainda, os comentários da Dra. Teresa Bago d'Uva, do Dr. Daniel Mota e do Dr. Daniel Santos.

At the same time, a study on the stationarity of the series used in the model was developed. The results obtained suggest that all the studied series have a simple unit root. Notably the series Gross Marriage Rate seems to have also some seasonal unit roots, a result that is rarely found in the literature.

KEY-WORDS:

- *Nuptiality, Economic Conjuncture, Stationarity, Seasonal Unit Roots.*

1. INTRODUÇÃO

O casamento desempenha um papel fundamental na estrutura da sociedade, uma vez que é a forma de criar o elemento mais importante que está na sua base: a família. Consequentemente, um estudo sobre o fenómeno nupcialidade revela-se de particular importância.

Com este trabalho, procuramos verificar se a conjuntura económica teve alguma influência nos casamentos em Portugal, durante o período 1980-1998.

Este estudo tem a seguinte organização. Na secção 2 fazemos um enquadramento do fenómeno nupcialidade, apresentando a definição legal de casamento e descrevendo os factores que, segundo a teoria, são apontados como determinantes para o crescimento ou diminuição do número de casamentos.

A análise da a série dos casamentos mensais permite concluir que ao longo da década de 70 existiu uma transformação lenta, tanto da tendência, como da sazonalidade, cujos determinantes são explicados na secção 3. Na nossa opinião, este processo só terminou em 1979. Este fenómeno, na Teoria de Sucessões Cronológicas, é geralmente designado por quebra de estrutura. Por este motivo e pelo facto de nos parecer que a relação entre nupcialidade e conjuntura económica só é estável durante um curto período, o nosso estudo incidiu sobre os anos 80 e 90.

Na secção 4 apresentamos um resumo da metodologia ARIMA e na secção 5 são focados alguns temas relacionados com a Teoria Económica, salientando alguns aspectos que, na nossa opinião, têm sido negligenciados na literatura.

Elaboramos um modelo econométrico a fim de poder saber qual o sentido da relação entre a conjuntura económica e os casamentos (positivo ou negativo) e se esta é significativa do ponto de vista estatístico. A qualidade das previsões deste modelo é avaliada através da comparação com as previsões de um modelo ARIMA. Este estudo é apresentado no capítulo 6.

Por último, apresentamos na secção 7 as conclusões que julgamos mais importantes e sugerimos algumas ideias para trabalhos futuros.

2. ENQUADRAMENTO

Ao iniciar um estudo desta natureza é necessário definir claramente o que é o casamento. Segundo o artigo 1577 do Código Civil (Dec.-Lei nº 496/77, de 25-11), *“casamento é um contrato celebrado entre duas pessoas de sexo diferente que pretendem constituir família mediante uma plena comunhão de vida, nos termos das disposições deste Código”*.

Existem duas modalidades de casamento em Portugal desde a assinatura da Concordata com a Santa Sé em 7 de Maio de 1940: o casamento católico e o casamento civil. A lei civil reconhece o valor e a eficácia do casamento católico. Não têm relevância jurídica os casamentos celebrados segundo as regras de qualquer outra denominação religiosa.

Para compreender integralmente a evolução do matrimónio ao longo do tempo, é necessário analisar as causas que norteiam o comportamento das pessoas relativamente a este assunto. Segundo Michel (1983), os principais factores que podem fazer variar as Taxas Brutas de Nupcialidade são:

1. a legislação do casamento e do divórcio;
2. factores sócio-económicos;
3. a população em idade de casar;
4. a classe social, a religião, o nível de educação e de qualificação profissional, etc..

No que respeita ao ponto 1, Roland Pressat (citado por Michel, 1983) refere o "carácter variável" da nupcialidade na União Soviética. Notou-se, segundo este autor, um aumento de casamentos nos jovens desde 1964, relacionado com uma legislação mais liberal do divórcio, pois tratando-se de uniões facilmente revogáveis havia uma maior possibilidade de se efectivarem do que na época em que a dissolução enfrentava verdadeiros obstáculos jurídicos.

Em relação ao ponto 2, é importante verificarmos se o ambiente social e económico é um factor que pesa na decisão do casamento. Historicamente, existe a ideia que a nupcialidade está relacionada de uma maneira muito estreita com a conjuntura económica. Michel (1983) cita dois autores da década de 50 que analisaram esta relação em França, em dois períodos distintos da sua história:

- Duplessis-Le Guelinel observou que durante o Antigo Regime (séculos XVI a XVIII) os casamentos em França eram frequentes quando o preço do trigo era baixo e diminuía quando a miséria se instalava.
- T.Caplow afirmou sobre o período de 1933 a 1953: "O número de casamentos é o melhor índice do andamento dos negócios, bem melhor do que os índices comerciais correntes".

A relação entre casamentos e o ambiente económico também é influenciada pelas mudanças culturais que vão surgindo na sociedade.

Um exemplo flagrante da forte relação que existe entre os casamentos que se realizam e a composição da população em idade de casar (referido no ponto 3) é o fenómeno demográfico denominado "*marriage squeeze*". Michel (1983) descreve este fenómeno da seguinte forma: No período após a Segunda Guerra Mundial assistiu-se a uma explosão demográfica em consequência de um forte aumento de natalidade geralmente designada por "*baby-boom*". É sabido que as mulheres, na sua generalidade, casam mais cedo que os homens. Quando as raparigas da geração do "*baby-boom*" atingiram a idade do casamento verificou-se que havia um excesso de jovens mulheres em idade de casar. Essas mulheres não encontravam para cônjuges, homens na idade preferencial do casamento. Glick (citado por Michel, 1983)

constatou algumas consequências deste fenómeno: o aumento da idade do casamento por parte das mulheres americanas nos anos 60 e um crescimento do número de celibatárias. Outro efeito deste fenómeno foi uma queda dramática do número de casamentos, que ocorreu em Inglaterra e no País de Gales, nos primeiros anos da década de 70 (Ermisch, 1981).

Passando ao ponto 4, a instrução é um factor a considerar para o aumento da idade do casamento, o que pode constatar-se em alguns países da América Latina: as mulheres que têm instrução secundária casam-se mais tarde do que as mulheres que não têm instrução. Nos Estados Unidos, o número de raparigas na Universidade triplicou entre 1960 e 1972, tendo-se verificado, tal como em França, o crescimento da actividade profissional feminina. Como consequência, assistiu-se ao aparecimento de movimentos feministas com uma opinião crítica relativamente ao casamento. O nascimento destes grupos constituídos por mulheres da classe média e com instrução mais elevada que a maioria levou a uma menor frequência do casamento acompanhada do aumento da idade média do mesmo (Michel, 1983).

O estatuto profissional é a variável com maior influência na idade do casamento, quanto mais baixo for o estatuto profissional, mais precoce é o casamento (para as pessoas não qualificadas o casamento é uma compensação para o carácter impessoal do seu papel profissional - Michel, 1983).

Há que ter em conta que a idade do primeiro casamento, sobretudo no caso das mulheres é um dos factores determinantes para a variação da taxa de natalidade. Vemos, assim, que o comportamento da população relativamente ao casamento é de extraordinária importância devido à sua influência nas taxas de crescimento da população (Michel, 1983).

Durante os anos 70, notou-se em França que as mulheres consideravam o casamento como um obstáculo para a sua promoção profissional e que os homens pelo contrário, tinham a perspectiva de que o casamento lhes trazia benefícios em termos de subida na hierarquia profissional. (Michel, 1983)

Becker (1981) explica a diminuição do número de casamentos nos países ocidentais desde a década de 60, com a crescente independência económica da mulher. Segundo o autor, este facto reduz o “ganho” ou a “atractividade” do casamento.

O casamento é mais atraente quando os parceiros têm diferentes atributos para trocar. Enquanto um dos cônjuges contribui com trabalho doméstico, cabe ao outro desempenhar um trabalho remunerado. Segundo o mesmo autor, quanto maior for a capacidade das mulheres para ganhar dinheiro, menor será o seu interesse em ocupar o tempo a cuidar dos filhos ou das tarefas caseiras.

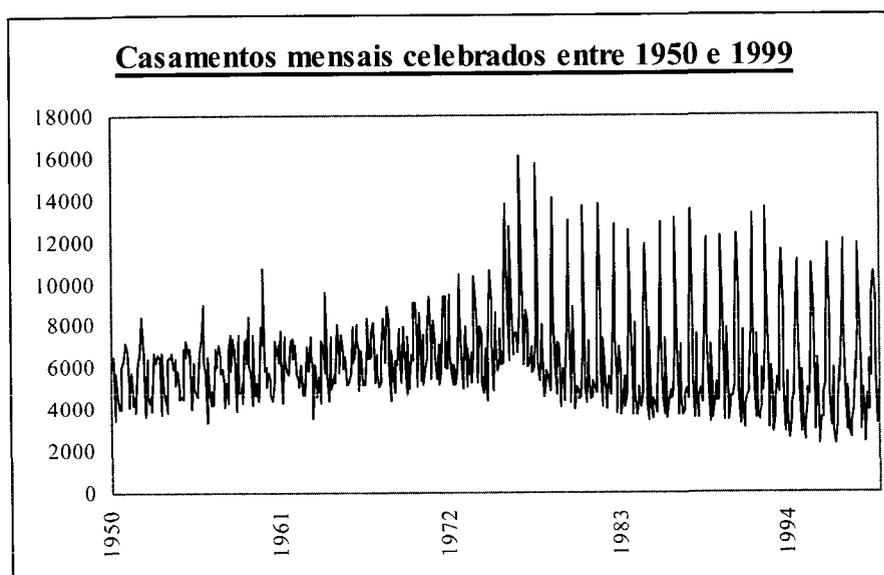
A existência de grandes custos emocionais associados com a dissolução do casamento leva ao surgimento das “uniões de facto”. Este tipo de relação tem vindo a aumentar nas sociedades contemporâneas substituindo os rígidos modelos de união do passado.

Não podemos esquecer que a introdução de novas ideologias levou a que as diversas comunidades passassem a encarar o casamento de uma forma mais flexível,

repudiando o modelo rígido defendido por várias religiões que o consideravam um laço indissolúvel e sagrado.

3. CASAMENTOS MENSAIS DE 1950 A 1999

Figura 1



A análise da Figura 1 permite verificar que a tendência dos casamentos mudou a partir de 1976. No período anterior, observamos um crescimento dos casamentos motivada principalmente pela política conservadora do Estado Novo baseada fortemente em ideais religiosos, embora acompanhado por um grande número de oscilações provocadas pelos fenómenos relacionados com as guerras coloniais. No período posterior, os matrimónios diminuem. Em relação à sazonalidade, esta começou a modificar-se ao longo da década de 70. Durante as décadas de 50, 60 e primeiros anos da década de 70, o mês em que se efectuaram a maior parte dos casamentos era Dezembro, nos anos seguintes e até 1996 as pessoas passaram a casar-se mais em Agosto. Entrando no campo da especulação, esta alteração dos meses preferenciais para casar, poderá dever-se ao facto das pessoas marcarem a data do casamento para uma época em que toda a família esteja reunida, aguardando a chegada dos seus parentes emigrados que vêm passar as férias no Verão em Portugal. A acrescentar a este facto, há que notar que, pela Lei do Contrato de Trabalho de 1969 (Dec.-Lei 49408, artº 62º), já existe uma referência, embora de carácter indicativo, ao estabelecimento do subsídio de férias. Em 1975, o Dec.-Lei 292/75 (artº 18º) garante aos trabalhadores o direito a férias e o direito a um subsídio equivalente ao da remuneração do respectivo período de férias (Santos, 1981). Talvez este facto explique a mudança de sazonalidade dos casamentos.

4. DESCRIÇÃO DA MODELIZAÇÃO ARIMA⁵⁷

Um dos modelos mais populares utilizados para descrever séries cronológicas é o modelo multiplicativo integrado sazonal inicialmente sugerido por Box e Jenkins. Um processo X_t segue tal modelo quando:

$$\phi_p(L)\Phi_P(L^S)(1-L)^d(1-L^S)^D X_t = \theta_q(L)\Theta_Q(L^S)\varepsilon_t$$

$$p \geq 0, q \geq 0, P \geq 0, Q \geq 0, d \geq 0, D \geq 0$$

em que S é o período da sazonalidade, ε_t é um ruído branco, $\phi_p(L)$ e $\theta_q(L)$ são polinómios finitos no operador de desfasamento L ⁵⁸ e são definidos da seguinte forma:

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

$$\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$$

Os polinómios $\Phi_P(L^S)$ e $\Theta_Q(L^S)$ são polinómios finitos no operador de desfasamento L^S e têm a definição que se segue:

$$\Phi_P(L^S) = 1 - \Phi_1 L^S - \Phi_2 L^{2S} - \dots - \Phi_P L^{PS}$$

$$\Theta_Q(L^S) = 1 + \Theta_1 L^S + \dots + \Theta_Q L^{QS}$$

O polinómio $(1-L)^d$ serve para remover a tendência e o operador $(1-L^S)^D$ serve para remover tanto a tendência como a componente sazonal. Estes polinómios são aplicados de forma a obter séries estacionárias.

Como a inclusão de uma constante pode piorar as previsões, uma vez que o modelo se torna pouco flexível, optou-se por só a incluir se for estatisticamente significativa. Observe-se que se num modelo em diferenças e em logaritmos existir uma constante, esta poderá ser interpretada como uma taxa de crescimento.

⁵⁷ Este texto apoia-se no livro de Murteira et al. (1993)

⁵⁸ O operador de desfasamento L transforma o valor de um processo no instante t no valor anterior $t-1$, ou seja: $LY_t = Y_{t-1}$

O modelo descrito anteriormente é designado abreviadamente por $SARIMA(p,d,q) \times (P,D,Q)_S$.

A metodologia Box - Jenkins é composta de três etapas:

- 1- Identificação;
- 2- Estimação;
- 3- Avaliação e diagnóstico.

A Identificação é constituída por dois passos:

- a) Estacionarização da sucessão cronológica. A identificação de **d**, **D** e **S** foi feita através da análise da Função de Autocorrelação estimada. Uma convergência lenta para zero indicia não estacionaridade, e uma convergência lenta para zero sobre os “lags” múltiplos de **S** sugere a aplicação de uma diferença sazonal.
- b) Identificação de **p**, **q**, **P**, **Q**. Esta análise foi feita através da Função de Autocorrelação (FAC) e da Função de Autocorrelação Parcial (FACP) estimadas.

O programa utilizado foi o TSP (Versão 4.5- Hall e Cummins, 1999). Este *software* estima os parâmetros do modelo ARIMA por máxima verosimilhança. Note-se que a função de verosimilhança dos modelos ARIMA depende da sequência passada dos resíduos. Se fixarmos estes a zero podemos afectar seriamente as estimativas pelo que, para ultrapassar este problema, o programa calcula os chamados “*back-forecasts*” dos resíduos.

Para avaliar a qualidade dos modelos, foram empregues testes de autocorrelação. Existem dois tipos de testes que assentam sobre a FAC e FACP estimadas dos resíduos : Individuais e de Conjunto.

Os testes individuais utilizados neste trabalho foram o teste de Bartlett e o teste de Jenkins e Daniels. O de Ljung-Box. foi o teste de conjunto calculado. Este último normalmente é considerado pouco potente⁵⁹. A descrição destes testes poderá ser encontrada em Murteira et al. (1993).

Para seleccionar o melhor modelo foi utilizado o critério de Schwartz (SBC) e as previsões. O critério de Akaike foi, também, calculado, verificou-se, todavia, que os modelos escolhidos através deste critério continham um grande número de parâmetros.

⁵⁹ Potência do teste é igual à probabilidade de rejeitar correctamente uma hipótese nula que é falsa.

5. ALGUNS TÓPICOS DE ECONOMETRIA

Neste ponto são feitas algumas considerações sobre os testes estatísticos utilizados, as medidas escolhidas para a detecção de observações influentes no Modelo de Regressão Linear e sobre um assunto que era muito dramatizado até há pouco tempo: a Multicolinearidade.

Não indicamos os procedimentos de todos os testes estatísticos uma vez que alguns já são muito conhecidos. Contudo, são referenciados os livros onde os leitores menos familiarizados com estes assuntos poderão obter essa informação.

5.1. ESTACIONARIDADE

5.1.1. PROCESSOS COM RAÍZES UNITÁRIAS SIMPLES

Quando se trabalha em séries cronológicas, temos de ter em conta um resultado obtido por Granger e Newbold (1974). Estes dois autores elaboraram várias vezes a seguinte experiência: simularam de forma independente dois passeios aleatórios X_t e Y_t :

$$Y_t = Y_{t-1} + \varepsilon_{1t}$$

$$X_t = X_{t-1} + \varepsilon_{2t}$$

em que ε_{1t} , ε_{2t} são ruídos brancos independentes.

Ao estimar o modelo

$$Y_t = \beta X_t + u_t$$

chegaram à conclusão de que a relação entre as variáveis Y_t e X_t era estatisticamente significativa em 74% dos casos. Contudo, este resultado não tinha sentido uma vez que as séries foram construídas de forma independente.

Verificamos, assim, que quando trabalhamos com séries temporais os resultados de uma regressão podem parecer bons, apesar de serem falsos na realidade. Consequentemente este tipo de regressões são espúrias.

Nos anos seguintes, estes resultados foram sendo explicados pela teoria: as “regressões espúrias” acontecem quando trabalhamos com séries não estacionárias⁶⁰. Dentro da classe das séries não estacionárias, existe a classe das séries integráveis. Estas séries tornam-se estacionárias após tomarmos as suas diferenças.

Para averiguar se é uma série é integrada ou não, é necessário fazer os chamados testes de raízes unitárias, sendo o mais famoso o teste Dickey-Fuller. Consideremos o modelo $AR(1)$:

$$(1) Y_t = \phi Y_{t-1} + \varepsilon_t.$$

Notemos que uma série tem uma raiz unitária se $\phi = 1$. Subtraindo Y_{t-1} em ambos os lados da equação obtemos:

$$(2) \Delta Y_t = (\phi - 1)Y_{t-1} + \varepsilon_t.$$

A hipótese nula do Teste Dickey-Fuller é a existência de uma raiz unitária. Assim, este teste consiste na estimação da equação (2) pelo método dos Mínimos Quadrados e na utilização da estatística t do parâmetro $(\phi - 1)$, geralmente designado por τ , como estatística de teste. Sob a hipótese nula de existência de uma raiz unitária esta estatística tem uma distribuição assintótica não standard.

Outro tipo de testes Dickey-Fuller consiste em utilizar como estatística de teste $z = n(\hat{\phi} - 1)$. Em que $\hat{\phi} - 1$ é obtido pelo método dos Mínimos Quadrados aplicado à equação (2). Tal como no caso anterior, a distribuição assintótica desta estatística é não standard.

As distribuições destas estatísticas encontram-se tabeladas em, por exemplo, Enders (1995). Note-se que se incluirmos na equação (1) um termo independente ou uma tendência determinística as distribuições assintóticas das estatísticas de teste alteram-se. Todavia, quando a equação (1) contém um termo independente, a inclusão de variáveis “dummy” sazonais não altera as distribuições assintóticas das estatísticas.

Quando o verdadeiro processo gerador de dados é $AR(p)$, estes testes não podem ser aplicados. Neste caso, os resíduos da regressão feita com a equação (2) irão apresentar sintomas de autocorrelação. Existem duas soluções possíveis para este problema.

Uma solução consiste na inclusão de um número suficiente de defasamentos de ΔY_t como regressores na equação (2) e na utilização da estatística t relativa ao termo Y_{t-1} . Estes testes têm o nome Augmented Dickey-Fuller (ADF) e a distribuição da estatística de teste é a mesma do que no caso em que os erros são i.i.d.. O Teste z não é válido se utilizarmos esta metodologia.

⁶⁰ Um processo é fracamente estacionário até à segunda ordem se tem média e variância independentes do tempo e as covariâncias entre os seus valores dependem apenas do defasamento entre eles.

Outra solução consiste na correção das estatísticas τ e z do facto de existir autocorrelação. Este teste tem o nome de Phillips-Perron. No estudo empírico desenvolvido neste trabalho apresentamos apenas os resultados referentes ao Teste Phillips-Perron utilizando a estatística z , uma vez que este tem mais potência que o teste baseado em τ ⁶¹.

Um teste de raízes unitárias que, segundo testes de simulação, parece ter maior potência que os o Teste ADF e o teste Phillips-Perron é o teste τ simetricamente ponderado de Pantula et al (1994).

Considere-se o seguinte modelo $AR(1)$ com deriva

$$Y_t = \mu(1 - \phi) + \phi Y_{t-1} + e_t$$

em que os e_t são variáveis aleatórias i.i.d. com média 0 e variância σ^2 .

Se este modelo é estacionário então $E(Y_t) = \mu$ e $Var(Y_t) = \sigma^2 / (1 - \phi^2)$. Assumindo que esta hipótese é verdadeira é fácil provar que

$$Y_t = \mu(1 - \phi) + \phi Y_{t+1} + u_t,$$

em que os u_t são variáveis aleatórias i.i.d. com média 0 e variância σ^2 .

Esta simetria leva-nos a considerar estimadores de ϕ que minimizem

$$Q_w(\phi) = \sum_{t=2}^N w_t (y_t - \phi y_{t-1})^2 + \sum_{t=1}^{N-1} (1 - w_{t+1}) (y_t - \phi y_{t+1})^2$$

em que os w_t , $t = 2, 3, \dots, N$ são pesos e $y_t = Y_t - \bar{Y}$ em que $\bar{Y} = \left(\sum_{i=1}^N Y_i \right) / N$

O teste τ simetricamente ponderado consiste na utilização da estatística t de significância do parâmetro $\phi - 1$ considerando como ponderadores $w_t = N^{-1}(t - 1)$.

No caso em que o verdadeiro processo gerador de dados é $AR(p)$, a função perda a minimizar é dada por:

$$\sum_{t=p+1}^N (t-p) \left(y_t - \phi_1 y_{t-1} - \sum_{i=2}^p \phi_i \Delta Y_{t-i} \right)^2 + \sum_{t=p+1}^{N-1} (N-t+p) \left(y_{t-p} - \phi_1 y_{t+1} + \sum_{i=2}^p \phi_i \Delta Y_{t-p+i} \right)^2.$$

⁶¹ Veja-se Davidson e MacKinnon (1993) para mais pormenores sobre este teste.

A estatística de teste é a mesma que no caso $AR(1)$.

Seguiu-se, neste trabalho, a sugestão de Dickey e Pantula (1987) de começar os testes pela ordem de integração mais elevada. Ou seja, testamos primeiro a existência de duas raízes unitárias antes de testarmos a hipótese de existência de apenas uma raiz unitária. Assim, para testar a existência de d raízes unitárias os testes anteriormente descritos foram aplicados à série $\Delta^{d-1}Y_t$.

Neste trabalho os testes de raízes unitárias simples foram realizados utilizando o procedimento automático do *software* TSP. Com este procedimento o processo gerador de dados de cada variável é aproximado por um modelo $AR(p)$ com $p = q + 2$ em que q é escolhido pelo critério de Akaike. No caso do teste Dickey-Fuller comparou-se o valor do p calculado automaticamente pelo TSP com o p escolhido utilizando o teste de autocorrelação Breusch-Godfrey (ver secção 5.3.). Uma vez que ambas as metodologias produziram valores semelhantes para p e levaram a conclusões idênticas optou-se por apresentar apenas os resultados obtidos pelo procedimento automático do TSP.

Apesar deste tema ter recebido a atenção de inúmeros autores, existem alguns problemas inerentes aos testes de raízes unitárias que até ao momento não foram ultrapassados:

- Em primeiro lugar a potência destes testes é muito baixa uma vez que não conseguem, geralmente, distinguir entre um processo quase não estacionário e um processo integrado de primeira ordem (Veja-se por exemplo os resultados de Pantula et al., 1994);
- Os testes de raízes unitárias são assintóticos pelo que têm problemas quando aplicados a séries de pequena dimensão (Harris, 1995);
- Uma série pode ser estacionária, mas num seu segmento um teste pode não rejeitar a hipótese de existência de uma raiz unitária. Vulgarmente, na prática, só trabalhamos com uma pequena parte da série cronológica pelo que é grande a probabilidade de estarmos a cometer este erro.

O estudo da estacionaridade de uma variável é um dos pontos mais delicados em Econometria, pelo que é necessário ter prudência.

5.1.2. PROCESSOS COM RAÍZES UNITÁRIAS SAZONAIS

Uma vez que os dados que são trabalhados neste estudo são mensais é necessário ter em conta as flutuações sazonais. Murteira et al.(1993) definem os movimentos sazonais de uma série temporal como “oscilações de ritmo forçado, que se repetem todos os anos (ou com uma periodicidade mais curta), nem sempre seguindo um padrão rígido; podem ter causas naturais ou sociais”.

Uma das formas mais frequentes de tratar a sazonalidade consiste na utilização de procedimentos de dessazonalização das séries. Contudo, tal como refere Harris (1995), estes procedimentos geralmente distorcem as propriedades dos dados. Segundo este autor, quando os dados são dessazonalizados a estatística Dickey-Fuller

está enviesada, pelo que a hipótese nula de não estacionaridade é rejeitada menos vezes do que se devia.

Assim, é necessário utilizar outras metodologias, nomeadamente os processos sazonais de sucessões cronológicas. Segundo Hylleberg et al.(1990), geralmente são utilizadas três classes de sucessões cronológicas para modelizar a sazonalidade⁶²:

- a) Processos sazonais determinísticos puros;
- b) Processos sazonais estacionários;
- c) Processos integrados sazonais.

Um processo sazonal determinístico puro é um processo gerado por “*dummies*” sazonais. Ou seja, para dados mensais, se X_t é um processo deste tipo então é gerado na seguinte forma:

$$X_t = m_1 + \sum_{k=2}^{12} m_k S_{kt}$$

em que S_{kt} são “*dummies*” sazonais. Neste caso o padrão sazonal é fixo.

Um processo sazonal estacionário pode ser gerado por um processo autorregressivo de ordem infinita:

$$\varphi(L)X_t = \varepsilon_t$$

em que $\varphi(L)$ é um polinómio no operador desfasamento com raízes fora do círculo unitário, podendo ser algumas pares complexos com periodicidades sazonais.

Num processo integrado sazonal algumas raízes do polinómio $\varphi(L)$ estão sobre o círculo unitário. Se o polinómio $\varphi(L)$ tem uma raiz unitária simples e onze raízes unitárias sazonais então pode ser decomposto da seguinte forma (para dados mensais):

$$\varphi(L) = (1 - L^{12})\phi(L)$$

em que $\phi(L)$ é um polinómio no operador desfasamento com raízes fora do círculo unitário.

Note-se que o polinómio $(1 - L^{12})$ tem a raiz 1 e a seguintes raízes sazonais:

$$-1; \pm i; \frac{-1 \pm \sqrt{3}i}{2}; \frac{1 \pm \sqrt{3}i}{2}; \frac{-\sqrt{3} \pm i}{2}; \frac{\sqrt{3} \pm i}{2}$$

⁶² Geralmente, neste tipo de estudo é utilizada a Análise Espectral, todavia, devido ao carácter introdutório do presente texto, optou-se por não empregar esta metodologia.

em que $i^2 = -1$

Recordando que o módulo de um número complexo $a + bi$ é $\sqrt{a^2 + b^2}$ verificamos, facilmente, que o módulo das raízes do polinómio $(1 - L^{12})$ são todos iguais à unidade.

Enquanto a raiz simples corresponde a 12 ciclos por ano, as raízes sazonais correspondem, respectivamente a 6, 3, 9, 8, 4, 2, 10, 7, 5, 1 e 11.

Neste caso, o processo X_t não é estacionário mas o processo $(1 - L^{12})X_t$ é estacionário.

As características das séries integradas sazonalmente são as seguintes:

- qualquer choque tem um efeito permanente sobre a série e pode mudar o seu padrão sazonal (como diz Osborn, 1993, a “Primavera” pode tornar-se “Verão”);
- A variância da série aumenta linearmente com o tempo.

Utilizando a notação de Osborn et al. (1988) uma série é dita integrada de ordem (d, D) e escreve-se $I(d, D)$ se a série fica estacionária depois de tomadas d diferenças simples e D diferenças sazonais, ou seja, $X_t \sim I(d, D)$ se e só se $(1 - L)^d (1 - L^S)^D X_t$ é estacionária.

Tal como é referido por Beaulieu e Miron (1993), quando trabalhamos com duas ou mais séries integradas sazonalmente podemos obter correlações espúrias.

Segundo Osborn (1993), as variáveis macroeconómicas geralmente podem ser descritas como um processo $I(1)$ com um padrão sazonal determinístico, ou seja, na prática não são muito frequentes séries com raízes unitárias sazonais.

O teste utilizado neste estudo baseia-se numa extensão do teste *HEGY* (Hylleberg, Engle, Granger e Yoo, 1990) a dados mensais feita por Beaulieu e Miron em 1993.

Seja X_t uma série gerada pelo modelo autorregressivo:

$$\varphi(L)X_t = \varepsilon_t.$$

Em que $\varphi(L)$ é um polinómio no operador de desfasamento.

O objectivo deste teste é verificar a hipótese das raízes do polinómio $\varphi(L)$ estarem sobre o círculo unitário.

Para isso os autores utilizam a seguinte proposição originalmente devida a Lagrange que geralmente é utilizada em Análise Numérica para aproximar funções:

Proposição: Qualquer polinómio $\varphi(L)$ (possivelmente infinito ou racional) que toma valores finitos nos pontos distintos, diferentes de zero e possivelmente complexos $\theta_1, \dots, \theta_p$ pode ser expresso em termos de polinómios elementares e um resto da seguinte forma:

$$\varphi(L) = \sum_{k=1}^p \lambda_k \frac{\Delta(L)}{\delta_k(L)} + \Delta(L)\varphi^{**}(L)$$

em que λ_k são constantes,

$\varphi^{**}(L)$ é um polinómio possivelmente infinito ou racional,

$$\delta_k(L) = 1 - \frac{L}{\theta_k} \quad \text{e} \quad \Delta(L) = \prod_{k=1}^p \delta_k(L)$$

É fácil verificar que θ_k é uma raiz de $\varphi(L)$ se e só se $\lambda_k = 0$. A demonstração desta proposição encontra-se em Hylleberg et al.(1990). A partir deste resultado Beaulieu e Miron (1993) constroem o seguinte ensaio de forma a testar a hipótese das raízes do polinómio serem também raízes do polinómio $\varphi(L)$:

Estimação pelo método dos mínimos quadrados da equação:

$$\varphi(L)^* y_{13t} = \sum_{k=1}^{12} \pi_k y_{k,t-1} + m_1 + m_0 t + \sum_{j=2}^{12} m_j S_{jt} + \varepsilon_t$$

em as variáveis S_{kt} são “*dummies*” sazonais, $\varphi(L)^*$ é um polinómio estacionário no operador de desfasamento que permite eliminar a eventual autocorrelação do termo de perturbação e

$$y_{1,t} = (1 + L + L^2 + L^3 + L^4 + L^5 + L^6 + L^7 + L^8 + L^9 + L^{10} + L^{11})X_t$$

$$y_{2,t} = -(1 - L + L^2 - L^3 + L^4 - L^5 + L^6 - L^7 + L^8 - L^9 + L^{10} - L^{11})X_t$$

$$y_{3,t} = -(L - L^3 + L^5 - L^7 + L^9 - L^{11})X_t$$

$$y_{4,t} = -(1 - L^2 + L^4 - L^6 + L^8 - L^{10})X_t$$

$$y_{5,t} = -\frac{1}{2}(1 + L - 2L^2 + L^3 + L^4 - 2L^5 + L^6 + L^7 - 2L^8 + L^9 + L^{10} - 2L^{11})X_t$$

$$y_{6,t} = \frac{\sqrt{3}}{2} (1 - L + L^3 - L^4 + L^6 - L^7 + L^9 - L^{10}) X_t$$

$$y_{7,t} = -\frac{1}{2} (1 - L - 2L^2 - L^3 + L^4 + 2L^5 + L^6 - L^7 - 2L^8 - L^9 + L^{10} + 2L^{11}) X_t$$

$$y_{8,t} = -\frac{\sqrt{3}}{2} (1 + L - L^3 - L^4 + L^6 + L^7 - L^9 - L^{10}) X_t$$

$$y_{9,t} = -\frac{1}{2} (\sqrt{3} - L + L^3 - \sqrt{3}L^4 + 2L^5 - \sqrt{3}L^6 + L^7 - L^9 + \sqrt{3}L^{10} - 2L^{11}) X_t$$

$$y_{10,t} = \frac{1}{2} (1 - \sqrt{3}L + 2L^2 - \sqrt{3}L^3 + L^4 - L^6 + \sqrt{3}L^7 - 2L^8 + \sqrt{3}L^9 - L^{10}) X_t$$

$$y_{11,t} = \frac{1}{2} (\sqrt{3} + L - L^3 - \sqrt{3}L^4 - 2L^5 - \sqrt{3}L^6 - L^7 + L^9 + \sqrt{3}L^{10} + 2L^{11}) X_t$$

$$y_{12,t} = -\frac{1}{2} (1 + \sqrt{3}L + 2L^2 + \sqrt{3}L^3 + L^4 - L^6 - \sqrt{3}L^7 - 2L^8 - \sqrt{3}L^9 - L^{10}) X_t$$

$$y_{13,t} = (1 - L^{12}) X_t$$

Uma vez que:

$$\lambda_1 = -\pi_1 \qquad \lambda_5 = (-\pi_5 + i\pi_6)/2 \qquad \lambda_9 = (-\pi_9 + i\pi_{10})/2$$

$$\lambda_2 = -\pi_2 \qquad \lambda_6 = (-\pi_5 - i\pi_6)/2 \qquad \lambda_{10} = (-\pi_9 - i\pi_{10})/2$$

$$\lambda_3 = (-\pi_3 + i\pi_4)/2 \qquad \lambda_7 = (-\pi_7 + i\pi_8)/2 \qquad \lambda_{11} = (-\pi_{11} + i\pi_{12})/2$$

$$\lambda_4 = (-\pi_3 - i\pi_4)/2 \qquad \lambda_8 = (-\pi_7 - i\pi_8)/2 \qquad \lambda_{12} = (-\pi_{11} - i\pi_{12})/2$$

Para verificar as hipóteses de existência de raízes unitárias sazonais, testamos:

- a hipótese $\pi_k = 0$ para $k=1,2$ através da respectiva estatística t . A hipótese alternativa é $\pi_k < 0$.
- a hipótese $\pi_{k-1} = \pi_k = 0$ para $k \geq 4$ através da estatística F . A hipótese alternativa é.

Para mostrar que não existem raízes unitárias sazonais basta que π_k seja diferente de zero para $k=2$ e que hipótese $\pi_{k-1} = \pi_k = 0$ seja rejeitada para pelo menos um dos seguintes conjuntos: {3,4}; {5,6}; {7,8}; {9,10}; {11,12}.

Inspeccionaram-se, também, os gráficos e as Funções de Autocorrelação das séries originais e diferenciadas, para confirmar os resultados dos testes.

Após o estudo da estacionaridade das variáveis e, caso estas tenham a mesma ordem de integração procura-se, geralmente, averiguar se existe cointegração entre as variáveis, ou seja, uma combinação linear entre elas que exiba uma ordem de integração inferior.

A interpretação económica de cointegração é a de que se duas ou mais séries estão ligadas de forma a existir uma relação de equilíbrio de longo prazo, mesmo que as variáveis não sejam estacionárias, então elas terão um movimento próximo uma da outra e existirá uma combinação linear entre as duas séries que será estável, ou seja, estacionária (Harris, 1995).

A análise do fenómeno nupcialidade leva-nos a chegar à conclusão que não tem sentido a existência de cointegração entre a Taxa Bruta de Nupcialidade e os indicadores de conjuntura económica. Em primeiro lugar, a relação entre casamentos e o ambiente económico é influenciada pelas mudanças culturais que vão surgindo na sociedade pelo que não existe uma relação estável de longo prazo. Em segundo lugar, algumas das variáveis utilizadas, como por exemplo as “Vendas de Gasolina”, um dia, no futuro, poderão acabar e não tem sentido que os casamentos acompanhem essa tendência. Mesmo que os testes estatísticos indicassem a existência de uma relação de cointegração, esta seria espúria. Assim, especificamos o modelo nas diferenças das variáveis, tal como se fazia, nos primeiros trabalhos imediatamente após descoberta de que a maior parte das séries não são estacionárias.

5.2. HETEROCEDASTICIDADE ⁶³

Greene (1993) mostra que se a heterocedasticidade não está correlacionada com as variáveis no modelo então o estimador dos Mínimos Quadrados, pelo menos em amostras grandes, não será enganador.

Assim, devemos-nos preocupar, principalmente, com a heterocedasticidade devida aos regressores. Por este motivo, pelo facto de exigir a normalidade dos erros e, ainda, por supor que sobre a hipótese alternativa as variâncias dos erros podem ser ordenadas de forma crescente ou decrescente, não foi utilizado o teste Goldfeld-Quandt.

⁶³ Se a variância dos erros varia de observação para observação, então existe heterocedasticidade. O estimador dos mínimos quadrados é centrado, é consistente, mas não é BLUE, nem é eficiente em termos assintóticos e produz estimadores dos desvios padrões dos estimadores dos parâmetros que são enviesados, assim os procedimentos de teste clássicos e os intervalos de confiança para os parâmetros não são válidos (Godfrey, 1988).

A hipótese nula de homocedasticidade foi testada através dos ensaios de White (poderá ver a descrição deste teste em Greene, 1993) e de Breusch-Pagan. Note-se que, em geral, o teste de White tem pouca potência.

Foi aplicado o teste Breusch-Pagan na versão que não depende da normalidade dos erros. Este teste consiste na realização da regressão auxiliar do quadrado dos resíduos estimados $\hat{\varepsilon}_t^2$ sobre uma constante e \hat{Y}_t^2 . A estatística de teste é $N \times R^2$ e segue assintoticamente a distribuição Qui-quadrado com um grau de liberdade⁶⁴.

Tradicionalmente nos estudos econométricos de sucessões cronológicas assume-se a hipótese de homocedasticidade. Contudo, Engle (1982) mostrou através de um estudo empírico que esta hipótese poderá ser irrealista. Este autor verificou que os erros de previsão grandes e pequenos aparecem, geralmente, em “clusters”, sugerindo que a variância dos erros de previsão depende de erros de períodos anteriores. Assim, Engle (1982) sugeriu modelizar a variância dos erros condicional na sua história passada como um processo $AR(p)$:

$$(3) E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$$

Este modelo é designado por ARCH(p) de Autoregressive Conditional Heteroscedasticity em que p é o número de defasamentos do quadrado dos erros que se inclui em (3). Com estas hipóteses o estimador usual dos Mínimos Quadrados é consistente, embora não seja eficiente. No caso em que não existem variáveis dependentes defasadas incluídas nos regressores, os estimadores dos desvios padrões usuais são, também, consistentes. Contudo, se esta hipótese for violada os desvios padrões dos estimadores dos parâmetros terão de ser estimados utilizando um estimador robusto à presença de heterocedasticidade pois os estimadores usuais são inconsistentes (Engle, 1982).

Engle (1982) sugeriu, também, um teste de *score* para verificar se o modelo é ARCH. Neste trabalho foi empregue este teste para verificar se o modelo era ARCH(12) ou ARCH(24).

É necessário salientar que não existe nenhuma justificação estatística ou económica para a variância condicional seguir a expressão definida por (3). Consequentemente o teste de Engle (1982) é apenas utilizado, neste trabalho, como teste de especificação. Note-se que os resultados deste teste só serão importantes se os regressores incluírem defasamentos da variável dependente pois nesse caso o estimador usual dos desvios padrões dos parâmetros é inconsistente.

⁶⁴ Veja-se Koenker (1981).

5.3. AUTOCORRELAÇÃO ⁶⁵

Uma das estatísticas mais utilizadas para testar a existência de autocorrelação é a Durbin-Watson. Greene (1993) faz duas críticas a esta estatística:

- a estatística Durbin-Watson só é estritamente válida se os regressores são não estocásticos.
- se o processo dos erros não é AR(1) então a estatística poderá indicar, de forma incorrecta, que não existem problemas de autocorrelação.

Assim, para superar estes problemas sugere a utilização das estatísticas Durbin-h, Durbin-h alternativa no caso em que incluímos no modelo uma variável dependente desfasada ou as estatísticas Breusch-Godfrey, Box-Pierce e Ljung-Box, que, segundo o autor, são aplicáveis em qualquer situação.

Contudo, um estudo de simulação feito por Kiviet (1986) mostrou que estes testes têm vários inconvenientes:

- as estatísticas Box-Pierce e Ljung-Box não devem ser aplicadas quando o modelo inclui regressores exógenos e variáveis dependentes desfasadas ao mesmo tempo;
- teste Durbin-h, bem como as suas generalizações têm propriedades pobres em amostras de pequena dimensão;
- teste Breusch-Godfrey tem problemas quando existem regressores redundantes, principalmente, em amostras pequenas.

Segundo este autor, o teste que apresentou melhores propriedades foi um teste semelhante ao teste Breusch-Godfrey sugerido por Harvey(1982). A hipótese nula é a não existência de autocorrelação. A hipótese alternativa é os erros seguirem um processo AR(p) ou MA(p).

O teste consiste na regressão dos resíduos do método dos mínimos quadrados $\hat{\varepsilon}_t$ sobre $x_t, \hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-p}$ (substituindo os valores que faltam dos resíduos desfasados por zero⁶⁶). A estatística de teste é a estatística F de significância dos resíduos desfasados. Com base no seu estudo Kiviet (1986) sugere que se deve ignorar o facto desta estatística não ter exactamente a distribuição F em modelos dinâmicos e devem usar-se os valores críticos da distribuição F com p e N-K-p graus de liberdade (em que

⁶⁵ Quando existe pelo menos uma covariância entre os erros desfasados no tempo diferente de zero, então estamos na presença de autocorrelação. Uma explicação para a autocorrelação é a de que os factores omitidos do modelo estão correlacionados entre períodos (Greene, 1993). Nesta situação, o estimador dos mínimos quadrados é centrado, é consistente, mas não é BLUE (Best Linear Unbiased Estimator), nem é eficiente em termos assintóticos e produz estimadores dos desvios padrões dos estimadores dos parâmetros que são enviesados, pelo que os procedimentos de teste clássico e os intervalos de confiança para os parâmetros não são válidos. Se o modelo incluir valores desfasados da variável dependente e existir autocorrelação então o estimador dos mínimos quadrados é, em geral, enviesado e inconsistente. (ver Godfrey, 1988).

⁶⁶ Segundo Godfrey(1988), a utilização destes valores iniciais é irrelevante em termos assintóticos mas pode influenciar o comportamento do teste em amostras pequenas.

N é o número de observações e K o número de regressores no modelo). Os resultados do estudo permitiram verificar que este teste tem probabilidade de erro tipo I⁶⁷ que é relativamente invariante com a dimensão da amostra, ordem de autocorrelação, verdadeiros valores dos coeficientes e regressores redundantes.

O teste Breusch-Godfrey calculado pelo software TSP baseia-se na mesma regressão do teste anterior e tem a mesma hipótese nula mas a estatística de teste é $p \times F$ que tem assintoticamente a distribuição Qui-quadrado com p graus de liberdade.

Contudo, Kiviet (1986) mostra (para N grande) que este teste tem uma maior probabilidade de erro tipo I que o teste F de Harvey. Tendo em atenção este resultado, no modelo em estudo sobre a Taxa Bruta de Nupcialidade foi utilizado o teste F de Harvey. O teste Breusch-Godfrey foi utilizado para estudar a veracidade da hipótese de ausência de autocorrelação nos testes de raízes unitárias sazonais.

5.4. ESTABILIDADE DOS PARÂMETROS

Para avaliar a estabilidade dos parâmetros, utiliza-se geralmente o teste de Chow. Contudo este teste é afectado pela violação da hipótese de normalidade e pela existência de autocorrelação e heterocedasticidade dependente dos regressores (ver Godfrey, 1988, para mais pormenores).

Existem inúmeros métodos que permitem efectuar testes de quebra de estrutura na presença de heterocedasticidade. Um método muito simples consiste nos seguintes passos:

- 1- construir uma variável “*dummy*” que toma o valor zero na primeira parte da amostra e um na segunda parte;
- 2- incluir no modelo de estimação variáveis adicionais que correspondem ao produto desta variável “*dummy*” por todos os regressores existentes no modelo original.
- 3- Utilizar a estatística de Wald de significância dos parâmetros adicionais incluídos em 2 utilizando uma matriz de covariâncias robusta a heterocedasticidade.

Note-se, contudo, que quando existe heterocedasticidade e a quebra de estrutura é pequena este teste revela muito pouca potência⁶⁸.

⁶⁷ Probabilidade de erro tipo I é igual à probabilidade de rejeitar a hipótese nula quando esta é verdadeira.

⁶⁸ Verificou-se através de simulações de Monte Carlo que este resultado é comum a vários testes de quebra de estrutura robustos à presença de heterocedasticidade. Estes resultados serão disponibilizados a eventuais interessados.

5.5. FORMA FUNCIONAL

Neste trabalho, empregamos a transformação das variáveis que proporcionaram melhores previsões: os logaritmos. No caso em que a variável corresponde a uma taxa, então tomamos o logaritmo da variável somada com o número um. Encontra-se nesta situação a Taxa Bruta de Nupcialidade.

Para verificar se a forma funcional aplicada foi correcta, decidimos utilizar o teste RESET de Ramsey. Este teste assenta no pressuposto de que se a forma funcional de um modelo de regressão linear não está adequada às variáveis intervenientes, então é de esperar que o quadrado, ou uma potência de ordem superior, de um ou mais regressores melhore significativamente a explicação do comportamento da variável dependente. Na prática, o teste consiste em fazer uma regressão auxiliar da variável dependente sobre todos os regressores do modelo e o quadrado do estimador das observações da variável dependente (\hat{Y}_i^2) e em seguida realizar um teste *t* ao seu coeficiente (Godfrey, 1988). Este teste poderá ser influenciado pela presença de heterocedasticidade.

5.6. DETECÇÃO DE OBSERVAÇÕES INFLUENTES

Observações influentes no Modelo de Regressão linear são aquelas que dão um contributo relativamente grande para a estimação dos parâmetros.

Existem dois tipos de observações influentes:

- “*Leverages*”: linha da matriz dos regressores X que tem grande influência na estimação correcta dos coeficientes de regressão. A medida de “leverage” utilizada foi o elemento da diagonal de ordem t da hat matrix $H_X = X(X'X)^{-1}X'$: h_t . Considerou-se influente a observação cujo h_t é superior a $2K/N$.
- “*Outliers*”: Observações que correspondem a erros elevados. Para detectar os “*Outliers*” foram calculados os resíduos studentizados e considerados aqueles cujo módulo é superior a 2.

Foi, também utilizada uma medida que tem em conta estes dois tipos de observações influentes: *DFFITs*. Considerou-se a observação t influente se

$$|DFFITs_t| > 2\sqrt{K/N}.$$

As fórmulas destas medidas poderão ser vistas em Judge et al. (1988) ou Greene (1993).

5.7. MULTICOLINEARIDADE

Segundo Judge et al. (1993), multicolinearidade exacta é definida como a existência de uma ou mais combinações lineares exactas das colunas da matriz dos regressores X . Neste caso a matriz $X'X$ é singular e o estimador dos mínimos quadrados $B = (X'X)^{-1} X'Y$ não é definido. Não existe um estimador único para β que minimiza a soma dos quadrados dos resíduos.

Note-se, contudo, que a multicolinearidade exacta resulta da má especificação do modelo (Greene, 1993).

Quando existe uma combinação linear aproximada entre as variáveis explicativas então existe multicolinearidade aproximada. Neste caso, Judge et al. (1993) apontam as seguintes consequências:

- Alguns elementos da matriz $(X'X)^{-1}$ podem ter valores muito elevados. Uma vez que a matriz de covariâncias do estimador dos Mínimos Quadrados é $Cov(B) = \sigma^2 (X'X)^{-1}$ então os erros padrão e as covariâncias do estimador dos mínimos quadrados podem ser muito altos. Isto significa que a variabilidade da amostra é alta, os intervalos de confiança são grandes e a informação obtida é relativamente imprecisa;
- Quando os erros-padrão são elevados a estatística t irá levar-nos à conclusão que os valores dos parâmetros não são significativamente diferentes de zero. O problema é as variáveis colineares não proporcionam informação suficiente para estimar os seus efeitos separados;
- As estimativas dos coeficientes podem variar muito com a adição ou subtracção de observações.

Greene (1993) ainda acrescenta que os sinais nos coeficientes podem apresentar o sinal errado.

As consequências apontadas são todas verdadeiras, mas o problema não é muito grave.

O ponto de vista de Goldberger (1991) sobre este assunto é extremamente importante pelo que é necessário chamar a atenção para alguns factos referidos pelo autor:

- Mesmo quando existe multicolinearidade aproximada o estimador dos mínimos quadrados continua a ser *BLUE* e os erros padrão e os intervalos de confiança continuam a ser válidos;
- A variância do estimador b_j é dada por

$$Var(b_j) = \frac{\sigma^2}{(1 - R_j^2) \sum_i (x_{ij} - \bar{x}_j)^2}$$

Onde R_j^2 é o coeficiente de determinação da regressão auxiliar do regressor x_j sobre todos os outros e σ^2 é a variância do erro. Logo, podemos concluir que para além da multicolinearidade (medida por R_j^2) existem outros factores que fazem aumentar a variância do estimador, como por exemplo alto σ^2 ou uma pequena variância do regressor. O que é desejável não é baixa multicolinearidade mas sim baixas variâncias dos coeficientes;

- Afirmar que a multicolinearidade faz aumentar os erros padrão, é sugerir que eles são grandes de uma forma espúria. Contudo, tem sentido que os erros padrão sejam altos porque as estimativas dos coeficientes variam muito de amostra para amostra;
- Quando se diz que um parâmetro é significativo mas é a multicolinearidade que nos leva a aceitar a hipótese da sua nulidade, estamos a confundir significância estatística com significância económica;

Por último, quando se afirma que a multicolinearidade leva a que os sinais apresentem sinais contrários aos esperados e, portanto, sem sentido, não se estará a levar em conta um facto. Como mostra Greene (1993), o sinal de um coeficiente de um regressor é igual ao sinal da correlação parcial entre a variável dependente e esse regressor⁶⁹. Isto significa que o sinal do parâmetro depende das outras variáveis explicativas que incluímos no modelo.

Tomando em atenção o que acabamos de afirmar é possível encontrar uma situação em que existe multicolinearidade, um dos sinais parece, à primeira vista, estar trocado mas tem sentido. O leitor imagine uma situação em que incluímos num modelo duas variáveis que servem para medir o mesmo efeito e estão correlacionadas (por exemplo taxa de juro e inflação). Imagine, ainda, que se espera que os sinais destas variáveis sejam positivos e, contudo, os resultados da estimação mostram que um dos sinais das variáveis é contrário. Uma justificação para esta situação é a de que o parâmetro de uma das variáveis estará a captar o efeito esperado (positivo) em excesso e, de forma a corrigir este facto, o coeficiente da outra variável apresenta um valor negativo.

⁶⁹ Correlação parcial entre a variável dependente e um regressor mede a correlação entre estas duas variáveis excluindo os efeitos que todas as outras variáveis explicativas incluídas no modelo têm sobre elas.

6. MODELO ECONOMÉTRICO PARA A TAXA BRUTA DE NUPCIALIDADE

6.1. CONJUNTURA ECONÓMICA E SUA INFLUÊNCIA NA DECISÃO DO CASAMENTO

Conjuntura económica é a sucessão de fluxos, de movimentos cíclicos de alta e baixa, que animam a economia em períodos mais ou menos longos. Estes movimentos detectam-se a partir de alguns indicadores como os preços, as produções, as vendas, os salários, etc.. Os movimentos conjunturais são em síntese dois:

- os de baixa, de contracção e de depressão;
- os de alta, de prosperidade e expansão.

Obviamente, as situações de expansão ou contracção da economia de um país têm uma grande influência nas condições financeiras das pessoas que desejam casar. Para contraírem os laços do matrimónio, a ocasião ideal para essas pessoas é aquela em que possam reunir em simultâneo um bom salário e um emprego estável que lhes permitam suportar os custos da compra/arrendamento de uma casa, dos móveis, dos electrodomésticos, talvez do automóvel, e também dos filhos que porventura irão nascer. É claro que esta situação é mais provável quando o custo de vida é mais baixo e quando existe um clima de prosperidade na economia: maiores níveis de produção, emprego, etc..

Assim, construiu-se um modelo econométrico com o objectivo de verificar de que forma alguns indicadores de conjuntura económica estão relacionados com os casamentos. Ao elaborar um estudo desta natureza é necessário ter em conta duas situações:

- Em primeiro lugar, existem indicadores de conjuntura económica que captam efeitos que só serão sentidos pela generalidade da população num período posterior;
- Em segundo lugar, de uma forma geral, entre a ocasião em que as pessoas se apercebem que têm as condições reunidas para casar e a realização do mesmo decorre algum tempo. O casal, na medida das suas possibilidades, começa por fazer a compra/arrendamento da casa, aquisição de móveis, electrodomésticos, roupas, etc..

Com estas duas afirmações, pretendemos chamar a atenção para o facto de que a relação entre os casamentos e os indicadores de conjuntura económica é desfasada no tempo e não é contemporânea. Neste trabalho, admitimos que o período máximo de desfasamento é de um ano.

Notemos, contudo, que este modelo não é um modelo causal uma vez que não é pelo facto de um indicador ser alto ou baixo que as pessoas casam ou deixam de casar. Por exemplo, as pessoas não se casam porque o índice de produção industrial é alto ou porque se está a vender muita gasolina (alguns indicadores de conjuntura económica

que vamos utilizar no modelo). As pessoas casam-se em primeiro lugar porque estão apaixonadas e em segundo lugar porque têm condições económicas para tal.

O período em análise é 1980-1998 e a qualidade do modelo é avaliada comparando as previsões para o ano de 1999 com as previsões obtidas com um modelo ARIMA.

Como o intervalo de tempo sobre o qual incide a nossa atenção é muito diminuto, optou-se por criar um modelo com dados mensais. Apesar de não ser muito frequente a utilização deste tipo de dados em modelos econométricos, esta parece-nos a melhor forma de não perder informação e de criar um modelo sobre este período com alguma credibilidade.

6.2. APRESENTAÇÃO DAS VARIÁVEIS

Neste modelo, a variável dependente é a Taxa Bruta de Nupcialidade (*TBNUPC*). Esta taxa dá-nos o número de casamentos ocorridos durante um certo período de tempo por cada indivíduo. Assim, em termos mensais:

$$TBNUPC_t = \frac{\text{casamentos realizados no mês } t}{\text{estimativas para a população no mês } t}$$

Os indicadores de conjuntura económica utilizados no modelo são os seguintes⁷⁰:

- Índice de Produção Industrial da Indústria Transformadora no mês *t* (base 1995 = 100) - *IPIT*_{*t*};
- Vendas de Automóveis no mês *t* (em unidades) - *VAUT*_{*t*};
- Vendas de Gasolina no mês *t* (em toneladas) - *VGAS*_{*t*};
- Ofertas de Emprego disponíveis nos Centros de Emprego ao longo do mês *t* - *OFEMP*_{*t*};
- Número de desempregados inscritos nos Centros de Emprego ao longo do mês *t* - *DES*_{*t*};
- Índice de Preços no Consumidor no mês *t* (base 1997=100) - *IPC*_{*t*};

Os indicadores “Vendas de Automóveis” e “Vendas de Gasolina” poderão ser interpretadas como indicadores de bem estar económico da sociedade.

Nenhum destes indicadores pode-se considerar perfeito. A utilização do “Índice de Produção Industrial” poderá ser criticada porque o sector com maior peso na

⁷⁰ Fonte dos dados: Instituto Nacional de Estatística.

economia portuguesa no período em causa foi o sector de serviços. Também existem algumas dúvidas sobre o facto da série “Vendas de Automóveis” detectar de forma correcta os movimentos que animam a economia. Este indicador é muito afectado por mudanças legislativas no ramo automóvel (nomeadamente relacionadas com alterações no imposto).

Note-se que a série “Vendas de Automóveis” foi fornecida pelo INE e que parte dela foi construída por retropolação, uma vez que antes de 1994 os dados originais incluíam os "todo-o-terreno". Em relação ao IPC uma vez que os índices se encontravam reportados a diferentes anos base (1980 a 1984- base 1976; 1984 a 1995 –base 1991; 1995 a 1999- base 1997), foi necessário proceder a uma mudança de base para converter os índices para a base 1997=100.

Com o intuito de testar a ideia de que os casamentos se realizam aos fins de semana e feriados foi também incluído no modelo a variável dias úteis mensais (*DU*).

Existem algumas superstições em volta dos casamentos. Por exemplo, algumas pessoas evitam casar em anos bissextos pensando que esse facto irá trazer azar ao casamento. Por nossa curiosidade, foi incluída no modelo uma variável “dummy” para investigar se esta crença das pessoas tem algum efeito sobre o número de casamentos que se realizam. Esta variável será denominada *BI* e toma o valor 1 durante os anos bissextos e 0 caso contrário

Os anos bissextos que ocorreram entre 1980 a 1998 foram 1980, 1984, 1988, 1992 e 1996.

Toda a legislação fundamental sobre casamento e divórcio em vigor durante período em análise foi criada até 1977 (inclusive). Nos anos 1980 a 1998 não existiu nenhuma alteração fundamental a essas leis. Assim, verificamos que não é necessário introduzir alguma variável que pudesse ter em conta as modificações do quadro jurídico que regula este domínio.

6.3. RESULTADOS DOS TESTES DE ESTACIONARIDADE

Neste ponto são expostos os resultados dos testes de raízes unitárias e raízes unitárias sazonais.

No Quadro nº 1 estão inscritos os valores observados das estatísticas dos testes de raízes unitárias. Nos modelos utilizados nestes testes, para além dos regressores habituais, foram incluídas uma constante, uma tendência determinística e 11 variáveis “dummy” sazonais. Note-se que *D* representa a ordem de integração das séries que se testou.

Iniciou-se a sequência de testes pela ordem de integração imediatamente superior ao que se supunha à priori.

Alguns autores como por exemplo Harris (1995) indicam que os logaritmos dos índices de preços no Reino Unido devem ser tratados como I(2) pelo que é necessário aferir da veracidade deste resultado para Portugal.

Quadro nº 1
Testes de Raízes unitárias
Período Jan-1980 a Dez- 1998

D	ADF			Phillips-Perron			Teste Sim. Ponderado		
	1	2	3	1	2	3	1	2	3
LOG(IPIT)	-2.268	-2.696	-	-110.296**	-289.334**	-	-3.994**	-26.735**	-
Desf.	17	15	-	17	15	-	24	2	-
LOG(VAUT)	-1.917	-6.416**	-	-15.405	-237.850**	-	-3.047*	-17.715**	-
Desf.	4	6	-	4	6	-	4	4	-
LOG(VGAS)	-1.892	-1.930	-	-365.131**	-275.309**	-	-4.235**	-12.415**	-
Desf.	24	23	-	24	23	-	24	24	-
LOG(OFEMP)	-1.981	-9.318**	-	-15.375	-256.177**	-	-3.057*	-16.441**	-
Desf.	4	3	-	4	3	-	3	2	-
LOG(DES)	-2.177	-6.409**	-	-83.536**	-235.935**	-	-5.115**	-15.181**	-
Desf.	9	8	-	6	8	-	6	15	-
LOG(1+TBNUP)	-1.526	-3.406*	-	-250.473**	-266.588**	-	-8.247**	-17.959**	-
Desf.	24	24	-	24	24	-	24	24	-
LOG(IPC)	-2.025	-3.928**	-5.303**	-0.451	-217.073**	-239.352**	2.685	-8.450**	-10.070**
Desf.	8	7	12	8	7	12	5	4	20

Notas: * indica as hipóteses nulas que são rejeitadas a um nível de significância de 10%.

** indica as hipóteses nulas que são rejeitadas a um nível de significância de 5%.

Desf. é a ordem de desfasamento máxima da variável dependente incluída no modelo.

Dos testes apresentados, aquele que apresentou um pior comportamento foi o teste ADF. O facto de não rejeitar a hipótese das séries log(VGAS) e log(IPIT) serem I(2) mostra a pouca potência deste teste. Note-se que os resultados dos outros testes indicam que estas variáveis são I(0).

Os resultados indicam ainda que log(OFEMP), log(VAUT), log(DES), e log(IPC) são variáveis I(1). O facto dos testes rejeitarem a hipótese de log(IPC) ser I(2) surpreende devido aos resultados obtidos por Harris(1995).

Segundo o teste ADF a série log(1+TBNUP) pode ser considerada com I(1). Por outro lado, os restantes testes rejeitam a hipótese de existência de qualquer raiz unitária simples nesta série.

O resultado dos testes HEGY são apresentados no Quadro nº 2. Nas regressões auxiliares foram incluídas uma constante, variáveis "dummy" sazonais e uma tendência determinística.

Quadro nº 2

Teste HEGY - Raízes unitárias sazonais

Período Jan-1980 a Dez- 1998

	t_1	t_2	$F_{3,4}$	$F_{5,6}$	$F_{7,8}$	$F_{9,10}$	$F_{11,12}$	Desf.
LOG(IPIT)	0.742	-3.233*	25.317**	24.827**	44.053**	28.531**	26.220**	0
LOG(VAUT)	1.453	-4.215**	43.204**	19.958**	32.400**	48.520**	47.164**	0
LOG(VGAS)	-0.512	-4.447**	32.810**	35.615**	26.341**	35.356**	24.313**	13
LOG(OFEMP)	1.907	-4.819**	33.902**	33.463**	55.424**	44.052**	54.128**	0
LOG(DES)	1.850	-3.788**	37.057**	32.714**	33.203**	66.603**	32.985**	0
LOG(1+TBNUUP)	0.969	-2.062	9.482**	27.653**	24.096**	8.451**	2.868	3
LOG(IPC)	1.490	-4.268**	56.204**	43.021**	37.660**	60.687**	41.243**	0

Notas: * indica as hipóteses nulas que são rejeitadas a um nível de significância de 10%.

** indica as hipóteses nulas que são rejeitadas a um nível de significância de 5%.

Desf. é a ordem de desfasamento máxima da variável dependente incluída no modelo.

Os valores críticos encontram-se tabelados em Beaulieu e Miron (1993) e dependem da normalidade dos erros. Aceitando os resultados destes testes verificamos que para estacionarizar a série do logaritmo de um mais a Taxa Bruta de Nupcialidade é necessário utilizar o filtro $(1 - L^{12})$. Os resultados indicam que todas as outras séries têm apenas uma raiz unitária.

Podemos concluir que só a série $\log(1+TBNUUP)$ apresentam raízes unitárias sazonais. Existem algumas dúvidas sobre a ordem de integração das séries $\log(IPIT)$ e $\log(VGAS)$: $I(0)$ ou $I(1)$. Todas as outras séries, na nossa opinião podem ser tratadas como $I(1)$.

Este facto poderia fazer crer que seria necessário aplicar diferenças sazonais à série $\log(1+TBNUUP)$ e diferenças simples às restantes. Contudo, um modelo econométrico tem de ter sentido e não podemos comparar variáveis que foram submetidas a diferentes tipos de tratamento. Assim, o modelo foi especificado nas diferenças sazonais das variáveis⁷¹. A aplicação deste filtro leva a que todas as variáveis sejam estacionárias mas torna algumas não invertíveis⁷².

6.4. RESULTADOS EMPÍRICOS DO MODELO

A estratégia utilizada na construção do modelo consistiu na inclusão dos indicadores de conjuntura económica em que o período de desfasamento máximo destes foi de doze meses, sendo o passo seguinte a identificação dos que eram significativos em termos estatísticos.

⁷¹ Numa primeira fase, ainda se construiu um modelo em diferenças simples e com “dummies” sazonais. Pelo facto de apresentar autocorrelações significativas, optou-se com a utilização de diferenças sazonais.

⁷² Um processo diz-se invertível quando é passivo de representação autoregressiva.

As variáveis “Dias Úteis” e BI foram incluídas no modelo sem qualquer período de desfasamento.

Para eliminar a eventual autocorrelação, optou-se por introduzir a variável dependente desfasada alguns períodos. Um modelo deste tipo é designado ARDL (autoregressive distributed lag). O método de estimação utilizado foi o dos mínimos quadrados. O estimador obtido com este método na presença de variáveis dependentes desfasadas é consistente e os procedimentos de inferência estatística são assintoticamente válidos (veja-se, por exemplo, White, 2000).

Segundo Goldberger (1991) quando trabalhamos com o modelo de regressão clássico o objectivo é estimar o vector de parâmetros β e não obter um bom ajustamento dos dados ou explicar a variação da variável dependente. Por este motivo, não utilizamos neste estudo nenhum critério de selecção de modelos ou medida da qualidade do ajustamento, como por exemplo o R^2 . A escolha do modelo é baseada, apenas, no bom senso.

A variável dependente é $Y_t = 10000 \times \Delta_{12} \text{LOG}(1 + \text{TBNUPC}_t)^{73}$. O modelo seleccionado está apresentado no Quadro nº 3. Os resultados dos testes estatísticos e das medidas de detecção das observações influentes estão expostos nos Quadros nºs 4 e 5. Verificamos que o teste Bresch-Pagan (robusto à violação da Normalidade) indica que podemos rejeitar a hipótese nula de homocedasticidade. Os resultados dos testes ARCH permitem obter a mesma conclusão. Consequentemente é necessário calcular os rácios t utilizando um estimador da matriz de covariâncias consistente quando existe heterocedasticidade. Por este motivo os rácios t apresentados no Quadro nº3 foram calculados utilizando o estimador *jackknife* da matriz de covariâncias, que, segundo estudos de simulação feitos por MacKinnon e White (1985) têm boas propriedades em amostras finitas. Pelo mesmo motivo o Teste RESET foi calculado utilizando este estimador da matriz de covariâncias e foi utilizado o teste de quebra de estrutura robusto à presença de heterocedasticidade. Os resultados destes testes mostram que a forma funcional utilizada no modelo é correcta e que não existe uma alteração de estrutura da primeira para a segunda metade da amostra.

⁷³ A multiplicação da variável dependente por 10000 deveu-se ao facto dos coeficientes das variáveis incluídas no modelo original apresentarem valores muito pequenos.

Quadro nº 3

Variáveis	Coeficientes		Erro Estatística t
	Estimados	Padrão	
Constante	-0.184	0.046	-3.520**
Y_{t-1}	-0.242	0.058	-3.449**
$\Delta_{12}\text{LOG}(\text{IPIT}_{t-12})$	2.840	0.560	4.080**
$\Delta_{12}\text{LOG}(\text{DES}_{t-3})$	-0.520	0.195	-2.235**
$\Delta_{12}\text{LOG}(\text{VGAS}_{t-2})$	1.308	0.428	2.769**
$\Delta_{12}\text{LOG}(\text{VGAS}_{t-4})$	1.447	0.434	3.632**
$\Delta_{12}\text{LOG}(\text{VAUT}_{t-3})$	-0.557	0.150	-3.626**
$\Delta_{12}\text{LOG}(\text{OFEMP}_{t-10})$	0.301	0.092	3.396**
$\Delta_{12}\text{LOG}(\text{DU}_t)$	-3.155	0.532	-6.189**
BI_t	-0.217	0.065	-3.131**

Estatística F = 17.0886**

- Notas:** * indica as hipóteses nulas que são rejeitadas a um nível de significância de 10%.
 ** indica as hipóteses nulas que são rejeitadas a um nível de significância de 5%.

Quadro nº 4

Testes de Especificação		Valor da estatística	Distribuição	p-value
Autocorrelação	Teste F - AR(1)/MA(1)	0.271	F(1,193)	0.603
	Teste F - AR(2)/MA(2)	0.512	F(2,192)	0.600
	Teste F - AR(3)/MA(3)	0.786	F(3,191)	0.503
	Teste F - AR(4)/MA(4)	0.611	F(4,190)	0.655
	Teste F - AR(5)/MA(5)	0.554	F(5,189)	0.735
	Teste F - AR(6)/MA(6)	0.486	F(6,188)	0.819
	Teste F - AR(12)/MA(12)	0.711	F(12,182)	0.740
	Teste F - AR(18)/MA(18)	0.823	F(18,176)	0.672
	Teste F - AR(24)/MA(24)	1.690	F(24,170)	0.030**
Heterocedasticidade	Teste de White	63.324	$\chi^2(54)$	0.157
	Breusch Pagan	4.552	$\chi^2(1)$	0.033**
	ARCH(12)	29.322	$\chi^2(12)$	0.004**
	ARCH(24)	33.931	$\chi^2(24)$	0.086*
Quebra de estrutura	Wald	3.846	$\chi^2(10)$	0.954
Forma Funcional	RESET	-0.998	N(0,1)	0.320

- Notas:** * indica as hipóteses nulas que são rejeitadas a um nível de significância de 10%.
 ** indica as hipóteses nulas que são rejeitadas a um nível de significância de 5%.

Quadro nº 5

Detecção de Observações influentes

Data	h_t	Data	Resíduos Studentizados	Data	DFITS
Junho-83	0.10207	Agosto-82	-2.01189	Agosto-82	-0.51228
Março-84	0.12109	Setembro-82	-2.19476	Setembro-82	-0.50704
Abril-84	0.12598	Dezembro-83	2.4775	Dezembro-83	0.47408
Maió-84	0.12666	Março-85	2.32757	Junho-84	0.58772
Junho-84	0.12155	Setembro-89	2.37743	Março-85	0.8145
Agosto-84	0.10468	Abril-90	-2.01472	Março-88	0.47094
Setembro-84	0.1199	Agosto-91	2.36264	Dezembro-88	-0.52329
Dezembro-84	0.10858	Maió-92	3.03046	Setembro-89	0.50058
Março-85	0.1091	Dezembro-92	-3.47966	Maió-92	0.81851
Abril-85	0.11997	Julho-93	2.14259	Dezembro-92	-0.70964
Junho-88	0.10611	Agosto-93	-2.85305	Julho-93	0.46826
Julho-88	0.14649	Abril-97	2.10189	Agosto-93	-1.09093
Agosto-88	0.11899	Maió-97	2.70787	Outubro-96	0.48277
Setembro-88	0.12318			Maió-97	0.54537
Novembro-88	0.099829				
Junho-92	0.11275				
Agosto-93	0.12756				
Setembro-93	0.12442				

6.5. INTERPRETAÇÃO DOS RESULTADOS

Dado que estamos a trabalhar em diferenças sazonais de logaritmos de uma variável, é necessário chamar a atenção para o facto deste valor ser aproximadamente igual à taxa de variação homóloga da variável (se esta estiver perto de zero)⁷⁴. A análise que vamos expor neste ponto terá este factor em conta. Para não tornar confuso o texto, quando explicitamos o nome de uma variável estamos sempre a referir-mo-nos à sua taxa de variação homóloga.

O estudo dos vários modelos estimados permitiu concluir que a relação entre o Índice de Preços no Consumidor e a Taxa Bruta de Nupcialidade não é significativa do ponto de vista estatístico.

É de notar que a Taxa Bruta de Nupcialidade depende de forma negativa da Taxa Bruta de Nupcialidade do período anterior. O sentido desta relação parece sugerir que em cada período existe um conjunto de pessoas que desejam casar. Se durante um mês, se realizam muitos casamentos, então no mês seguinte a dimensão desse conjunto irá diminuir pelo que existem muito menos pessoas que irão casar. O efeito contrário também acontece, se num mês se casam poucas pessoas, então o número de pessoas, que se casam no mês seguinte, irá aumentar.

⁷⁴ Note-se que: $\Delta^{12} \log(X_t) = \log(X_t / X_{t-12}) = \log(1 + \Delta^{12} X_t / X_{t-12}) \cong \Delta^{12} X_t / X_{t-12}$

Verificamos que nos meses com maior número de dias úteis se realizam menos casamentos, confirmando-se a hipótese de que estes se efectuam mais aos fins de semana e feriados.

A variável “número de desempregados inscritos” é significativa apresentando o sinal negativo, o que já era esperado. A “oferta de emprego”, também, é significativa e apresenta sinal positivo. Vemos, assim que os casamentos se realizam quando existe uma maior estabilidade no emprego.

Os coeficientes do “Índice de Produção Industrial da Indústria Transformadora” e das “Vendas de Gasolina” apresentam sinais positivos, mostrando que, quando a situação económica é favorável, os casamentos aumentam. À primeira vista, o sinal das “Vendas de Automóveis” parece contradizer o sinal dos outros indicadores. A correlação simples entre a Taxa Bruta de Nupcialidade e esta variável apresentou o mesmo sinal que o seu coeficiente. Assim, concluímos que não são as outras variáveis que influenciam o sinal do parâmetro. Os problemas apontados no ponto 6.2. poderão ser os responsáveis por este facto. Todavia, o sinal deste coeficiente pode ter sentido. Existem algumas pessoas que desejam casar e que não têm dinheiro suficiente para a compra da casa e do carro ao mesmo tempo. Sendo a casa mais importante para a vida futura do casal, estes optam por não adquirirem o automóvel. Esta será uma hipótese para justificarmos o sinal negativo das vendas de automóveis.

Outra justificação possível, embora menos credível, é o facto de nem todas as pessoas terem a mesma atitude quando dispõem de dinheiro. Existem pessoas que quando tem dinheiro decidem casar e outras que preferem manter uma certa independência e gozar a sua “liberdade”. Ao comprarem um carro estes indivíduos poderão, implicitamente, estar a tomar a decisão de não se casarem a curto-prazo.

O facto de um ano ser bissexto leva, também, leva a uma diminuição dos casamentos como pode ser confirmado pelo sinal do coeficiente da variável BI_t .

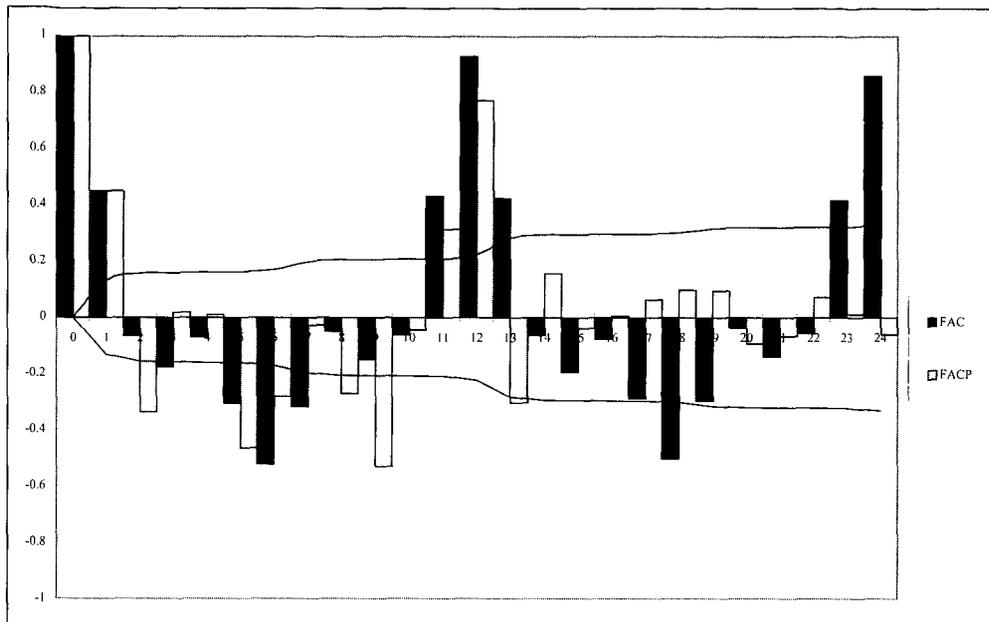
6.6. UM MODELO ARIMA

Para termos uma melhor percepção da qualidade das previsões do modelo econométrico, elaboramos um modelo ARIMA para o período 1980-1998 para a Taxa Bruta de Nupcialidade.

Uma vez que esta taxa toma valores extremamente pequenos, então $TBNUPC_t \cong \log(1 + TBNUPC_t)$. Logo a aplicação da transformação logarítmica à série não terá efeitos práticos. No modelo que expomos neste ponto a série $TBNUPC_t$ não foi submetida a nenhuma transformação.

A Figura nº 2 mostra a FAC e a FACP estimadas da Taxa Bruta de Nupcialidade.

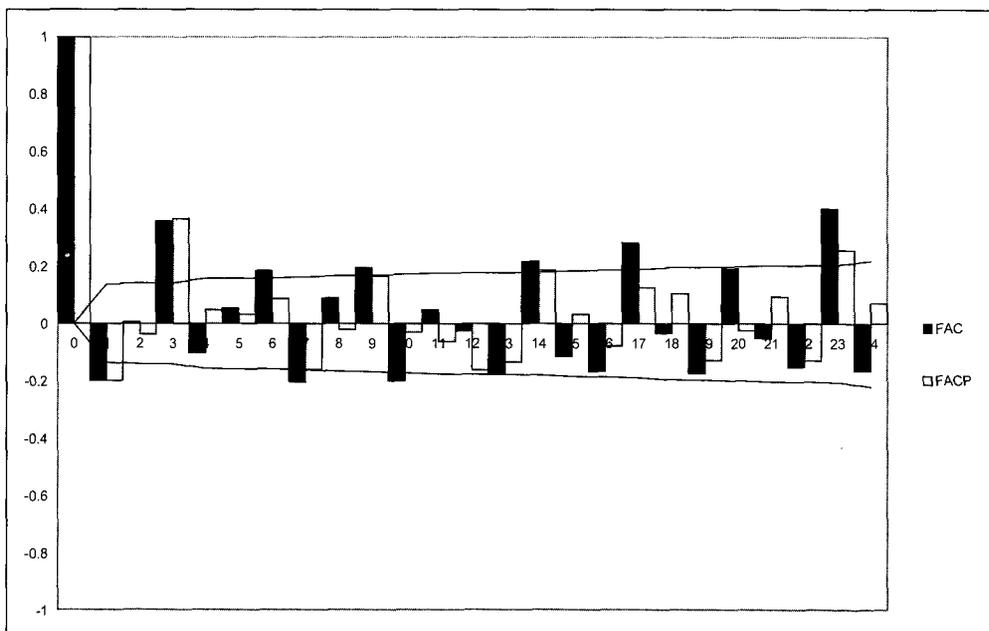
Figura nº 2



Nota: as bandas assinaladas na figura correspondem a $\pm 2 \times$ erro padrão das autocorrelações

A análise da FAC estimada permite chegar à conclusão que é necessário aplicar uma diferença sazonal para estacionarizar a série. A Figura nº 3 representa a FAC e a FACP estimadas da diferença sazonal da Taxa Bruta de Nupcialidade.

Figura nº 3



Nota: as bandas assinaladas na figura correspondem a $\pm 2 \times$ erro padrão das autocorrelações

O melhor modelo segundo o critério SBC foi um modelo muito complexo. Este modelo foi o $SARIMA(3,0,9) \times (0,1,1)_{12}$ sem constante e com as seguintes restrições:

$$\phi_2 = \theta_1 = \theta_2 = \theta_3 = \theta_5 = \theta_6 = \theta_7 = \theta_8 = 0$$

O modelo passou os testes de autocorrelação, contudo, produziu previsões muito más. Retirando o parâmetro de médias móveis do "lag" 9 ficamos com um modelo que produz melhores previsões, embora o resíduo não seja um ruído branco. Este modelo é o $SARIMA(3,0,4) \times (0,1,1)_{12}$ sem constante e apresentando o segundo parâmetro autorregressivo e os três primeiros parâmetros de médias móveis iguais a zero. Os valores estimados dos parâmetros deste modelo são apresentados no Quadro nº 6.

Quadro nº 6

Parâmetro	Estimativas	Erro	
		Padrão	Estatística t
ϕ_1	-0.275077	0.061923	-4.44224
ϕ_3	0.455231	0.060507	7.52359
θ_4	-0.162444	0.074447	-2.18201
Θ_1	0.18462	0.070991	2.60061

No Quadro nº 7 são apresentadas a FAC e a FACP estimadas dos resíduos do modelo, os valores da estatística Ljung-Box e os "p-values" desta última.

Quadro nº 7

Lag	Estatística			
	FAC	Ljung-Box	p-value	FACP
1	0.040	0.349	.	0.040
2	0.036	0.640	.	0.036
3	-0.083	2.176	.	-0.089
4	0.013	2.213	.	0.017
5	0.005	2.219	0.136	0.014
6	-0.063	3.110	0.211	-0.075
7	-0.078	4.485	0.214	-0.078
8	0.088	6.243	0.182	0.106
9	0.194	14.784	0.011	0.198
10	-0.031	14.998	0.020	-0.073
11	-0.082	16.558	0.020	-0.101
12	0.016	16.617	0.034	0.072
13	-0.045	17.083	0.047	-0.053
14	0.104	19.606	0.033	0.089
15	-0.133	23.720	0.014	-0.100
16	-0.014	23.765	0.022	0.003
17	0.189	32.217	0.002	0.211
18	0.040	32.592	0.003	-0.050
19	-0.118	35.902	0.002	-0.167
20	-0.069	37.061	0.002	0.004
21	0.023	37.184	0.003	0.071
22	-0.032	37.432	0.005	-0.106
23	0.304	59.986	0.000	0.372
24	-0.066	61.066	0.000	0.017

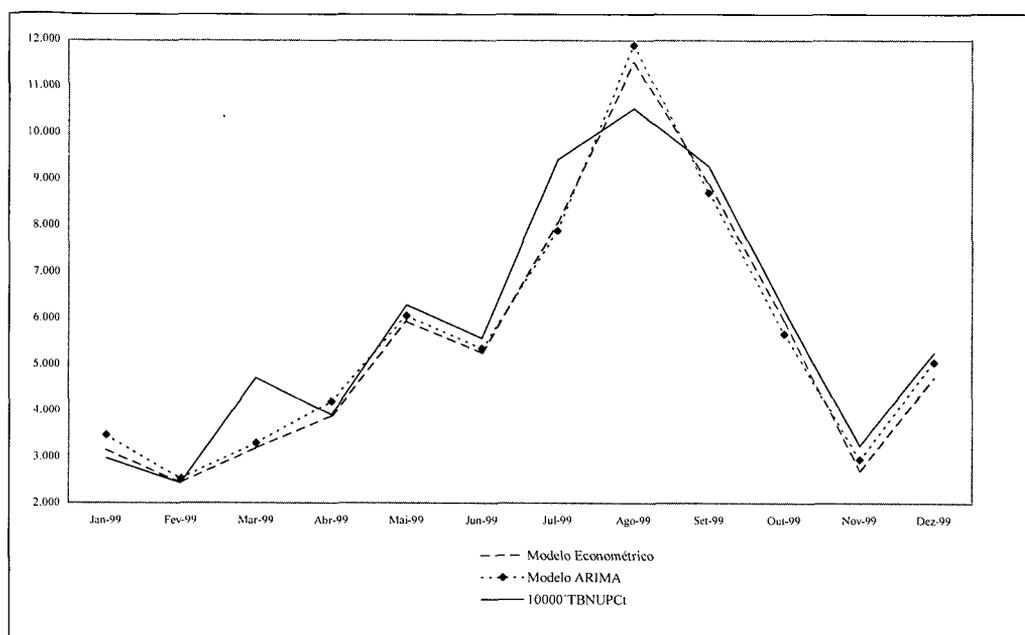
Limite crítico: $\pm 0,136$.

Verificamos que o "lag" 9 é o mais problemático em termos dos testes de autocorrelação. Esta situação poderá a ficar a dever-se à existência de alguns "Outliers".

6.7. COMPARAÇÕES ENTRE AS PREVISÕES DAS DUAS METODOLOGIAS

Na Figura nº 4 e Quadro nº 8 apresentamos as previsões para $10000 \times TBNUPC_t$ dadas pelo modelo econométrico e pelo modelo ARIMA. Podemos ver através das medidas de avaliação das previsões (Raiz do Erro Quadrático Médio -REQM - e Erro Absoluto Médio -EAM) constantes no Quadro nº 9 que as previsões obtidas através do modelo econométrico são melhores que as previsões calculadas utilizando o modelo ARIMA.

Figura nº4
Comparações entre as previsões do modelo econométrico e do modelo ARIMA



Quadro nº 8

Data	Modelo Econométrico	Modelo ARIMA	10000 x TBNUPC _t
Jan-99	3.151	3.470	2.972
Fev-99	2.431	2.531	2.430
Mar-99	3.186	3.292	4.713
Abr-99	3.883	4.206	3.907
Mai-99	5.932	6.057	6.286
Jun-99	5.252	5.345	5.571
Jul-99	8.053	7.892	9.426
Ago-99	11.517	11.895	10.530
Set-99	8.894	8.709	9.282
Out-99	5.944	5.668	6.189
Nov-99	2.704	2.938	3.237
Dez-99	4.703	5.050	5.262

Quadro nº 9

Modelo	REQM	EAM
Modelo Econométrico	0.722	0.541
Modelo ARIMA	0.786	0.616

CONCLUSÕES

Pretendemos através deste estudo fazer uma análise sobre a influência da conjuntura económica nos casamentos em Portugal no período 1980-1998.

Os resultados obtidos levam-nos a crer que existe uma relação entre os casamentos e os indicadores de conjuntura económica, isto é, os casamentos aumentam quando a situação económica do país é boa. Todavia, o indicador “Vendas de Automóveis” parece ser uma excepção neste quadro, o que nos leva a interrogar sobre o motivo de tal situação. Seria interessante analisar mais profundamente este tema. As respostas indicadas no ponto 6.5. são algumas entre as muitas possíveis.

Uma vez que partimos do pressuposto que a relação entre os casamentos e os indicadores de conjuntura económica não é estável a longo prazo, a metodologia utilizada parece ser a mais correcta. Assim, os resultados obtidos só são válidos numa perspectiva de curto-prazo.

Voltamos a chamar a atenção para o facto de nenhum dos indicadores de conjuntura económica utilizados no modelo poder ser considerado perfeito pelos motivos expostos no ponto 6.2..

A mudança de sazonalidade da série dos casamentos durante a década de 70 é um fenómeno sociológico importante que procurámos justificar mas que carece de uma explicação mais desenvolvida e pormenorizada e poderá ser objecto de um futuro estudo.

Seria interessante, também, estudar a relação entre os casamentos e outras variáveis demográficas. Um estudo importante consistiria na investigação da existência de uma relação de longo prazo entre os casamentos e nascimentos utilizando a teoria da cointegração. Contudo, seria necessário ter em conta alguns factores como a sazonalidade e as quebras de estrutura.

REFERÊNCIAS BIBLIOGRÁFICAS

- BEAULIEU, J.J. e MIRON, J. A. (1993), “Seasonal unit roots in aggregate U.S. data”, *Journal of Econometrics*, 55, pp. 305-328.
- BECKER, G. S. (1981), *A Treatise on the Family*, Harvard University Press.
- DAVIDSON, R. e MACKINNON, J. (1993), *Estimation and Inference in Econometrics*, Oxford University Press.
- DICKEY, D. A. e PANTULA, S. G.(1987), “Determining the Order of Differencing in Autoregressive Processes”, *Journal of Business & Economic Statistics*, 5(4), pp. 455-461.

- ENDERS, W. (1995), *Applied Econometric Time Series*, John Wiley & Sons, Inc..
- ENGLE, R.F., (1982), “Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom Inflation”, *Econometrica*, Vol. 50, N0. 4, pp. 987-1007.
- ERMICH, J.F. (1981), Economic Opportunities, Marriage Squeezes and Propensity to Marry: An Economic Analysis of Period Marriage Rates in England and Wales, *Population Studies*, 35 (3),pp. 347-356.
- GODFREY, L.G. (1988), *Misspecification Tests in Econometrics*, Econometric Society Monographs.
- GOLDBERGER, ARTHUR S. (1991), *A Course in Econometrics*, Harvard University Press.
- GRANGER, C.W.J. e NEWBOLD, P. (1974), “Spurious Regressions in Econometrics”, *Journal of Econometrics*, 2 (1974), pp. 111-120.
- GREENE, WILLIAM (1993), *Econometric Analysis*, Prentice Hall International Editions.
- HALL, B.H. e CUMMINS, C. (1999) *Times Series Processor, Version 4.5, Reference Manual*, TSP International.
- HARRIS, R.I.D.(1995), *Using Cointegration Analysis in Econometric Modelling*, Prentice Hall/ Harvester Wheatsheaf.
- HARVEY, A.C., (1982) *The Econometric Analysis of Time Series*, Philip Allan Publishers Limited.
- HYLLEBERG, S.; ENGLE, R.F.; GRANGER,C.W.J. e YOO, B.S.(1990), “Seasonal Integration and Cointegration”, *Journal of Econometrics*, 44, pp. 215-238 .
- JUDGE, G.; HILL, R.; GRIFFITHS, W.; LUTKEPOHL H. e LEE, T.C. (1988), *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, Inc..
- JUDGE, G.; HILL, R. e GRIFFITHS, W. (1993), *Learning and Practicing Econometrics*, John Wiley & Sons, Inc..
- KIVIET, J. F. (1986), “On the Rigour of Some Misspecification Tests for Modelling Dynamic Relationships”, *Review of Economic Studies*, 53, pp. 241-261.
- KOENKER, R. (1981), “A note on studentizing a test for heteroscedasticity”, *Journal of Econometrics*, 17, pp. 107-112.
- MACKINNON, J.G. e WHITE, H. (1985), “Some Heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties”, *Journal of Econometrics*, 29, pp. 305-325.
- MICHEL, A. (1983), *Sociologia da Família e do Casamento*, Rés-Editora, Lda. (Tradução do original em francês Sociologie de la famille et du mariage, Presses Universitaires de France).
- MURTEIRA, B. J.F.; MULLER, D. A. e TURKMAN, K. F. (1993) , *Análise de Sucessões Cronológicas*, McGraw-Hill.

OSBORN, D. R.; CHUI, A.P.L.; SMITH, J. P. e BIRCHENHALL, C. R. (1988), "Seasonality and the order of integration for consumption", *Oxford Bulletin of Economics and Statistics*, 50, pp. 361-377.

OSBORN, D. R. (1993) "Discussion, Seasonal cointegration", *Journal of Econometrics*, 55, pp. 299-303.

PANTULA, S.G.; GONZALEZ-FARIAS, G. e FULLER, W.A. (1994), "A Comparison of Unit-Root Test Criteria", *Journal of Business & Economic Statistics*, Vol. 2 No. 4., pp. 449-459.

SANTOS, J. D. (1981), *O Direito a Férias e ao Subsídio de Férias*, Multinova.

WHITE, H. (2000), *Asymptotic Theory for Econometricians*, Academic Press.

Equações salariais e rendibilidade da educação em Portugal

Autor:
Daniel Mota

VOLUME I

1º QUADRIMESTRE DE 2001

EQUAÇÕES SALARIAIS E RENDIBILIDADE DA EDUCAÇÃO EM PORTUGAL

A CONDITIONAL QUANTILE REGRESSION APPROACH TO RETURNS TO EDUCATION IN PORTUGAL

Autor: Daniel Mota

- Técnico Superior de Estatística no Gabinete de Estudos e Conjuntura –
Núcleo de Apoio e Desenvolvimento - INE – Instituto Nacional de
Estatística

RESUMO:

- É um facto reconhecido que a desigualdade salarial aumentou nos países ocidentais no período que vai dos anos 80 até meados da década de 90, sendo comumente aceite que uma sociedade com maiores níveis de escolaridade seria uma sociedade com níveis de desigualdade salarial mais esbatidos. Usando dados do Inquérito ao Emprego dos anos 1998 a 2000, pretendo mostrar que não existe evidência empírica que fundamente esta opinião.

PALAVRAS-CHAVE:

- *Rendibilidade da educação, regressão de quantis, desigualdade salarial.*

ABSTRACT:

- There is wide evidence that wage inequality has grown in western countries for the period covering the eighties and the first half of the nineties. A very common belief rests on the assumption that a more educated society would strongly reduce this inequality. Using data from the Portuguese Employment Survey covering the years 1998 to 2000, I intend to show that there is no empirical evidence supporting this view.

KEY-WORDS:

- *Returns to education, quantile regression, wage inequality.*

VOLUME I

1º QUADRIMESTRE DE 2001

1. INTRODUÇÃO

A desigualdade salarial crescente em muitos países industrializados nos últimos anos está bem documentada. O tema é controverso e tem proporcionado uma discussão acesa sobre quais os reais efeitos de um maior investimento na educação na redução desta desigualdade.

A teoria económica fornece, basicamente, duas teorias alternativas mas simultaneamente não dissociáveis entre si para explicar este fenómeno. A primeira insere-se no contexto da procura e oferta de trabalho. Uma notável evolução no campo tecnológico tem induzido uma procura crescente de mão de obra especializada, não correspondida no aumento da oferta. Deste modo, é criada uma pressão para a subida relativa dos salários dos trabalhadores mais qualificados. Esta subida dá-se, habitualmente, pela subida real dos salários destes trabalhadores mas pode também acontecer através da descida real dos salários dos trabalhadores menos qualificados. A segunda tese reflecte o aumento do grau de abertura das economias mundiais ao exterior. A adesão de Portugal à União Europeia teve como consequências mais directas o crescimento da competição externa e a consequente pressão para a melhoria da competitividade das empresas nacionais. Largamente financiado por fundos comunitários, assistiu-se a um forte investimento na modernização do parque industrial português, já que a competição com mercados onde a mão de obra era incomparavelmente mais barata se tornou insustentável. A aquisição de maquinaria moderna e novas tecnologias provocaram, mais uma vez, o aumento da procura de trabalhadores qualificados.

O conhecimento das causas deste aumento da desigualdade salarial é fundamental. Persegue-se frequentemente o objectivo político de reduzir ou, pelo menos, não aumentar este índice de desigualdade. É comum ouvir-se dizer que a concretização deste objectivo passa pela aposta numa maior educação da população. Embora esta solução seja de louvar, tal não implica que a prossecução dessa política dê frutos. Usando a técnica de regressão de quantis, pretende-se mostrar com este estudo que, se não se implementarem políticas alternativas, um maior investimento na educação não conduz necessariamente a uma redução da desigualdade salarial. Pelo contrário, é muito provável que tal medida, sendo bem sucedida, conduza a um aumento dessa mesma desigualdade.

Habitualmente, utilizam-se resultados da regressão na média (cujo estimador é, por norma, o dos mínimos quadrados) para estudar o impacto das variações de um conjunto de variáveis nos salários. Contudo, esta técnica apenas permite conhecer os seus efeitos na média da distribuição condicional dos salários. Não é natural que um dado conjunto de variáveis tenha o mesmo impacto em pessoas situadas nos extremos opostos da distribuição salarial.

Deste modo, são usadas as técnicas de regressão de quantis para documentar a heterogeneidade do impacto nos salários de variáveis consideradas fundamentais numa equação salarial. Esta faz depender o logaritmo do salário horário do grau académico atingido, da experiência e antiguidade de cada pessoa e da região (NUTS II) onde vive. A análise presente vai permitir responder a questões como a seguinte:

tomando uma amostra de indivíduos idênticos e atribuindo-lhes mais um ano de educação, será que os salários se vão tornar mais dispersos?

O artigo encontra-se organizado da seguinte forma. Na secção seguinte, é feita uma breve descrição dos dados utilizados. Estimam-se as densidades salariais pelo método kernel e oferecem-se algumas explicações simples para as diferenças que se encontram entre as amostras dos homens e das mulheres. Na terceira secção, introduz-se a técnica de regressão de quantis e um exemplo prático da sua utilidade é exposto. Na quarta secção, os resultados empíricos da estimação das regressões são apresentados. Estes são discutidos para diversos quantis, com particular ênfase para o primeiro e último decis e a mediana. O impacto da educação na desigualdade salarial é, de igual forma, amplamente discutido. Finalmente, a última secção conclui.

2. UMA ANÁLISE PRELIMINAR: BREVE DESCRIÇÃO DOS DADOS

Os dados utilizados são retirados do Inquérito ao Emprego (INE) e referem-se ao primeiro trimestre dos anos 1998 a 2000. Este inquérito é realizado trimestralmente, tendo sido adoptado o esquema de rotação de 1/6 da amostra em cada trimestre. Deste modo, é natural que as amostras em cada dois anos consecutivos sejam parcialmente idênticas. Restringiu-se a *amostra* para a estimação de equações salariais a estes anos pois a utilização de inquéritos anteriores colocava problemas de difícil resolução ao nível da estimação, já que a variável rendimento está dividida por categorias. Assim, de uma base com cerca de 50 000 indivíduos para cada ano, foram seleccionados todos os trabalhadores por conta de outrem cujo salário horário reportado fosse superior a 200 escudos. A não inclusão de indivíduos com salário inferior a esse nível justifica-se pelo facto de os valores em causa implicarem um salário mensal muito abaixo do salário mínimo nacional sendo os dados, portanto, pouco fiáveis. Por outro lado, os trabalhadores por conta própria não são incluídos pois existem vários factores não observáveis, tais como a habilidade ou o nível de riqueza herdado, que influenciam decisivamente os rendimentos deste grupo e que podem tornar irrelevantes variáveis como o nível de escolaridade atingida ou a experiência.

Apesar das restrições impostas, a base de dados constituída tem uma dimensão considerável. A tabela seguinte mostra a distribuição dos indivíduos por regiões e o número total de observações.

TABELA 1

	HOMENS			MULHERES		
	1998	1999	2000	1998	1999	2000
NORTE	2515	2372	2224	1956	1892	1804
CENTRO	785	764	660	702	635	560
LVT	1647	1417	1388	1496	1325	1325
ALENTEJO	668	619	550	486	466	449
ALGARVE	717	721	684	628	625	622
MADEIRA	692	707	716	438	439	458
AÇORES	523	444	411	368	346	319
Total	7547	7044	6633	6074	5728	5537

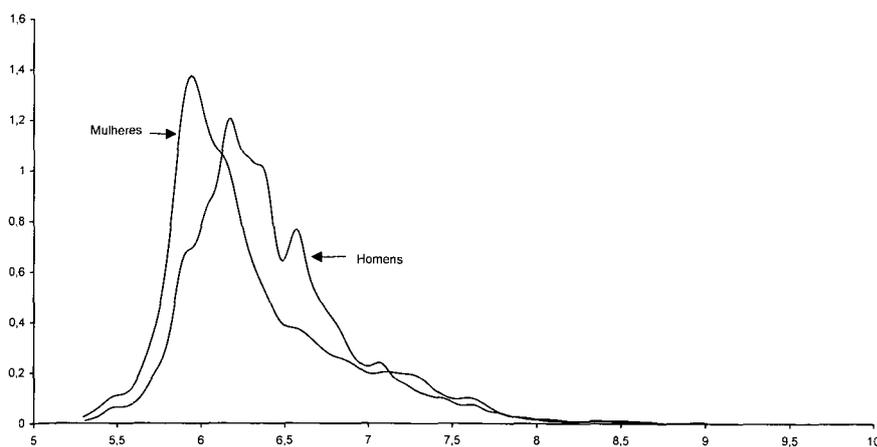
Seguidamente, apresenta-se uma tabela resumo de estatísticas descritivas dos salários horários. Como seria previsível, os homens ganham, em média, mais do que as mulheres. Por outro lado, a variabilidade do salário é menor nos homens, o que sugere que as mulheres, além de ganharem menos, estão sujeitas a uma maior desigualdade salarial. Este facto pode ser confirmado pelos valores do *inter-quartile range* (IQR – diferença entre o 3º e o 1º quartis) e da tabela 3 apresentada adiante.

TABELA 2

	Salário Horário								(unid.: escudos)
	média	d. p.	mín	1º q	med	3º q	máx	iqr	
HOMENS 1998	644	568	204	405	503	694	19048	290	
1999	649	532	204	417	515	714	25740	298	
2000	687	466	202	446	549	756	7937	310	
MULHERES 1998	623	567	201	351	423	662	15476	310	
1999	619	713	200	357	435	672	41005	314	
2000	656	556	200	381	476	714	8242	333	

A tabela 2 oculta características que uma análise gráfica pode revelar. Para aprofundar o estudo da distribuição salarial não condicional, as densidades salariais foram estimadas pelo método kernel gaussiano. A janela (*bandwidth*) é a ideal tal como definida por Silverman (1986). Foi usado um factor de alisamento de 80% da janela, obtendo-se as densidades da figura 1.

FIGURA 1



Densidades do logaritmo do salário horário

As densidades estimadas têm, no geral, a forma esperada. As diferenças anuais dentro de cada grupo não são significativas, pelo que se apresentam apenas as curvas para o ano 2000. O que é surpreendente é o facto de as mulheres, no topo da distribuição, ganharem mais do que os homens ou, alternativamente, para o mesmo nível de salários, a proporção de mulheres é superior a 50%.

Este facto pode ser explicado por um conjunto diverso de factores, mas é, principalmente, reflexo de que existe menos discriminação sexual nas remunerações dos trabalhadores por conta de outrem. Após uma análise da variável educação, as causas para esta disparidade tornam-se evidentes. No último quintil da distribuição salarial, as mulheres revelam ter um nível de educação bem mais elevado do que os homens. Assim, 25, 50 e 75% das mulheres têm, respectivamente, 17, 15 e 12 ou mais anos de escolaridade. No caso dos homens, estes números baixam para 15, 9 e 6 ou mais anos.

Finalmente, apresenta-se uma tabela com as medidas de desigualdade mais habituais. Uma análise da evolução da desigualdade é possível, mas é complicado ter uma ideia da tendência desta série. Quer o índice de Gini quer o quociente dos decis extremos indicam uma evolução no sentido da redução da desigualdade salarial, embora não se revelem concordantes quando cada grupo é analisado separadamente⁷⁵.

TABELA 3

	Índice de Gini			decil 9 / decil 1		
	homens	mulheres	total	homens	mulheres	total
1998	0.302	0.348	0.324	3.22	4.00	3.64
1999	0.287	0.331	0.309	3.05	3.75	3.34
2000	0.281	0.333	0.307	3.15	3.62	3.33

Numa nota final, e antes de avançar para a construção do modelo teórico, atente-se nas taxas de actividade baseadas na amostra original do Inquérito ao Emprego – 68% para as mulheres contra 84% para os homens⁷⁶. Com uma taxa de participação feminina tão elevada, não seriam de esperar problemas graves ao nível da selectividade da amostra, por ocasião de uma eventual estimação do modelo teórico. No entanto, esta questão não se chega a colocar pois, como é explicado adiante, a estimação do modelo é feita separadamente para homens e para mulheres.

⁷⁵ É preciso não esquecer que estas medidas são específicas à amostra em estudo e que, portanto, não é representativa do país.

⁷⁶ A taxa de actividade utilizada é igual a total de activos / pessoas com idade compreendida entre os 15 e os 64 anos.

3. REGRESSÃO DE QUANTIS: TEORIA E PRÁTICA

“Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture of a single distribution. (...) what the regression curve does is give a grand summary for the averages of the distributions correspondingly to the set of x 's. We could go one step further and compute several regression curves corresponding to the various percentage points of the distribution and thus get a more complete description of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture” (Mosteller e Tukey, 1977, 266)

A regressão clássica estima os efeitos de variáveis exógenas na *média* da distribuição condicional da variável dependente. Isto pode ser uma simplificação forçada, já que as variáveis exógenas podem não só determinar o comportamento da variável dependente na média mas também exercer influência em parâmetros de interesse de toda a distribuição condicional da mesma variável. De qualquer modo, em caso de homocedasticidade (que se pode definir ‘grosseiramente’ como a igualdade das variâncias de cada distribuição condicional individual), as linhas de regressão serão paralelas entre si – ver figura A1, no anexo. Não faz sentido, neste caso, estimar mais do que uma regressão, seja ela da média condicional ou de um qualquer quantil já que as linhas de regressão diferem apenas no termo independente. No entanto, a existência de heterocedasticidade é frequente em modelos com dados seccionais. Nesta situação, as linhas de regressão deixam de ser paralelas e aqui reside o interesse em estimar o modelo por regressão de quantis. As mesmas variáveis independentes do modelo influenciam de forma diferenciada cada quantil da variável dependente.

Tome-se um simples exemplo, discutido em Machado e Mata (2001). Suponha-se que se estima uma simples equação de salários em que se considera a informação de ambos os sexos conjuntamente. Os regressores são apenas uma constante e uma *dummy* tomando o valor 1 se a observação em causa corresponder a uma mulher. A regressão na média dar-nos-ia apenas informação sobre a diferença salarial média entre os sexos. Assim, imagine-se que, em média, as mulheres ganham menos 20% que os homens. A grande vantagem da regressão de quantis é que nos permite tirar conclusões semelhantes para qualquer quantil ao longo da distribuição. É natural que a diferença salarial no primeiro decil da distribuição salarial seja menos acentuada e no último decil mais acentuada. Ou seja, é possível que as mulheres que se situam no primeiro decil da distribuição salarial ganhem apenas menos 10% que os homens nas mesmas condições mas que, de forma semelhante, no último decil, a diferença seja já de 30%. Desta forma, se se perseguir o objectivo de não discriminação salarial é necessário um maior esforço junto dos trabalhadores melhor remunerados. Este exercício revela que a diferença salarial entre sexos vai mais além do que o simples facto de que os homens ganham, em média, mais 25% que as mulheres.

De seguida, introduz-se a técnica de regressão de quantis. A variável dependente é o logaritmo do salário horário e x representa o vector de regressores. O índice i diz respeito aos indivíduos e β é o vector de parâmetros a estimar. Deste modo, o modelo de regressão pode ser escrito como:

$$\ln sh_i = x_i' \beta_0 + u_i \quad \text{com} \quad \text{Quant}_\theta(\ln sh_i | x_i) = x_i' \beta_\theta$$

$\text{Quant}_\theta(\ln sh_i | x_i)$ corresponde ao quantil θ de $\ln sh_i$ condicional em x . A regressão do quantil θ , para $0 < \theta < 1$, é definida como sendo a solução do seguinte problema de minimização

$$\min \left\{ \sum_{i: \ln w_i \geq x_i' \beta} \theta |\ln s_i - x_i' \beta_\theta| + \sum_{i: \ln w_i \leq x_i' \beta} (1 - \theta) |\ln sh_i - x_i' \beta_\theta| \right\}$$

que, no fundo, é o mesmo que minimizar em β

$$\sum_{i=1}^n \rho_\theta(\ln sh_i - x_i' \beta)$$

onde a função ρ , denominada de *check function*, é definida por

$$\rho_\theta(u) = \begin{cases} \theta u & \text{para } u \geq 0 \\ (\theta - 1)u & \text{para } u < 0 \end{cases}$$

O problema não pode ser tratado analiticamente mas, em contrapartida, pode ser resolvido por métodos de programação linear relativamente simples, como o demonstram Koenker e Bassett (1978). O estimador LAD (*least absolute deviation*) é um caso particular e é obtido quando se tem $\theta = 0.5$. Como os erros serão, por hipótese, sempre heterocedásticos, as estimativas para o desvio padrão de cada coeficiente têm de ser obtidas através de métodos *bootstrapping*.

Para testar a forma funcional do modelo, utilizou-se o teste RESET. A sua aplicação em regressão de quantis é idêntica à regressão linear e tem exactamente a mesma interpretação. Se a forma funcional não é adequada dadas as variáveis intervenientes, então é possível que o quadrado ou uma ordem superior de $x'\beta$ melhore substancialmente a explicação do comportamento da variável dependente. A estatística teste é um simples teste t (ou F) à(s) variável(eis) adicional(ais). Obviamente, os desvios-padrão têm de ser, também, estimados por *bootstrap*.

Por outro lado, para testar se existe realmente heterocedasticidade intrínseca ao modelo foi feito o teste sugerido por Machado e Santos Silva (2000). Na verdade, se os coeficientes para as mesmas variáveis diferirem muito de quantil para quantil então é de esperar que o modelo seja mesmo heterocedástico. Remete-se o leitor mais interessado na origem do teste e sua implementação prática para o artigo citado.

Finalmente, há que esclarecer que a estimação foi feita separadamente para homens e mulheres. A estimação conjunta pode apresentar problemas de selectividade da amostra. Em causa está o facto de que muitos indivíduos (principalmente mulheres) se auto-excluírem do mercado de trabalho visto que ou não lhes compensa ou não pretendem trabalhar. Seria necessário ter esse facto em conta na estimação de uma

equação conjunta. Com uma taxa de participação feminina tão elevada, optou-se pela estimação separada condicionando-se a análise dos resultados apenas às pessoas que trabalham por conta de outrem.

4. RESULTADOS EMPÍRICOS

Nesta secção discute-se os resultados da análise do impacto das variáveis independentes na distribuição do logaritmo dos salários horários. Uma versão da equação Minceriana foi adoptada neste artigo, tendo como forma funcional

$$\ln sh_i = \alpha + \sum_{i=1}^9 \beta_i educ_i + \sum_{i=10}^{18} \beta_i edexp_{i-9} + \beta_{19} exper_i + \beta_{20} exper_i^2 + \beta_{21} ten_i + \beta_{22} ten_i^2 + \sum_{i=23}^{28} \beta_i reg_{i-22} + u_i \quad (1)$$

Os regressores considerados são 10 *dummies* para a educação (omite-se a categoria das pessoas sem qualquer escolaridade), variáveis de interacção entre o nível de educação e o número de anos de experiência (*edexp*), a experiência e a antiguidade e os respectivos quadrados e, finalmente, *dummies* para as regiões NUTS II, tendo sido excluída a região de Lisboa e Vale do Tejo.

Esta não é a forma funcional mais habitual na literatura, que trata, muitas vezes, a variável educação como variável contínua. Em vez de *dummies*, é utilizada uma variável definida pelo número mínimo de anos necessário para completar o grau académico que o indivíduo revela ter. Trata-se de uma hipótese bastante forte pois assume uma relação de linearidade entre os diferentes graus académicos. Assim, cada ano de educação teria o mesmo impacto quer uma pessoa atingisse o mais baixo nível de escolarização quer atingisse o mais alto. Para evitar este tipo de suposições, preferiu-se a especificação já apresentada em que se privilegia o impacto do grau académico atingido. Por seu turno, a variável experiência é normalmente uma variável *proxy* definida por (idade – educação – 6). Como a base de dados do Inquérito ao Emprego tem informação sobre a data de início do primeiro emprego de cada pessoa, foi adoptada uma definição que se aproxima mais da verdadeira definição de experiência, atribuindo à variável em causa o valor (ano actual – ano de início do primeiro emprego). Os outros regressores são comumente adoptados na literatura.

Os resultados, devido à sua extensão, são apresentados em anexo⁷⁷. A dimensão da amostra foi referida em secção anterior. Para calcular os erros padrão de

⁷⁷ Ainda assim, só se apresentam parcialmente. Todas as Tabelas estarão disponíveis no Infoline (www.ine.pt).

cada estimador utilizou-se o método *bootstrap* em que se fizeram 800 réplicas de cada sub-amostra, de dimensão igual à amostra original. O número de réplicas foi escolhido de forma a respeitar as condições propostas por Andrews e Buchinsky (2000) para que o *percentage deviation bound* (pdb) fosse, no máximo, de 10%. Assim, as estimativas dos desvios-padrão não se afastam mais de 10% dos seus valores “correctos”⁷⁸.

Primeiro que tudo, os resultados dos testes de especificação dão boas indicações. A forma funcional, dadas as variáveis utilizadas, parece bem especificada. O RESET, na média, aceita a hipótese nula de boa especificação em 1998 e 2000 a um nível de significância de 5%, para as equações dos homens (no caso das mulheres, apenas para o ano 2000 a hipótese nula é aceite e para um nível de significância de 1%). Em relação às regressões de quantis, como é necessário estimar o desvio-padrão do coeficiente utilizado pelo teste por *bootstrap*, o RESET só foi calculado para a mediana. Nestes casos, a hipótese de boa especificação foi quase sempre aceite a um nível de 5%. As excepções são o ano de 1998, no caso dos mulheres, e o ano de 2000, no caso dos homens, em que H_0 é aceite apenas a um nível de significância de 1%.

Os testes de heterocedasticidade feitos para a regressão da mediana não deixam quaisquer dúvidas. Os dados são mesmo heterocedásticos, o que é um resultado importante pois, caso contrário, não se justificaria a análise de equações salariais por regressão de quantis. Com resultados tão conclusivos, considerou-se desnecessário o cálculo do valor da estatística teste para mais decis. Finalmente, relativamente à qualidade de ajustamento, medida tal como sugerido por Koenker e Machado (1999), há que dizer que, aparentemente, o modelo se ajusta relativamente bem à realidade. Como esta medida de ajustamento ainda foi pouco utilizada é difícil ajuizar os valores obtidos. Caso sejam comparados com os valores para R^2 que se obtêm habitualmente em estudos micro-económicos, então a especificação adoptada pode ser considerada uma boa aproximação à realidade.

As tabelas B permitem analisar o impacto das variáveis relativas à antiguidade e região onde se trabalha. A antiguidade tem um impacto positivo na formação de salários mas que se desvanece ao longo do tempo. Esse impacto parece ser uniforme ao longo da distribuição de salários. Os coeficientes são de ordem relativamente pequena mas estão bem definidos. Por seu turno, a variável relativa à região onde se trabalha não é sempre significativa. Algumas vezes, os coeficientes não são bem definidos mas, mesmo assim, há padrões que emergem da análise cuidada dos resultados. Assim, trabalhar na região de Lisboa e Vale do Tejo (LVT) é, em termos salariais, mais vantajoso do que em qualquer outra região. Pelo contrário, trabalhar na região Norte, ao invés daquilo que normalmente se pensa, implica receber em média menos 17%, no caso dos homens, ou 5%, no caso das mulheres, que em LVT⁷⁹.

⁷⁸ Buchinsky (1995) mostrou, por simulação de Monte Carlo, que o melhor método para estimar a matriz de variâncias-covariâncias é o método *bootstrap* dos pares em detrimento do método kernel, já provado ser consistente.

⁷⁹ Estes valores são relativos ao ano 2000. Para o caso dos homens chega mesmo a verificar-se uma tendência crescente nos últimos anos. Em 98, este coeficiente era de cerca de 11.5%

Ao contrário da antiguidade, o impacto de viver na região Norte varia ao longo da curva de distribuição de salários. Assim, para as pessoas que se situam nos decis inferiores da distribuição, a diferença salarial é menos expressiva. As pessoas que se situam nos decis superiores é que sentem mais as diferenças a nível salarial. No entanto, esta situação não significa que haja declarada disparidade no nível de vida das regiões. Segundo o estudo de Bago d' Uva e Parente (2001), a região Norte é uma das regiões em que o custo de vida é mais baixo. Por outro lado, é preciso não esquecer que a amostra utilizada é condicionada aos trabalhadores por conta de outrem. Seria natural, havendo dados fiáveis, que a inclusão de outros trabalhadores na amostra levasse a resultados diferentes, encurtando esta diferença.

Para as outras regiões, as conclusões não diferem muito das da região Norte. Ganha-se menos que em LVT mas as diferenças são menores. Na região do Alentejo, os coeficientes estimados são, de forma sistemática, estatisticamente nulos. Para a região Centro, essas diferenças não são influenciadas pela posição do trabalhador na distribuição salarial.

O resultado mais surpreendente acaba por ser a evolução dos próprios salários na região LVT. De 1998 a 2000, os trabalhadores desta região alargaram a diferença salarial para os trabalhadores do resto do país, *ceteris paribus* (controlando-se as restantes características dos indivíduos).

A análise da rendibilidade da educação e da experiência deve ser feita de forma mais cuidada. Primeiro, porque a interação entre estas duas variáveis impede a leitura directa das tabelas de resultados das regressões. Segundo, porque a análise da rendibilidade da educação é o principal objectivo deste estudo.

Comece-se por explicitar o significado da expressão taxa de rendibilidade. Esta refere-se tão somente ao efeito da variável educação no quantil condicional do logaritmo do salário horário, não se procurando uma interpretação causa-efeito.

Assim, para a experiência, tal como acontece com a antiguidade, a taxa de rendibilidade é positiva mas decrescente no tempo. No entanto, estes só são estatisticamente diferentes de zero para os homens. Além disso, o impacto é notoriamente baixo rondando frequentemente o valor de 1% - ver tabelas D. Também neste caso, não existem diferenças claras entre quantis.

Por seu turno, com uma breve manipulação dos resultados de base, pode obter-se a taxa de rendibilidade directa da educação. Para tal basta fazer

$$\text{Taxa rendibilidade} = \frac{\partial Q_0(\ln sh | x)}{\partial educ} \times 100$$

Os resultados deste exercício são apresentados em anexo, tabelas C. Uma conclusão óbvia e que confirma o ponto de vista gerado pelo senso comum é a de que

quanto maior o nível de educação, maiores são os ganhos salariais. Contudo, a análise destes resultados requer alguma cautela. Estes são condicionais às variáveis explicativas. Assim, o ganho salarial para quem se situa no *primeiro decil da distribuição salarial dos trabalhadores licenciados* é superior ao ganho para quem se situa no *primeiro decil da distribuição salarial dos trabalhadores com o 4º ano de escolaridade*. Estes trabalhadores não se situam necessariamente no mesmo decil da distribuição salarial incondicional. Na verdade, é muito provável que não façam parte do mesmo decil – ver tabelas A.

A leitura destas tabelas não é imediata quando se pretende responder à questão introduzida na secção introdutória. Será que um maior investimento por parte das entidades responsáveis para elevar o nível de educação da população conduz a uma redução do índice de desigualdade salarial? Para responder, construíram-se duas novas tabelas – as tabelas 4 e 5 – que apresentam a taxa de rentabilidade de um ano adicional de escolaridade. Ou seja

$$TR \text{ ano adicional} = \left[\left(1 + TR_{gact} - TR_{gant} \right)^{\frac{1}{\text{anos } gact - \text{anos } gant}} - 1 \right] \times 100$$

em que *TR* é o acrónimo para taxa de rentabilidade, *gact* denomina o nível académico realmente atingido e *gant* representa o nível académico imediatamente anterior. Para interpretar estes resultados, é necessário notar que a taxa de rentabilidade se traduz num ganho (ou perda, caso o sinal seja negativo) em relação ao que se pode obter tendo o grau académico imediatamente inferior. Tome-se como *benchmark* o ano 2000. Para uma pessoa que possui o 4º ano de escolaridade e se situa no primeiro decil da distribuição salarial, o ganho médio em ter estudado mais um ano iria rondar os 13,7% se essa pessoa tiver 5 anos de experiência ou 14,25% se tiver 30 anos de experiência. No entanto, se essa pessoa se situar no último decil da distribuição, os ganhos seriam substancialmente menores (2,35% e 8%, respectivamente⁸⁰). Neste caso, a educação contribui para diminuir a desigualdade salarial.

Se, por outro lado, se considerar pessoas com o 9º ano de escolaridade até pessoas licenciadas, verificar-se-á que o fenómeno não se repete. Pelo contrário, nestas situações, as pessoas que se situam em decis inferiores da distribuição salarial teriam menos vantagens em termos de ganhos médios em relação às pessoas que se situam mais acima na distribuição. Para estes níveis de escolaridade, mais educação significa maior desigualdade na distribuição do rendimento.

⁸⁰ Estes valores são obtidos fazendo uma média ponderada das taxas constantes das tabelas 4 e 5.

TABELA 4

<i>Taxa de rentabilidade de um ano adicional de escolaridade (homens)</i>						
1998						
Educação (em anos)	5 anos de experiência			30 anos de experiência		
	decil 1	mediana	decil 9	decil 1	mediana	decil 9
3	4.82	7.96	1.43	2.41	3.59	0.64
4	2.03	-0.74	5.14	9.20	9.35	14.14
6	1.06	0.57	-3.23	3.80	5.56	7.14
9	1.72	3.28	6.57	5.89	8.69	11.01
12	4.63	6.18	10.35	8.03	9.55	10.20
15	15.23	16.32	11.79	10.53	8.56	3.56
17	9.15	6.48	14.84	12.23	14.06	27.31
19	18.36	8.81	15.45	-1.41	4.62	21.23
22	-14.36	-0.21	-6.98	12.42	2.92	-53.30
1999						
3	2.49	3.55	1.39	4.10	3.53	3.04
4	11.48	7.22	4.79	6.78	9.76	11.68
6	2.12	1.42	1.14	4.45	5.08	7.97
9	3.10	3.30	7.31	2.81	7.87	12.38
12	3.96	6.72	8.80	7.12	8.96	5.53
15	12.64	15.70	11.10	14.02	10.82	11.65
17	10.83	5.10	10.95	9.13	9.20	10.01
19	18.60	7.73	0.59	4.75	17.82	26.94
22	-26.10	-9.00	15.03	8.37	-3.08	-13.92
2000						
3	-0.30	0.23	2.18	1.06	0.71	1.78
4	14.47	5.75	3.12	11.94	9.16	9.19
6	2.50	0.55	0.99	4.31	5.91	6.37
9	2.97	3.50	4.84	2.89	6.39	10.57
12	2.96	5.60	8.21	6.94	7.78	6.25
15	12.48	14.46	14.46	11.16	10.56	6.18
17	6.29	9.09	5.74	12.63	13.35	19.17
19	12.06	10.58	7.45	6.85	10.07	17.19
22	-7.87	-10.58	8.16	-14.28	1.26	-9.33

Daqui se pode inferir que não há necessariamente benefícios, em termos de redução da desigualdade salarial, em fazer um maior investimento na educação das

peças. Este resultado acaba por não ser tão surpreendente como poderia parecer numa primeira instância. Ao elevar o nível de educação de uma sociedade, está-se implicitamente a contribuir para uma abertura do leque salarial pois a dispersão salarial é mais acentuada dentro de níveis de escolaridade mais altos. É sabido que uma pessoa licenciada tanto pode ganhar cerca de 200 mil escudos como 2 milhões. Por seu turno, já não é natural que pessoas com o 4º ano de escolaridade possam auferir salários tão díspares.

TABELA 5

<i>Taxa de rentabilidade de um ano adicional de escolaridade (mulheres)</i>						
1998						
Educação (em anos)	5 anos de experiência			30 anos de experiência		
	decil 1	mediana	decil 9	decil 1	mediana	decil 9
3	-6.83	3.51	-2.72	-0.63	3.07	0.36
4	25.17	0.93	10.11	10.11	4.22	3.23
6	1.20	-0.19	1.98	3.43	6.56	8.62
9	2.20	3.31	6.05	6.09	12.76	15.58
12	4.39	6.25	8.74	6.29	9.00	9.71
15	13.62	17.77	16.20	17.84	12.11	3.50
17	10.72	9.23	14.44	11.74	10.98	20.52
19	-4.60	6.26	9.60	13.88	10.46	16.25
22	40.73	25.11	9.62	-7.46	-12.72	-43.72
1999						
3	5.72	-1.89	-3.96	0.91	0.01	0.16
4	21.06	6.17	7.53	18.79	8.85	10.36
6	2.29	0.87	0.45	3.09	5.59	7.37
9	2.00	3.38	7.35	4.57	12.29	15.39
12	2.68	4.63	8.71	10.06	9.05	5.49
15	12.56	16.88	16.37	18.50	13.37	11.37
17	13.24	11.88	11.34	4.46	4.19	7.63
19	-20.09	1.64	-9.68	-8.13	12.15	34.57
22	22.16	-2.34	-12.57	5.27	-6.18	-21.08
2000						
3	-1.64	3.43	0.97	-4.49	0.72	0.16
4	12.70	4.90	1.25	17.54	9.13	6.28
6	3.15	-0.03	0.25	3.64	3.80	9.97
9	1.20	2.93	7.44	5.60	12.24	12.29
12	3.99	6.05	8.35	4.33	7.07	6.08
15	13.79	17.27	14.04	15.02	14.60	12.02
17	6.73	4.25	7.60	20.76	10.82	21.70
19	17.87	9.01	14.68	6.36	5.83	-7.51
22	-16.89	12.22	4.40	-3.57	-9.33	-17.69

É ainda necessário fazer alguns reparos aos valores apresentados nas tabelas anteriores. Primeiro, visto que as variáveis educação e experiência são interdependentes, não é possível apresentar uma tabela com a taxa de rendibilidade da educação sem ter em conta a experiência da pessoa em causa. Nestas tabelas apresentam-se taxas para pessoas recém-empregadas e para pessoas com elevada experiência profissional⁸¹. Em segundo lugar, é necessária grande cautela para analisar os números para pessoas que possuam o grau de mestre ou doutor. Essa análise é bastante condicionada pelo facto de o número de observações dentro de cada um destes grupos ser extremamente reduzido. No entanto, de um modo geral, há indicações de que os benefícios salariais de ter um doutoramento não são superiores aos de ter um mestrado, facto que pode ser uma consequência do tipo de profissões desempenhadas pelos trabalhadores de cada grupo. Uma pessoa que possua um MBA tem grandes possibilidades de auferir mais que uma pessoa doutorada que é, por norma, professor universitário. Em terceiro lugar, é de frisar que as taxas de rendibilidade variam muito de quantil para quantil, o que favorece uma abordagem de regressão de quantis em detrimento da habitual regressão clássica na média, mesmo que os resultados sejam mais fáceis de interpretar. Finalmente, as taxas têm variado muito de ano para ano, o que torna a análise de resultados muito complicada. Seria ideal repetir este exercício apenas para anos mais espaçados temporalmente.

Para complementar as conclusões das tabelas anteriores, estimaram-se algumas regressões auxiliares. Assumiu-se, neste caso, uma forma reduzida da equação (1) amplamente utilizada na literatura. Desta forma, o logaritmo do salário horário apenas depende da educação (variável agora assumida contínua), da experiência e respectivo quadrado. Os resultados não são apresentados mas serão disponibilizados pelo autor a eventuais interessados. A vantagem desta forma funcional é que permite uma análise directa da taxa de rendibilidade da educação que não é condicionada a um nível académico específico.

Os resultados são inequívocos e coadunam-se com os obtidos em estudos já realizados anteriormente para o caso português (veja-se, por exemplo, Machado e Mata, 2001). A taxa de rendibilidade é sempre superior no decil imediatamente acima. Significa isto que a rendibilidade da educação é mais que proporcional nos decis superiores. Para quem se situa no decil um, a variável educação exhibe um coeficiente de cerca de 6%, mas para quem se situa no último decil, a mesma variável já exhibe um coeficiente a rondar os 10%.

De qualquer modo, sob esta forma funcional, há que ter em atenção duas possibilidades diferentes do modo como a variável educação influencia a desigualdade salarial: desigualdade intra- e inter-grupo. Desigualdade intra-grupo apresenta-se mais como uma questão de dispersão salarial dentro de um determinado escalão de escolaridade enquanto desigualdade inter-grupo diz respeito a diferenças nos salários médios em cada grupo. O que acontece, neste ensaio, é que a taxa de rendibilidade da educação cresce à medida que se avança na distribuição condicional dos salários. Em termos gráficos, esta questão pode ser visualizada da seguinte forma. À medida que se consideram níveis de educação cada vez mais elevados, a distribuição condicional dos salários torna-se *mais dispersa* (caso assim não fosse, o coeficiente estimado para o

⁸¹ No apêndice considera-se um caso adicional – pessoas com 15 anos de experiência – e apresentam-se valores para todos os decis – tabelas E.

último decil seria também 6% ou inferior) – ver exemplo na figura A2. Por outro lado, o efeito inter-grupo deve ser sempre positivo pois até hoje esta taxa sempre foi positiva.

Assim, o efeito intra-grupo (mantendo a distribuição do nível de escolaridade da população constante) ou reforça o efeito inter-grupo ou o enfraquece. Deste modo, taxas de rendibilidade crescentes com os quantis implicam que mais educação conduz realmente a uma maior desigualdade – o caso presente. Por outro lado, taxas decrescentes são apenas uma condição necessária mas não suficiente para a redução de desigualdade. Se o efeito inter-grupo for suficientemente forte, então mais educação continua a implicar maior desigualdade apesar das taxas de rendibilidade decrescentes.

A discussão decorrente da especificação de uma forma funcional diferente para as equações salariais conduz a uma conclusão em todo idêntica à anterior mas onde os resultados são mais fáceis de analisar embora a própria definição da variável educação não seja tão coerente. Ainda assim, é deveras interessante que este pequeno exercício corrobore as primeiras ilações.

5. CONCLUSÃO

A caracterização da distribuição condicional dos salários exclusivamente pela sua média condicional constitui uma séria limitação no estudo de populações heterogéneas. Na verdade, a média condicional pode ocultar características importantes dessa mesma distribuição. Actualmente, o conhecimento da influência, em média, de uma dada variável numa outra já não é satisfatório. É preciso conhecer mais.

Neste artigo, estimaram-se equações salariais do tipo Minceriano pelo método habitual dos mínimos quadrados e utilizando a técnica de regressão de quantis. Desta forma, obtém-se um retrato mais fiel da distribuição condicional dos salários. Especificamente, a rendibilidade da educação e a influência de variáveis como a experiência, antiguidade e região onde se trabalha foram examinadas para todos os decis relevantes.

Dos resultados obtidos há alguns que merecem uma pequena referência. Primeiro, trabalhar na região de Lisboa e Vale do Tejo revela ser vantajoso em termos salariais. Estes benefícios são, geralmente, mais acentuados para quem se situa no extremo direito da distribuição. Por outro lado, as variáveis experiência e antiguidade, apesar de bem definidas, têm pouca influência na formação dos salários. Ainda assim, o impacto é positivo mas desvanece-se ao longo do tempo. Não há diferenças assinaláveis entre os diversos decis. Finalmente, mostrou-se que as taxas de rendibilidade da educação variam bastante de decil para decil. É de prever um aumento da desigualdade salarial caso o nível educacional da população aumente significativamente. A variável educação contribui para uma contracção da desigualdade se se tiver 6 ou menos anos de escolaridade visto que o seu impacto é menor nos decis superiores. Pelo contrário, dado que se tem pelo menos 9 anos de

escolaridade, o impacto da variável é superior para os últimos decis. Contudo, não se deve confundir um aumento da desigualdade medida em termos salariais com um aumento medido em termos sociais. No presente estudo, não se faz qualquer referência a essa possibilidade que não é, portanto, comentada.

Uma questão fundamental é a da complementaridade das variáveis educação e habilidade. O coeficiente da variável educação pode ser sobrestimado se, por exemplo, indivíduos mais hábeis beneficiarem proporcionalmente mais de um investimento adicional em educação. Apesar desta questão não poder ser descurada, este é um problema inerente às equações salariais de difícil resolução.

Não se perseguiu, aqui, o objectivo de explicar as causas para o crescimento da desigualdade salarial nos últimos anos. Preferiu-se dar uma imagem da distribuição condicional dos salários para os anos de 1998 a 2000. Dada a proximidade temporal dos anos em estudo, a análise da evolução desta mesma distribuição é relegada para segundo plano, já que as diferenças encontradas em termos anuais são, principalmente, devidas à amostragem. De qualquer modo, parte da literatura actual tem atribuído estas alterações na estrutura salarial quer a evoluções tecnológicas e institucionais quer ao menor envolvimento dos sindicatos ou do governo.

Numa nota final, existem dois grandes temas que podem constituir a base para pesquisa futura em regressões de quantis e equações salariais. Estes são a estimação conjunta para homens e mulheres utilizando uma correcção para a selectividade da amostra e, num campo distinto, a decomposição contra-factual que permite responder a questões como o que aconteceria hoje se alguns aspectos da população em estudo se mantivessem constantes nos anos de referência.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDREWS, D. e BUCHINSKY, M. (2000), "A Three-Step Method for Choosing the Number of Bootstrap Repetitions", *Econometrica*, 68 (1), 23-51.
- BAGO D' UVA, T. e PARENTE, P. (2001), "Índices de custo de vida: um estudo sobre a realidade portuguesa", *Revista de Estatística INE*, 1º Quadrimestre de 2001.
- BUCHINSKY, M. (1994), "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression", *Econometrica*, 62 (2), 405-458.
- BUCHINSKY, M. (1995), "Estimating the asymptotic covariance matrix for quantile regression models. A Monte Carlo study", *Journal of Econometrics*, 68 (2), 303-338.
- BUCHINSKY, M. (1998), "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research", *The Journal of Human Resources*, 33; 88-126.
- BUCHINSKY, M. (2001), "Quantile Regression with Sample Selection: Estimating women's return to education in the U.S.", *Empirical Economics*, 26 (1), 87-113.

- GARCÍA, J.; HERNÁNDEZ, P. J. e LÓPEZ-NICOLÁS, A. (2001), "How wide is the gap? An investigation of gender wage differences using quantile regression", *Empirical Economics*, 26, 149-167.
- KOENKER, R. e BASSETT, G. (1978), "Regression Quantiles", *Econometrica*, 46 (1), 33-50.
- KOENKER, R. e BASSETT, G. (1982), "Robust Tests for Heteroscedasticity Based on Regression Quantiles", *Econometrica*, 50 (1), 43-61.
- KOENKER, R. e MACHADO, J. (1999), "Goodness of Fit and Related Inference Processes for Quantile Regression", *Journal of American Statistical Association*, 94, 1296-1310.
- MACHADO, J. e MATA, J. (2000), "Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression", *não publicado*.
- MACHADO, J. e MATA, J. (2001), "Earning Functions in Portugal 1982-1994: Evidence from Quantile Regressions", *Empirical Economics*, 26 (1), 115-134.
- MACHADO, J. e SANTOS SILVA, J. (2000), "Glejser's Test Revisited", *Journal of Econometrics*, 97 (1), 189-202.
- MOSTELLER, F. e TUKEY, J. (1977), *Data analysis and Regression*, Addison-Wesley, Reading, MA
- SILVERMAN, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, Londres.

APÊNDICE

FIGURA A1
Distribuição condicional individual dos salários
Caso homocedástico

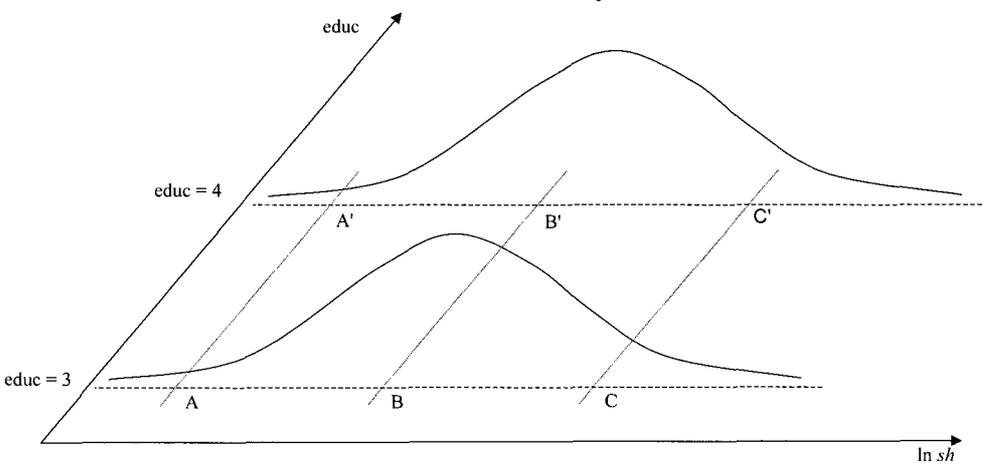


FIGURA A2
Distribuição condicional individual dos salários
Caso heterocedástico

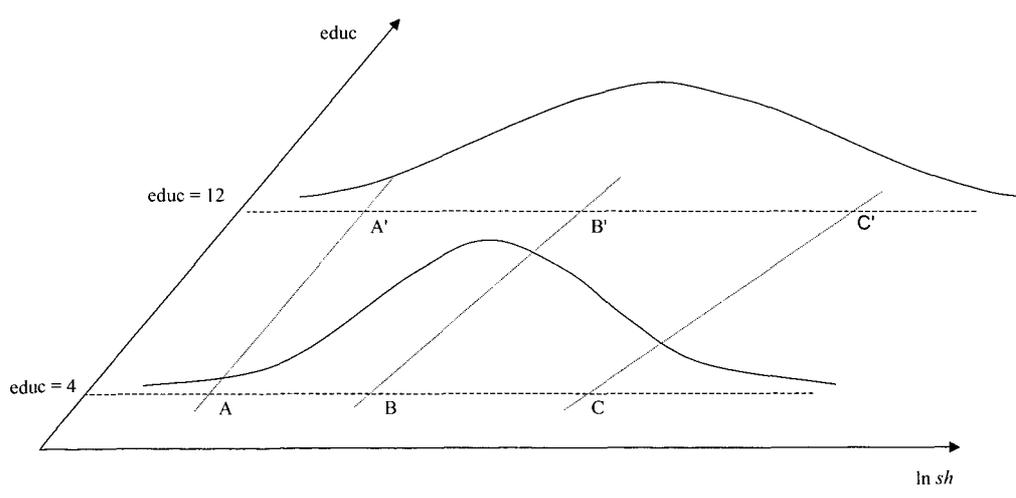


TABELA A1 - Distribuição dos salários horários segundo o nível de educação

HOMENS											
1998											
	decil 1	decil 2	decil 3	decil 4	decil 5	decil 6	decil 7	decil 8	decil 9	decil 10	total (%)
educ = 0	55	34	36	28	21	13	13	5	6	1	2.81
educ = 3	46	50	50	36	36	23	27	13	5	6	3.87
educ = 4	270	290	306	397	362	367	371	330	234	63	39.62
educ = 6	237	237	236	177	191	205	159	146	105	34	22.88
educ = 9	115	93	100	78	95	96	110	139	189	140	15.30
educ = 12	31	51	23	35	48	48	65	103	152	160	9.49
educ = 15	0	0	1	2	1	1	3	5	21	59	1.23
educ = 17	0	0	3	1	1	2	6	13	42	261	4.36
educ = 19	0	0	0	0	0	0	0	1	0	19	0.27
educ = 22	0	0	0	0	0	0	0	0	1	12	0.17

1999											
	decil 1	decil 2	decil 3	decil 4	decil 5	decil 6	decil 7	decil 8	decil 9	decil 10	total (%)
educ = 0	63	30	24	13	16	15	10	9	2	0	2.58
educ = 3	42	47	56	39	44	28	27	19	7	0	4.39
educ = 4	250	273	307	335	349	326	325	280	201	66	38.50
educ = 6	219	229	191	211	176	184	174	149	116	33	23.88
educ = 9	100	90	90	81	79	94	85	134	159	129	14.78
educ = 12	25	35	36	25	39	52	69	97	150	157	9.72
educ = 15	2	0	0	0	0	2	4	4	21	63	1.36
educ = 17	3	0	1	0	1	4	10	13	44	230	4.34
educ = 19	0	0	0	0	0	0	0	0	2	20	0.31
educ = 22	0	0	0	0	0	0	0	0	2	7	0.13

2000											
	decil 1	decil 2	decil 3	decil 4	decil 5	decil 6	decil 7	decil 8	decil 9	decil 10	total (%)
educ = 0	31	27	26	20	15	14	11	15	1	1	2.43
educ = 3	41	45	37	44	34	23	14	21	8	3	4.07
educ = 4	230	247	287	291	327	277	279	250	192	59	36.77
educ = 6	217	204	191	183	165	170	152	156	115	47	24.12
educ = 9	110	99	76	85	74	105	110	105	137	104	15.15
educ = 12	32	36	45	40	44	67	84	97	134	146	10.93
educ = 15	1	2	0	0	2	2	5	9	34	64	1.79
educ = 17	1	3	1	1	2	5	9	8	37	206	4.12
educ = 19	0	0	0	0	0	0	0	1	3	21	0.38
educ = 22	0	0	0	0	0	0	0	1	2	13	0.24

TABELA A2 - Distribuição dos salários horários segundo o nível de educação

MULHERES											
1998											
	decil 1	decil 2	decil 3	decil 4	decil 5	decil 6	decil 7	decil 8	decil 9	decil 10	total (%)
educ = 0	34	15	8	18	13	11	5	4	4	2	1.88
educ = 3	40	37	19	41	19	25	12	11	5	4	3.51
educ = 4	262	238	303	285	260	221	217	142	39	12	32.58
educ = 6	153	192	185	162	145	150	119	74	34	9	20.14
educ = 9	83	73	66	56	103	107	109	133	114	35	14.47
educ = 12	31	47	27	44	64	83	127	176	160	70	13.65
educ = 15	4	1	0	0	1	7	6	32	105	167	5.32
educ = 17	0	4	0	1	2	3	12	35	141	290	8.03
educ = 19	0	0	0	0	0	1	0	1	5	14	0.35
educ = 22	0	0	0	0	0	0	0	0	0	5	0.08

1999											
	decil 1	decil 2	decil 3	decil 4	decil 5	decil 6	decil 7	decil 8	decil 9	decil 10	total (%)
educ = 0	39	9	23	14	10	16	9	4	2	0	2.20
educ = 3	47	29	29	22	20	27	5	8	4	1	3.35
educ = 4	251	222	253	259	254	221	191	146	35	9	32.14
educ = 6	138	204	168	158	148	143	128	68	25	17	20.90
educ = 9	56	78	63	76	81	87	101	136	118	39	14.58
educ = 12	39	30	36	40	52	71	126	163	166	73	13.90
educ = 15	0	0	1	2	5	4	7	21	87	162	5.05
educ = 17	2	1	0	2	2	4	4	27	132	261	7.59
educ = 19	0	0	0	0	0	0	2	0	3	7	0.21
educ = 22	0	0	0	0	0	0	0	0	1	4	0.09

2000											
	decil 1	decil 2	decil 3	decil 4	decil 5	decil 6	decil 7	decil 8	decil 9	decil 10	total (%)
educ = 0	17	17	8	9	11	9	12	3	1	0	1.57
educ = 3	51	44	25	24	25	14	17	10	3	3	3.90
educ = 4	227	220	231	250	216	206	194	128	33	12	31.01
educ = 6	137	165	183	155	153	127	97	67	32	14	20.41
educ = 9	84	63	74	72	89	92	97	120	116	42	15.33
educ = 12	36	43	32	40	57	93	115	159	152	65	14.30
educ = 15	0	0	1	2	2	6	8	25	80	138	4.73
educ = 17	1	2	0	1	1	7	12	42	133	256	8.22
educ = 19	0	0	0	0	0	0	0	0	4	14	0.33
educ = 22	0	0	0	0	0	0	1	0	0	10	0.20

TABELA B1

Mínimos quadrados				
1998 (homens)				
Variável	coef. est.	desvio p.	est. t	p value
C	5.689	0.053	106.549	0.00
Educ = 3	0.262	0.072	3.656	0.00
Educ = 4	0.217	0.052	4.169	0.00
Educ = 6	0.196	0.053	3.708	0.00
Educ = 9	0.294	0.054	5.438	0.00
Educ = 12	0.481	0.058	8.366	0.00
Educ = 15	1.050	0.089	11.831	0.00
Educ = 17	1.221	0.066	18.366	0.00
Educ = 19	1.586	0.186	8.505	0.00
Educ = 22	1.360	0.213	6.374	0.00
Edexp 3	-0.006	0.002	-2.987	0.00
Edexp 4	-0.001	0.001	-0.642	0.52
Edexp 6	0.004	0.002	2.247	0.02
Edexp 9	0.010	0.002	5.950	0.00
Edexp 12	0.013	0.002	6.057	0.00
Edexp 15	0.003	0.004	0.811	0.42
Edexp 17	0.009	0.003	3.511	0.00
Edexp 19	-0.001	0.009	-0.134	0.89
Edexp 22	0.008	0.007	1.105	0.27
Exper	0.019	0.002	9.589	0.00
Exper^2	0.000	0.000	-11.427	0.00
Tenure	0.014	0.001	11.571	0.00
Tenure^2	0.000	0.000	-5.458	0.00
Norte	-0.116	0.012	-10.049	0.00
Centro	-0.080	0.015	-5.185	0.00
Alentejo	-0.020	0.016	-1.227	0.22
Algarve	-0.047	0.016	-2.951	0.00
Açores	-0.058	0.017	-3.474	0.00
Madeira	-0.112	0.015	-7.671	0.00
obs			7547	
teste F			288.11	
RESET			1.33	

Nota: DP's robustos a heterocedasticidade

TABELA B2

Mínimos quadrados				
1998 (mulheres)				
Variável	coef. est.	Desvio p.	est. t	p value
C	5.754	0.104	55.242	0.00
Educ = 3	-0.079	0.117	-0.674	0.50
Educ = 4	0.050	0.101	0.492	0.62
Educ = 6	0.015	0.103	0.143	0.89
Educ = 9	0.099	0.104	0.953	0.34
Educ = 12	0.289	0.105	2.758	0.01
Educ = 15	0.854	0.110	7.759	0.00
Educ = 17	1.112	0.108	10.278	0.00
Educ = 19	1.179	0.239	4.938	0.00
Educ = 22	2.315	0.234	9.876	0.00
Edexp 3	0.004	0.003	1.340	0.18
Edexp 4	0.001	0.003	0.519	0.60
Edexp 6	0.008	0.003	2.670	0.01
Edexp 9	0.017	0.003	5.571	0.00
Edexp 12	0.020	0.003	6.235	0.00
Edexp 15	0.014	0.003	4.289	0.00
Edexp 17	0.014	0.004	3.937	0.00
Edexp 19	0.022	0.011	1.960	0.05
Edexp 22	-0.028	0.010	-2.755	0.01
Exper	0.002	0.003	0.678	0.50
Exper^2	0.000	0.000	-1.020	0.31
Tenure	0.021	0.002	12.730	0.00
Tenure^2	0.000	0.000	-7.360	0.00
Norte	-0.047	0.012	-3.878	0.00
Centro	-0.077	0.015	-5.092	0.00
Alentejo	-0.005	0.018	-0.299	0.77
Algarve	-0.006	0.016	-0.386	0.70
Açores	0.041	0.020	2.053	0.04
Madeira	-0.064	0.019	-3.332	0.00
obs			6074	
teste F			381.31	
RESET			3.78	

Nota: DP's robustos a heterocedasticidade

TABELA B3

Mínimos quadrados 1999 (homens)				
Variável	coef. est.	Desvio p.	est. t	p value
C	5.749	0.055	104.545	0.00
Educ = 3	0.091	0.062	1.479	0.14
Educ = 4	0.157	0.053	2.962	0.00
Educ = 6	0.182	0.054	3.395	0.00
Educ = 9	0.322	0.056	5.740	0.00
Educ = 12	0.502	0.058	8.635	0.00
Educ = 15	0.946	0.087	10.891	0.00
Educ = 17	1.167	0.067	17.505	0.00
Educ = 19	1.265	0.121	10.482	0.00
Educ = 22	1.218	0.330	3.693	0.00
Edexp 3	0.001	0.002	0.339	0.73
Edexp 4	0.002	0.001	1.167	0.24
Edexp 6	0.004	0.002	2.695	0.01
Edexp 9	0.008	0.002	4.422	0.00
Edexp 12	0.011	0.002	5.251	0.00
Edexp 15	0.009	0.003	2.804	0.01
Edexp 17	0.007	0.003	2.505	0.01
Edexp 19	0.019	0.008	2.392	0.02
Edexp 22	0.017	0.014	1.159	0.25
Exper	0.019	0.002	8.987	0.00
Exper^2	0.000	0.000	-11.388	0.00
Tenure	0.012	0.001	9.539	0.00
Tenure^2	0.000	0.000	-4.148	0.00
Norte	-0.129	0.012	-10.940	0.00
Centro	-0.074	0.016	-4.520	0.00
Alentejo	-0.015	0.016	-0.910	0.36
Algarve	-0.065	0.015	-4.261	0.00
Açores	-0.107	0.016	-6.537	0.00
Madeira	-0.126	0.016	-7.935	0.00
obs			7044	
teste F			256.92	
RESET			2.03	

Nota: DP's robustos a heterocedasticidade

TABELA B4

Mínimos quadrados 1999 (mulheres)				
Variável	coef. est.	Desvio p.	est. t	p value
C	5.770	0.080	72.064	0.00
Educ = 3	-0.066	0.093	-0.704	0.48
Educ = 4	0.044	0.077	0.564	0.57
Educ = 6	0.052	0.080	0.653	0.51
Educ = 9	0.138	0.080	1.728	0.08
Educ = 12	0.317	0.083	3.841	0.00
Educ = 15	0.853	0.089	9.541	0.00
Educ = 17	1.119	0.085	13.173	0.00
Educ = 19	0.922	0.246	3.752	0.00
Educ = 22	1.203	0.290	4.151	0.00
Edexp 3	0.003	0.002	1.057	0.29
Edexp 4	0.002	0.002	1.236	0.22
Edexp 6	0.007	0.002	2.829	0.01
Edexp 9	0.015	0.002	6.594	0.00
Edexp 12	0.018	0.003	6.663	0.00
Edexp 15	0.017	0.003	5.614	0.00
Edexp 17	0.011	0.003	3.616	0.00
Edexp 19	0.025	0.014	1.760	0.08
Edexp 22	0.011	0.016	0.704	0.48
Exper	0.005	0.003	1.737	0.08
Exper^2	0.000	0.000	-3.085	0.00
Tenure	0.017	0.001	11.744	0.00
Tenure^2	0.000	0.000	-5.502	0.00
Norte	-0.066	0.012	-5.461	0.00
Centro	-0.063	0.015	-4.287	0.00
Alentejo	0.020	0.018	1.128	0.26
Algarve	-0.015	0.014	-1.043	0.30
Açores	-0.010	0.019	-0.524	0.60
Madeira	-0.072	0.020	-3.668	0.00
obs			5728	
teste F			360.96	
RESET			2.72	

Nota: DP's robustos a heterocedasticidade

TABELA B5

Mínimos quadrados 2000 (homens)				
Variável	coef. est.	desvio p.	est. t	p value
C	5.998	0.076	79.396	0.00
Educ = 3	-0.080	0.091	-0.881	0.38
Educ = 4	0.010	0.073	0.143	0.89
Educ = 6	0.041	0.075	0.552	0.58
Educ = 9	0.158	0.076	2.079	0.04
Educ = 12	0.285	0.077	3.701	0.00
Educ = 15	0.804	0.095	8.465	0.00
Educ = 17	0.951	0.085	11.202	0.00
Educ = 19	1.194	0.232	5.155	0.00
Educ = 22	0.944	0.259	3.644	0.00
Edexp 3	0.003	0.002	1.382	0.17
Edexp 4	0.003	0.002	1.963	0.05
Edexp 6	0.006	0.002	3.102	0.00
Edexp 9	0.009	0.002	4.215	0.00
Edexp 12	0.013	0.002	5.846	0.00
Edexp 15	0.006	0.003	1.802	0.07
Edexp 17	0.011	0.003	3.477	0.00
Edexp 19	0.011	0.012	0.923	0.36
Edexp 22	0.016	0.013	1.186	0.24
Exper	0.017	0.002	6.741	0.00
Exper^2	0.000	0.000	-11.297	0.00
Tenure	0.014	0.001	9.858	0.00
Tenure^2	0.000	0.000	-5.048	0.00
Norte	-0.170	0.013	-13.464	0.00
Centro	-0.140	0.017	-8.324	0.00
Alentejo	-0.061	0.016	-3.759	0.00
Algarve	-0.100	0.016	-6.068	0.00
Açores	-0.156	0.017	-9.167	0.00
Madeira	-0.077	0.017	-4.622	0.00
obs			6633	
teste F			216.27	
RESET			-1.62	

Nota: DP's robustos a heterocedasticidade

TABELA B6

Mínimos quadrados 2000 (mulheres)				
Variável	coef. est.	Desvio p.	est. t	p value
C	5.761	0.084	68.374	0.00
Educ = 3	0.095	0.117	0.814	0.42
Educ = 4	0.128	0.082	1.571	0.12
Educ = 6	0.124	0.083	1.492	0.14
Educ = 9	0.205	0.084	2.438	0.02
Educ = 12	0.396	0.085	4.646	0.00
Educ = 15	0.929	0.092	10.052	0.00
Educ = 17	1.023	0.089	11.504	0.00
Educ = 19	1.290	0.185	6.986	0.00
Educ = 22	1.498	0.323	4.642	0.00
Edexp 3	-0.003	0.003	-1.115	0.27
Edexp 4	-0.001	0.002	-0.608	0.54
Edexp 6	0.003	0.002	1.109	0.27
Edexp 9	0.011	0.002	4.590	0.00
Edexp 12	0.012	0.003	4.576	0.00
Edexp 15	0.010	0.003	3.218	0.00
Edexp 17	0.017	0.003	5.368	0.00
Edexp 19	0.010	0.007	1.418	0.16
Edexp 22	-0.004	0.014	-0.262	0.79
Exper	0.009	0.003	3.003	0.00
Exper^2	0.000	0.000	-3.382	0.00
Tenure	0.016	0.002	9.758	0.00
Tenure^2	0.000	0.000	-4.254	0.00
Norte	-0.083	0.013	-6.410	0.00
Centro	-0.053	0.018	-2.968	0.00
Alentejo	-0.018	0.018	-1.013	0.31
Algarve	-0.033	0.016	-2.072	0.04
Açores	-0.014	0.020	-0.713	0.48
Madeira	-0.097	0.022	-4.451	0.00
obs			5537	
teste F			308.31	
RESET			2.27	

Nota: DP's robustos a heterocedasticidade

TABELA B7

Regressão Quantis				
1998 (homens) decil 1				
Variável	coef. est.	Desvio p.	est. t	p value
C	5.418	0.086	62.725	0.00
Educ = 3	0.167	0.110	1.518	0.13
Educ = 4	0.173	0.084	2.053	0.04
Educ = 6	0.183	0.086	2.131	0.03
Educ = 9	0.209	0.086	2.423	0.02
Educ = 12	0.331	0.092	3.613	0.00
Educ = 15	0.897	0.164	5.472	0.00
Educ = 17	1.075	0.117	9.161	0.00
Educ = 19	1.561	0.211	7.394	0.00
Educ = 22	1.031	0.289	3.562	0.00
Edexp 3	-0.003	0.003	-0.990	0.32
Edexp 4	0.000	0.002	-0.104	0.92
Edexp 6	0.002	0.003	0.764	0.44
Edexp 9	0.007	0.003	2.833	0.00
Edexp 12	0.012	0.004	3.055	0.00
Edexp 15	0.005	0.006	0.781	0.43
Edexp 17	0.008	0.004	2.099	0.04
Edexp 19	-0.010	0.015	-0.657	0.51
Edexp 22	0.022	0.011	1.966	0.05
Exper	0.015	0.003	5.098	0.00
Exper^2	0.000	0.000	-7.154	0.00
Tenure	0.015	0.002	8.062	0.00
Tenure^2	0.000	0.000	-5.160	0.00
Norte	-0.076	0.018	-4.301	0.00
Centro	-0.066	0.019	-3.462	0.00
Alentejo	-0.018	0.025	-0.709	0.48
Algarve	-0.074	0.031	-2.353	0.02
Açores	-0.061	0.023	-2.609	0.01
Madeira	0.013	0.021	0.653	0.51
obs			7547	
R1			0.18	

TABELA B8

Regressão Quantis				
1998 (mulheres) decil 1				
Variável	coef. est.	Desvio p.	est. t	p value
C	5.441	0.122	44.615	0.00
Educ = 3	-0.226	0.161	-1.402	0.16
Educ = 4	0.056	0.114	0.494	0.62
Educ = 6	0.071	0.118	0.603	0.55
Educ = 9	0.113	0.122	0.930	0.35
Educ = 12	0.238	0.120	1.993	0.05
Educ = 15	0.671	0.157	4.266	0.00
Educ = 17	0.892	0.139	6.399	0.00
Educ = 19	0.725	0.263	2.752	0.01
Educ = 22	2.911	1.261	2.310	0.02
Edexp 3	0.007	0.005	1.525	0.13
Edexp 4	0.001	0.003	0.286	0.78
Edexp 6	0.003	0.004	0.737	0.46
Edexp 9	0.008	0.004	1.737	0.08
Edexp 12	0.010	0.004	2.701	0.01
Edexp 15	0.017	0.006	3.086	0.00
Edexp 17	0.018	0.006	3.069	0.00
Edexp 19	0.033	0.015	2.291	0.02
Edexp 22	-0.046	0.053	-0.880	0.38
Exper	0.007	0.005	1.467	0.14
Exper^2	0.000	0.000	-2.889	0.00
Tenure	0.021	0.003	8.499	0.00
Tenure^2	0.000	0.000	-6.507	0.00
Norte	0.000	0.017	-0.009	0.99
Centro	-0.077	0.026	-2.929	0.00
Alentejo	0.016	0.028	0.556	0.58
Algarve	0.006	0.022	0.277	0.78
Açores	0.002	0.028	0.082	0.93
Madeira	-0.007	0.031	-0.233	0.82
obs			6074	
R1			0.23	

TABELA B9

Regressão Quantis

1998 (homens) decil 5

Variável	coef. est.	desvio p.	est. t	p value
C	5.640	0.072	78.527	0.00
Educ = 3	0.287	0.074	3.869	0.00
Educ = 4	0.260	0.068	3.807	0.00
Educ = 6	0.251	0.070	3.565	0.00
Educ = 9	0.316	0.073	4.330	0.00
Educ = 12	0.489	0.075	6.499	0.00
Educ = 15	1.122	0.094	11.964	0.00
Educ = 17	1.222	0.080	15.252	0.00
Educ = 19	1.424	0.316	4.511	0.00
Educ = 22	1.398	0.359	3.893	0.00
Edexp 3	-0.006	0.002	-3.109	0.00
Edexp 4	-0.002	0.002	-1.123	0.26
Edexp 6	0.002	0.002	1.178	0.24
Edexp 9	0.010	0.002	4.688	0.00
Edexp 12	0.014	0.003	5.494	0.00
Edexp 15	0.003	0.005	0.557	0.58
Edexp 17	0.009	0.003	2.976	0.00
Edexp 19	0.006	0.014	0.390	0.70
Edexp 22	0.009	0.014	0.693	0.49
Exper	0.019	0.002	7.904	0.00
Exper^2	0.000	0.000	-9.882	0.00
Tenure	0.014	0.001	11.073	0.00
Tenure^2	0.000	0.000	-5.934	0.00
Norte	-0.126	0.012	-10.770	0.00
Centro	-0.101	0.015	-6.707	0.00
Alentejo	-0.014	0.018	-0.752	0.45
Algarve	-0.029	0.018	-1.623	0.10
Açores	-0.088	0.017	-5.151	0.00
Madeira	-0.114	0.014	-7.894	0.00
obs			7547	
R1			0.30	
t. FM - SS			209.87	
RESET			0.97	

TABELA B10

Regressão Quantis

1998 (mulheres) decil 5

Variável	coef. est.	desvio p.	est. t	p value
C	5.685	0.080	71.287	0.00
Educ = 3	0.112	0.086	1.296	0.19
Educ = 4	0.115	0.079	1.459	0.14
Educ = 6	0.083	0.079	1.049	0.29
Educ = 9	0.119	0.079	1.513	0.13
Educ = 12	0.300	0.084	3.563	0.00
Educ = 15	0.979	0.098	10.017	0.00
Educ = 17	1.164	0.086	13.538	0.00
Educ = 19	1.275	0.319	4.000	0.00
Educ = 22	2.492	0.988	2.521	0.01
Edexp 3	-0.001	0.003	-0.228	0.82
Edexp 4	0.001	0.002	0.342	0.73
Edexp 6	0.006	0.003	2.528	0.01
Edexp 9	0.020	0.002	7.846	0.00
Edexp 12	0.023	0.003	7.304	0.00
Edexp 15	0.014	0.003	4.414	0.00
Edexp 17	0.016	0.003	4.613	0.00
Edexp 19	0.020	0.015	1.321	0.19
Edexp 22	-0.032	0.041	-0.781	0.43
Exper	0.003	0.003	0.930	0.35
Exper^2	0.000	0.000	-0.839	0.40
Tenure	0.016	0.002	8.597	0.00
Tenure^2	0.000	0.000	-4.853	0.00
Norte	-0.059	0.012	-4.990	0.00
Centro	-0.068	0.016	-4.142	0.00
Alentejo	0.001	0.018	0.083	0.93
Algarve	0.005	0.017	0.274	0.78
Açores	0.035	0.019	1.823	0.07
Madeira	-0.030	0.016	-1.826	0.07
obs			6074	
R1			0.43	
t. FM - SS			250.25	
RESET			2.03	

TABELA B11

Regressão Quantis				
1998 (homens) decil 9				
Variável	coef. est.	desvio p.	est. t	p value
C	6.200	0.124	49.982	0.00
Educ = 3	0.048	0.157	0.308	0.76
Educ = 4	0.082	0.124	0.661	0.51
Educ = 6	-0.024	0.125	-0.193	0.85
Educ = 9	0.155	0.125	1.239	0.22
Educ = 12	0.500	0.136	3.682	0.00
Educ = 15	0.954	0.216	4.427	0.00
Educ = 17	1.212	0.131	9.261	0.00
Educ = 19	1.518	0.576	2.634	0.01
Educ = 22	1.464	0.337	4.348	0.00
Edexp 3	-0.001	0.004	-0.279	0.78
Edexp 4	0.003	0.003	0.949	0.34
Edexp 6	0.011	0.003	3.282	0.00
Edexp 9	0.017	0.003	5.139	0.00
Edexp 12	0.017	0.005	3.661	0.00
Edexp 15	0.006	0.009	0.627	0.53
Edexp 17	0.018	0.004	4.002	0.00
Edexp 19	0.023	0.031	0.747	0.46
Edexp 22	-0.005	0.013	-0.383	0.70
Exper	0.018	0.004	4.234	0.00
Exper^2	0.000	0.000	-6.008	0.00
Tenure	0.008	0.003	2.362	0.02
Tenure^2	0.000	0.000	0.363	0.72
Norte	-0.172	0.024	-7.083	0.00
Centro	-0.073	0.032	-2.298	0.02
Alentejo	-0.061	0.029	-2.084	0.04
Algarve	-0.052	0.032	-1.650	0.10
Açores	-0.065	0.040	-1.599	0.11
Madeira	-0.161	0.042	-3.860	0.00
obs			7547	
R1			0.40	

TABELA B12

Regressão Quantis				
1998 (mulheres) decil 9				
Variável	coef. est.	desvio p.	est. t	p value
C	6.030	0.308	19.604	0.00
Educ = 3	-0.098	0.334	-0.292	0.77
Educ = 4	0.017	0.305	0.057	0.96
Educ = 6	0.030	0.309	0.095	0.92
Educ = 9	0.152	0.308	0.493	0.62
Educ = 12	0.431	0.309	1.395	0.16
Educ = 15	1.092	0.313	3.488	0.00
Educ = 17	1.373	0.311	4.423	0.00
Educ = 19	1.544	0.894	1.727	0.08
Educ = 22	2.089	0.761	2.746	0.01
Edexp 3	0.004	0.009	0.426	0.67
Edexp 4	0.001	0.008	0.114	0.91
Edexp 6	0.007	0.008	0.806	0.42
Edexp 9	0.021	0.008	2.513	0.01
Edexp 12	0.022	0.008	2.641	0.01
Edexp 15	0.004	0.008	0.419	0.68
Edexp 17	0.009	0.009	1.072	0.28
Edexp 19	0.015	0.044	0.343	0.73
Edexp 22	-0.030	0.030	-1.011	0.31
Exper	0.007	0.009	0.768	0.44
Exper^2	0.000	0.000	-0.388	0.70
Tenure	0.018	0.004	4.884	0.00
Tenure^2	0.000	0.000	-1.849	0.06
Norte	-0.083	0.023	-3.589	0.00
Centro	-0.082	0.030	-2.762	0.01
Alentejo	-0.049	0.035	-1.407	0.16
Algarve	-0.054	0.030	-1.778	0.08
Açores	-0.019	0.034	-0.573	0.57
Madeira	-0.061	0.036	-1.689	0.09
obs			6074	
R1			0.49	

TABELA B13

Regressão Quantis				
1999 (homens) decil 1				
Variável	coef. est.	desvio p.	est. t	p value
C	5.422	0.068	80.074	0.00
Educ = 3	0.066	0.087	0.762	0.45
Educ = 4	0.191	0.067	2.868	0.00
Educ = 6	0.224	0.068	3.320	0.00
Educ = 9	0.322	0.068	4.742	0.00
Educ = 12	0.424	0.075	5.650	0.00
Educ = 15	0.842	0.141	5.966	0.00
Educ = 17	1.078	0.096	11.276	0.00
Educ = 19	1.547	0.192	8.075	0.00
Educ = 22	0.776	0.547	1.420	0.16
Edexp 3	0.002	0.003	0.782	0.43
Edexp 4	0.000	0.002	0.086	0.93
Edexp 6	0.002	0.002	0.885	0.38
Edexp 9	0.002	0.003	0.654	0.51
Edexp 12	0.006	0.003	1.850	0.06
Edexp 15	0.008	0.005	1.663	0.10
Edexp 17	0.007	0.005	1.301	0.19
Edexp 19	-0.006	0.013	-0.443	0.66
Edexp 22	0.029	0.028	1.038	0.30
Exper	0.018	0.003	6.151	0.00
Exper^2	0.000	0.000	-7.183	0.00
Tenure	0.012	0.002	7.239	0.00
Tenure^2	0.000	0.000	-3.876	0.00
Norte	-0.110	0.018	-6.116	0.00
Centro	-0.055	0.022	-2.536	0.01
Alentejo	-0.021	0.025	-0.856	0.39
Algarve	-0.068	0.024	-2.775	0.01
Açores	-0.108	0.022	-4.839	0.00
Madeira	-0.011	0.027	-0.412	0.68
obs				7044
R1				0.17

TABELA B14

Regressão Quantis				
1999 (mulheres) decil 1				
Variável	coef. est.	desvio p.	est. t	p value
C	5.145	0.269	19.148	0.00
Educ = 3	0.213	0.311	0.683	0.50
Educ = 4	0.428	0.264	1.619	0.11
Educ = 6	0.471	0.267	1.766	0.08
Educ = 9	0.515	0.267	1.928	0.05
Educ = 12	0.548	0.268	2.041	0.04
Educ = 15	0.926	0.278	3.330	0.00
Educ = 17	1.247	0.272	4.577	0.00
Educ = 19	0.844	0.464	1.820	0.07
Educ = 22	1.799	2.439	0.738	0.46
Edexp 3	-0.006	0.007	-0.864	0.39
Edexp 4	-0.007	0.006	-1.222	0.22
Edexp 6	-0.006	0.006	-1.070	0.29
Edexp 9	-0.003	0.006	-0.503	0.62
Edexp 12	0.007	0.007	1.037	0.30
Edexp 15	0.016	0.007	2.215	0.03
Edexp 17	0.009	0.007	1.300	0.19
Edexp 19	0.017	0.030	0.559	0.58
Edexp 22	-0.009	0.098	-0.095	0.93
Exper	0.018	0.007	2.702	0.01
Exper^2	0.000	0.000	-5.219	0.00
Tenure	0.014	0.002	7.596	0.00
Tenure^2	0.000	0.000	-4.278	0.00
Norte	-0.033	0.016	-2.029	0.04
Centro	-0.062	0.021	-2.939	0.00
Alentejo	0.009	0.021	0.450	0.65
Algarve	0.004	0.024	0.158	0.88
Açores	-0.037	0.022	-1.717	0.09
Madeira	-0.031	0.022	-1.378	0.17
obs				5728
R1				0.24

TABELA B15

Regressão Quantis				
1999 (homens) decil 5				
Variável	coef. est.	desvio p.	est. t	p value
C	5.749	0.059	97.730	0.00
Educ = 3	0.111	0.073	1.513	0.13
Educ = 4	0.178	0.054	3.262	0.00
Educ = 6	0.191	0.057	3.366	0.00
Educ = 9	0.263	0.058	4.542	0.00
Educ = 12	0.463	0.065	7.153	0.00
Educ = 15	1.049	0.098	10.759	0.00
Educ = 17	1.136	0.065	17.470	0.00
Educ = 19	1.251	0.149	8.407	0.00
Educ = 22	0.973	0.639	1.522	0.13
Edexp 3	0.000	0.002	-0.015	0.99
Edexp 4	0.001	0.002	0.586	0.56
Edexp 6	0.004	0.002	2.050	0.04
Edexp 9	0.010	0.002	4.634	0.00
Edexp 12	0.013	0.003	5.203	0.00
Edexp 15	0.006	0.004	1.294	0.20
Edexp 17	0.009	0.002	3.916	0.00
Edexp 19	0.018	0.011	1.744	0.08
Edexp 22	0.025	0.028	0.867	0.39
Exper	0.018	0.002	7.728	0.00
Exper^2	0.000	0.000	-10.480	0.00
Tenure	0.013	0.001	9.330	0.00
Tenure^2	0.000	0.000	-4.434	0.00
Norte	-0.141	0.013	-10.569	0.00
Centro	-0.077	0.018	-4.379	0.00
Alentejo	-0.022	0.018	-1.204	0.23
Algarve	-0.052	0.018	-2.921	0.00
Açores	-0.139	0.018	-7.877	0.00
Madeira	-0.139	0.017	-8.134	0.00
obs			7044	
R1			0.30	
t. FM - SS			219.14	
RESET			0.26	

TABELA B16

Regressão Quantis				
1999 (mulheres) decil 5				
Variável	coef. est.	desvio p.	est. t	p value
C	5.849	0.081	72.075	0.00
Educ = 3	-0.067	0.086	-0.775	0.44
Educ = 4	-0.010	0.079	-0.131	0.90
Educ = 6	-0.012	0.080	-0.154	0.88
Educ = 9	0.030	0.081	0.378	0.71
Educ = 12	0.146	0.085	1.718	0.09
Educ = 15	0.770	0.092	8.374	0.00
Educ = 17	1.055	0.083	12.644	0.00
Educ = 19	1.043	0.370	2.819	0.00
Educ = 22	0.996	2.165	0.460	0.65
Edexp 3	0.002	0.003	0.891	0.37
Edexp 4	0.003	0.002	1.435	0.15
Edexp 6	0.007	0.003	2.795	0.01
Edexp 9	0.020	0.003	7.354	0.00
Edexp 12	0.026	0.003	8.735	0.00
Edexp 15	0.020	0.003	6.344	0.00
Edexp 17	0.013	0.003	4.640	0.00
Edexp 19	0.022	0.024	0.920	0.36
Edexp 22	0.018	0.089	0.205	0.84
Exper	0.001	0.003	0.437	0.66
Exper^2	0.000	0.000	-1.628	0.10
Tenure	0.015	0.001	12.304	0.00
Tenure^2	0.000	0.000	-6.288	0.00
Norte	-0.079	0.012	-6.772	0.00
Centro	-0.058	0.015	-3.878	0.00
Alentejo	0.017	0.016	1.022	0.31
Algarve	-0.005	0.015	-0.322	0.75
Açores	-0.022	0.016	-1.394	0.16
Madeira	-0.072	0.016	-4.592	0.00
obs			5728	
R1			0.43	
t. FM - SS			236.05	
RESET			0.77	

TABELA B17

Regressão Quantis				
1999 (homens) decil 9				
Variável	coef. est.	desvio p.	est. t	p value
C	6.157	0.140	44.005	0.00
Educ = 3	0.032	0.148	0.217	0.83
Educ = 4	0.066	0.137	0.483	0.63
Educ = 6	0.060	0.139	0.435	0.66
Educ = 9	0.260	0.139	1.864	0.06
Educ = 12	0.570	0.147	3.881	0.00
Educ = 15	0.937	0.201	4.662	0.00
Educ = 17	1.172	0.151	7.787	0.00
Educ = 19	1.064	0.209	5.090	0.00
Educ = 22	1.763	0.733	2.407	0.02
Edexp 3	0.002	0.004	0.561	0.58
Edexp 4	0.005	0.003	1.464	0.14
Edexp 6	0.011	0.004	2.881	0.00
Edexp 9	0.018	0.004	4.664	0.00
Edexp 12	0.013	0.005	2.863	0.00
Edexp 15	0.014	0.007	2.013	0.04
Edexp 17	0.013	0.006	2.325	0.02
Edexp 19	0.037	0.010	3.573	0.00
Edexp 22	0.002	0.038	0.051	0.96
Exper	0.019	0.004	4.368	0.00
Exper^2	0.000	0.000	-6.904	0.00
Tenure	0.009	0.003	2.905	0.00
Tenure^2	0.000	0.000	-0.659	0.51
Norte	-0.154	0.025	-6.271	0.00
Centro	-0.078	0.035	-2.240	0.03
Alentejo	-0.030	0.033	-0.884	0.38
Algarve	-0.073	0.028	-2.572	0.01
Açores	-0.097	0.031	-3.143	0.00
Madeira	-0.089	0.036	-2.504	0.01
obs				7044
R1				0.39

TABELA B18

Regressão Quantis				
1999 (mulheres) decil 9				
Variável	coef. est.	desvio p.	est. t	p value
C	6.151	0.221	27.835	0.00
Educ = 3	-0.138	0.219	-0.629	0.53
Educ = 4	-0.068	0.209	-0.327	0.74
Educ = 6	-0.088	0.216	-0.408	0.68
Educ = 9	0.089	0.219	0.407	0.68
Educ = 12	0.396	0.218	1.820	0.07
Educ = 15	1.011	0.235	4.293	0.00
Educ = 17	1.267	0.220	5.768	0.00
Educ = 19	0.884	0.427	2.070	0.04
Educ = 22	0.587	2.078	0.283	0.78
Edexp 3	0.005	0.005	0.901	0.37
Edexp 4	0.006	0.005	1.240	0.22
Edexp 6	0.012	0.006	2.102	0.04
Edexp 9	0.024	0.006	4.076	0.00
Edexp 12	0.019	0.006	3.227	0.00
Edexp 15	0.011	0.007	1.686	0.09
Edexp 17	0.008	0.006	1.335	0.18
Edexp 19	0.048	0.030	1.583	0.11
Edexp 22	0.041	0.086	0.474	0.64
Exper	0.001	0.007	0.154	0.88
Exper^2	0.000	0.000	-0.603	0.55
Tenure	0.014	0.004	4.071	0.00
Tenure^2	0.000	0.000	-0.530	0.60
Norte	-0.090	0.020	-4.506	0.00
Centro	-0.088	0.033	-2.705	0.01
Alentejo	-0.032	0.032	-0.999	0.32
Algarve	-0.069	0.023	-2.977	0.00
Açores	-0.057	0.030	-1.925	0.05
Madeira	-0.078	0.033	-2.351	0.02
obs				5728
R1				0.50

TABELA B19

Regressão Quantis				
2000 (homens) decil 1				
Variável	coef. est.	desvio p.	est. t	p value
C	5.546	0.117	47.316	0.00
Educ = 3	-0.017	0.158	-0.109	0.91
Educ = 4	0.132	0.117	1.131	0.26
Educ = 6	0.176	0.119	1.480	0.14
Educ = 9	0.268	0.119	2.257	0.02
Educ = 12	0.333	0.120	2.785	0.01
Educ = 15	0.766	0.165	4.634	0.00
Educ = 17	0.868	0.153	5.667	0.00
Educ = 19	1.147	0.489	2.347	0.02
Educ = 22	0.960	0.364	2.640	0.01
Edexp 3	0.002	0.004	0.386	0.70
Edexp 4	0.001	0.003	0.194	0.85
Edexp 6	0.002	0.004	0.615	0.54
Edexp 9	0.002	0.004	0.525	0.60
Edexp 12	0.007	0.004	1.921	0.06
Edexp 15	0.005	0.005	1.006	0.31
Edexp 17	0.011	0.006	1.696	0.09
Edexp 19	0.006	0.032	0.193	0.85
Edexp 22	0.000	0.021	0.008	0.99
Exper	0.018	0.004	4.533	0.00
Exper^2	0.000	0.000	-7.679	0.00
Tenure	0.012	0.002	6.008	0.00
Tenure^2	0.000	0.000	-3.902	0.00
Norte	-0.132	0.019	-6.943	0.00
Centro	-0.097	0.022	-4.433	0.00
Alentejo	0.008	0.028	0.282	0.78
Algarve	-0.078	0.032	-2.403	0.02
Açores	-0.152	0.025	-6.185	0.00
Madeira	0.037	0.021	1.806	0.07
obs			6633	
R1			0.17	

TABELA B20

Regressão Quantis				
2000 (mulheres) decil 1				
Variável	coef. est.	desvio p.	est. t	p value
C	5.479	0.230	23.876	0.00
Educ = 3	-0.032	0.277	-0.116	0.91
Educ = 4	0.085	0.229	0.373	0.71
Educ = 6	0.147	0.228	0.644	0.52
Educ = 9	0.155	0.229	0.678	0.50
Educ = 12	0.278	0.231	1.203	0.23
Educ = 15	0.741	0.251	2.957	0.00
Educ = 17	0.817	0.233	3.499	0.00
Educ = 19	1.257	0.388	3.240	0.00
Educ = 22	0.767	0.625	1.228	0.22
Edexp 3	-0.003	0.007	-0.498	0.62
Edexp 4	-0.001	0.005	-0.249	0.80
Edexp 6	-0.001	0.005	-0.166	0.87
Edexp 9	0.005	0.006	0.863	0.39
Edexp 12	0.005	0.006	0.878	0.38
Edexp 15	0.007	0.009	0.795	0.43
Edexp 17	0.020	0.006	3.360	0.00
Edexp 19	0.010	0.013	0.734	0.46
Edexp 22	0.023	0.025	0.899	0.37
Exper	0.013	0.006	2.133	0.03
Exper^2	0.000	0.000	-4.029	0.00
Tenure	0.013	0.002	5.177	0.00
Tenure^2	0.000	0.000	-2.991	0.00
Norte	-0.017	0.019	-0.880	0.38
Centro	-0.010	0.024	-0.431	0.67
Alentejo	0.013	0.034	0.389	0.70
Algarve	0.043	0.029	1.494	0.14
Açores	-0.043	0.030	-1.441	0.15
Madeira	-0.010	0.033	-0.297	0.77
obs			5537	
R1			0.21	

TABELA B21

Regressão Quantis				
2000 (homens) decil 5				
Variável	coef. est.	desvio p.	est. t	p value
C	5.980	0.079	75.933	0.00
Educ = 3	0.004	0.102	0.039	0.97
Educ = 4	0.055	0.079	0.695	0.49
Educ = 6	0.044	0.079	0.552	0.58
Educ = 9	0.133	0.079	1.676	0.09
Educ = 12	0.296	0.081	3.670	0.00
Educ = 15	0.825	0.127	6.516	0.00
Educ = 17	0.996	0.085	11.698	0.00
Educ = 19	1.221	0.284	4.293	0.00
Educ = 22	0.872	0.389	2.239	0.03
Edexp 3	0.001	0.003	0.224	0.82
Edexp 4	0.002	0.002	0.966	0.33
Edexp 6	0.006	0.002	2.822	0.00
Edexp 9	0.010	0.002	4.195	0.00
Edexp 12	0.013	0.003	5.090	0.00
Edexp 15	0.007	0.005	1.420	0.16
Edexp 17	0.011	0.003	4.192	0.00
Edexp 19	0.011	0.013	0.815	0.42
Edexp 22	0.024	0.019	1.208	0.23
Exper	0.015	0.003	5.867	0.00
Exper^2	0.000	0.000	-9.839	0.00
Tenure	0.013	0.001	9.222	0.00
Tenure^2	0.000	0.000	-4.694	0.00
Norte	-0.180	0.013	-13.841	0.00
Centro	-0.151	0.018	-8.231	0.00
Alentejo	-0.051	0.019	-2.765	0.01
Algarve	-0.074	0.020	-3.762	0.00
Açores	-0.183	0.020	-9.390	0.00
Madeira	-0.083	0.022	-3.787	0.00
obs			6633	
R1			0.29	
t. FM - SS			163.83	
RESET			-2.30	

TABELA B22

Regressão Quantis				
2000 (mulheres) decil 5				
Variável	coef. est.	desvio p.	est. t	p value
C	5.760	0.116	49.485	0.00
Educ = 3	0.124	0.128	0.963	0.34
Educ = 4	0.164	0.113	1.449	0.15
Educ = 6	0.148	0.116	1.276	0.20
Educ = 9	0.174	0.115	1.508	0.13
Educ = 12	0.360	0.116	3.111	0.00
Educ = 15	0.994	0.127	7.802	0.00
Educ = 17	1.053	0.118	8.899	0.00
Educ = 19	1.255	0.284	4.421	0.00
Educ = 22	1.801	0.404	4.462	0.00
Edexp 3	-0.003	0.003	-1.100	0.27
Edexp 4	-0.002	0.003	-0.634	0.53
Edexp 6	0.001	0.003	0.478	0.63
Edexp 9	0.014	0.003	4.769	0.00
Edexp 12	0.016	0.003	4.617	0.00
Edexp 15	0.011	0.004	3.104	0.00
Edexp 17	0.017	0.004	4.727	0.00
Edexp 19	0.014	0.012	1.232	0.22
Edexp 22	-0.012	0.020	-0.631	0.53
Exper	0.007	0.003	2.134	0.03
Exper^2	0.000	0.000	-2.148	0.03
Tenure	0.011	0.002	6.751	0.00
Tenure^2	0.000	0.000	-2.491	0.01
Norte	-0.098	0.012	-8.103	0.00
Centro	-0.072	0.016	-4.394	0.00
Alentejo	-0.020	0.018	-1.135	0.26
Algarve	-0.026	0.015	-1.752	0.08
Açores	-0.038	0.021	-1.778	0.08
Madeira	-0.074	0.021	-3.507	0.00
obs			5537	
R1			0.40	
t. FM - SS			234.13	
RESET			0.12	

TABELA B23

Regressão Quantis				
2000 (homens) decil 9				
Variável	coef. est.	desvio p.	est. t	p value
C	6.333	0.134	47.351	0.00
Educ = 3	0.069	0.160	0.432	0.67
Educ = 4	0.088	0.127	0.697	0.49
Educ = 6	0.086	0.133	0.646	0.52
Educ = 9	0.198	0.135	1.464	0.14
Educ = 12	0.479	0.143	3.339	0.00
Educ = 15	1.039	0.174	5.988	0.00
Educ = 17	1.097	0.152	7.230	0.00
Educ = 19	1.208	0.606	1.994	0.05
Educ = 22	1.577	0.514	3.066	0.00
Edexp 3	-0.001	0.004	-0.135	0.89
Edexp 4	0.002	0.003	0.662	0.51
Edexp 6	0.006	0.004	1.687	0.09
Edexp 9	0.014	0.004	3.877	0.00
Edexp 12	0.012	0.005	2.412	0.02
Edexp 15	0.000	0.005	-0.093	0.93
Edexp 17	0.012	0.005	2.293	0.02
Edexp 19	0.020	0.027	0.769	0.44
Edexp 22	0.000	0.020	-0.021	0.98
Exper	0.020	0.005	3.959	0.00
Exper^2	0.000	0.000	-5.652	0.00
Tenure	0.008	0.003	2.701	0.01
Tenure^2	0.000	0.000	0.209	0.83
Norte	-0.225	0.029	-7.757	0.00
Centro	-0.183	0.034	-5.391	0.00
Alentejo	-0.151	0.038	-4.010	0.00
Algarve	-0.133	0.034	-3.876	0.00
Açores	-0.167	0.036	-4.675	0.00
Madeira	-0.175	0.035	-4.944	0.00
obs			6633	
R1			0.37	

TABELA B24

Regressão Quantis				
2000 (mulheres) decil 9				
Variável	coef. est.	desvio p.	est. t	p value
C	6.126	0.204	30.024	0.00
Educ = 3	0.034	0.283	0.121	0.90
Educ = 4	0.037	0.196	0.187	0.85
Educ = 6	0.001	0.203	0.004	1.00
Educ = 9	0.206	0.203	1.013	0.31
Educ = 12	0.494	0.212	2.328	0.02
Educ = 15	0.992	0.225	4.410	0.00
Educ = 17	1.085	0.206	5.281	0.00
Educ = 19	1.493	0.342	4.364	0.00
Educ = 22	1.747	0.267	6.536	0.00
Edexp 3	-0.001	0.007	-0.153	0.88
Edexp 4	0.001	0.005	0.228	0.82
Edexp 6	0.009	0.006	1.682	0.09
Edexp 9	0.016	0.006	2.895	0.00
Edexp 12	0.013	0.007	1.974	0.05
Edexp 15	0.010	0.007	1.425	0.15
Edexp 17	0.023	0.006	3.752	0.00
Edexp 19	0.005	0.013	0.337	0.74
Edexp 22	-0.019	0.012	-1.542	0.12
Exper	0.008	0.006	1.274	0.20
Exper^2	0.000	0.000	-1.711	0.09
Tenure	0.016	0.003	4.895	0.00
Tenure^2	0.000	0.000	-1.074	0.28
Norte	-0.138	0.025	-5.645	0.00
Centro	-0.100	0.034	-2.957	0.00
Alentejo	-0.071	0.039	-1.819	0.07
Algarve	-0.084	0.030	-2.843	0.00
Açores	-0.037	0.037	-0.987	0.32
Madeira	-0.092	0.036	-2.523	0.01
obs			5537	
R1			0.47	

VOLUME I

1º QUADRIMESTRE DE 2001

INFORMAÇÕES

VOLUME I

1º QUADRIMESTRE DE 2001

ACTIVIDADES E PROJECTOS IMPORTANTES DO SISTEMA ESTATÍSTICO NACIONAL

IMPORTANT ACTIVITIES AND PROJECTS OF THE NATIONAL STATISTICAL SYSTEM

PRINCIPAIS OBJECTIVOS E ACÇÕES NO ÂMBITO DO SISTEMA ESTATÍSTICO NACIONAL 2001

O processo de planeamento da actividade estatística nacional abrange: a definição dos objectivos estratégicos, através das Linhas Gerais da Actividade Estatística Nacional definidas pelo Conselho Superior de Estatística, de periodicidade quadrienal; a concretização destes objectivos e a apresentação das principais acções a desenvolver a médio prazo, no Programa Estatístico de Médio Prazo, com a mesma periodicidade; e a planificação anual das acções a executar, no Plano de Actividades do INE e das Entidades com delegação de competências do INE.

As Linhas Gerais da Actividade Estatística Nacional e respectivas prioridades para 1998-2002 foram aprovadas pelo Conselho Superior de Estatística em Maio de 1997 (Deliberação nº 125/97). O Programa Estatístico de Médio Prazo 1998-2002, que obteve parecer favorável do Conselho Superior de Estatística em Novembro de 1997 (Deliberação nº 135/97), concretiza os principais objectivos estratégicos e acções de desenvolvimento para os vários domínios da actividade estatística nacional.

Na sequência dos dois documentos supracitados, foi elaborado o Plano de Actividades do INE e das Entidades com delegação de competências do INE para 2001, sobre o qual o Conselho Superior de Estatística emitiu parecer favorável em Dezembro de 2000.

1. PRINCIPAIS OBJECTIVOS A DESENVOLVER EM 2001

A. Plano de Acção relativo às necessidades estatísticas da União Económica e Monetária

Em 5 de Julho de 2000 o ECOFIN aprovou o segundo relatório de progresso sobre as necessidades de informação estatística da União Económica e Monetária (UEM) e solicitou à Comissão (EUROSTAT) que, em estreita colaboração com o Banco Central Europeu, apresentasse um Plano de Acção sobre a produção das estatísticas nacionais necessárias à compilação, pelo EUROSTAT, de um conjunto de estatísticas fiáveis e atempadas para a área do Euro e da União Europeia.

O Plano de Acção, aprovado em 29 de Setembro de 2000 pelo ECOFIN, inventaria as principais estatísticas a desenvolver face às prioridades políticas relevantes para todos os Estados-membros da UEM, e identifica, para cada um deles e por área estatística, o conjunto de acções a desenvolver no âmbito dos respectivos Planos de Acção nacionais, agrupadas nos seguintes domínios:

- Principais agregados das Contas Nacionais Trimestrais (SEC 95)
- Estatísticas Trimestrais das Finanças Públicas
- Estatísticas do Mercado de Trabalho
- Estatísticas do Comércio Externo
- Difusão de dados estatísticos

Um dos objectivos a atingir, a médio prazo, é a convergência para o padrão correspondente à média dos três Estados-membros com melhor desempenho relativamente a cada um dos indicadores, tanto em termos de prazos como de qualidade.

Sendo muito exigente no que se refere às obrigações estatísticas a assumir por cada Estado-membro, que decorrerão de legislação estatística comunitária em fase de preparação, a implementação deste Plano de Acção para a UEM e dos correspondentes Planos de Acção nacionais dependerá dos recursos financeiros que forem disponibilizados para apoiar o seu cumprimento.

Estas novas obrigações estatísticas constituem um desafio ao qual o Sistema Estatístico Nacional terá de ser capaz de responder nos próximos anos, com sérias implicações ao nível dos Planos de Actividades para 2001 e anos seguintes. Nestas condições, o INE considerou dever propôr ao CSE uma nova categoria de prioridades para as actividades estatísticas - a prioridade absoluta.

As entidades responsáveis pela produção dos dados estatísticos necessários à UEM são o INE, o Banco de Portugal, ambos em estreita articulação com o Ministério das Finanças enquanto fornecedor de informação de base absolutamente essencial, e o Departamento de Estatística do Trabalho, Emprego e Formação Profissional do Ministério do Trabalho e da Solidariedade.

Não obstante a ainda não existência de base legal Comunitária adequada que altere os calendários em vigor, o nível a que foram tomadas as decisões relativas ao Plano de Acção determina que se assumam, desde já, como uma “obrigação” a preparação do Sistema Estatístico Nacional para rápida resposta a este novo quadro de referência.

Assim, está em curso no INE um trabalho de análise das estatísticas incluídas no Plano de Acção para a UEM, no que respeita à sua cobertura, prazos de disponibilidade, responsabilidades, recursos necessários e principais problemas a resolver de modo a tornar possível o envio atempado ao EUROSTAT das estatísticas nacionais, de acordo com os padrões definidos neste Plano de Acção.

A elaboração das contas nacionais, trimestrais e anuais, faz apelo a um vasto conjunto de informação de base de carácter infra-anual e anual, sem a qual ela se torna impossível ou, no mínimo, se degrada.

Assim, para além de algumas estatísticas de base que o próprio Plano de Acção identifica claramente e que se constituem também como fontes essenciais para a elaboração das contas nacionais, existe um outro conjunto, mais vasto, de estatísticas de base, as fontes das contas nacionais, directamente afectadas pela aprovação daquele Plano de Acção, embora nele não directamente explicitadas.

Os objectivos na área das contas nacionais limitam-se, o que é já extremamente ambicioso, ao cumprimento do Plano de Acção no prazo de dois anos e ao cumprimento dos calendários do Regulamento SEC 95 na área das contas regionais.

A realização sistemática dos trabalhos de compatibilização das componentes financeira e não financeira das contas nacionais portuguesas, a realizar conjuntamente com o Banco de Portugal, é outro dos importantes objectivos a reter.

A concretização daqueles objectivos repercute-se sobre as mais importantes operações das estatísticas económicas e sociais e depende absolutamente de essa repercussão ser inteiramente assumida pelas entidades envolvidas.

Contas Nacionais

Pela sua relevância no âmbito do Plano de Acção e pelas importantes alterações metodológicas verificadas, as contas nacionais justificam que se faça aqui um ponto da situação.

Estabilizado o cumprimento dos calendários relativos aos denominados exercícios dos défices orçamentais excessivos e recursos próprios PNB e IVA e de contas trimestrais, importa agora assegurar que o mesmo acontecerá com as contas anuais provisórias e definitivas.

Os calendários constantes do Regulamento SEC 95, cujo cumprimento se previa garantir no decurso de 2000, estão em vias de ser revistos na sequência das decisões do ECOFIN relativas ao Plano de Acção para a UEM.

Razões relacionadas primeiro com o processo de superação das “reservas” levantadas no contexto da análise do Sistema de Contas Nacionais Portuguesas (SCNP) para determinação do recurso próprio PNB, depois com a sua repercussão em toda a série de contas até 1988 e, em simultâneo, com a implementação de um novo sistema de contas e a obrigatória adopção de 1995 como ano-base, determinaram uma situação de transição / dificuldade só susceptível de superação em 2000.

Uma das consequências mais evidentes do processo referido foi o abandono da elaboração de contas nacionais provisórias situação que, associada ao inevitável período de carência na difusão de contas definitivas, criou algumas dificuldades aos utilizadores de dados detalhados da contabilidade nacional.

Decorre do Regulamento SEC 95 um calendário de disponibilização da informação que, no caso das contas anuais, implica a produção de contas provisórias até Setembro de cada ano seguinte ao de referência dos dados. Este conjunto de contas distingue-se das definitivas unicamente pela utilização de fontes estatísticas de carácter provisório e, eventualmente, não cobrindo ainda a totalidade do ano de referência, e não pelo detalhe da informação a apresentar, a qual coincide nas diferentes versões.

No caso concreto do SCNP, dadas as datas de disponibilização e qualidade das fontes estatísticas, o disposto no Regulamento SEC 95 determina o recurso às contas trimestrais como principal suporte para a elaboração das contas provisórias ou, dito de outro modo, para a disponibilização da primeira versão das contas nacionais anuais.

Contudo, os métodos de elaboração das contas trimestrais e das contas anuais definitivas são e continuarão a ser, pela própria natureza da informação infra-anual disponível, substancialmente diferentes, assentando as primeiras no recurso generalizado a “métodos indirectos / econométricos” e, as segundas, essencialmente no recurso a “métodos directos / contabilísticos”. Ou seja, a compilação das primeiras faz apelo à utilização de taxas de evolução de um conjunto seleccionado de indicadores quantitativos, enquanto que as segundas são determinadas em nível a partir de fontes estatísticas assumidas como exaustivas.

Não obstante a conciliação feita, no contexto da compilação das contas trimestrais, entre estas e as anuais logo que disponíveis, tal não é suficiente para assegurar que as revisões que daí decorrem se situem a um nível mínimo. Só uma tendencial convergência de métodos entre um e outro conjunto de contas, ou seja, a utilização tão generalizada quanto possível no processo de compilação das contas trimestrais de “métodos directos” e, nas contas nacionais provisórias de “métodos indirectos” sempre que necessário, permitirá caminhar para a sistemática coerência de resultados, ou seja, para a atenuação das diferenças entre as sucessivas versões das contas nacionais.

B. Outros Objectivos e Principais Acções a Desenvolver

Tendo sido desenvolvidos, com um maior detalhe, os objectivos relacionados com o Plano de Acção relativo às necessidades estatísticas da União Económica e Monetária, apresentam-se seguidamente outros objectivos a atingir no âmbito do Plano de Actividades, bem como as principais acções a desenvolver em 2001.

Como objectivos gerais, salientam-se:

- Melhorar a qualidade da informação estatística oficial, designadamente através do reforço da sua fiabilidade, actualidade e coerência;
- Assegurar a disponibilização de resultados de acordo com os calendários e padrões de qualidade estabelecidos;
- Intensificar e melhorar a utilização de suportes electrónicos na recolha da informação estatística de base;
- Reduzir a carga estatística sobre os inquiridos, através da procura de soluções para o acesso a fontes administrativas para fins estatísticos oficiais.

No sentido de alcançar estes objectivos, o Plano de Actividades para 2001 tem inscritos 333 modelos estatísticos¹, dos quais 229 da responsabilidade do INE (68,8%), e 104 de outras entidades intervenientes na produção estatística oficial (31,2%).

Dos 333 modelos estatísticos, 29 são de prioridade absoluta (8,7%), 258 são de 1ª prioridade (77,5%), e 46 são de 2ª prioridade (13,8%); do total dos modelos, 125 são obrigatórios por Legislação Nacional ou Comunitária (37,5%).

Apresentam-se ainda 21 modelos estatísticos⁸² novos (6,3%), considerando-se os restantes 312 de desenvolvimento corrente.

O Plano de Actividades para 2001 apresenta a caracterização de todos os modelos estatísticos nele inscritos, ventilados por 29 áreas estatísticas, a seguir enumeradas:

- | | |
|--------------------------------------|--|
| ⇒ Administrações Públicas | ⇒ Agricultura, Produção Animal e Silvicultura |
| ⇒ Ambiente | ⇒ Ciência e Tecnologia e Sociedade da Informação |
| ⇒ Comércio Internacional | ⇒ Comércio Interno e Outros Serviços |
| ⇒ Condições de Vida das Famílias | ⇒ Conjuntura Económica |
| ⇒ Contas Nacionais e Regionais | ⇒ Cultura, Desporto e Recreio |
| ⇒ Deficiência e Reabilitação | ⇒ Demografia |
| ⇒ Educação | ⇒ Emprego e Salários |
| ⇒ Empresas | ⇒ Estatísticas Gerais |
| ⇒ Formação Profissional | ⇒ Habitação, Construção e Obras Públicas |
| ⇒ Indústria e Energia | ⇒ Iniciativas de Produção e Estudos Regionais |
| ⇒ Instituições Financeiras e Seguros | ⇒ Justiça |
| ⇒ Pesca | ⇒ Preços |
| ⇒ Protecção Social | ⇒ Relações e Condições do Trabalho |
| ⇒ Saúde | ⇒ Transportes e Comunicações |
| ⇒ Turismo e Restauração | |

⁸² Modelo Estatístico - Conjunto de características metodológicas que definem uma ou mais operações estatísticas. É constituído pelos atributos permanentes, com base nos quais se executam as operações estatísticas.

2. PRINCIPAIS OBJECTIVOS A DESENVOLVER EM 2001

2.1. ESTATÍSTICAS ECONÓMICAS

Principais objectivos:

- Consolidar os calendários das contas nacionais e, em particular, a disponibilização da primeira versão de contas nacionais definitivas do ano t, em t+2, sem degradação de qualidade;
- Desenvolver as estatísticas das empresas e os indicadores de curto prazo relacionados com a União Económica e Monetária e o Mercado Único;
- Alargar a cobertura das estatísticas dos serviços;
- Consolidar o processo de transferência electrónica da dados do Comércio Internacional.

Acções a desenvolver:

- Transformação dos dados das contas nacionais segundo o SEC 1995 em SEC 79, para cumprimento do disposto no nº 1 do artigo 8º do Regulamento (CE) nº 2223/96 do Conselho, de 25 de Junho (utilização do PNB, na determinação do 4º recurso próprio comunitário);
- Prossecução, em articulação com as unidades de produção estatística sectorial, dos esforços que permitam contribuir para garantir o objectivo de disponibilizar as contas definitivas do ano t em Julho de t+2;
- Prossecução dos trabalhos que são objecto de protocolo com o Banco de Portugal, no sentido de garantir a adequada integração das contas financeiras nas contas nacionais portuguesas;
- Prossecução das acções conjuntas com a Direcção-Geral do Orçamento, no sentido de vir a assegurar a obtenção directa dos dados da contabilidade pública na óptica da contabilidade nacional;
- Disponibilização das Contas Nacionais Trimestrais no mês seguinte ao final de cada trimestre;
- Reforço do processo de integração dos trabalhos de elaboração das contas regionais com os das contas nacionais;
- Ajustamento dos calendários de disponibilização das contas regionais aos das contas nacionais, visando o objectivo da disponibilização dos dados do ano t no ano t+2;
- Ultimeção da concretização da informatização das contas nacionais;
- Continuação dos trabalhos de aplicação do SEC 95 às Contas Económicas da Agricultura, nacionais e regionais, da Silvicultura e da Pesca;

- Continuação dos trabalhos de aplicação do SEC 95 aos Ramos Agricultura, Silvicultura e Pescas, nacionais e regionais, em articulação com as contas nacionais;
- Finalização dos trabalhos de mudança de base (ano:1995) para os Preços e Índices de Preços na Agricultura, decorrentes das necessidades do SEC 95;
- Conclusão dos trabalhos relativos ao Recenseamento Geral da Agricultura 1999;
- Revisão das metodologias dos Indicadores de Curto Prazo e criação de novos Indicadores na Indústria, Construção e Obras Públicas, Comércio a Retalho e Outros Serviços;
- Mudança de base dos Indicadores de Curto Prazo já actualmente existentes, de 1995 para 2000;
- Definição de novos sistemas de recolha de informação estatística do Tratado CECA;
- Desenvolvimento de um subsistema estatístico da construção e da habitação;
- Execução do Inquérito ao Alojamento Particular (Permanência de Hóspedes e Capacidade de Alojamento e Pessoal ao Serviço), no âmbito da Directiva 95/57/CE – Estatísticas do Turismo;
- Preparação e realização dos Inquéritos na área dos Serviços (Investigação e Desenvolvimento, Arquitectura, Engenharia, Outras Actividades de Serviços);
- Reformulação e adequação das operações estatísticas às disposições dos Regulamentos sobre Transportes Aéreos e Transportes Ferroviários, decorrentes da sua eventual aprovação;
- Definição do protótipo da nova aplicação informática do Inquérito às Empresas 2002;
- Alargamento da cobertura sub-sectorial do inquérito de conjuntura aos serviços;
- Definição do protótipo da nova aplicação informática do Comércio internacional e realização de testes;
- Conclusão da implementação de formulários Web do Intrastat;
- Consolidação dos novos procedimentos automatizados de crítica e validação da informação-base do Comércio Internacional;
- Participação na preparação de eventuais simplificações adicionais na recolha de informação Intrastat (iniciativa SLIM), bem como na reformulação global do sistema, através do Programa Edicom;
- Desenvolvimento dos trabalhos de substituição da inquirição directa às entidades do Sistema Financeiro;
- Estudo metodológico e de viabilidade sobre Demografia de Empresas, na sequência dos trabalhos promovidos pelo Eurostat em 2000;
- Definição do suporte de divulgação de resultados do Inquérito Trimestral às Empresas Não Financeiras;
- Início do projecto harmonizado sobre Demografia de Empresas;

- Desenvolvimento dos subsectores da área financeira – Serviços de Seguros, Fundo de Pensões, Instituições de Crédito, Outra Intermediação Financeira, Auxiliares de Intermediação Financeira;
- Introdução de diversas melhorias no Índice de Preços no Consumidor, ao nível dos processos de recolha, análise das novas formas de comércio, análise da representatividade e revisão do cabaz de produtos;
- Início dos trabalhos da nova base (2002=100) do IPC (modelo de cálculo) e realização de estudos associados;
- Revisão do modelo de difusão mensal do IPC e introdução de indicadores complementares para interpretação da inflação;
- Introdução dos ajustamentos de cobertura e de regras de observação e tratamento da informação sobre preços que decorrem da regulamentação comunitária (IPCH).

2.2. ESTATÍSTICAS SOCIAIS

Principais objectivos:

- Garantir a execução técnica dos Censos 2001 em condições de qualidade;
- Disponibilizar todos os dados preliminares dos Censos 2001 até ao final de Junho de 2001;
- Desenvolver o Sistema de Informação Geográfica do INE;
- Conceber um sistema integrado de informação para o conjunto das estatísticas das famílias, em articulação com as necessidades específicas de cada área estatística.

Acções a desenvolver:

- Realização dos Recenseamentos Gerais da População e Habitação 2001, centrando-se os trabalhos nas fases de recolha e tratamento informático dos questionários;
- Implementação da ligação dos ficheiros estatísticos existentes ao Sistema de Informação Geográfica decorrente dos Censos 2001;
- Implementação da ligação geográfica de cada edifício no Sistema de Informação Geográfica, com início nas Secções Estatísticas da Amostra-Mãe e posterior generalização com base nos dados dos Censos 2001;
- Concepção da nova Amostra-Mãe, com base nos resultados dos Censos 2001;
- Continuação dos trabalhos de concepção relativos às Contas Satélites do Ambiente e do Trabalho;
- Concepção e desenvolvimento de um sub-sistema de informação estatística para as Instituições Particulares sem fins Lucrativos, com repercussões nas áreas do ambiente, cultura, protecção social, saúde, etc;
- Desenvolvimento dos Indicadores agro-ambientais;

- Revisão do Painel de Famílias, em articulação com a evolução do dossier a nível comunitário, em ligação com o estudo do futuro do Inquérito aos Orçamentos Familiares e do eventual Inquérito Anual ao Consumo, no quadro da concepção de um sistema de informação sobre as Estatísticas das Famílias;
- Concepção e desenvolvimento das Estatísticas do Livro;
- Aplicação de nova metodologia de recolha de dados sobre o Saneamento básico;
- Reformulação do Inquérito aos Museus, no âmbito do Protocolo assinado com o Instituto Português dos Museus (IPM) e o Observatório das Actividades Culturais, tendo em consideração o recente recenseamento do sector, promovido pelo IPM;
- Reformulação da metodologia das Estimativas da População, a aplicar após a disponibilização dos resultados dos Censos 2001;
- Reformulação da metodologia do Inquérito aos movimentos migratórios de saída;
- Elaboração de indicadores que permitem avaliar a qualidade do emprego;
- Concepção de um sistema regular de informação estatística sobre o mercado de formação profissional, na óptica da oferta e da procura, abrangendo a nível nacional a totalidade da formação efectuada;
- Caracterização através de indicadores apropriados, dos diferentes domínios de observação do mercado de trabalho – demografia, emprego, desemprego, educação e formação, estrutura empresarial, remunerações – quer do ponto de vista nacional, quer regional;
- Conhecimento das diferentes formas de emprego que estruturam o mercado de trabalho português;
- Aproveitamento e tratamento complementar de dados do Inquérito Nacional às Incapacidades, Deficiências e Desvantagens, e do projecto Acessibilidade do meio edificado, em combinação com estatísticas de outras fontes;
- Realização de um novo Inquérito às Instituições e Programas de Reabilitação;
- Realização de novos inquéritos sobre recursos humanos, população escolar e resultados escolares;
- Introdução de uma nova abordagem na produção de dados da Segurança Social, prevendo-se a disponibilização de dados sobre Desemprego e Apoio ao Emprego, prestações familiares e pensões ou subsídio de doença.

2.3. ESTATÍSTICAS DA CIÊNCIA E TECNOLOGIA E DA SOCIEDADE DA INFORMAÇÃO

Principais objectivos:

- Desenvolver a produção de indicadores relacionados com a sociedade e a economia da informação;
- Implementar um sistema integrado de informação de ciência e tecnologia;

- Melhorar o conhecimento do sistema de inovação das empresas portuguesas.

Acções a desenvolver:

- Realização de diversos inquéritos relacionados com a sociedade e a economia da informação, designadamente os Inquéritos à utilização de tecnologias da informação e da comunicação nas empresas e na administração pública, e o Inquérito à utilização de tecnologias e serviços de comunicação e informação nas famílias;
- Analisar a possibilidade de inclusão de questões sobre a Sociedade da Informação nos inquéritos a realizar nas seguintes áreas:
 - Educação;
 - Emprego e Salários;
 - Empresas;
 - Formação Profissional;
 - Comércio Interno e Internacional;
 - Conjuntura Económica.
- Realização do Inquérito à inovação na indústria e nos serviços;
- Validação e divulgação dos resultados do Inquérito ao potencial científico e tecnológico nacional;
- Preparação de outros indicadores de ciência e tecnologia, tais como doutoramentos, produção científica, formação avançada e projectos de investigação;
- Prossecução do estudo e preparação de indicadores sobre o emprego científico.

3. ESTUDOS

Principais objectivos:

- Alcançar resultados significativos no processo de integração do Sistema Estatístico Nacional (SEN), promovendo uma forte ligação entre a investigação e a produção estatística e a cooperação entre o INE e outras instituições do SEN e as Universidades;
- Contribuir para a melhoria da qualidade da informação estatística, através da análise crítica dos métodos adoptados nas diversas fases da produção estatística, e da consistência longitudinal e transversal da informação produzida;
- Criar uma linha de investigação aplicada.

Acções a desenvolver:

- Intensificação do processo de investigação de indicadores económicos, coincidentes e avançados;
- Constituição de um protótipo de Matriz de Contabilidade Social (MCS), para o ano económico de 1995, no quadro do SEC 95, e desenvolvimento de uma MCS no quadro do SEC 95;
- Análise económica das Contas Nacionais, Trimestrais e Síntese de Conjuntura;
- Análise e estudo de novas metodologias para a produção de informação sobre as migrações;
- Desenvolvimento de estudos temáticos no campo demográfico e das condições de vida dos indivíduos e famílias, nomeadamente a exclusão social;
- Análise da exclusão, pobreza e desigualdades sociais em Portugal;
- Análise da dinâmica demográfica em Portugal (séries longas de população), em colaboração com o Observatório Demográfico Europeu (ODE – França);
- Harmonização intercensitária dos dados do Inquérito ao Emprego (emprego por sectores de actividade e por situação perante a profissão).

4. INFRA-ESTRUTURAS DE APOIO À PRODUÇÃO

Principais objectivos:

- Prosseguir os esforços para assegurar o normal acesso às fontes administrativas adequadas para permanente actualização do ficheiro de empresas e estabelecimentos;
- Assegurar a intervenção no processo de revisão internacional das grandes nomenclaturas económicas.

Acções a desenvolver:

- Disponibilizar na INTERNET e estimular a utilização no âmbito do SEN, das principais nomenclaturas e conceitos estatísticos;
- Continuar a assegurar a criação, em suporte automático, de um Sistema Integrado de Nomenclaturas Económicas, sua gestão permanente e difusão;
- Prosseguir a realização dos inquéritos de qualidade e de actualização do ficheiro central de empresas e estabelecimentos.

5. TECNOLOGIAS DE INFORMAÇÃO

Principais objectivos:

- Consolidar a arquitectura do Sistema de Informação do INE;
- Definir um ambiente de desenvolvimento de aplicações, que contemple o paradigma dos objectos, adequando-o às actuais realidades tecnológica e de mercado, bem como às diferentes necessidades do INE;
- Definir um modelo e políticas de segurança informáticas;
- Definir ferramentas para utilizadores finais, de acordo com os diferentes paradigmas de computação necessários, de modo a existir um só software por cada tipo de utilização;
- Potenciar a utilização do DataWarehouse e de DataMarts;
- Adequar a infra-estrutura tecnológica, em particular de comunicações e servidores às novas necessidades;
- Consolidar os vários servidores, de forma a diminuir os custos de gestão envolvidos e otimizar a utilização dos recursos.

Acções a desenvolver:

- Conclusão dos apuramentos do RGA/99 e colaboração nas soluções para difusão electrónica dos dados;
- No âmbito da preparação da operação Censos 2001:
 - Ajustes finais à cadeia de reconhecimento, testes dos Subsistemas de Correções Automáticas, de Cálculo e Especialização e de Gestão de Processos;
 - Projecto e implementação de toda a infra-estrutura tecnológica para suportar o processo de captura e correcção da informação (digitalização, correcções automáticas e manuais);
 - Finalização do *software* para apuramentos e preparação de soluções de difusão.
- Concepção, desenvolvimento e manutenção das aplicações que suportam a carteira de projectos proposta pelas diferentes Unidades Orgânicas no seu plano de actividades, nomeadamente:
 - Sistema Integrado de Gestão de Inquéritos de Ramo;
 - Sistema Informatizado de Contas;
 - Intrastat;
 - Sistema Integrado de Gestão de Inquéritos Agrícolas.
- Disponibilização do *Datawarehouse* para um número mais significativo de utilizadores;
- Implementação de *Data Marts*;

- Implementação de uma *Extranet* no âmbito do protocolo com o Ministério da Ciência e da Tecnologia;
- Reformulação da infra-estrutura de comunicações, visando aumentar a largura de banda efectiva, o nível de serviço e a segurança;
- Criação de condições para a progressiva utilização de meios complementares de recolha da informação;
- Implementação na *web* do sistema de consulta de informação estatística do INE.

6. REGIONALIZAÇÃO

Principais objectivos:

- Recorrer a novas abordagens que permitam evidenciar as características que integram o desenvolvimento das regiões;
- Favorecer a articulação com os diferentes saberes, nomeadamente das Universidades e Centros de Investigação, por forma a potenciar novas perspectivas de observação das regiões à luz da informação estatística;
- Difundir e promover novos estudos, publicações e outras iniciativas regionais.

Acções a desenvolver:

- Promover a necessária articulação com as Autarquias e Associações de Municípios, tendo em vista a realização dos Censos 2001;
- Dinamização de iniciativas de apreensão de fenómenos de crescente preocupação social;
- Desenvolvimento das estatísticas transfronteiriças, no contexto das preocupações comunitárias com a revitalização das regiões de fronteira;
- Desenvolvimento de um sistema de estatísticas Urbanas;
- Continuação do Estudo de caracterização socio-económica da área de influência do empreendimento para fins múltiplos do Alqueva;
- Elaboração de uma monografia sobre a região de Lisboa e Vale do Tejo.

VOLUME I

1º QUADRIMESTRE DE 2001

ACÇÕES DESENVOLVIDAS PELO INE NO ÂMBITO DA COOPERAÇÃO

ACTIONS ACHIEVED BY NSI IN THE SCOPE OF COOPERATION

(DE 1 DE JANEIRO A 30 DE ABRIL DE 2001)

- **COOPERAÇÃO DESENVOLVIDA COM OS PALOP:**

No âmbito do projecto comum aos PALOP sobre “Classificações, Conceitos e Nomenclaturas”, teve lugar em Cabo Verde, de 5 a 9 de Março, o 3º Seminário Regional entre representantes dos respectivos Institutos Nacionais de Estatística, com o apoio financeiro do Eurostat, do Instituto da Cooperação Portuguesa, do Gabinete de Assuntos Europeus e Relações Externas do Ministério do Planeamento e do próprio INE português. O evento teve como objectivos principais confrontar experiências dos diversos países sobre a concepção e aplicação das suas Classificações Económicas, bem como analisar as implicações das novas Classificações nos respectivos Sistemas Estatísticos Nacionais. Os trabalhos decorreram com o maior sucesso, tendo sido animados pelo coordenador do projecto no INE de Portugal, Dr. Hermínio Saraiva Aguiar, e contado com a participação de dois delegados de cada um dos INE beneficiários, para além de representantes de instituições Caboverdianas, tais como o Banco de Cabo Verde e o Conselho Nacional de Estatística.

Ainda no quadro do referido projecto comum, foi realizado no INE um estágio sobre “Classificação do Consumo das Famílias”, destinado a um técnico do INE Caboverdiano, tendo sido também editada a publicação da Classificação Nacional de Bens e Serviços de Cabo Verde, impressa no INE com o apoio financeiro da Cooperação Portuguesa.

No período em apreço, e no âmbito da **Cooperação Bilateral com os PALOP**, foram realizadas as seguintes acções:

ANGOLA

No âmbito dos projectos "Comércio Externo" e "Base Cartográfica Censitária" foram realizadas duas missões de assistência técnica ao INE-Angola, tendo sido concluída uma visita de trabalho ao INE-Portugal no âmbito do projecto "Tecnologias de Informação e Informática".

A missão de assistência técnica no âmbito das “Estatísticas do Comércio Externo”, incidiu no diagnóstico da situação das estatísticas do Comércio Externo de Angola e na preparação de um programa de trabalho para desenvolvimento do respectivo Sub-sistema de Informação.

A missão de assistência técnica relativa à “Base Cartográfica Censitária”, teve como principais objectivos a avaliação da implementação do projecto e a sua revisão,

incluindo a reorganização dos serviços responsáveis, tendo visado igualmente a avaliação do projecto-piloto para a implementação do Sistema de Informação Geográfica (SIG) e utilização da cartografia digital para preparação da base cartográfica.

A visita de trabalho ao INE-Portugal, no âmbito do projecto "Tecnologias de Informação e Informática", teve como principais objectivos a análise do sistema de bases de dados estatísticos, nas perspectivas da produção e da difusão.

CABO VERDE

Foram realizadas nove acções no âmbito dos projectos "Apoio Institucional", "Coordenação Estatística", "Estatísticas da Construção", "Estatísticas Demográficas e Sociais", "Difusão Estatística", "Tecnologias de Informação e Informática", "Índices do Comércio Externo" e "Estatísticas de Conjuntura".

No domínio do "Apoio Institucional" foi realizada uma missão de assistência técnica e uma visita de trabalho. A missão no INE-CV, efectuada pelo Dr. Ferreira da Cunha, teve como objectivo principal efectuar a Avaliação do Estado do Sistema Estatístico Nacional, trabalho que foi efectuado em conjunto com os principais dirigentes do INE-CV e com os representantes dos organismos com delegação de competências no âmbito do SEN. O relatório daquela avaliação deverá ser submetido pelo INE-CV ao Conselho Nacional de Estatística, para apreciação e eventual aprovação. No que se refere à visita de trabalho ao INE-P, esta incidiu essencialmente sobre a análise da execução do programa de cooperação, neste domínio, e das suas perspectivas de desenvolvimento.

No âmbito do projecto "Coordenação Estatística" realizou-se um estágio no INE-Portugal para análise do sistema implementado no INE, tendo em vista a organização de serviços similares no INE de Cabo Verde.

Relativamente ao projecto "Estatísticas da Construção", foi realizada no INE-CV uma missão de identificação tendo em vista o lançamento de um sistema de informação neste domínio. No decurso da missão, o Prof. Paulo Gomes, Director Regional do Norte do INE, animou uma palestra subordinada ao tema "O Estado da Arte das Estatísticas Oficiais Portuguesas na área da Construção e Habitação".

Foi realizada no INE-P uma visita de trabalho na área das "Estatísticas Demográficas e Sociais" com o objectivo de analisar as metodologias e a organização do inquérito aos orçamentos familiares.

No domínio da "Difusão Estatística" foi efectuado um estágio no INE-P sobre técnicas para desenvolvimento do "web site" do INE de Cabo Verde na Internet.

No âmbito do projecto "Tecnologias de Informação e Informática" teve lugar um estágio no INE-P para desenvolvimento das técnicas de manutenção de hardware.

Em continuidade de acções anteriores no quadro do projecto "Índices do Comércio Externo", foi organizada uma missão de assistência técnica ao INE-CV que

visou a análise dos primeiros cálculos efectuados, bem como o apoio à equipa do INE-CV na redacção do manual metodológico e de procedimentos.

Relativamente ao projecto “Estatísticas de Conjuntura”, realizou-se uma missão de assistência técnica ao INE-CV que teve como principais objectivos o diagnóstico técnico do projecto, bem como o apoio à preparação de um programa de trabalho.

S. TOMÉ E PRÍNCIPE

Foi realizada uma acção no quadro do projecto “Tecnologias de Informação e Informática” que consistiu no fornecimento pelo INE-P de equipamento e material para manutenção de hardware.

- COOPERAÇÃO DESENVOLVIDA COM OS PAÍSES DA EUROPA CENTRAL E ORIENTAL, NO QUADRO DO PROGRAMA PHARE:

Realizaram-se sete acções de cooperação ao abrigo do programa PHARE da União Europeia, de assistência técnica aos Países da Europa Central e Oriental, tendo incidido todas na área das estatísticas agrícolas e das pescas.

No âmbito do projecto-piloto “*Farm Structure Survey*”, foram realizadas duas missões de cooperação, tendo uma delas sido levada a cabo no Instituto Nacional de Estatística e Estudos Económicos da **Roménia**, de 23 a 28 de Março, e a segunda sido realizada no Instituto Nacional de Estatística da **Bulgária**.

No quadro do projecto-piloto “Emprego no Sector das Pescas”, foram efectuadas duas acções de cooperação, nomeadamente uma missão de assistência técnica aos INE da **Estónia** e da **Letónia**, bem como a participação do INE num *Workshop* que teve lugar a 21 de Fevereiro no Luxemburgo.

O INE tem vindo também a participar em projectos de cooperação com o INE da **Bulgária** no domínio do “*acquis communautaire*”, na área das estatísticas agrícolas. Esta assistência desenvolve-se através de missões, tendo a última ocorrido de 7 a 14 de Fevereiro, sendo a representação do INE, neste domínio, assegurada pelo Engº António Macedo, Director do Departamento de Estatísticas da Agricultura e Pescas.

O INE participou igualmente em dois *Workshops* na área das estatísticas agrícolas. Um deles, subordinado ao tema “*Balance Sheets*”, teve lugar em 7 de Março, no Luxemburgo. O outro, subordinado à temática “*Agricultural Price Statistics – Eurostat PHARE pilot project*”, realizou-se em 15 e 16 de Março, tendo a representante do INE, Dra Maria José Correia, apresentado comunicações sobre “A Actualização da Base nos Índices dos Preços Agrícolas” e “A Experiência Portuguesa da adesão à União Europeia”.

- **COOPERAÇÃO DESENVOLVIDA NO ÂMBITO DO PROGRAMA MEDSTAT:**

Ao abrigo do Programa MEDSTAT - Assistência técnica aos Países do Mediterrâneo - da União Europeia, realizaram-se, durante o período em referência, oito acções de cooperação no domínio do projecto MED-IS (Sistemas de Informação Estatística – componente de difusão).

Assim, no seguimento do trabalho desenvolvido no ano anterior, o INE continuou a participar em missões com o objectivo de avaliar os sistemas de informação estatística dos países beneficiários do Programa MEDSTAT, designadamente a Marrocos, Argélia, Líbano, Egipto, Síria, Jordânia e Turquia. A reunião de "*debriefing*" que ocorreu no Luxemburgo, em 19 e 20 de Março, concluiu a 1ª fase deste projecto.

- **COOPERAÇÃO DESENVOLVIDA NO ÂMBITO DO PROGRAMA UNIÃO EUROPEIA —
MERCOSUL:**

Ao abrigo do Programa da União Europeia de assistência técnica aos Países do MERCOSUL e ao Chile, realizaram-se, no período em apreço, duas acções de cooperação.

No âmbito do grupo de trabalho sobre "Produtividade e Competitividade das Empresas", o INE participou na 4ª reunião do projecto, que teve lugar na Argentina, de 25 a 26 de Março.

O INE também esteve presente na 4ª reunião do grupo de trabalho sobre "Nomenclaturas e Classificações", que se realizou no Chile, em 28 e 29 de Março.

CONGRESSOS, SEMINÁRIOS, COLÓQUIOS E CONFERÊNCIAS

CONGRESS, SEMINARS AND CONFERENCES

2001

□ 11-12 May

International Conference on Survey Research Methods - The Challenge of the Internet organised by ASC, ONS, MRS, SRA. To be held at the Latimer Conference Centre, London, UK.

Information: Diana Elder, Administrator ASC, PO Box 60, Chesham, Bucks, UK HP5 3QH tel/fax: +44 (0)1494 793033
E-mail: admin@asc.org.uk
URL: <http://www.asc.org.uk/>.

□ 14-15 May

International Conference on Quality in Official Statistics to be held in Stockholm, Sweden.

Information: Conference Chair: Lars Lyberg
E-mail: lars.lyberg@scb.se
or
q2001@scb.se
URL: <http://www.q2001.scb.se>.

□ 14-18 May

33rd Annual Conference of the French Statistical Society (Société Française de statistique), to be held in Nantes-France.

Informações: Email: sfds2001@enitiaa-nantes.fr
URL: <http://www.enitiaa-nantes.fr/sfds2001/>

□ 14-19 May

2001 A DATA ODYSSEY, Collaborative Working in the Social Science Cyberspace, to be held in Amsterdam, The Netherlands.

Information: International Association for Social Science Information Services and Technology (ASSIST) together with the International Federation of Data Organizations
E-mail: jassist2001@niwi.knaw.nl
URL: <http://www.niwi.knaw.nl/ia2001.htm>

□ 5-7 June

CAESAR 2001: Conference on Agricultural and Environmental Statistical Application, to be held in Rome, Italy at the Palazzo dei Congressi.

Informações: URL: <http://www.istat.it/caesar/caesar2>

- 10-14 June
Annual Meeting of the Statistical Society of Canada, Burnaby, British Columbia, Canada.
Informações: Local Arrangements Chair: Tim Schwartz, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada..
 E-mail: tim@cs.sfu.ca
- 12-15 June
ASMDA 2001 10th International Symposium on Applied Stochastic Models And Data Analysis, to be held in Compiègne (France).
Informações: URL: <http://www.ciemat.es/convocatorias/eventos/samo2001>
- 13-16 June
Interface 2001: The Interface Foundation of North America 33rd Symposium on the Interface of Computer Science and Statistics in Costa Mesa, Orange County, California. The conference theme is: "*Frontiers of Data Mining and Bioinformatics*".
Information: URL: <http://www.ics.uci.edu/interface/>.
- 16-22 June
ETK "2001 – NTTS" 2001, Exchange of Technology and Know-how, New Techniques and Technologies for Statistics. The European Commission is pleased to announce two major conferences on methods and tools for Official Statistics in Europe: ETK and NTTS 2001, to be held jointly in Hermissos, Crete.
Information: E-mail: lars.lyberg@scb.se
 URL: <http://www.jrc.cec.eu.int/etk-ntts>
- 18-20 June
Third International Symposium on Sensitivity Analysis of Model Output, to be held in Madrid.
Informações: Ana García Triviño- CIEMAT - Instituto de Estudios de la Energía, Edificio 1 despacho 49 - Avda. Complutense, 22 - 28040 MADRID.
 Tel +34 - 91 346 64 86,
 Fax: +34 - 91 346 60 05.
 E-mail trivi@ciemat.es
 URL: <http://www.ciemat.es/convocatorias/eventos/samo2001>
- 18-30 June
IASC-IASS Joint European Summer School. Knowledge Discovery and Large Marketing Survey: Design and Analysis. The 6th Summer School of the European Regional Section of IASC, organised jointly with IASS, will be hosted by the Department of Mathematics and Statistics of the University of Naples, Italy.
Information: Simona Balbi and Vincenzo Esposito, Department of Mathematics and Statistics, University of Naples "Federico II", Italy.
 E-mail: binci@unina.it.

- 19-22 June
23rd International Conference on "Information Technology Interfaces" ITI 2000 to be held in Pula, Croatia.
Informações: Conference Secretariat ITI 2000 Tel. +385 1 616 55 97; +385 1 616 55 99 Fax: +385 1 616 55 91
E-mail iti@srce.hr
Or
Damir Kalpic, Chairman of the International Program Committee,
Faculty of Electrical Engineering and Computing, University of Zagreb. Tel. +385 1 612 99 19 , Fax + 385 1 612 99 15.
E-mail iti@srce.hr

- 19-22 June
International ICSC Congress **COMPUTATIONAL INTELLIGENCE; METHODS AND**, University of Wales, Bangor, Wales (UWB) U.K.
Informações: General Chair: Ludmila I. Kuncheva, UWB. General Co-chair: Tim Porter, UWB.
URL: <http://www.icsc.ab.ca/cima2001.htm> (updated weekly).
E-mail iti@srce.hr

- 20-30 June
IASC-IASS Joint European Summer School on: "Knowledge Discovery & Large Surveys: Design and Analysis", to take place on the Island of Capri (Italy).
Organisation: Department of Mathematics and Statistics of the University of Naples, Italy.
Information: Dr. Vincenzo Esposito; telephone: +39 081 67 5112
E-mail: binici@unina.it
URL: <http://www.dms.unina.it/School2001.html>.

- 23-27 June
VIII Seminar of IASI on Applied Statistics will be held in Panama. Selected theme: Statistical Methods in Quality and Productivity.
Information: Departamento de Estadística, Universidad de Panamá, Estafeta Universitaria, Panamá, Panamá. Tel. (507) 263-6133 ext 189, Fax (507) 263-6000 or Secretariat of IASI.
E-mail: viiiisest@ancon.up.ac.pa
or
efabb@mecon.gov.ar
URL: <http://www.up.ac.pa/UnidadesAcademicas/facultad/FCNET/VIII-SEM-EST/VIIISem.htm>

- 25-27 June
International Seminar on Methodologies for Sampling Household Surveys, to be held in Rio de Janeiro, RJ, Brazil, jointly organized by IASI and the Brazilian Institute of Geography and Statistics (IBGE).
Information: Local organizing committee: Carmen Aparecida Feijó or the Secretariat of IASI.
E-mail: cfeijo@ibge.gov.br
or
efabb@indec.mecon.gov.ar
URL: <http://www.ibge.gov.br/>.
- 26-29 June
Fourth International ICSC Symposia on SOFT COMPUTING AND INTELLIGENT SYSTEMS FOR INDUSTRY, Paisley, Scotland, U.K.
Informações: ICSC International Computer Science Conventions 5101C - 50 Street, Wetaskiwin, Alberta T9A 1K1 / Canada. Phone: +1-780-352-1912 / Fax: +1-780-352-1913
E-mail operating@icsc.ab.ca
or
E-mail planning@icsc.ab.ca
- 26-30 June
International Conference and Workshop on Recent Developments in Statistics and Its Applications to be held in Kuala Lumpur, Malaysia.
Information: Mokhtar B. Abdullah, School of Mathematical Sciences, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor D.E., Malaysia.
Tel. +603-8293857, Fax +603-8264816.
E-mail: admacs@admacs.com.my
URL: <http://www.admacs.com.my/conference>
- July
VIII Seminario Interamericano de Estadística Aplicada - Panamá.
Informações: Tel. (507) 263-6133 ext. 189.
Fax: (507) 253-6000.
E-mail: estadist@ancon.up.ac.pa
- 2-6 July
New Trends in Statistical Modelling, to be held in Odense, Denmark.
Informações: Bent Jorgensen, Department of Statistics and Demography, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. Phone: +45 65 50 33 97, fax: +45 65 95 77 66.
Email: IWSM@statdem.sdu.dk
or
URL: <http://www.statdem.sdu.dk/IWSM/>

- 3-6 July
RSS2001: Spatial Modelling: The Department of Statistics, University of Glasgow is organising the 2001 Royal Statistical Society Theme Conference on Spatial Modelling.
Informações: URL: <http://www.statdem.sdu.dk/IWSM/>
or
E-mail: RSS2001@statistics.glasgow.ac.uk
- 5-6 July
CLADAG 2001, the Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society, to be held in Palermo, Italy.
Information: Antonella Plaia, Department of Statistical and Mathematical Sciences "Silvio Vianelli", tel +39 091 6626244, Fax +39 091 485726.
E-mail: cladag@unipa.it
URL: <http://statistica.economia.unipa.it/cladag2001>.
- 6-8 July
STATISTICS 2001 CANADA - Fourth Canadian Conference on Applied Statistics, organized by the Department of Mathematics and Statistics and the Department of Decision Sciences and MIS, to be held in Montreal.
Informações: Yogendra P. Chaubey, Chair, Scientific Committee. Phone: (514) 848-3258 Fax: (514) 848-2831.
E-mail: chaubey@alcor.concordia.ca
or
URL: <http://alcor.concordia.ca/~chaubey/stat2001canada.html>.
- 9-13 July
SPA'01: 27th International Conference on Stochastic Processes and their Applications, to be held in Cambridge, U.K.
Information: J.R. Norris.
E-mail: spa27@statslab.cam.ac.uk
- 18-21 July
Statistical Challenges in Modern Astronomy III to be held at The Penn State University Park campus, Pennsylvania, USA.
Information: Co-Chairs: Dr. G.Jogesh Babu, Statistics Co-chair, 814-863-2837; Dr. Eric D. Feigelson, 814-865-0162; Astronomy Co-chair.
E-mail: babu@stat.psu.edu
or
edf@astro.psu.edu
URL: <http://www.outreach.psu.edu/C&I/SCMA/>
- 18-21 July
ICORS 2001 - International Conference on Robust Statistics 2001 (sponsored by the European Commission), held from 23-27 July 2001 in Vorau, Austria.
Informações: Peter Filzmoser, Dept. of Statistics, Vienna University of Technology, A-1040 Vienna, Austria.
URL: <http://www.statistik.tuwien.ac.at/icors2001>
or
E-mail: icors2001@statistik.tuwien.ac.at

- 23-28 July
Mixtures 2001- Recent Developments in Mixture Modelling, Universität der Bundeswehr, Hamburg, Germany, sponsored by the German Research Foundation.
Informações: Wilfried Seidel (local organizer), FB Wirtschafts- und Organisationswissenschaften, Universität der Bundeswehr Hamburg, D-22039 Hamburg, Germany.
 E-mail: mixtures@unibw-hamburg.de
 URL: <http://bruce.unibw-hamburg.de/mix01>

- 2-3 August
Tenth International Workshop on Matrices and Statistics, to be held in Voorburg, The Netherlands under the auspices of the International Statistical Institute.
Information: Patrick Groenen
 E-mail: matrix@fsw.leidenuniv.nl
 URL: <http://matrix.fsw.leidenuniv.nl/>

- 4-7 August
Third International Conference on Multiple Comparisons, to be held at Bethesda, Maryland, USA
Information: Ajit Tamhane, Peter Westfall.
 E-mail: ajit@iems.northwestern.edu
 or
westfall@ba.ttu.edu
 URL: <http://www.ba.ttu.edu/isqs/westfall/mcp2002.htm>

- 4-8 August
The European Young Statisticians Meeting Date, to be held in Jánška Dolina, Slovakia.
Information: Radoslav Harman.
 E-mail: Radoslav.Harman@fmph.uniba.sk
 URL: <http://www.uniba.sk/~ktpms/eysm/main.htm>

- 5-9 August
Joint Statistical Meetings, Atlanta, Georgia. Atlanta Marriott Marquis and the Atlanta Hilton & towers. Sponsored by ASA, ENAR, WNAR, IMS and SSC.
Informações: ASA, 1429 Duke Street, Alexandria VA 22314-3415
 (703) 684-1221;
 Email meetings@amstat.org

- 5-10 August
International Conference on Transport Survey Quality and Innovation, to be held in Kruger National Park, South Africa.
Information: ISCTSC, Institute of Transport Studies, University of Sydney, NSW 2006, Australia. Telephone: +61 2 9351 0010, Fax: +61 2 9351 0088.
 E-mail: peters@its.usyd.edu.au
 URL: www.its.usyd.edu.au/conferences_main.htm

- 6-10 August
International Conference on "Extremes in Theory and Practice", Leuven, Belgium.
Informações: Jan Beirlant, University Center of Statistics, Katholieke Universiteit Leuven, De Croylaan 52B, 3001 Heverlee, Belgium.
 Tel.: +32-16-322789;
 Fax: +32-16-322831;
 E-mail: jan.beirlant@wis.kuleuven.ac.be

- 12-14 August
International Conference on Industrial Mathematics, to be held in Madras, India.
Informações: Prof. S.N. Mahji, Secretary ICIM-2001, Department of Mathematics, Indian Institute of Technology, Madras, Chennai - 600 036, India.
 Tel: +91-44-4458476 (O), +91-44-4459476 (R)
 Fax: +91-44-2350509.
 E-mail: icim2001@iitm.ac.in

- 12-18 August
Fourth International Conference on Statistical Data Analysis based on the L₁-Norm and Related Methods - to be held at the University of Neuchâtel, Switzerland.
Informações: Prof. Yadolah Dodge, Conference Organizer Statistics Group, Case Postale 1825, CH-2002 Neuchatel.
 Tel: +41 32 718 13 80
 Fax: +41 32 718 13 81
 Email: yadolah@seco.unine.ch

- 13-18 August
23rd European Meeting of Statisticians, Funchal, Island of Madeira, Portugal.
Informações: Dinis Pestana, University of Lisbon and Rita Vasconcelos, University of Madeira.
 E-mail: dinis.pestana@fc.ul.pt
 rita@dragoeiro.uma.pt
 URL: <http://www.fc.ul.pt/cea/ems2001>

- 15-20 August
SRTL-2: The Second International Research Forum on Statistical Reasoning, Thinking, and Literacy, to be hosted by the Centre for Cognition Research in Learning and Teaching and the School of Curriculum Studies in the University of New England, Armidale, Australia.
Informações: Dr Chris Reading, Department of Curriculum Studies, University of New England, Armidale NSW 2351, Australia,
 Tel: +02-67735060,
 Fax: +02-67735078,
 Email: creading@metz.une.edu.au,
 URL: <http://www.beeri.org.il/SRTL/>

- 19-23 August
22nd Annual Conference of ISCB (The International Society for Clinical Biostatistics) will be held in Stockholm, Sweden, August 19 - 23, 2001.
Information: Scientific Secretariat.
 E-mail: Theresa.Westerstrom@iscb.stockholm2001.org
 URL: <http://www.iscb.stockholm2001.org/>.

- 21-25 August
ICANN 2001, International Conference on Artificial Neural Networks of the European Network Society, to be held at the Vienna University of Technology, Austria.
Informações: Conference Secretariat: ICANN 2001, Austrian Research Institute for Artificial Intelligence, Schottengasse 3, A-1010 Vienna, Austria.
 Email: icann@ai.univie.ac.at

- 22-29 August
International Statistical Institute, 53rd Biennial Session (includes meetings of the Bernoulli Society, The International Association for Statistical Computing, The International Association of Survey Statisticians, The International Association for Official Statistics, The International Association for Statistical Education), Seoul, Korea.
Informações: ISI Permanent Office, Prinses Beatrixlaan 428,
 P.O. Box 950, 2270 AZ Voorburg, The Netherlands.
 Tel.: +31-70-337-5737;
 Fax: +31-70-386-0025;
 E-mail: isi@cbs.nl
 or visit the Session website at <http://www.nso.go.kr/isi2001>

- 30-31 August
IAOS Satellite Meeting on Statistics for Information Society, to be held in Tokyo, Japan.
Informações: Akihito ITO, Japan Statistical Association, 2-4-6 Hyakunin-cho, Shinjuku-ku, Tokyo 169-0073, Japan.
 Tel: +81-3-5332-3151;
 Fax: +81-3-5389-0691;
 Email: jsa@jstat.or.jp or Ito@jstat.or.jp

- 30 August-1 September
International Conference on Statistical Challenges in Environmental Health Problems, to be held at the Soft Research Park, Fukuoka City, Japan.
Information: The Chairman, Organizing Committee, Takashi Yanagawa, Graduate School of Mathematics, Kyushu University, Fukuoka 812-8581, Japan.
 E-mail: yanagawa@math.kyushu-u.ac.jp

- 30 August-1 September
ICNCB – International Conference on New Trends in Computational Statistics with Biomedical Applications (ISI 2001 Satellite Meeting, co-sponsored by IASC), to be held at the Osaka University Convention Center, Osaka, Japan.
Informações: ICNCB Office, Division of Mathematical Science, Graduate School of Engineering Science, Osaka University. 1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan; Fax +81(6)6850-6496.
 Email: ICNCB@jscs.or.jp
 URL: <http://www.jscs.or.jp/ICNCB/>
- 1-4 September
The annual meeting of Japan Statistical Society will be held at Seinan Gakuin University.
Informações: URL: <http://sunyht2.ism.ac.jp>
- 6-12 September
International Association for MATHEMATICAL GEOLOGY 6th Int'l Conference Cancún, Mexico.
Informações: Gina Ross, Kansas Geological Survey.
 Email: aspiazu@kgs.ukans.edu
 URL: <http://www.kgs.ukans.edu/Conferences/IAMG>
- 17-19 September
Methodology and Statistics, to be held in Ljubljana, Slovenia at the Faculty of Social Science, University of Ljubljana, Kardeljeva pl. 5, Ljubljana.
Information: Anuska Ferligoj.
 E-mail: anuska.ferligoj@uni-lj.si
 URL: <http://vlado.fmf.uni-lj.si/trubar/preddvor/2001/>.
- 24-25 September
Statistical methods in biopharmacy, 4th international meeting: "Integrating issues of efficacy, safety and cost-effectiveness", to be held in Paris, France.
Informações: Jean Auclair, IRI Servier, 6 place des pléiades, 92415 Courbevoie cedex, France. Fax: 33 1 55 72 68 27.
 Email: sfds.2001@curie.net
 URL: <http://www.sfds.asso.fr/groupe/congresbiophar/congres2001.htm>
- 25-29 September
32nd European Mathematical Psychology Group Meeting, to be held in Lisbon, Portugal. Includes a workshop on Teaching and Training Mathematical Psychology in an Interdisciplinary and International Context. An Introductory Course on "Mathematical Psychology and Data Analysis" will be held on September 25th.
Information: Prof. Dr. Helena Bacelar-Nicolau, Tel: +351 21 793 45 54; Fax: +351 21 793 34 08.
 E-mail: hbacelar@fc.ul.pt
 or
empg2001@fpce.ul.pt
 URL: <http://correio.cc.fc.ul.pt/~cladlead/EMPG01.html>.

- 29-31 October

Statistics as bases of creation the economic policy and the economic development in the South-East Europe, to be held in Skopje, Republic of Macedonia.

Information: Mr. Sasho Kjosev - Faculty of Economics, University " Sts. Cyril and Methodius", Skopje, Republic of Macedonia or Mrs. Biljana Apostolovska - State Statistical Office of the Republic of Macedonia.

E-mail: skosev@eccf.ukim.edu.mk
or
biljanaa@stat.gov.mk

- 1-4 November

Euroworkshop on Statistical Modelling - Nonparametric Models, to be held in Schloss Hoehenried, Bernried, near Munich, Germany.

Information: Göran Kauermann (coordinator of the project) University of Glasgow, Dep of Statistics & Robertson Centre, Boyd Orr Building, Glasgow G12 8QQ.

E-mail: goeran@stats.gla.ac.uk
URL: <http://www.stat.uni-muenchen.de/euroworkshop>.

- 12-16 November

VIII Latin-American Congress in Probability and Mathematical Statistics, to be held at the University of Havana, Cuba.

Information: Gonzalo Perera (Chairman Program Committee), Pablo Olivares (Chairman Local Organizing Committee).

E-mail: gperera@fing.edu.uy
or
clapem@matcom.uh.cu
URL: <http://www.uh.cu/eventos/clapem/ehome.htm>.

- 14-16 November

The Federal Committee on Statistical Methodology, which is composed of the senior statisticians from several U.S. federal statistical agencies and is sponsored by the U.S. Office of Management and Budget is planning a research conference in Arlington, Virginia.

Information: The conference will feature papers and software demonstrations on topics related to a broad range of government statistical research interests.

URL: <http://www.fcsm.gov/>

- 21-22 November

9th Conference on National Accounting: the measurement of the new economy; Paris, France. Simultaneous translation French-English.

Information: Michel Boëda (INSEE) - Simultaneous translation French-English
E-mail: michel.boeda@insee.fr
URL: http://www.insee.fr/fr/av_service/colloques/cnat_accueil.html
or
http://www.insee.fr/en/av_service/colloques/cnat_accueil.html

□ 7-9 December

International Conference on "Characterization Problems and Applications", tentative Venue: Antalya, Turkey.

Information: Omer L. Gebizlioglu, Ankara, Turkey; N. Balakrishnan, McMaster University, Canada; Ismihan Bayramov, Ankara, Turkey.

E-mail: Omer.L.Gebizlioglu@science.ankara.edu.tr

bala@mcmail.cis.mcmaster.ca

Ismihan.Bayramov@science.ankara.edu.tr

□ 19-22 December

International Conference on Statistics, Combinatorics and Related Areas and The Eighth International Conference of the Forum for Interdisciplinary Mathematics, to be held at the University of Wollongong, Australia.

Information: Chandra M. Gulati, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia. Telephone:+61-2-4221-3836, fax:+61-2-4221-4845.

E-mail: chandra_gulati@uow.edu.au

or

cmg@uow.edu.au

URL: <http://www.uow.edu.au/informatics/maths/statconference>.

□ 20-22 December

Statistical Analysis for Global Environment, to be held at the Siam Intercontinental Hotel, Bangkok, Thailand.

Information: Dr.Supol Durongwatana.

E-mail: fcomsdu@phoenix.acc.chula.ac.th

□ 20-23 December

International Conference on History of Mathematical Sciences, to be held in Delhi, India.

Information: Dr. Y. P. SABHARWAL, Department of Mathematics & Statistics, Ramjas College, University of Delhi, Delhi 110 007, India; Tel : (011) 294 1119.

E-mail: ypsabharwal@yahoo.com

or

ichm2001rjc@yahoo.com

2002

- 15-18 January
First International ICSC Congress on Neuro-Fuzzy NF'2002 to be held at The Capitolio de la Habana, Cuba.
Informações: INTERNATIONAL COMPUTER SCIENCE CONVENTIONS
Head Office: 5101C-50 Street, Wetaskiwin AB, T9A 1K1, Canada
(Phone: +1-780-352-1912 / Fax: +1-780-352-1913)
Email: operating@icsc.ab.ca, or planning@icsc.ab.ca
URL: <http://www.icsc.ab.ca/NF2002.htm>, or <http://www.icsc.ab.ca/>

- 16-18 January
Food-Industry and Statistics, to be held in Villeneuve d'Ascq (LILLE), France.
Bât. EUDIL IAAL - Cité Scientifique, F 59655.
Information: E-mail: agrostat2002@eudil.fr
URL: <http://www.eudil.fr/~agrostat>.

- 4-8 February
ProbaStat 2002, the 4th International Conference on Mathematical Statistics, to be held at Smolenice Castle, Smolenice, Slovak Republic.
Information: E-mail: probastat@savba.sk
URL: http://www.um.savba.sk/lab_15/probastat.html.

- 2-5 June
Annual Meeting of the Statistical Society of Canada, Hamilton, Ontario, Canada.
Informações: Peter Macdonald, Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada.
E-mail: pdmamac@mcmaster.ca

- 17-20 June
MMR 2002, Third International Conference on Mathematical Methods in Reliability, to be held at the Norwegian University of Science and Technology, Trondheim, Norway.
Informações: Professor Bo Lindqvist, Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway. Tel.: +47-73 59 35 20 - Fax: +47-73 59 35 24.
E-mail: mmr2002@math.ntnu.no
URL: <http://www.math.ntnu.no/mmr2002/>

- 23-29 June
The 8th International Vilnius Conference on Probability Theory and Mathematical Statistics, Vilnius, Lithuania.
Informações: Professor Vytautas Statulevicius, Institute of Mathematics and Informatics, Akademijos str. 4, 2600 Vilnius, Lithuania.
E-mail: conf@ktl.mii.lt

- 2-5 July
MCQT'02 - First Madrid Conference on Queueing Theory, to be held at the Department of Statistics and OR, Faculty of Mathematics, University Complutense of Madrid, Spain.
Information: Jesus R. Artalejo.
 E-mail: mc_qt@mat.ucm.es
 URL: <http://www.mat.ucm.es/deptos/es/mcqt/conf.html>.
- 7-12 July
The Sixth International Conference on Teaching Statistics (ICOTS6), to be held in Durban, South Africa.
Information: Maria-Gabriella Ottaviani - IPC Chair; Brian Phillips - International Organizer; , Dani Ben-Zvi - IPC Scientific Secretary.
 E-mail: mariagabriella.ottaviani@uniroma1.it;
bphillips@swin.edu.au;
dani.ben-zvi@weizmann.ac.il.
 URL: <http://icots.itikzn.co.za/>.
- 21-26 July
IBC 2002 - International Biometric Conference 2002, to be held at the University of Freiburg, Germany.
Information: Chair: Robert Curnow; Chair Local Organizing Committee: Martin Schumacher.
 E-mail: r.n.curnow@reading.ac.uk
ms@imbi.uni-freiburg.de
 URL: <http://www.ibc2002.uni-freiburg.de/>.
- 4-9 August
Fourth International Conference on Statistical Data Analysis based on the L₁-Norm and Related Methods - to be held at the University of Neuchâtel, Switzerland.
Information: Prof. Yadolah Dodge, Conference Organizer Statistics Group, Case Postale 1825, CH-2002 Neuchatel. Phone +41 32 718 13 80 Fax +41 32 718 13 81.
 E-mail: Yadolah.Dodge@unine.ch
- 11-15 August
Joint Statistical Meetings, New York, Hilton and Sheraton New York.
 Sponsored by ASA, ENAR, WNAR, IMS, and SCC.
Informações: ASA, 1429 Duke St., Alexandria, VA 22314-3415;
 Tel. (703) 684-1221;
 Email meetings@amstat.org
- 19-23 August
24th European Meeting of Statisticians, Prague, Czech Republic.
Informações: Martin Janzura, Institute of Information Theory and Automation, POB 18, 182 08 Praha 8, Czech Republic.
 Tel: 420 2 6605 2572.
 Fax: 420 2 688 4903.
 Email: janzura@utia.cas.cz

- 25-28 August
International Conference on Improving Surveys (ICIS-2002), to be held at the University of Copenhagen.
Information: International Conference Services, P.O. box 41, Strandvejen 171, DK-2900 Hellerup, Copenhagen, Denmark. Telephone: +45 3946 0500, Fax +45 3946 0515.
 E-mail: ICIS2002@ics.dk

- 2-6 September
RSS 2002 Conference to be held at the University of Plymouth, Plymouth, England.
Information: The 2002 Conference of the Royal Statistical Society (4-6 September) will be preceded by short courses (2-3 September).
 E-mail: J.Stander@plymouth.ac.uk

- 13-17 November
International Conference on Questionnaire Development, Evaluation, and Testing, probably to be held in the southeastern United States.
Information: URL: <http://www.jpsm.umd.edu/>

- 28-30 December
International Conference on "Ranking and Selection, Multiple Comparisons, Reliability, and Their Applications". Tentative Venue: Hotel Savera, Chennai, Tamilnadu, India.
Organizers: bala@mcmaster.ca; NKannan@utsa.edu; H. N. Nagaraja, Ohio State University, <mailto:hnn@stat.ohio-state.edu>
Information: N. Balakrishnan, McMaster University; N. Kannan, University of Texas at San Antonio; H. N. Nagaraja, Ohio State University.
 E-mail: bala@mcmaster.ca
NKannan@utsa.edu
<mailto:hnn@stat.ohio-state.edu>

2003

- 10-20 August
International Statistical Institute, 54th Biennial Session (includes meetings of the Bernoulli Society, The Intern. Assoc. for Statistical Computing, The Intern. Assoc. of Survey Statisticians, The Intern. Assoc. for Official Statistics and The Intern. Assoc. for Statistical Education), to be held in Berlin, Germany.
Informações: ISI Permanent Office, Prinses Beatrixlaan 428, P.O. Box 950, 2270 AZ Voorburg, The Netherlands.
 Tel.: +31-70-337-5737;
 Fax: +31-70-386-0025;
 E-mail: isi@cbs.nl
 or visit the Session website at <http://www.isi-2003.de>

FUNDAMENTO, OBJECTO E ÂMBITO DA REVISTA

O INE, consciente de como uma cultura estatística é essencial para a compreensão da maioria dos fenómenos do mundo actual, e da sua responsabilidade na divulgação do conhecimento estatístico, fazendo-o chegar ao maior número possível de leitores, tendo reconhecido a necessidade de dar um passo nesse sentido, passou a editar quadrimestralmente a presente *Revista de Estatística* destinada a divulgar:

- a) Numa perspectiva científica, artigos originais sobre temas especializados da estatística, tanto pura como aplicada, bem como sobre estudos e análises nos domínios económico, social e demográfico;
- b) Informações sobre actividades e projectos importantes do Sistema Estatístico Nacional;
- c) Informações sobre acções desenvolvidas pelo INE no âmbito da cooperação.
- d) Informações sobre congressos, seminários, colóquios e conferências de interesse estatístico ou afim;

Para tal, são adoptadas as seguintes formas de contribuição para publicação na Revista:

- Quanto aos artigos referidos em a), contribuições da *iniciativa* dos próprios autores e por *convite* do Conselho Editorial, pertencentes ou não ao INE;
- Quanto às informações referidas em b), c) e d), contribuições dos departamentos do INE.

As contribuições de artigos por iniciativa dos próprios autores serão objecto de avaliação de mérito científico pelo Conselho Editorial, que decidirá ou não pela sua publicação.

Para a elaboração e envio das contribuições de artigos para publicação na Revista são adoptadas as *Normas de Apresentação de Originais* que figuram na última página.

Os autores dos artigos publicados, a que se refere a alínea a), receberão uma contribuição financeira paga pelo INE, de montante a fixar por despacho da Direcção mediante proposta do Director da Revista.

**OS PONTOS DE VISTA EXPRESSOS PELOS AUTORES DOS ARTIGOS PUBLICADOS NA REVISTA
NÃO REFLECTEM NECESSARIAMENTE A POSIÇÃO OFICIAL DO INE.**

FOUNDATION, SUBJECT MATTER AND SCOPE OF THE REVIEW

INE is conscious of how statistical awareness is essential to the understanding of the majority of phenomena in the present world and is aware of its responsibility to disseminate statistical knowledge, making it available to the widest possible range of readers. INE has recognised the need to take a step in that direction and will begin publication of this *Statistical Review* three times yearly, designed to provide the following:

- a) Within a scientific perspective, original articles on specialised areas of statistics, both pure and applied, as well as studies and analyses within the sphere of economics, social issues and demographics;
- b) Information on activities and projects of the National Statistical System;
- c) Information on activities developed by INE within the scope of co-operation;
- d) Information on congresses, seminars and conferences of a statistical or related nature;

The following approaches for contributing material for publication in the review have been adopted:

- In relation to the articles referred to in section a), contributions are made by the authors themselves and by invitation of the Editorial Committee, whether they are employees of INE or not;
- In relation to the information referred to in section b), c) and d); contributions are from departments of INE.

The Editorial Committee who has sole discretion in deciding whether or not the material will be published will assess the scientific merit of contributions made on the initiative of the authors themselves.

The preparation and delivery of material for publication in the Review are subject to the *Rules for Submitting Originals* presented on the last page.

The authors of the published articles referred to in section a) will receive pecuniary compensation from INE in an amount to be determined by resolution of the Board on the recommendation of the Director of the Review.

**THE VIEWPOINTS EXPRESSED BY THE AUTHORS OF THE ARTICLES PUBLISHED IN THE REVIEW
DO NOT NECESSARILY REFLECT THE OFFICIAL POSITION OF I.N.E.**

NORMAS DE APRESENTAÇÃO DE ORIGINAIS

Nos termos do *Regulamento da Revista de Estatística*, o Conselho Editorial aprovou as seguintes **Normas de Apresentação de Originais**:

1. Os originais dos artigos serão enviados ao Director da Revista pelos respectivos autores, devendo ser escritos em *português* e não terem sido ainda totalmente publicados, ou estar em processo de edição em outra publicação.
2. Poderão também ser apresentados artigos escritos em *inglês*, cabendo ao Director da Revista a decisão sobre a sua aceitação.
3. Quanto à *avaliação do mérito científico* dos artigos:
 - a) Os artigos apresentados por *iniciativa* dos respectivos autores serão submetidos à avaliação do mérito científico pelo Conselho Editorial, com garantia do anonimato, tanto do autor como dos avaliadores;
 - b) Os autores receberão a informação sobre o resultado da avaliação num prazo máximo de trinta dias, com indicação, nos casos de avaliação positiva, do número da *Revista* em que serão publicados, e nos casos de avaliação negativa com a devolução do original apresentado.
4. Os artigos aceites para publicação na *Revista de Estatística* serão igualmente divulgados no *site* do INE na *Internet*.
5. Os originais, com uma extensão não superior a trinta páginas, serão processados em *Word for Windows*, integralmente a preto e branco, com indicação do(s) software(s) adicional(ais) eventualmente utilizado(s) na produção do documento original, e entregues em suporte papel acompanhado da respectiva *disquette*, ou enviados por E-mail para o seguinte endereço: liliana.martins@ine.pt
6. Na apresentação dos originais, os autores respeitarão ainda as seguintes normas:
 - 6.1. Quanto à *estrutura*:
 - a) O texto deve ser processado em formato *A4*, com utilização do tipo de letra *Times New Roman 11*, espaçamento *at least 12*, e com as seguintes margens: *top*: 4 cm, *bottom*: 3 cm, *left*: 2,5 cm, *right*: 5 cm, *header*: 1,25cm, *footer*: 1,25cm;
 - b) A primeira página conterà exclusivamente o título do artigo, bem como o nome, morada e telefone, fax e E-mail do autor, com indicação das funções exercidas e da instituição a que pertence, devendo, no caso de vários autores, ser indicado a quem deverá ser dirigida a correspondência da Revista;
 - c) A segunda página conterà, em português e inglês, unicamente o título e um *resumo* do artigo, com um máximo de 100 palavras,

seguido de um parágrafo com indicação de *palavras-chave* até ao limite de 15;

- d) Na terceira página começará o texto do artigo, sendo as suas eventuais secções ou capítulos numeradas sequencialmente;

6.2. Quanto a *referências bibliográficas*:

- a) Os autores eventualmente citados no texto do artigo serão indicados entre parênteses curvos pelo seu nome seguido da data da respectiva publicação e, se for caso disso, do número de página (p. ex.: Malinvaud, 1989, 23);
- b) As Referências Bibliográficas serão listadas, por ordem alfabética dos apelidos dos respectivos autores, imediatamente a seguir ao final do texto, de acordo com a fórmula seguinte:

GREENE, W. H., “*Econometric Analysis*”, Prentice-Hall, New Jersey, 1993.

6.3. Quanto à *revisão de provas e publicação*:

- a) Uma vez aceite o artigo e antes da sua publicação, receberá o autor provas para revisão, as quais serão devolvidas ao Director da Revista no prazo máximo de uma semana contado da data da sua recepção;
- b) Serão da responsabilidade dos respectivos autores as consequências de eventuais modificações da versão inicial aceite, bem como de atrasos na revisão das provas, que impossibilitem a publicação no número da Revista previsto, reservando-se o Director o direito de decidir a data da sua publicação futura;
- c) Uma vez publicado o artigo, o autor receberá vinte exemplares da sua versão impressa e um exemplar do respectivo número da *Revista*.

7. Para *informações adicionais* contactar o Secretariado de Redacção:

Euarda Liliana Martins
Instituto Nacional de Estatística
Av.^a. António José de Almeida, n.º 5 – 9.^o.
1000-043 Lisboa - Portugal

- ☐ Tel.: +351 21 842 62 05
- ☐ Fax.: +351 21 842 63 66
- ☐ e-mail: liliana.martins@ine.pt

RULES FOR SUBMITTING ORIGINALS

Within the terms of the *Regulation of the Statistical Review*, the Editorial Committee has approved the following **Rules for Submitting Originals**:

1. The original articles will be sent to the Review Director by the respective authors. They should be written in *Portuguese*, they should not have already been published in their entirety nor should they be in the process of being published in any other publication.
2. Articles may also be submitted in *English* to the Review's Director who will decide whether to accept them.
3. In relation to the *evaluation of the scientific merit* of the articles:
 - a) The Editorial Committee will assess the articles submitted on the initiative of the authors on the basis of their scientific merit. The identity of both the author and the Committee members will be strictly confidential;
 - b) The authors will receive information regarding the results of the evaluation of scientific merit within a maximum period of 30 days. If the article is accepted, the Committee will indicate the issue number of the *Review* in which the article will be published. If the article is not accepted, the original will be returned to the author.
4. The articles accepted for publication in the *Statistical Review* will also be made public on the Internet site of the INE.
5. The original articles having no more than thirty pages must be processed in *Word for Windows*, completely at black and white, with the information on the additional(s) software(s) eventually used in the production of the original document, and they will be delivered in hard copy as well as on diskette, or sent by E-mail to: liliana.martins@ine.pt
6. With the presentation of the original articles, the authors must also respect the following rules:
 - 6.1 In relation to the *structure*:
 - a) The text shall be printed on A4 format paper utilising the font *Times New Roman* size 11, spacing at least 12, and with the margins: *top* 4cm, *bottom* 3cm, *left* 2,5cm, *right* 5cm, *header* 1,25cm, *footer* 1,25cm;
 - b) The first page shall contain only the title of the article as well as the name, address and telephone, fax and E-mail number of the author, indicating the position held and the institution that he/she belongs to. In the case of various authors, it is necessary to indicate the person to whom all correspondence received should be forwarded;

- c) The second page shall contain in *Portuguese* and *English* only the *title* and an *abstract* of the article with the maximum of 100 words followed by a paragraph indicating *key words* up to the limit of 15;
- d) The third page will begin the text of the article with its respective sections or chapters sequentially numbered;

6.2 Regarding *Bibliographical References*:

- a) Authors who are cited in the text of the article shall be indicated in parentheses with their name followed by the date of the respective publication and, if necessary, the page number (ex.: Malinvaud, 1989, 23);
- b) All bibliographical references will be listed in alphabetical order by the surnames of the respective authors, immediately following the end of the text, as in the following example:

GREENE, W. H., "*Econometric Analysis*", Prentice-Hall, New Jersey, 1993.

6.3 Regarding *proof-reading and publication*:

- a) Once the article is accepted and prior to its publication, the author will receive a copy for review. These copy will be returned to the Director of the Review within a maximum period of one week from the date of its reception;
- b) The consequences of subsequent changes to the accepted first version are the responsibility of the respective authors as well as any delays in proof-reading that make its publication in the planned issue of the Review impossible. The Director reserves the right to decide upon the date for future publication;
- c) Once the article is published, the author will receive twenty copies of his/her printed version and a copy of the respective issue of the *Review*.

7. For *further information* kindly contact the Editorial Secretary:

Eduarda Liliana Martins
Instituto Nacional de Estatística
Av^a. António José de Almeida, nº. 5 – 9º.
1000-043 Lisbon - Portugal

- ☐ Tel.: +351 1 21 842 62 05
- ☐ Fax.: +351 1 21 842 63 66
- ☐ e-mail: liliana.martins@ine.pt