



INSTITUTO NACIONAL DE ESTATÍSTICA

PORTUGAL

REVISTA DE ESTATÍSTICA

STATISTICAL REVIEW



CATALOGAÇÃO RECOMENDADA:

REVISTA DE ESTATÍSTICA. Lisboa, 1996-
Revista de estatística / ed. Instituto Nacional de
Estatística. - Vol. 1, 1º quad. 1996-
Lisboa : I.N.E., 1996- . - 30 cm
Quadrimestral
ISSN 0873-4275

FICHA TÉCNICA

- | | |
|--|---|
| - DIRECTOR | - EDITOR |
| - <i>Adrião Simões Ferreira da Cunha</i> | - <i>Instituto Nacional de Estatística</i>
<i>Av. António José de Almeida, n.º 2</i>
<i>1 000-043 LISBOA</i>
<i>Tel.: 21 842 61 00</i>
<i>Fax: 21 842 63 64</i> |
| - DIRECTOR-ADJUNTO | - CAPA |
| - <i>Pedro Jorge Nunes da Silva Dias</i> | - <i>Design de Mário Bouçadas sobre o vitral do INE da</i>
<i>autoria do pintor Abel Manta</i> |
| - CONSELHO EDITORIAL | - LAYOUT E MAQUETAGEM |
| - <i>Adrião Simões Ferreira da Cunha</i> | - <i>Mário Bouçadas</i> |
| - <i>António Daniel Correia dos Santos</i> | - IMPRESSÃO |
| - <i>Dinis Duarte Ferreira Pestana</i> | - <i>Instituto Nacional de Estatística</i>
<i>Secção de Artes Gráficas</i> |
| - <i>Francisco José Neto Melro</i> | - TIRAGEM |
| - <i>João António Branco</i> | - <i>750 exemplares</i> |
| - <i>João Ferreira do Amaral</i> | - DEPÓSITO LEGAL |
| - <i>Óscar Soares Barata</i> | - <i>N.º 99514/96</i> |
| - <i>Pedro Jorge Nunes da Silva Dias</i> | |
| - <i>Pedro Miguel Girão Nogueira Ramos</i> | |
| - <i>Sérgio Manuel Bacelar e Silva</i> | |
| - SECRETARIADO DE REDACÇÃO | |
| - <i>Eduarda Liliana Marques Martins</i> | |

PREÇO (IVA 5% incluído)

- | | | |
|--------------------|-----------|---------|
| - N.º avulso | 2 600\$00 | € 129,7 |
| - Assinatura anual | 6 200\$00 | € 30,9 |

© INE, Lisboa, Portugal, 2000* Reprodução autorizada, excepto para fins comerciais, com indicação da fonte bibliográfica.

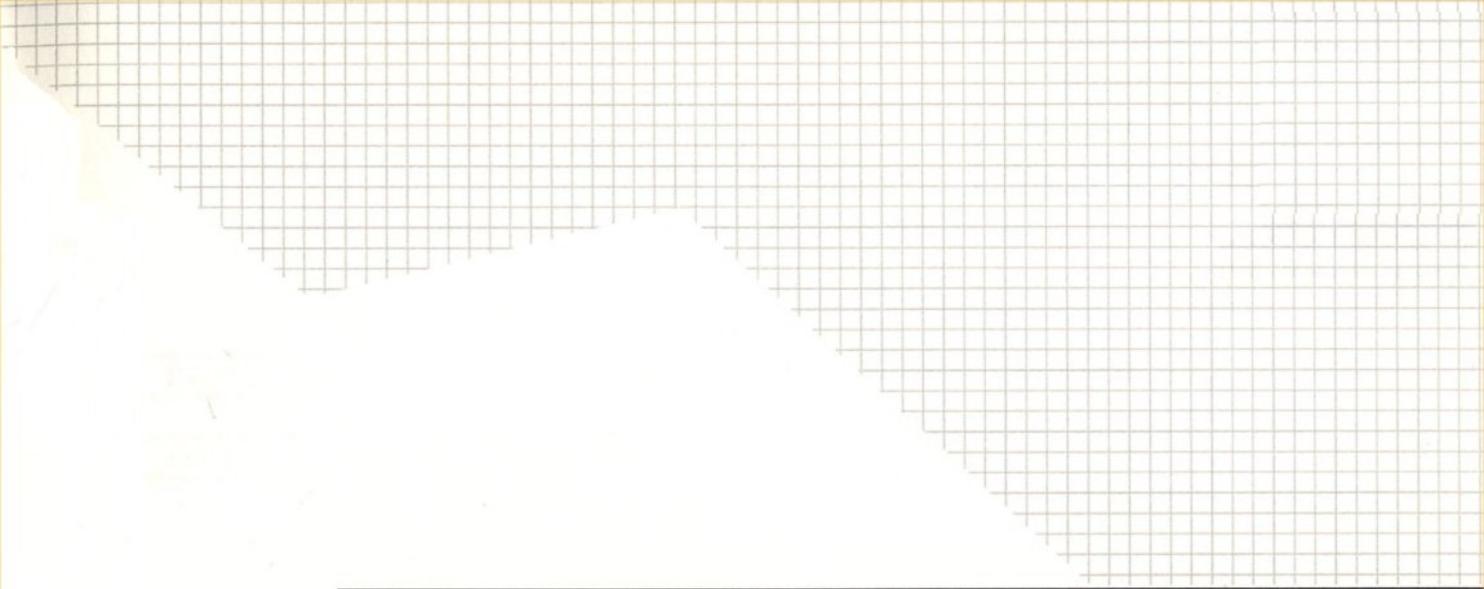
CREDITS

- | | |
|--|---|
| - DIRECTOR | - PUBLISHER |
| - <i>Adrião Simões Ferreira da Cunha</i> | - <i>Instituto Nacional de Estatística</i>
<i>Av. António José de Almeida, n.º 2</i>
<i>1 000-043 LISBOA</i>
<i>Tel.: 21 842 61 00</i>
<i>Fax: 21 842 63 64</i> |
| - ASSISTANT DIRECTOR | - COVER DESIGN |
| - <i>Pedro Jorge Nunes da Silva Dias</i> | - <i>Designed by Mário Bouçadas on the stain glass</i>
<i>window at INE by the painter Abel Manta</i> |
| - EDITORIAL BOARD | - LAYOUT AND GRAPHIC DESIGN |
| - <i>Adrião Simões Ferreira da Cunha</i> | - <i>Mário Bouçadas</i> |
| - <i>António Daniel Correia dos Santos</i> | - PRINTING |
| - <i>Dinis Duarte Ferreira Pestana</i> | - <i>Instituto Nacional de Estatística</i>
<i>Secção de Artes Gráficas</i> |
| - <i>Francisco José Neto Melro</i> | - EDITION |
| - <i>João António Branco</i> | - <i>750 exemplares</i> |
| - <i>João Ferreira do Amaral</i> | - LEGAL DEPOSIT REGISTRATION |
| - <i>Óscar Soares Barata</i> | - <i>N.º 99514/96</i> |
| - <i>Pedro Jorge Nunes da Silva Dias</i> | |
| - <i>Pedro Miguel Girão Nogueira Ramos</i> | |
| - <i>Sérgio Manuel Bacelar e Silva</i> | |
| - EDITORIAL SECRETARIAT | |
| - <i>Eduarda Liliana Marques Martins</i> | |

PRICE (IVA 5% included)

- | | | |
|-----------------------|-----------|---------|
| - Single issue | PTE 2,600 | € 129,7 |
| - Annual subscription | PTE 6,200 | € 30,9 |

© INE, Lisbon, Portugal, 2000* Reproduction authorised, except for commercial purposes by indicating the source.



REVISTA DE ESTADÍSTICA

STATISTICAL REVIEW

ÍNDICE

INDEX

- ARTIGOS	
ARTICLES	5
TAXA NATURAL DE DESEMPREGO, UMA VARIÁVEL CÍCLICA? - UMA APLICAÇÃO A ALGUMAS ECONOMIAS EUROPEIAS <i>THE NATURAL RATE OF UNEMPLOYMENT, A CYCLICAL VARIABLE? - A STUDY APPLIED TO SOME EUROPEAN ECONOMIES</i>	
Por/By: <i>Maria Adelaide Silva Duarte e João Sousa Andrade</i>	7
DECOMPOSIÇÃO E EVOLUÇÃO DA DESIGUALDADE SALARIAL <i>THE SOURCES OF INTERREGIONAL WAGE DIFFERENTIALS</i>	
Por/By: <i>Filipe Santos e Paulino Teixeira</i>	35
MAJOR GUIDELINES TOWARDS THE DEVELOPMENT OF A PORTUGUESE SAM <i>PRINCIPAIS ORIENTAÇÕES PARA O DESENVOLVIMENTO DE UMA MATRIZ DE CONTABILIDADE SOCIAL</i>	
Por/By: <i>António Machado Lopes</i>	71
ESTIMAÇÃO DE ESTADO EM MODELOS LINEARES POR BOCADOS EM TEMPO DISCRETO: COMPARAÇÃO NUMÉRICA DE DOIS PROCEDIMENTOS <i>DISCRETE-TIME PIECEWISE LINEAR FILTERING: NUMERICAL COMPARATION BETWEEN TWO PROCEDURES</i>	
Por/By: <i>Rita Pinto de Sousa</i>	85
UM ESTUDO SOBRE A SATISFAÇÃO NO TRABALHO - UMA APLICAÇÃO DO MODELO LOGIT ORDENADO <i>ORDERED LOGIT MODELS: A STUDY ON JOB SATISFACTION</i>	
Por/By: <i>Paulo Costa Parente</i>	115
INFORMAÇÕES	
INFORMATIONS	137
ACTIVIDADES E PROJECTOS IMPORTANTES DO SISTEMA ESTATÍSTICO NACIONAL <i>IMPORTANTS ACTIVITIES AND PROJECTS OF THE NATIONAL STATISTICAL SYSTEM</i>	139

ACÇÕES DESENVOLVIDAS PELO INE NO ÂMBITO DA COOPERAÇÃO <i>ACTIONS ACHIEVED BY NSI IN THE SCOPE OF COOPERATION</i>	145
CONGRESSOS, SEMINÁRIOS, COLÓQUIOS E CONFERÊNCIAS <i>CONGRESS, SEMINARS AND CONFERENCES</i>	149
FUNDAMENTO, OBJECTO E ÂMBITO DA REVISTA <i>FOUNDATION, SUBJECT MATTER AND SCOPE OF THE REVIEW</i>	159
NORMAS DE APRESENTAÇÃO DE ORIGINALS PARA A REVISTA <i>RULES FOR SUBMITTING ORIGINALS TO THE REVIEW</i>	161

TAXA NATURAL DE DESEMPREGO, UMA VARIÁVEL CÍCLICA? - UMA APLICAÇÃO A ALGUMAS ECONOMIAS EUROPEIAS

THE NATURAL RATE OF UNEMPLOYMENT, A CYCLICAL VARIABLE? - A STUDY APPLIED TO SOME EUROPEAN ECONOMIES

Autores: Maria Adelaide Silva Duarte

- Professora Auxiliar na Faculdade de Economia da Universidade de Coimbra

João Sousa Andrade

- Professor Catedrático na Faculdade de Economia da Universidade de Coimbra

RESUMO:

- Este artigo apresenta algumas das conclusões que obtivemos no âmbito de uma investigação mais vasta sobre a Taxa Natural de Desemprego (NRU). O nosso objectivo principal é determinar as propriedades da NRU como um indicador económico e testámo-las para a França, Espanha, Itália, Áustria e Dinamarca.

Primeiro revemos o conceito de NRU. Em segundo lugar comparámos os dois tipos mais importantes de modelos da NRU: o modelo de Phillips, actualizado por Gordon (1997) e o modelo de Layard, Nickell and Jackman (1991) na versão de Blanchard (1999). Os factores económicos e sociais que conduzem à existência de uma NRU de médio prazo variável ou à inexistência de uma NRU de longo prazo são realçados.

Em terceiro lugar efectuamos a análise econométrica para testar o comportamento cíclico da NAIRU. Depois de justificarmos a escolha da amostra de países europeus, utilizamos uma versão não linear do método de Elmeskov. Como suspeitávamos, obtivemos modelos cíclicos robustos da NAWRU.

PALAVRAS-CHAVE:

- *Curva de Phillips; NRU; NAWRU de curto e longo prazo; histerese; indicador de política económica; método de Elmeskov.*

ABSTRACT:

- This paper presents some of the conclusions reached by us in the framework of a broader research on the Natural Rate of Unemployment (NRU). Our main aim is to ascertain the NRU properties as an economic indicator, and we have tested its properties for France, Spain, Italy, Austria and Denmark.

First we review the concept of NRU. Secondly we compare the two most important types of NRU models: the Phillips model, updated by Gordon's model (1997) and Layard, Nickell and Jackman's model (1991) updated by Blanchard's version (1999). The economic and social factors leading to the existence of a short or medium-run varying NRU or to the non-existence of a long-run NRU are emphasized. Thirdly, we make the econometric analysis to confirm the cyclical behavior of NAIRU. After justifying the choice of the European countries sample, we use a non-linear version of Elmeskov method. As we suspected, we have obtained robust cyclical models for the NAWRU.

KEY-WORDS:

- *Phillips curve; NRU; short and long run NAWRU; hysteresis; policy indicator; Elmeskov method*

Este é o terceiro estudo¹ que resulta da nossa investigação sobre o conceito de taxa natural de desemprego. Esta investigação, que apresenta três componentes, a análise teórica, a análise empírica e a análise normativa, privilegiou em termos de objectivo de pesquisa, a questão normativa. Quais são as condições que devem estar presentes para podermos considerar a taxa natural de desemprego um “bom” (passe a tautologia) indicador de política económica?

Pareceu-nos que este objectivo era tanto mais importante quanto a maior parte de estudos de política económica em Portugal continua a considerar, directa ou indirectamente, a taxa natural de desemprego um indicador de política económica pertinente, partindo do pressuposto que as condições para o ser estão preenchidas². Ora é preciso provar que essas condições são cumpridas, e nós chegámos à conclusão que o não são no caso da economia portuguesa³.

Valerá como desculpa à posição dominante o facto de Portugal apresentar uma das taxas de desemprego mais baixas da União Europeia (UE) e da dita taxa natural de desemprego, por factores de inércia e histerese, ser mais baixa do que deveria ser na ausência desses fenómenos, justamente o oposto do que se passa na maioria dos Estados-membros da UE?

Não o cremos. Não conhecemos o valor da NAIRU, mas se esse valor for independente do ciclo económico, poderemos dizer que variações da taxa de desemprego efectivo à esquerda de dado valor estarão associadas a um aumento da inflação e a uma diminuição no caso oposto. Mas quando a taxa natural de desemprego depende da componente cíclica do PIB real e apresenta mesmo um comportamento contra-cíclico, uma diminuição da taxa de desemprego efectivo estará associada a uma taxa natural de desemprego que também diminuiu. Nestas condições, a variação da taxa de desemprego efectivo não nos pode dar uma indicação sobre o sinal da variação da inflação. Assim, uma política económica restritiva visando a estabilidade dos preços pode ter custos reais muito elevados, porque a estabilidade dos preços antes da intervenção não estava ameaçada.

Neste terceiro estudo alargámos a nossa investigação a um conjunto de Países da União Europeia⁴ e chegámos à mesma conclusão que para Portugal - a taxa natural de desemprego depende também da componente cíclica do PIB real, ou seja, varia de forma contra-cíclica.

Começaremos por apresentar o conceito de taxa natural de desemprego (I) e procuraremos inserir o seu cálculo e definição em modelos estilizados (II). Faremos depois o estudo empírico (III) das economias seleccionadas, apresentaremos a metodologia seguida e os resultados para cada uma delas. Finalmente concluiremos (IV).

I. O CONCEITO DE TAXA NATURAL DE DESEMPREGO

A primeira questão a colocar sobre a taxa natural de desemprego deve ser como defini-la. Começemos pela longa definição de Friedman (1968):

¹ Duarte, A. e Andrade, J. 1999 e 2000.

² Apesar da contestação a nível conceptual de Amaral, João F. 1995 e dos resultados de Modesto, L. e J. C. Neves 1993.

³ Outros autores chegaram a essa conclusão para o caso das outras economias da União Europeia.

⁴ França, Espanha, Áustria, Dinamarca e Itália.

“ The natural rate of unemployment is the level which would be ground out by the walrasian system of the general equilibrium equations, provided that there is in them the actual structural characteristics of the labour and commodities market, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on.”

As características do mercado de trabalho fazem com que este apresente sempre muitos indivíduos em trânsito, daí decorre que pleno-emprego não signifique ausência de desemprego. E é assim que temos o desemprego friccional e o desemprego estrutural.

O desemprego friccional é o desemprego que está associado à rotação do trabalho. Há rotação permanente de indivíduos em dados empregos, surgem novas oportunidades de emprego; há empregos que deixam de existir, há um fluxo de entrada e saída permanente de indivíduos na população activa. Mesmo supondo que o número de postos de trabalho oferecidos é igual ao número de postos de trabalho procurados, a procura desses postos de trabalho é uma actividade consumidora de tempo, o que significa que seja qual for o momento, haverá trabalhadores desempregados à procura de emprego.

O desemprego estrutural resulta da não adequação entre a estrutura da oferta de trabalho e a estrutura da procura de trabalho, supondo que a oferta de trabalho é igual à sua procura. Este tipo de desemprego pode ocorrer porque existe uma oferta de qualificações inadequada do ponto de vista geográfico, apesar de poder ser adequada a nível nacional.

O conceito de taxa natural de desemprego emana de uma teoria real do emprego que se baseia na tese clássica da dicotomia entre as variáveis reais e as variáveis nominais de equilíbrio de longo-prazo. O conceito de taxa natural de desemprego foi integrado pela primeira vez no seio de um modelo aceleracionista com antecipações inflacionistas adaptáveis (Humphrey, T. 1985 e 1985a e Solow, R. 1998). Neste modelo, a curva de Phillips de curto-prazo aumentada exprime uma relação de desequilíbrio entre a taxa de desemprego efectiva e a taxa natural de desemprego, o que se traduz numa taxa de inflação instável. Neste contexto fica claro o acrónimo escolhido para taxa natural de desemprego - NAIRU, taxa de desemprego que não acelera a inflação, embora o acrónimo mais correcto tivesse sido, CIRU, taxa de desemprego que mantém constante a taxa de inflação.

Consideraremos, a partir daqui, que a taxa natural de desemprego é equivalente à NAIRU. Refira-se no entanto que a NAIRU pode ser considerada como apenas tendo existência no quadro da concorrência perfeita (Lipsey, R. e Chrystal, A. 1995).

As questões que podem e devem ser colocadas a propósito do conceito de NAIRU têm uma natureza diferente, o que não quer dizer que sejam independentes entre si. A primeira questão releva da análise teórica e respeita ao quadro em que a NAIRU é definida (Hahn, F. 1995). A segunda questão releva da análise empírica e é relativa à robustez dos resultados empíricos a que conduzem os vários métodos da sua estimação. Finalmente, a terceira releva da análise normativa e é relativa à consideração ou não da NAIRU como indicador de política económica.

Ponhamos em destaque algumas das interrogações que devem ser colocadas no âmbito da primeira questão acima apontada. Qual é o quadro teórico no qual é definida a NAIRU e como é que definido o equilíbrio a ela associado? Trata-se de um equilíbrio estacionário globalmente estável para o qual convergirá a taxa de desemprego efectivo? Ou pelo contrário, trata-se de um equilíbrio de longo-prazo a taxa constante (equilíbrio de steady state), localmente estável, para o qual converge a taxa de desemprego efectiva? E a dicotomia clássica é assegurada sob aquelas condições? Ou seja, trata-se de uma taxa de desemprego cuja trajetória é independente das variáveis nominais que influenciam a procura global?

A extensão da análise da NAIRU através da consideração do quadro da concorrência imperfeita nos mercados de bens e do trabalho exigirá a consideração de fenómenos não-lineares (Amable, B.; Henry, J.; Lordon, F.; et Topol, R. 1995) e a hipótese de agentes heterogéneos (Caballero, R. e Engel, E. 1993)? E sendo assim, será que a filiação da NAIRU à teoria clássica se perderá?

Na ausência de choques de oferta, de choques de procura e de choques de salários, as determinantes da NAIRU são de três tipos: os factores de oferta que determinam o nível de actividade de equilíbrio; os factores de histerese; e os factores de inércia. A existência de factores de histerese pode por em causa o próprio conceito de NAIRU (Cahuc, P. 1993)⁵.

No caso da histerese parcial, são definidos dois conceitos de NAIRU, um conceito de curto-prazo e um conceito de longo-prazo. A pertinência do último conceito dependerá da existência de um mecanismo de convergência da NAIRU de curto-prazo à NAIRU de longo-prazo. Mas do ponto de vista normativo, do ponto de vista da política económica, é necessário também que a duração dos efeitos transitórios da taxa de desemprego efectivo sobre a taxa de desemprego de longo-prazo seja curta.

No caso da histerese total, a taxa de salário nominal deixa de depender da taxa de desemprego, a NAIRU deixa de poder ser definida como uma variável de equilíbrio, e coincide mesmo com a taxa de desemprego do período anterior. Como provámos (Duarte, A., e Andrade, J. 1999), as séries do desemprego e da taxa de desemprego em Portugal apresentam raiz unitária, pelo que a determinação da NAIRU nestas condições é impossível.

Mesmo que se considere que ambos os conceitos de histerese não exprimem senão um fenómeno de persistência, a duração dos efeitos, bastante longa, pode invalidar, quer do ponto de vista teórico, quer do ponto de vista empírico, a hipótese da NAIRU.

Blinder considera que uma teoria que permite que a taxa natural de desemprego siga de forma clara a taxa de desemprego efectivo não poderá ser considerada uma teoria da taxa natural de desemprego (Blinder, A. 1987). Rod Cross considera que a aplicação do verdadeiro conceito de histerese à NAIRU supõe o abandono desta hipótese:

“The present chapter has argued, along such lines, that the presence of hysteresis produces a time path for equilibrium unemployment which is inconsistent with the natural rate hypothesis. This conclusion follows from the recognition that hysteresis is a property of non-linear systems with heterogeneous micro elements. Such systems retain a memory of the non-dominated extremum values of shocks experienced, including temporary shocks to aggregate demand. Thus equilibrium unemployment rates are shaped, inter alia, by nominal variables, which contradicts the natural rate hypothesis” (Cross, R. 1995, p. 197)

II. MODELOS TEÓRICOS DE DETERMINAÇÃO DA NAIRU

A determinação da NAIRU poderá ser feita a partir de várias abordagens, no entanto as mais divulgadas são aquelas que se baseiam numa relação de Phillips aumentada e no modelo de Layard, Nickell e Jackman (1991). Um tipo de abordagem foi desenvolvida por Gordon (1997), e o seu modelo é conhecido pelo nome do modelo triangular de inflação (inflation triangle model) porque nele a inflação é explicada por três grupos de variáveis: as da procura, definidas pelo desvio entre a taxa de desemprego efectiva e a taxa natural de

⁵ Veja-se também a abundante bibliografia aí citada.

desemprego; as da oferta, que representam choques de oferta; e as da inércia, que são representadas pelos efeitos com desfasamentos da taxa de inflação. Um segundo modelo, é por vezes denominado o modelo da batalha das taxas de margem (the battle of the mark-up model), supõe um contexto de concorrência imperfeita no mercado dos bens e do trabalho. A taxa de desemprego é a variável estratégica que reconcilia o nível do preço de produção, pré-estabelecido pelas empresas, com a taxa de salário nominal. Neste modelo, a NAIRU é definida como sendo a taxa de desemprego quando os efeitos de inércia sobre o equilíbrio dos dois mercados se esgotam. A versão mais conhecida deste modelo não considera os efeitos temporários (mais ou menos longos) das alterações tecnológicas sobre a NAIRU, já que a produtividade tendencial é considerada com sinais contrários nas equações de preços e do salário. Podemos no entanto introduzir a influência dessa variável na análise se seguirmos a via proposta por Blanchard et Katz (1999). Aliás, essa via vai-lhes permitir não só reconciliar este segundo modelo com a abordagem dos autores da curva de salário (Blanchflower, D., e Oswald, A. 1994), como também interpretar com o auxílio das equações de preços e salário deduzidas, os resultados obtidos com a curva de Phillips para os EUA e para a UE. Assim, através da consideração dos efeitos directos positivos da produtividade do trabalho sobre o salário de reserva e sobre o salário, a influência dos choques de oferta sobre a taxa natural de desemprego pode ser considerada.

Analisemos o modelo mais simples de NAIRU de Layard, Nickell e Jackman. O modelo pode ser apresentado sob a forma de um sistema de três equações, a do preço, a do salário e a da procura, estando todas as variáveis definidas em logaritmos⁶.

$$p - w = a_0 - a_1 \cdot u + a_2 \cdot (p - p^e) - a_3 \cdot (k - 1) \quad (1)$$

$$w - p = b_0 - b_1 \cdot u + b_2 \cdot (p - p^e) + b_3 \cdot (k - 1) + z \quad (2)$$

$$y = v - p \quad (3)$$

A equação (1) representa o comportamento de mark-up do preço (do valor acrescentado) em relação ao salário nominal e que depende negativamente da taxa de desemprego, positivamente das surpresas dos preços e negativamente do coeficiente capital-trabalho (tomado aqui como uma “proxy” da produtividade tendencial).

A equação (2) representa a taxa de margem do salário nominal relativamente ao preço da produção e que depende negativamente da taxa de desemprego, positivamente das surpresas dos preços, do coeficiente capital trabalho e das variáveis exógenas que fazem pressão sobre a taxa de salário real, por exemplo, o grau de sindicalização dos trabalhadores e os subsídios de desemprego.

Finalmente a equação (3) retrata o comportamento da procura onde v representa as variáveis que fazem pressão sobre a procura.

⁶ Usaremos a seguinte simbologia para as variáveis em logaritmos: p - preços; w - salário; u - taxa de desemprego; k - capital; l - trabalho; ε - ruído branco; z - conjunto de variáveis tais como o grau de sindicalização dos trabalhadores e o nível do subsídio de desemprego; v - conjunto de variáveis que fazem pressão sobre a procura global; e o expoente 'e' representa valor antecipado.

Neste modelo, a taxa de inflação segue um processo aleatório (random walk) e as antecipações inflacionistas são definidas a partir desse processo.

$$\Delta p = \Delta p_{-1} + \varepsilon \quad (4)$$

$$p^e = p_{-1} + \Delta p_{-1}, \text{ com } p - p^e = \Delta^2 p \quad (5)$$

A NAIRU (u^*) corresponde à taxa de desemprego compatível com as realizações das antecipações dos agentes sobre os preços e salários (6).

$$u = u^* \Rightarrow \varepsilon = 0 \text{ e } \Delta^2 p = 0 \quad (6)$$

$$u^* = \frac{a_0 + b_0 + z}{a_1 + b_1} \quad (7)$$

Segundo a fórmula assim obtida (7), a NAIRU depende exclusivamente dos factores que fazem pressão sobre o salário real. É certo que a NAIRU pode variar, mas a sua variação não depende das variáveis nominais que agem sobre a procura, ou seja, não depende de “v”.

Se o preço e o salário dependem simultaneamente da taxa de desemprego e da variação da taxa de desemprego, então as equações de preço e do salário são rescritas e incluem como variável, a variação da taxa de desemprego ((8), (9)). Neste caso, pode ser definida uma taxa natural de desemprego de curto-prazo que depende quer da taxa natural de desemprego de longo-prazo, quer da taxa de desemprego efectiva.

$$p - w = a'_0 - a'_1 \cdot u + a_{11} \cdot \Delta u + a'_2 \cdot (p - p^e) - a'_3 \cdot (k - 1) \quad (8)$$

$$w - p = b'_0 - b'_1 \cdot u - b_{11} \cdot \Delta u + b'_2 \cdot (p - p^e) + b'_3 \cdot (k - 1) + z \quad (9)$$

A taxa de equilíbrio do sistema (u_{cp}^*) está também associada a um estado de equilíbrio de expectativas (10),

$$u = u_{cp}^* \Rightarrow \varepsilon = 0, \Delta^2 p = 0 \text{ e } \Delta u \neq 0 \quad (10)$$

vindo assim:

$$u_{cp}^* = \frac{(a'_1 + b'_1) \cdot u^* + (a_{11} + b_{11}) \cdot u_{-1}}{a'_1 + b'_1 + a_{11} + b_{11}} \quad (11)$$

A NAIRU de curto-prazo u_{cp}^* , em (11), vai depender simultaneamente da NAIRU e da taxa de desemprego do período anterior. Em suma, a NAIRU de curto-prazo dependerá da história do desemprego. Na presença de histerese parcial, a taxa de desemprego de curto-prazo convergirá para a taxa de desemprego de longo-prazo, mas os efeitos dos choques (inovações) podem durar bastante tempo, tanto mais quanto o valor do coeficiente u_{-1} em (11) for elevado (mas inferior à unidade).

Se os efeitos directos da NAIRU sobre o preço e o salário nominal são diminutos (vizinhos de zero), então a NAIRU de curto-prazo é variável e dependerá das variáveis nominais que determinam a taxa de desemprego do período anterior, (12)

$$u_{cp}^* = u_{-1} \quad (12)$$

Neste caso, estamos em presença de histerese total porque a taxa de salário nominal não depende da taxa de desemprego e nestas condições a NAIRU irá depender da taxa de desemprego do período anterior. Em suma, a série temporal da taxa de desemprego é uma série de raiz unitária. Nestas condições não é possível definir a NAIRU de longo-prazo.

As explicações micro-económicas avançadas para a explicação do fenómeno baseiam-se em modelos de determinação do salário do tipo insiders-outsiders. Escusado será dizer que a subida da NAIRU devido aos efeitos de histerese vai depender da grandeza do choque inicial, do grau de persistência do choque e da sucessão de choques com o mesmo sinal.

A hipótese extrema de histerese forte segundo a qual os desempregados de longa-duração não irão influenciar o nível da taxa de salário nominal poderá ser reconsiderada em face das diferentes instituições de trabalho. Tal significa que a importância do fenómeno da histerese na explicação da subida da taxa natural de desemprego na Europa poderá ser mantida ainda que seja excluída a hipótese extrema que admite que os desempregados de longa-duração não exercem uma pressão sobre o nível da taxa de desemprego.

Com efeito, a tomada em linha de conta das diferentes instituições de trabalho, pode condicionar a amplitude do fenómeno. Tomemos duas instituições do mercado de trabalho que sejam bastante favoráveis aos trabalhadores, como as leis de protecção do emprego e os subsídios de desemprego. A duração do desemprego será mais elevada num mercado de trabalho com estas instituições, por essa razão, um choque negativo que aumente o desemprego, será ampliado por estas instituições, porque uma parte acrescida do desemprego criado (acrescida relativamente a um mercado de trabalho sem aquelas instituições) transformar-se-á em desemprego de longa-duração (Blanchard, O. 1999).

Blanchard e Katz constroem uma curva de salário que contem uma curva de Phillips. A comparação entre as duas curvas apresenta várias vantagens quer do ponto de vista da análise teórica quer do ponto de vista da análise aplicada quer ainda pela explicação que é avançada para os diferentes resultados da estimação da NAIRU obtidos para os EUA e para a UE.

Assim, a curva de oferta de trabalho, porque baseada numa curva de salário vai permitir: a) que ao nível empírico sejam incluídas as contribuições comuns de vários modelos

de determinação do salário real, a saber que o salário real antecipado vai depender do salário de reserva e da produtividade do trabalho; b) a tomada em linha de conta dos efeitos permanentes dos choques de oferta (preço da energia), taxa de juro real e quotizações sociais (payroll taxes), quando os efeitos directos da produtividade do trabalho sobre o salário real e sobre o salário de reserva são positivos; c) possam ser controladas as diferenças entre as instituições do mercado de trabalho nos EUA e na UE. A curva de Phillips conduz a bons resultados econométricos para os dados dos EUA e a maus para os dados da UE. O que poderá ser explicado pela ausência de efeitos directos de produtividade de trabalho sobre o salário real e sobre o salário de reserva no caso dos EUA. Seja (13) a curva de Phillips.

$$w_t - p_t^e = a_w + (w_{t-1} - p_{t-1}) - \beta \cdot U_t + \varepsilon_t \quad (13)$$

As variáveis estão definidas em logaritmos excepto a taxa de desemprego. O salário real antecipado depende negativamente do salário real com desfasamento e negativamente da taxa de desemprego⁷. A curva de Phillips pode ser definida como uma relação negativa entre a taxa de variação do salário real e a taxa de desemprego. Já a curva de salário de Blanchflower e Oswald (1994) é definida como uma relação negativa entre o salário real antecipado e a taxa de desemprego, sendo dadas a taxa de salário de reserva (b_t) e a produtividade do trabalho (y_t)⁸.

$$w_t - p_t^e = \delta \cdot b_t + (1 - \delta) \cdot y_t - \beta \cdot U_t + \varepsilon_t \text{ com } 0 < \delta < 1 \quad (14)$$

Por sua vez, o salário de reserva (15) irá depender em última instância de dois factores, do salário real e da produtividade do trabalho. A equação do salário de reserva é homogénea de grau 1 exprimindo a inexistência de um efeito de longo-prazo da produtividade sobre a taxa natural de desemprego⁹.

$$b_t = a'' + \lambda \cdot (w_{t-1} - p_{t-1}) + (1 - \lambda) \cdot y_t \text{ com } 0 < \lambda < 1 \quad (15)$$

Substituindo na equação (14) obtém-se:

$$w_t - p_t^e = a'' \cdot \delta + \delta \cdot \lambda \cdot (w_{t-1} - p_{t-1}) + (1 - \delta \cdot \lambda) \cdot y_t - \beta \cdot U_t + \varepsilon_t \quad (16)$$

Se compararmos as equações (13) e (16) vemos que (13) é um caso particular de (16) justamente quando δ e λ são iguais à unidade:

$$\delta = 1 \Rightarrow (1 - \delta) = 0 \wedge \lambda = 1 \Rightarrow (1 - \lambda) = 0 \quad (17)$$

O efeito directo da produtividade do trabalho sobre o salário real antecipado é nulo sendo dado o salário de reserva e o efeito directo da produtividade do trabalho sobre o salário de reserva é também nulo. Aparentemente a robustez da estimação da curva de Phillips para os EUA poderá ser interpretada do ponto de vista dos resultados a que conduzem certos

⁷ a_w é uma constante e ε_t continua a ser um termo de erro.

⁸ $(1 - \delta)$ representa o efeito directo da produtividade sobre o salário real.

⁹ $(1 - \lambda)$ representa o efeito directo da produtividade sobre o salário de reserva.

modelos de salário de eficiência, segundo os quais o salário de reserva não depende senão do subsídio de desemprego.

A fim de estudar o efeito da produtividade do trabalho sobre a NAIRU, é preciso completar o modelo com uma equação de preços

$$w_t - p_t = y_t - x_t \quad (18)$$

onde x representa a taxa de margem de equilíbrio do mercado dos bens de concorrência imperfeita sendo dado o nível tecnológico.

Substituindo (16) em (18), supondo que as antecipações sobre os preços se realizam e que não há acontecimentos imprevisíveis, obtém-se uma NAIRU de curto-prazo:

$$U_t^* = \frac{a'' \cdot \delta + (1 - \delta \cdot \lambda) \cdot x_{t-1} + \Delta x_t - \delta \cdot \lambda \cdot \Delta y_t}{\beta} \quad (19)$$

Segundo a equação (19), a NAIRU de curto-prazo depende positivamente dos factores que actuam sobre a taxa de margem do período anterior (se os efeitos directos da produtividade sobre o salário real e sobre o salário de reserva forem positivos) e da variação da taxa de margem, e negativamente da variação da produtividade do trabalho do período. Se se considerar que a taxa de margem é constante, assim como a produtividade do trabalho, a influência da taxa de margem sobre a NAIRU de curto-prazo mantém-se se os efeitos directos de produtividade forem positivos:

$$U_t^* = \frac{a'' \cdot \delta + (1 - \delta \cdot \lambda) \cdot x_{t-1}}{\beta} \quad (20)$$

Mas se os efeitos directos são nulos, a NAIRU só dependerá de factores reais,

$$U_t^* = \frac{a'' \cdot \delta}{\beta} \quad (21)$$

Esta via de pesquisa é sem dúvida frutuosa no que respeita à NAIRU, os autores mostram que devido às características diferentes dos mercados de trabalho, a taxa de margem determina a NAIRU na UE e não nos Estados Unidos.

III. ESTUDO EMPÍRICO DO COMPORTAMENTO DA NAIRU

Staiger, Stock et Watson (1996 e 1997) consideram que existe um problema muito sério que decorre da estimação econométrica da NAIRU, qualquer que seja o tipo de modelo estimado, quaisquer que sejam as hipóteses auxiliares tomadas relativamente à formação das expectativas e quaisquer que sejam as séries consideradas - a estimação é imprecisa.

Mas estes resultados, por si, não significam que não se possam considerar as variações da taxa de desemprego como um indicador de variações futuras da inflação. O que a autoridade monetária não pode fazer é elaborar uma regra fina de política monetária que seja accionada quando a taxa de desemprego efectiva se aproximar da NAIRU.

Como veremos, iremos mais longe na nossa crítica à utilização do conceito de taxa natural de desemprego.

III.1 ECONOMIAS ESTUDADAS

O nosso estudo aplica-se à França, Espanha, Áustria, Dinamarca e Itália. Escolhemos um conjunto de economias europeias que apresentam não só valores passados da inflação bastante diferentes, como evoluções relativamente distintas. Seja quanto à inflação propriamente dita ou quanto à inflação salarial. Também a história mais recente de cada uma destas economias quanto aos valores da taxa de desemprego é bastante particular.

Por outro lado, pareceu-nos que seria interessante juntar no mesmo estudo três países de cultura latina com um país de cultura germânica e outro nórdico. Desta forma estamos a estudar economias com instituições bastante diferentes.

Nos Quadros 1, 2, e 3 podemos observar algumas particularidades do período considerado e também características das diferentes economias. O valor DP respeita ao desvio padrão de todos os valores em cada trimestre a que se refere o respectivo Quadro. Em Anexo indicamos de forma precisa as variáveis utilizadas.

Quadro 1
Taxas de Inflação (%) para o período 1961:I a 1999:IV

	TAXA MÉDIA	VALOR MÍNIMO	VALOR MÁXIMO	DP
França	5,4	0,26	14,0	3,6
Espanha	8,7	0,31	24,1	5,1
Itália	7,8	0,67	22,1	5,2
Áustria	3,9	0,36	9,5	2,0
Dinamarca	6,0	0,97	15,1	3,5
Portugal *	11,6	1,78	27,8	7,3

Quadro 2
Taxas de Inflação Salarial (%)

	TAXA MÉDIA	VALOR MÍNIMO	VALOR MÁXIMO	DP	PERÍODO
França	5,5	1,6	19,2	4,6	60:1-99:2
Espanha	9,1	2,3	30,8	7,3	64:1-99:2
Itália	7,8	1,8	29	7,0	60:1-99:3
Áustria	3,9	0,6	17,2	3,4	60:1-99:2
Dinamarca	7,0	1,8	23,7	4,8	71:1-93:4
Portugal *	12,5	0,3	32,7	6,6	78:1-98:4

Quadro 3
Taxas de Desemprego (%)

	TAXA MÉDIA	VALOR MÍNIMO	VALOR MÁXIMO	DP	PERÍODO
França	9,7	4,8	12,5	2,1	78:1-99:4
Espanha	17,1	4,6	24,5	5,4	76:3-99:2
Itália	8,3	3,6	12,4	2,8	60:1-99:2
Áustria	3,9	1,4	7,3	1,9	60:1-99:4
Dinamarca	7,4	0,8	12,5	3,2	70:1-93:3
Portugal *	7,4	3,9	11,1	1,9	78:1-98:4

* 1978:1 a 1998:4

A ideia geral a reter destes quadros é a da relativa diversidade de situações para estas economias.

III.2 METODOLOGIA DA CONFIRMAÇÃO EMPÍRICA

O nosso estudo empírico vai limitar-se a uma determinante real da NAIRU. Assim, o objectivo primeiro da aplicação que nos propusemos consiste em procurar detectar um comportamento cíclico da NAIRU e em seguida saber se para além desse comportamento ainda existirá uma componente tendencial que possa explicar a sua evolução. Para testar a evolução tendencial preferimos fazer uso do produto tendencial em vez de um simples *trend*. A razão prende-se com o simples facto de a economia, e a sua taxa de desemprego de equilíbrio, não deverem ser indiferentes a diferentes crescimentos do produto.

Para determinarmos os valores da taxa de desemprego que não aumente a taxa de inflação salarial, NAWRU, e a taxa de inflação propriamente dita, NAIRU, seguimos o método proposto por Elmeskov (1993). Esta metodologia tem como principal mérito a sua facilidade de aplicação e isenção em termos de modelos explicativos da taxa de desemprego. No entanto, faremos uma pequena alteração ao cálculo sugerido por este autor, uma vez que supomos uma relação não-linear entre a variação da inflação e o diferencial da taxa de desemprego relativamente ao seu valor de equilíbrio. Assim, para o caso dos salários, tomaremos,

$$\Delta w = \alpha \cdot (\bar{u} - u) \quad (22)$$

onde w é a inflação salarial, \bar{u} a taxa de desemprego de equilíbrio e u a taxa de desemprego. Neste caso, \bar{u} representa a NAWRU. O método sugerido por Elmeskov consiste em supor, numa primeira hipótese, que a taxa de desemprego de equilíbrio não se altera de um período para o outro. O que nos conduz à fórmula,

$$\alpha = \frac{-\Delta^2 w}{\Delta u} \quad (23)$$

que por substituição adequada nos leva ao valor da taxa de desemprego de equilíbrio, a partir de valores nossos conhecidos:

$$\bar{u} = u - \frac{\Delta u}{\Delta^2 w} \cdot \Delta w \quad (24)$$

Devemos no entanto ter em conta que aquela fórmula pode levar à obtenção de valores que correspondem a alterações bruscas daquela taxa de equilíbrio. Uma vez porque Δw pode tomar o valor zero e outras porque a sua diferença, $\Delta^2 w$, pode tomar valores reduzidos (ou mesmo zero) levando a valores nulos ou muito elevados da taxa de desemprego de equilíbrio. Estas variações bruscas devem ser corrigidas para que o próprio conceito de taxa de equilíbrio possa ter significado e ainda ter um carácter operacional.

A metodologia econométrica seguida na parte restante do estudo foi a seguinte:

- a) depois de obtido o valor da NAWRU ou NAIRU, pelo método de Elmeskov, procurámos corrigir¹⁰ alguns valores que resultam do facto de a primeira ou segunda diferenças da inflação (ou inflação salarial) poderem ter valores muito reduzidos;
- b) corrigimos o problema do *end-point* final associado ao método de Hodrick-Prescott através de modelos ARMA¹¹ aplicados às diferenças sazonais do produto real; no caso da Dinamarca, por exiguidade desta série, usámos os valores em volume das vendas;
- c) com base naqueles valores passámos a pesquisar um modelo da taxa de desemprego de equilíbrio com inclusão dos valores da componente cíclica da economia, tendo retido o modelo que minimizava o critério de Akaike, respeitando a exclusão da hipótese nula dos coeficientes e a ausência de auto-correlação de ordem (até) quatro;
- d) depois de obtido aquele modelo ensaiávamos a obtenção de um outro modelo que difere dele pela inclusão dos valores de tendência do produto, para além de respeitar os mesmos critérios.

Os modelos são identificados por M e um número que corresponde à sua ordem de apresentação. A variável dependente, assim como o período de observações efectivamente usado, é indicado para identificar o modelo apresentado. A coluna “Desf” indica o número de defasamentos da respectiva variável. Assim, “1-4” para “x” refere-se aos valores de “x₋₁” a x₄”. A estatística do Chi-quadrado, do lado esquerdo da parte de baixo de cada caixa refere-se a um teste LM à inclusão no modelo sem tendência do valor não desfasado da tendência do produto. O valor de Q respeita à estatística de Ljung-Box, com graus de liberdade que têm em conta as observações e os parâmetros do modelo. LM(4) é o valor do Chi-quadrado para o teste LM à auto-correlação de ordem 1 a 4 dos erros do modelo. \bar{R}^2 , é o R² ajustado.

III.3 RESULTADOS POR ECONOMIAS ESTUDADAS

Passamos a apresentar os modelos obtidos para cada uma das economias, chamando a atenção para os aspectos mais salientes que estes resultados nos sugerem.

¹⁰ Obtendo um novo valor que resulta da média aritmética simples entre o valor anterior e o imediatamente a seguir.

¹¹ Em Anexo indicamos os respectivos modelos.

Portugal

Em Duarte e Andrade (2000) estudámos esta metodologia para o período de 1977:1 a 1998:4, pelo que aqui apenas retomamos as conclusões da aplicação desta metodologia. No caso da nossa economia não podemos excluir a hipótese de um comportamento contra-cíclico da taxa natural de desemprego¹². E se num modelo parcimonioso admitíssemos a presença de uma tendência, uma investigação de um modelo mais geral levava à sua exclusão.

França

Os valores dos modelos que obtivemos não só retratam o comportamento contra-cíclico da NAWRU e da NAIRU, como os bastante bons ajustamentos obtidos (M1 e M3. Para ambos os modelos justifica-se a inclusão de uma tendência do produto (5,48% e 0,27%). Os modelos M2 e M4 traduzem o relacionamento cíclico e de tendência daquelas duas variáveis.

Podemos pois concluir que não só aquelas taxas naturais apresentam um comportamento cíclico com também apresentam uma clara tendência crescente com o produto. Esta última conclusão acaba por ir ao encontro de algumas outras análises que apontam para o crescimento da NAIRU em França.

NAWRU

Variáveis	M1, 1980:2-1999:2			M2, 1979:1-1999:2		
	Desf	Coefficiente	T	Desf	Coefficiente	T
Constante		0,531	3,66		-1,340	2,16
NAWRU	1-8	0,779	12,48	1-3	0,494	4,39
LY_FC	1	-3,667	2,46	1	-6,080	3,75
LY_FT				1	0,559	3,03

$\chi^2 = 3,69$ (5,48%) $\sigma=0,091$ Q(19)=12,41 LM(4)=7,51 $\sigma=0,096$ Q(20)=15,41 LM(4)=2,43
 $\overline{R^2} = 0,767$ $\overline{R^2} = 0,801$

NAIRU

Variáveis	M3, 1979:3-1999:4			M4, 1978:4-1999:4		
	Desf	Coefficiente	T	Desf	Coefficiente	T
Constante		0,242	3,25			
NAWRU	1-5	0,899	27,84	1-2	0,849	24,34
LY_FC	1	-1,508	2,00	1	-1,740	2,53
LY_FT				1-2	0,089	4,66

$\chi^2 = 9,00$ (0,27%) $\sigma=0,051$ Q(20)=15,37 LM(4)=3,77 $\sigma=0,048$ Q(21)=15,11 LM(4)=2,48
 $\overline{R^2} = 0,936$ $\overline{R^2} = 0,952$

Espanha

Os ajustamentos obtidos são também muito bons e comprovam a dependência dos valores da NAWRU e da NAIRU da evolução cíclica da economia (M5 e M8). Embora se obtenham ajustamentos razoáveis com a inclusão de uma variável de tendência (M6, M7 e

¹² É interessante verificar que por diferentes motivos Allan Larsson (1999) chegou a conclusão idêntica.

M9), de facto, a sua inclusão nos modelos apenas com ciclos não se justifica. Os modelos com tendência apresentam valores para o R^2 ajustado, e para o critério de Akaike, que são inferiores aos dos modelos sem essa variável. O próprio sinal da variável de tendência acaba por ser ou negativo (M6 e M9) ou positivo (M7) em modelo sem constante.

Como conclusão, devemos insistir no carácter cíclico do comportamento da NAWRU e NAIRU e na ausência de uma tendência clara na evolução dos seus valores.

NAWRU

VARIÁVEIS	M5, 1978:1-1999:2			M6, 1977:4-1999:2		
	DESF	COEFICIENTE	T	DESF	COEFICIENTE	T
Constante		0,318	4,04		0,741	3,33
NAWRU	1-5	0,893	33,38	1-3	0,951	30,82
LY_SC	1-2	-1,677	2,43	1	-1,533	2,29
LY_ST				1	-0,117	2,03

$\chi_1^2 = 0,34$ (55,90%) $\sigma=0,057$ Q(21)=11,07 LM(4)=5,46 $R^2 = 0,969$

$\sigma=0,056$ Q(21)=10,98 LM(4)=6,20 $R^2 = 0,973$

NAWRU

VARIÁVEIS	M7, 1977:2-1999:2		
	DESF	COEFICIENTE	T
NAWRU	1-2	0,878	37,64
LY_SC	1-2	-1,868	2,72
LY_ST	1	0,081	5,49

$\sigma=0,067$ Q(22)=21,79 LM(4)=9,23 $R^2 = 0,968$

NAIRU

VARIÁVEIS	M8, 1977:2-1999:1			M9, 1979:1-1999:1		
	DESF	COEFICIENTE	T	DESF	COEFICIENTE	T
Constante		0,307	6,47		1,529	3,34
NAWRU	1-2	0,898	53,22	1-9	0,922	19,40
LY_SC	1-2	-1,249	2,08	1	-2,934	2,36
LY_ST				1-2	-0,256	2,74

$\chi_1^2 = 2,47$ (11,63%) $\sigma=0,059$ Q(22)=16,25 LM(4)=4,04 $R^2 = 0,977$

$\sigma=0,056$ Q(20)=8,92 LM(4)=3,38 $R^2 = 0,951$

Itália

O caso da Itália é bastante interessante, porque apesar de uma clara relação negativa entre a NAWRU (e a NAIRU) e o produto cíclico, como podemos ver nas figuras em baixo, não obtemos nenhum modelo de comportamento cíclico para a primeira daquelas variáveis. Uma possível razão reside no facto de a relação NAWRU/produto cíclico não ser linear.

No entanto, não deixa de ser esclarecedor o facto de podermos obter um modelo, como M10, onde apenas incluímos valores desfasados da NAWRU e a tendência do produto. O coeficiente da tendência à positivo e o ajustamento bastante bom.

No que respeita à NAIRU, o ajustamento obtido, M11, confirma o seu comportamento cíclico e embora este modelo exclua a presença de uma tendência, o facto é que um modelo com tendência, M12, apresenta melhores valores globais de ajustamento.

NAWRU

M10, 1971:2-1999:2

VARIÁVEIS	DESF	COEFICIENTE	T
Constante		-8,794	2,77
NAWRU	1-3	0,699	6,98
LY_IT	1-4	0,658	2,83

$\sigma=0,090$ $Q(28)=23,92$ $LM(4)=1,02$

$\overline{R^2} = 0,910$

NAIRU

M11, 1970:3-1999:2

M12, 1970:2-1999:2

VARIÁVEIS	DESF	COEFICIENTE	T	DESF	COEFICIENTE	T
Constante		0,087	1,45 ¹³		-4,121	2,80
NAWRU	1-4	0,967	35,41	1-3	0,759	9,76
LY_IC	1-2	-1,714	2,74	1	-1,541	2,69
LY_IT				1	0,329	2,86

$\chi_1^2 = 1,39$ (23,75%) $\sigma=0,088$ $Q(29)=24,34$ $LM(4)=3,95$ $\sigma=0,086$ $Q(29)=28,86$ $LM(4)=3,16$

$\overline{R^2} = 0,917$

$\overline{R^2} = 0,922$

Áustria

A evolução dos salários e dos preços na Áustria sofre uma clara ruptura a partir de 1982. Como se as suas taxas de crescimento subissem um patamar mais elevado. A taxa de câmbio acompanhou este movimento de salários e preços. Impusemos assim a presença no nossos modelos de uma variável muda que vai até 1981:4 com o valor 1. O modelo da NAWRU, M13, apresenta um comportamento cíclico e não justifica a presença de uma tendência. Aliás, o modelo com tendência, M14, apresenta um ajustamento pior. O mesmo não se passa com a NAIRU. A presença de uma variável de tendência não pode claramente ser excluída. Retemos assim o comportamento cíclico para as duas variáveis e um comportamento cíclico e de tendência para a NAIRU.

¹³ N.S. de 14,9%

NAWRU

VARIÁVEIS	M13, 1965:2-1999:2 ¹⁴			M14, 1965:2-1999:2 ¹⁵		
	DESF	COEFICIENTE	T	DESF	COEFICIENTE	T
Constante		0,395	3,42			
NAWRU	1-15	0,787	12,74	1-15	0,816	12,23
LY_AC	1	-1,501	2,03	1	-1,498	2,39
LY_AT				1	0,757	2,79
D1		-0,241	3,34		-0,172	2,64

$\chi_1^2 = 0,49$ (48,39%) $\sigma=0,093$ Q(34)=33,70 LM(4)=26,13 $\overline{R^2} = 0,968$

$\sigma=0,094$ Q(34)=27,55 LM(4)=28,28 $\overline{R^2} = 0,967$

D1, até 1981:4 = 1,0

NAIRU

VARIÁVEIS	M15, 1964:3-1999:3			M16, 1961:1-1999:3		
	DESF	COEFICIENTE	T	DESF	COEFICIENTE	T
Constante		0,410	5,63		-0,458	2,34
NAWRU	1-12	0,778	18,74	1-14	0,940	31,70
LY_AC	1	-3,329	4,63	1-2	-2,830	3,00
LY_AT				1	0,127	2,45
D1		-0,249	5,62			

$\chi_1^2 = 18,02$ (0,002%) $\sigma=0,102$ Q(35)=32,27 LM(4)=7,26 $\overline{R^2} = 0,962$

$\sigma=0,108$ Q(34)=26,27 LM(4)=3,51 $\overline{R^2} = 0,957$

D1, até 1981:4 = 1,0

Dinamarca

Devido à disponibilidade da série sobre de salários, o estudo da NAWRU termina em 1993:4. Também por razões de disponibilidade estatística, em vez que obtermos valores de tendência e cíclicos com o PIB, usámos um índice do volume de vendas.

Como podemos verificar, M17 e M19, a NAWRU e a NAIRU apresentam um comportamento cíclico. Em ambos os casos não podemos excluir a presença de uma variável de tendência naqueles modelos (6,9% e 0,8%). As estimações, M18 e M20, demonstram a presença de um comportamento cíclico e de um comportamento de tendência de ambas as variáveis.

NAWRU

VARIÁVEIS	M17, 1975:2-1993:4			M18, 1977:4-1993:4		
	DESF	COEFICIENTE	T	DESF	COEFICIENTE	T
Constante		0,669	2,40			
NAWRU	1-16	0,707	5,49	1	0,375	2,76
LV_DC	1-4	-2,960	2,02	1-15	-7,539	2,13
LV_DT				1-2	0,352	4,51

¹⁴ Com correcção da matriz das variâncias pela imposição de uma estrutura dos erros MA de ordem quatro.

¹⁵ O mesmo que na nota anterior.

$$\chi_1^2 = 3,30 \text{ (6,9\%)} \quad \frac{\sigma=0,290 \text{ Q(19)=12,6 LM(4)=3,49}}{\overline{R^2} = 0,360} \quad \frac{\sigma=0,212 \text{ Q(16)=12,60 LM(4)=7,52}}{\overline{R^2} = 0,379}$$

NAIRU

VARIÁVEIS	M19, 1977:2-1999:3			M20, 1978:1-1999:3 ¹⁶		
	DESF	COEFICIENTE	T	DESF	COEFICIENTE	T
Constante		0,819	3,06			
NAWRU	1-2	0,622	5,07	1-2	0,728	4,94
LV_DC	1-13	-5,448	2,10	1-16	-6,526	2,09
LV_DT				1-2	0,151	2,25

$$\chi_1^2 = 6,90 \text{ (0,8\%)} \quad \frac{\sigma=0,236 \text{ Q(22)=18,16 LM(4)=8,65}}{\overline{R^2} = 0,419} \quad \frac{\sigma=0,240 \text{ Q(21)=20,44 LM(4)=23,96}}{\overline{R^2} = 0,410}$$

IV. CONCLUSÃO

Apesar das dúvidas postas ao conceito de NAIRU, esta não deixou de dominar a prática da política económica. Mankiw (1998) atribui comportamentos semelhantes ao que designou por "wishfull thinking". Neste caso, podemos dizer que a existência de uma taxa conhecida com razoabilidade facilitaria, e sobretudo justificaria, o comportamento da política económica: acções restritivas quando a taxa de desemprego se aproximasse e ultrapassasse à esquerda o seu valor e acções expansionistas com tal se verificasse à direita. Simples de fazer e ainda mais de compreender.

O desemprego na Europa foi uma fonte de investigação do conceito, porque afinal este não apresentava o comportamento que se lhe reconhecia nos Estados Unidos. O estudo de modelos bastantes simples permite esclarecer que a taxa natural de desemprego pode, em certas condições, depender não só da própria evolução do desemprego, como ainda de variáveis nominais. A presença de histerese, assim como de inéfrica, nos valores da taxa de desemprego pode pois ser o resultado de comportamentos do mercado de trabalho e ainda do mercado dos bens.

Depois de havermos constatado que não fazia sentido tomar o "velho" conceito de uma taxa natural de desemprego para a nossa economia, quisemos estudar se o mesmo comportamento se verificava em algumas economias europeias. Para isso escolhemos economias que apresentam um conjunto de instituições e culturas diversificadas e aplicámos a metodologia de Elmeskov.

Constatámos que durante o ciclo podemos reter dois fenómenos com efeitos contraditórios sobre a taxa de desemprego de equilíbrio:

- à medida que nos aproximamos do cume, a pressão sobre a inflação aumenta, por simples efeito de rendimentos decrescentes, pelo que a taxa de desemprego de equilíbrio poderá aumentar;
- mas por outro lado, o maior uso da capacidade de produção e as expectativas optimistas geradas pela subida do ritmo de actividade contribuem para reduzir o valor da taxa de desemprego de equilíbrio.

¹⁶ O mesmo que na nota anterior.

Estes dois efeitos são contraditórios. O primeiro efeito pode ser reduzido, ou mesmo eliminado, por políticas monetárias adequadas, enquanto que o segundo dificilmente o será. Pensamos pois que daqueles efeitos, o segundo será dominante. O que significa que a taxa de desemprego de equilíbrio tem um comportamento contra-cíclico.

A nossa insistência naqueles dois efeitos destina-se a lembrar que a estrutura da relação entre a taxa de desemprego de equilíbrio e a evolução cíclica do produto é bastante mais complexa que o sugerido por aqueles modelos simples que propusemos. No entanto, o facto de se tratarem de modelos simples não invalida as nossas principais conclusões.

Ficou claro que a taxa de desemprego de equilíbrio tem um comportamento cíclico, melhor contra-cíclico. Significa isto que a evolução da taxa de desemprego pode dar-nos falsas indicações sobre o estado da conjuntura no que se refere à inflação. Na fase ascendente do ciclo a taxa de desemprego reduzir-se-á, mas o mesmo acontecerá à taxa de desemprego de equilíbrio. Actuar de forma restritiva porque a taxa de desemprego caiu rapidamente pode assim ter custos económicos e sociais elevados sem que se possa afirmar que tal é necessário para manter a estabilidade dos preços¹⁷. O mesmo haverá a dizer da leitura das variações da taxa de desemprego.

ANEXO: VALORES PUBLICADOS PELA OECD

Quando o período é indicado com numeração árabe é porque a periodicidade é mensal. No caso de periodicidade trimestral o trimestre é indicado com numeração romana. A transformação de dados mensais em trimestrais é feita por média dos meses referentes a cada trimestre. Algumas das variáveis estão corrigidas da sazonalidade. No caso da Dinamarca usámos o índice de vendas para determinarmos os valores cíclicos da economia.

PAÍS	VARIÁVEL	DES-SAZONALIZADA	OBS	COBERTURA
<i>França</i>				
	p	não	IPC total para todo o país	1960:I-1999:10
	w	"	Salário horário na indústria	1960:I - 1999:II
	y	sim	PIBr em índice (base 1990)	1960:I - 1999:III
	u	não	Taxa de desemprego	1960:I - 1999:II
<i>Espanha</i>				
	p	"	IPC total para todo o país	1960:I - 1999:10
	w	"	salário horário na economia	1964:I - 1999:II
	y	sim	PIBr em índice (base 1990)	1970:I - 1999:II
	u	"	Taxa de desemprego	1976:III - 1999:II
<i>Áustria</i>				
	p	não	IPC total para todo o país	1960:1 - 1990:10
	w	"	Rendim. salariais mensais	1960:1 - 1999:6
	y	Sim	PIBr em índice (base 1990)	1964:I - 1999:II
	u	"	Taxa de desemprego	1960:1 - 1999:10

¹⁷ Referimos o caso da fase ascendente do ciclo, mas podíamos falar da fase descendente. Embora com as necessárias adaptações.

Dinamarca

P	Não	IPC total para todo o país	1960:1 - 1999:10
w	"	Salário horário na indústria	1971:1 - 1993:12
y	Sim	PIBr em índice (base 1990)	1988:I - 1999:II
v	"	Volume de vendas	1974:1 - 1999:9
u	Sim	Taxa de desemprego	1970:1 - 1999:9

Itália

p	Não	IPC total para todo o país	1960:1 - 1999:11
w	"	Salário horário	1960:1 - 1999:8
y	Sim	PIBr a preços de 1990	1970:I - 1999:II
u	Sim	Taxa de desemprego	1960:I - 1999:II

Como os valores cíclicos forma obtidos através do filtro de Holdrick-Prescott, resolvemos previamente o problema do end-point final através da previsão do produto para 12 trimestres posteriores ao fim da série. O modelo usado para cada economia foi um ARMA em diferenças sazonais, cuja ordem se encontra no Quadro em baixo.

ECONOMIA	ORDEM DO ARMA EM DIFERENÇA SAZONAL
França	AR=6 e MA=4, com constante
Espanha	AR=6 e MA=4, com constante
Áustria	AR=5 e MA=4, com constante
Dinamarca	AR=1 e MA=4, com constante
Itália	AR=6 e MA= 4 , sem constante

REFERÊNCIAS BIBLIOGRÁFICAS

- AMABLE, BRUNO, JÉRÔME HENRY, FRÉDÉRIC LORDON e RICHARD TOPOL (1995), "Hysteresis revisited: a methodological approach" em Rod Cross (1995), pp. 153-180
- AMARAL, JOÃO F. (1995), "A Falácia da Taxa Natural de Desemprego", em *Ensaio de Homenagem a Francisco pereira de Moura*, Lisboa, ISEG/UTL, pp. 231-7
- BLANCHARD, OLIVIER e LAWRENCE F. KATZ, (1999), "Wage Dynamics: Reconciling Theory and Evidence", *The American Economic Review*, 89, May, pp. 69-74
- BLANCHARD, OLIVIER (1999), "European Unemployment: The Role of Shocks and Institutions", Séminaire du CREFE
- BLANCHFLOWER, DAVID et ANDREW OSWALD (1994), *The Wage Curve*, Cambridge, Ma., MIT Press
- BLINDER, A. S. (1987), "Keynes, Lucas and Scientific Progress", *The American Economic Review*, 77, 2, pp. 130-6
- CABALLERO, RICARDO e Eduardo Engel (1993), "Microeconomic Adjustments Hazards and Aggregate Dynamics", *The Quarterly Journal of Economics*, CVIII, 2, pp. 359-83
- CAHUC, PIERRE (1993), "Les Fondements Théoriques", em Pierre-Yves Henin, ed., (1993) pp. 1-48

- CROSS, ROD (1995), "Is the natural rate hypothesis consistent with hysteresis ?", em Rod Cross (1995), pp. 181-200
- CROSS R. (1995), ed., *The Natural Rate of Unemployment. Reflections on the 25 years of the Hypothesis*, Cambridge, Cambridge University Press
- DUARTE, ADELAIDE e ANDRADE, JOÃO (1999), "Histerese da Taxa de Desemprego de Equilíbrio, um aplicação ao caso português", *Notas Económicas*, no prelo.
- DUARTE, ADELAIDE e ANDRADE, JOÃO (2000), "Le Taux de Chômage Naturel comme un indicateur de Politique Economique? Une application à l'économie portugaise", texto apresentado na 6ª Conf. do Cemapre e nas 17^{èmes} Journées Internationales d'Economie Monétaire et Financière.
- ELMESKOV, J. (1993), "High and Persistent Unemployment: assessment of the problem and its causes", OECD, WP 132
- FRIEDMAN, MILTON, (1968), "The role of monetary policy", *The American Economic Review*, 58, Abril, pp. 1-21
- GORDON, R. (1997), "The Time-Varying NAIRU and its Implications for Economic Policy", *Journal of Economic Perspectives*, 11, 1, pp. 11-32
- HAHN, FRANK (1995), "Theoretical Reflexions on the 'Natural Rate of Unemployment'", em Rod Cross (1995), pp. 43-56
- HENIN, PIERRE-YVES, ed., (1993), *La Persistance du Chômage*, Paris, Economica
- HUMPHREY, THOMAS (1985), "The Evolution and Policy Implications of Phillips Curve Analysis", Federal Reserve Bank of Richmond Economic Review, 71(2), pp. 3-22
- HUMPHREY, THOMAS (1985a), "The Early History of the Phillips Curve", Federal Reserve Bank of Richmond Economic Review, 71(5), pp. 17-24
- LARSSON, ALLAN (1999), "The European Employment Strategy, Towards the Helsinki Summit - Putting Europe to Work -", Dezembro, policopiado
- LAYARD, R., S. NICKELL e R. JACKMAN, (1991), *Unemployment: macroeconomic performance and the labour market*, Oxford, Oxford University Press
- LIPSEY, RICHARD e ALEC CHRYSTAL (1995), *Positive Economics*, Oxford, Oxford University Press
- MANKIW, GREGORY, "Comments" em Solow, R. e Taylor, J. (1998)
- MODESTO, LEONOR e JOÃO CÉSAR DAS NEVES, (1993), "Hysteresis and Sluggishness in Portuguese Unemployment, 1977-88, *Internacional Review of Applied Economics*, 7, 2, pp. 197-207
- STAIGER, DOUGLAS, JAMES STOCK e MARK WATSON (1996), "How Precise are Estimates of the Natural Rate of Unemployment", NBER, WP5477
- STAIGER, DOUGLAS JAMES STOCK e MARK WATSON (1997), "The NAIRU, Unemployment and Monetary Policy", *The Journal of Economic Perspectives*, 11, 1, pp. 33-49
- SOLOW, R. (1998), "How Cautious must the Fed be ?", em Solow, R. e Taylor, J. (1998), pp. 1-28
- SOLOW, R. e TAYLOR, J. (1998), eds., *Inflation, Unemployment, and Monetary Policy*, MIT Press, 1998

DECOMPOSIÇÃO E EVOLUÇÃO DA DESIGUALDADE SALARIAL

THE SOURCES OF INTERREGIONAL WAGE DIFFERENTIALS

Autores: Filipe Santos

- Assistente da Universidade Católica Portuguesa - Viseu

Paulino Teixeira

- Professor Associado da Faculdade de Economia da Universidade de Coimbra

RESUMO:

- Neste estudo, procurou-se identificar as principais componentes da dispersão salarial entre o Litoral e o Interior da região Centro. Em particular, discute-se se a dispersão observada se deve a diferentes características de trabalhadores, estabelecimentos e empregos ou à diferença entre taxas de rendibilidade e prémios salariais para trabalhadores e estabelecimentos com idênticas características. De acordo com a nossa análise, a diferença salarial resulta principalmente da primeira componente, destacando-se o contributo da variável Indústria, o que demonstra que a especialização sectorial é o factor decisivo na explicação do comportamento salarial inter-regional. A análise dinâmica, por seu turno, permite concluir que a evolução da dispersão salarial ao longo da última década se explica sobretudo com base nas alterações ocorridas nas taxas de rendibilidade, mais do que na variação das características dos trabalhadores ou dos empregos.

PALAVRAS-CHAVE:

- *Funções-salário; desigualdade salarial; diferenças salariais inter-indústrias; desigualdade salarial inter-regional.*

ABSTRACT:

- In this study, we discuss the wage differences between two areas of the Central Region of Portugal. In particular, we examine whether the observed interregional wage differentials are due to differences in the characteristics of individuals, firms, and occupations or to differences in the rates of return of these characteristics. We found that firm and industry effects are the key explanatory variables. The analysis of wage differentials over time shows that individuals' and firms' characteristics have been relatively stable and that the modest increase in inequality, observed in a period of almost a decade, is mostly due to changes in the rates of return.

KEY-WORDS:

- *Earnings functions; inter-industry wage differences; regional wage differentials.*

1. INTRODUÇÃO

A adesão à Comunidade Económica Europeia, em 1986, intensificou a modernização do tecido empresarial e a melhoria da qualificação da mão-de-obra, permitindo e criando as condições económicas e políticas para que o país se aproximasse de forma decisiva dos padrões médios europeus. De forma porventura mais visível, assistiu-se a um desenvolvimento sem precedentes da rede de infraestruturas, que conduziu a uma acrescida mobilidade dos recursos, em particular do factor trabalho, condição indispensável para uma melhoria sustentada do bem-estar. Ao mesmo tempo, e sem pôr em causa o objectivo central de alcançar um conjunto de metas macroeconómicas que nos pudesse colocar a par dos nossos parceiros europeus na construção, primeiro, do mercado único e, depois, da união económica e monetária, os volumosos fluxos financeiros canalizados para as grandes políticas estruturantes, repartidos por vários fundos europeus e fixados nos Quadros Comunitários de Apoio, procuraram não perder de vista o objectivo de reduzir as disparidades regionais. No entanto, e sem negar o visível sucesso em relação aos grandes alvos macroeconómicos, não restam dúvidas de que o objectivo de redução de assimetrias se revelou de mais difícil concretização, em especial, e é esse o assunto que nos ocupa, no que respeita à convergência dos salários médios.

Neste estudo, discutiremos o caso particular da Região Centro, um exemplo de disparidade litoral/interior que nos servirá para ilustrar como as reformas económicas introduzidas em Portugal na primeira década após a adesão à Comunidade Económica Europeia se repercutiram na malha das suas regiões e, em particular, nos desequilíbrios salariais regionais. Com efeito, na região Centro, em 1988, isto é, dois anos após a integração de Portugal na CEE, os ganhos mensais dos trabalhadores do Centro Litoral eram, em média, 14.8% mais elevados do que os dos trabalhadores do Centro Interior, enquanto em 1991, ou seja, no início do primeiro Quadro Comunitário de Apoio, esta diferença já se situava nos 16.9%, para, no início do Segundo Quadro, atingir os 17.8%. Não se tratando, é óbvio, de assimetrias médias extremas, parece-nos ainda assim que a sua dimensão, por um lado, e evolução, por outro, justificam um tratamento atento que tenha em conta, nomeadamente, as características individuais dos trabalhadores, empregos e estabelecimentos que compõem as duas sub-regiões em análise.

Que factores poderão então explicar as diferenças salariais observadas e o seu relativo agravamento? Em primeiro lugar, tendo em conta os valores médios para as duas sub-regiões, são evidentes as diferenças nas variáveis afectas à oferta de trabalho. Com efeito, não só os trabalhadores da região Litoral apresentam mais anos de escolaridade, mais anos de antiguidade e um maior peso de tarefas de maior complexidade e responsabilidade, como também estas variáveis tendem a apresentar um crescimento mais forte no Litoral. Na mesma direcção parecem actuar as características relativas às variáveis de procura de trabalho, com os estabelecimentos do Litoral a revelarem, em média, maior dimensão e especialização sectorial mais favorável, isto é, maior presença na fabricação de produtos metálicos, máquinas, equipamento e material de transporte e produtos minerais não metálicos, por exemplo, em contraste com o claro predomínio da indústria têxtil, do vestuário e do couro no Centro Interior.

É, portanto, neste contexto de características sub-regionais de procura e de oferta de trabalho bem diferenciadas que se pretende discutir a desigualdade salarial. Procurar-se-á assim, por um lado, identificar as fontes de dispersão salarial e, por outro, o contributo de cada uma dessas fontes para o agravamento da desigualdade salarial. Neste último ponto, importará determinar se a evolução na dispersão salarial se deve a alterações ocorridas nas taxas de rendibilidade para idênticas características ou à variação nas características dos trabalhadores/empregos/ estabelecimentos, residindo na primeira a verdadeira fonte de

desigualdade salarial. Assim, o nosso trabalho começa por, na secção 2, estimar a função salário para a região Centro, seguindo-se, na secção 3, a derivação da decomposição da dispersão total dos salários segundo Oaxaca (1973) e a explicitação do modelo de análise dinâmica de Farber e Newman (1987). O primeiro modelo permitir-nos-á identificar as principais componentes da dispersão total, enquanto o segundo analisa a evolução da dispersão salarial ao longo de um período de quase uma década, isto é, entre 1988 e 1996. A secção 4 sintetiza as principais conclusões do estudo.

2. GANHOS SALARIAIS NA REGIÃO CENTRO

A estimação do contributo das diferentes variáveis representativas das características do trabalhador e da procura de trabalho é feita a partir da seguinte função semi-logarítmica dos salários:

$$(1) \quad \ln W_i = \beta_0 + \beta_1' H_i + \beta_2' Q_i + \beta_3' I_i + \beta_4' O_i + \varepsilon_i ,$$

onde a variável explicada **ln W** (omitindo o índice *i*) representa o logaritmo natural dos ganhos nominais mensais, **H** o vector das variáveis de capital humano (nível de escolaridade, experiência, antiguidade, experiência e antiguidade ao quadrado, logaritmo das horas trabalhadas¹ e vários termos de interdependência), **Q** o vector das variáveis mudas correspondentes aos diferentes níveis de qualificação do emprego, **I** o vector das variáveis mudas representativas dos sectores de actividade e da dimensão dos estabelecimentos e **O** o vector que inclui outras variáveis relevantes como o sexo e a localização geográfica dos trabalhadores. ε representa o termo de erro e β_i os coeficientes a estimar.

Estas variáveis foram recolhidas a partir da base de dados *Quadros de Pessoal*, do Departamento de Estatística do Trabalho, Emprego e Formação Profissional, Ministério do Trabalho e da Solidariedade, e encontram-se descritas em pormenor nos Anexos 1 e 2 em apêndice. Esta base contém informação sobre todos os trabalhadores por conta de outrem, a tempo completo, que exercem funções em estabelecimentos com mais de um trabalhador, à excepção dos trabalhadores do sector agrícola, silvicultura, caça e pesca. Também não estão incluídos os trabalhadores da função pública, nem os membros das forças armadas.²

Mais de 60 % dos trabalhadores por conta de outrem na região Centro estão localizados em apenas três NUTS III: Baixo Vouga, Baixo Mondego e Pinhal Litoral. Designaremos esta área de Centro Litoral. As restantes sete formam a região Centro Interior e são constituídas por Dão-Lafões, Pinhal Interior Norte, Pinhal Interior Sul, Serra da Estrela, Cova da Beira, Beira Interior Norte e Beira Interior Sul. O número total de trabalhadores incluídos na nossa base de dados em 1988 e 1996 é de 158 338 e 221 690, respectivamente.

O Quadro 1 contém os resultados do modelo (1), estimado separadamente para o Litoral e para o Interior e para cada um dos anos da amostra (no Quadro 1 encontram-se indicados apenas os anos de 1988, 1991, 1994 e 1996). Todos os coeficientes estimados, com raríssimas excepções, são estatisticamente significativos a um nível de 1% na explicação (do

¹ Adopta-se o procedimento habitual de combinar as variáveis logaritmo das horas trabalhadas e antiguidade. Presume-se, neste caso, que a formação profissional específica num local de trabalho é tanto mais elevada quanto maior for o número de horas de trabalho.

² Informação detalhada sobre a base de dados utilizada encontra-se em Santos (1999).

logaritmo) dos ganhos mensais. A qualidade de ajustamento do modelo, ligeiramente maior na regressão estimada para o Interior, é sempre superior a 55% .

Destaque-se, em primeiro lugar, que as diferenças salariais em favor dos indivíduos com maior habilitação literária se acentuaram ao longo do período. Em 1988, a diferença de ganhos variava, no modelo restrito (isto é, na Região Centro) entre menos 13.3% por parte dos trabalhadores com o ensino básico preparatório e mais 34.7% por parte dos bacharéis comparativamente ao logaritmo dos ganhos dos trabalhadores com o nível de escolaridade mínima obrigatória (categoria omissa). Em 1996, a diferença salarial acentou-se, variando entre menos 9.5% por parte dos assalariados com o ensino primário e mais 42.6% por parte dos licenciados, de novo relativamente à categoria omissa. Esta tendência verificou-se também nas regressões estimadas para o Litoral e para o Interior, se bem que no último caso as diferenças fossem, em geral, de menor amplitude.

Como seria de esperar, mais anos de experiência e de antiguidade conduzem a ganhos mais elevados, verificando-se, no entanto, alguma discrepância inter-regional entre os coeficientes estimados. Para testar a hipótese de igualdade dos coeficientes estimados nas duas sub-regiões recorreu-se ao teste de Chow. Os resultados da aplicação deste teste mostram que, para qualquer dos anos, os coeficientes estimados para as duas sub-regiões são diferentes, confirmando-se portanto a existência de taxas de rendibilidade diferenciadas auferidas para trabalhadores com características (observadas) idênticas³.

A inclusão, no modelo restrito, da variável muda INTERIOR permitiu estimar que, em média, os trabalhadores do Interior auferem um ganho nominal mensal entre -5.5% e -6.5% do que os trabalhadores homólogos do litoral. Também as mulheres, apesar da maior escolaridade média, auferem um ganho inferior aos homens, tendo esta diferença vindo a aumentar: -14.4% em 1988, -16.2% em 1996. (As mulheres têm no entanto menos anos de experiência e de antiguidade, para além de trabalharem em média menos horas).

Por seu turno, a análise das variáveis afectas à procura de trabalho revela que, para a totalidade dos trabalhadores, o ganho mensal aumenta com o nível de complexidade ou responsabilidade da ocupação e com a dimensão (medida pelo número de pessoas ao serviço nos estabelecimentos), embora a evolução das diferenças salariais entre os vários níveis de qualificação (do emprego) e as diferentes dimensões dos estabelecimentos pareça não ser exactamente coincidente: a dispersão dos ganhos entre os níveis de qualificação, para ambos os mercados, aumentou entre 1988 e 1991, apresentando uma tendência para diminuir nos anos de 1994 e 1996. Comparativamente, a dispersão dos ganhos em relação à dimensão dos estabelecimentos, face à variável omissa, diminuiu entre 1988 a 1991, para aumentar em 1994 e voltar a diminuir em 1996. Em qualquer dos anos, o diferencial de salários associado à dimensão é maior no Litoral, enquanto o que diz respeito aos níveis de qualificação é maior no Interior.

Testou-se, por último, a hipótese de as indústrias (sectores de actividade) explicarem a variação dos ganhos entre trabalhadores com idênticas características observadas (v.g. Krueger e Summers, 1988, e Dickens e Katz, 1987, entre outros). No caso da região Centro, constata-se de facto que existem diferenças notórias inter-sectorais, com os trabalhadores do sector dos bancos e outras instituições financeiras, dos seguros e da electricidade, gás e água a auferirem ganhos muito superiores aos dos seus congéneres da indústria dos têxteis, do vestuário e do couro (variável omissa). Controlando para as restantes características, um trabalhador no sector dos Seguros auferia, em 1988, mais 88,9% do que na da indústria dos têxteis, a indústria simultaneamente mais empregadora e de salários mais baixos.

³ Os valores do F estatístico obtidos no teste de Chow foram de 94.04, 84.19, 49.93 e 76.48 para os anos de 1988, 1991, 1994 e 1996 . Estes valores são superiores ao F crítico para um nível de significância de 1%, o que nos permite rejeitar a hipótese da igualdade dos coeficientes no Centro Litoral e no Centro Interior.

3. DECOMPOSIÇÃO DA DISPERSÃO SALARIAL

Seguindo Jackson (1986), Farber e Newman (1987) e Beeson e Groshen (1991), podemos distinguir duas parcelas na dispersão total dos salários inter-regiões, sendo a primeira a que resulta da existência de diferentes características por parte dos trabalhadores e/ou estabelecimentos, e a segunda a que deriva da existência de diferentes taxas de rendibilidade ou prémios salariais para um mesmo conjunto de características. Em geral, define-se desigualdade salarial inter-regional como sendo apenas a diferença salarial que persiste entre trabalhadores e estabelecimentos com características semelhantes, ou seja, como aquela que decorre exclusivamente da segunda parcela.

Esta definição conduz, desde logo, a que a preocupação central seja não tanto a de determinar por que é que trabalhadores com diferentes características individuais auferem salários distintos, mas antes a de identificar quais os factores que contribuem para que trabalhadores de diferentes regiões com idênticas características auferam salários diferenciados⁴. Para os Estados Unidos, por exemplo, onde a discussão sobre diferenças salariais Norte-Sul e entre áreas metropolitanas é recorrente, os resultados não são conclusivos. Com efeito, enquanto alguns trabalhos mostram que a maior parcela da dispersão pode ser atribuída à existência de diferentes taxas de rendibilidade (Hanushek, 1973; Sahling e Smith, 1983; Farber e Newman, 1987; e Beeson e Groshen, 1991), outros realçam sobretudo a parcela resultante da existência de diferentes características dos trabalhadores/estabelecimentos (Coelho e Ghali, 1971; Bellante, 1979; e Gerking e Weirick, 1983).

Maior motivo de preocupação tem sido porventura a persistência de dispersão salarial inter-regional, o que tem conduzido a remodelações nos pressupostos dos modelos puros de ajustamento dos mercados, de forma a evidenciar que, por um lado, os “*nonwage factors*” podem influenciar as preferências dos trabalhadores e que, por outro, as empresas não baseiam as suas decisões de localização exclusivamente em custos salariais. O trabalhador pode fazer depender, por exemplo, o seu local de trabalho de factores de ordem familiar, como a existência de oportunidades de emprego para a esposa/esposo, o que poderia conduzir a um comportamento menos reivindicativo do ponto vista salarial; ou, ao contrário, perante condições de trabalho relativamente adversas, reclamar um salário superior ao salário corrente como forma de obter compensação pelas características desfavoráveis do emprego (existência de elevada densidade populacional, custo de vida mais elevado, difícil acesso ou inexistência de determinados serviços públicos, por exemplo). As empresas, por seu turno, ao ponderarem as acessibilidades aos mercados de factores, de bens finais e serviços ou a proximidade aos centros de inovação industrial, poderiam localizar-se em áreas geográficas com salários acima da média (para idênticas características). Importa, portanto, ter presente que, mesmo na hipótese de existência de mecanismos concorrenciais na fixação dos salários, a mobilidade dos factores capital e trabalho pode não conduzir necessariamente à igualização dos salários entre as diferentes regiões.

Nos pontos seguintes deste trabalho, é nosso propósito identificar as fontes de dispersão salarial entre o Litoral e o Interior da região Centro. Procurar-se-á assim evidenciar se a dispersão total dos salários entre as regiões do Centro Litoral e o Centro Interior se deve à existência de diferentes características dos trabalhadores e dos estabelecimentos/empregos,

⁴ Neste contexto, espera-se que a ocorrência de diferentes taxas de rendibilidade ou prémios salariais associados a idênticas características possam ser atribuídas a práticas ligadas à tradição, condições de vida, grau de sindicalização e outros factores institucionais dificilmente mensuráveis, não incluídos em regra como regressores na função salário utilizada na secção 2.

ou, ao invés, à existência de diferentes taxas de rendibilidade e à prática de diferentes prémios salariais relativamente a trabalhadores e estabelecimentos com idênticas características observáveis.

3.1 O MODELO DE OAXACA (1973)

Para análise da diferença salarial total entre o Centro Litoral e o Centro Interior, recorreremos ao modelo desenvolvido por Oaxaca (1973). Este modelo, concebido para analisar a questão da discriminação sexual e racial, tem sido amplamente utilizado no estudo da dispersão salarial entre diferentes espaços geográficos. Em Portugal, concretamente, foi aplicado na questão da discriminação sexual (Martins, 1991; e Vieira, 1992), mas na área da dispersão salarial inter-regional e da desigualdade salarial, não cremos que tenha sido alguma vez aplicado.

Seguindo a apresentação de Oaxaca, podemos formalizar a noção de desigualdade salarial entre o Litoral e o Interior da região Centro recorrendo ao coeficiente de desigualdade D:

$$D = \frac{W_L / W_I - (W_L / W_I)^0}{(W_L / W_I)^0},$$

ou, logaritmando:

$$(2) \quad \ln(D+1) = \ln(W_L / W_I) - \ln(W_L / W_I)^0,$$

onde W_L / W_I é a razão dos salários médios mensais observados no Centro Litoral e Centro Interior e $(W_L / W_I)^0$ a razão dos salários médios mensais entre o Litoral e o Interior na ausência de desigualdades salariais.

A razão dos salários entre o Litoral e o Interior, na ausência de desigualdade salarial inter-regional, só poderá ser estimada se se admitir que a actual estrutura salarial do Centro Interior é aplicável ao Centro Litoral (e vice-versa). Na primeira (segunda) hipótese pressupõe-se que, no caso de não se registar desigualdade, os trabalhadores do Centro Interior (Litoral) recebem em média o mesmo que actualmente recebem, mas que, no caso de se verificar desigualdade, os trabalhadores do Litoral (Interior) recebem mais (menos) do que receberiam num mercado onde não existisse desigualdade.

Considerando as funções salário estimadas separadamente para o Centro Litoral e para o Centro Interior pelo método dos mínimos quadrados, temos:

$$\ln(\bar{W}_L) = \bar{X}_L \hat{\beta}_L,$$

$$(3) \quad \ln(\bar{W}_I) = \bar{X}_I \hat{\beta}_I,$$

onde \bar{W}_L e \bar{W}_I são os salários mensais nominais médios no Litoral e no Interior, \bar{X}_L e \bar{X}_I os vectores dos valores médios das variáveis explicativas para o Litoral e o Interior e $\hat{\beta}_L$ e $\hat{\beta}_I$ os vectores de coeficientes estimados.

Seja a dispersão total G dada por:

$$G = \frac{\bar{W}_L - \bar{W}_I}{\bar{W}_I}$$

$$(4) \quad \ln(G + 1) = \ln(\bar{W}_L) - \ln(\bar{W}_I).$$

Substituindo a equação (3) em (4) e assumindo

$$(5) \quad \Delta\bar{X} = \bar{X}_L - \bar{X}_I$$

e

$$(6) \quad \Delta\hat{\beta} = \hat{\beta}_L - \hat{\beta}_I \Leftrightarrow \hat{\beta}_L = \Delta\hat{\beta} + \hat{\beta}_I,$$

temos que a diferença salarial entre as regiões será igual a

$$(7) \quad \ln(G + 1) = \Delta\bar{X}\hat{\beta}_L + \bar{X}_I\Delta\hat{\beta},$$

de onde se pode retirar

$$(8) \quad \ln(\bar{W}_L) - \ln(\bar{W}_I) = (\bar{X}_L - \bar{X}_I)\hat{\beta}_L + \bar{X}_I(\hat{\beta}_L - \hat{\beta}_I).$$

Repare-se que, de acordo com a expressão (2), e partindo da segunda hipótese (ou seja, de que a estrutura salarial existente no Centro Litoral se aplica ao Centro Interior num mercado não desigual), se obtém:

$$\ln\left(\frac{\hat{W}_L}{\hat{W}_I}\right)^0 = \Delta\bar{X}\hat{\beta}_L,$$

e logo

$$\ln(\hat{D} + 1) = [\Delta\bar{X}\hat{\beta}_L + \bar{X}_I\Delta\hat{\beta}] - \Delta\bar{X}\hat{\beta}_L = \bar{X}_I\Delta\hat{\beta}.$$

Ou seja, a desigualdade salarial é medida pelo segundo termo da equação (8). Esta última expressão permite-nos portanto decompor a diferença salarial entre regiões em dois efeitos. Um primeiro, correspondente ao primeiro termo da equação (8), é designado de efeito *médias* e decorre da existência de diferentes características dos trabalhadores e/ou dos estabelecimentos; e um segundo, denominado de efeito *retorno*, que resulta da existência de diferenças na estrutura salarial dos mercados de trabalho e que leva a que trabalhadores com idênticas características tenham taxas de rendibilidade diferenciadas consoante a região ou mercado onde exercem a sua actividade profissional. A estimação deste último efeito, captado, como se disse, pelo segundo termo da equação (8), permitir-nos-á medir a amplitude da desigualdade dos salários inter-regiões.

3.2 O MODELO DE FARBER E NEWMAN PARA UMA ANÁLISE INTERTEMPORAL DA DISPERSÃO SALARIAL

Embora a equação (8) nos permita avaliar a amplitude da diferença salarial total entre o Litoral e o Interior da região Centro para qualquer ano e em qualquer ponto no tempo, estamos igualmente interessados em identificar as fontes das alterações ocorridas nos salários relativos entre as duas regiões ao longo do período amostral. Para este fim, seguiremos o modelo apresentado por Farber e Newman (1987).

Seja o logaritmo natural dos ganhos mensais do indivíduo i , na região j no momento t dado por $\ln W_{j,t}^i$. Assume-se que os ganhos nominais mensais são determinados por um conjunto de h características representadas por um vector $X_{j,t}^i = (X_{lj,t}^i, \dots, X_{hj,t}^i)$ e que as observações relativas aos trabalhadores se encontram agrupadas numa base de dados com a seguinte forma: $i = 1, \dots, k$ no momento $t-1$ e as restantes por $i = k+1, \dots, m$ no momento t . Desta forma, suprimindo a referência relativa à identificação das observações, i , podemos representar a função salário de uma forma totalmente interactiva entre as regiões e no tempo:

$$(9) \quad \ln W_{j,t} = X_{j,t} \beta + L X_{j,t} \beta^L + (T-1) X_{j,t} \beta^{T-1} + L(T-1) X_{j,t} \beta^{(L,T-1)} + \varepsilon,$$

onde L é um vector constituído por um conjunto de valores igual a um se os indivíduos observados estão a exercer a sua actividade no Litoral, ou por um conjunto de zeros no caso contrário. Por sua vez, $T-1$ é um vector constituído por um conjunto de valores iguais a um se os indivíduos são observados no momento $t-1$, ou por um conjunto de valores iguais a zero no caso contrário. O coeficiente do vector X é um vector $\beta = (b_0, b_1, \dots, b_h)$, associado aos trabalhadores do Litoral no momento t . β^L representa a diferença nas taxas e prémios para idênticas características do Litoral e do Interior no momento t – ou seja, $(\beta_L - \beta_I)_t$ – e β^{T-1} representa a diferença nas taxas e prémios no Centro Litoral, ocorridas entre o momento $t-1$ e t .

Com base na equação (9), podemos derivar o logaritmo natural dos salários para a região Centro Litoral ($L=1$) e Centro Interior ($L=0$) no momento $t-1$ ($T-1=1$):

Com base na equação (9), podemos derivar o logaritmo natural dos salários para a região Centro Litoral ($L=1$) e Centro Interior ($L=0$) no momento $t-1$ ($T-1=1$):

$$\begin{aligned} \ln \bar{W}_{L,t-1} &= \bar{X}_{L,t-1} \hat{\beta} + \bar{X}_{L,t-1} \hat{\beta}^L + \bar{X}_{L,t-1} \hat{\beta}^{T-1} + \bar{X}_{L,t-1} \hat{\beta}^{L,T-1} \\ \ln \bar{W}_{I,t-1} &= \bar{X}_{I,t-1} \hat{\beta} + \bar{X}_{I,t-1} \hat{\beta}^{T-1} \end{aligned}$$

onde $\bar{X}_{L,t-1}$ e $\bar{X}_{I,t-1}$ representam os valores médios dos determinantes dos ganhos mensais na região do Litoral e do Interior no momento t-1.

O logaritmo natural das diferenças salariais no momento t-1 virá então igual a:

$$(10) \quad \ln \bar{W}_{L,t-1} - \ln \bar{W}_{I,t-1} = (\bar{X}_L - \bar{X}_I)_{t-1} \hat{\beta} + (\bar{X}_L - \bar{X}_I)_{t-1} \hat{\beta}^{T-1} \\ + \bar{X}_{L,t-1} (\hat{\beta}^L + \hat{\beta}^{L,T-1})$$

Ao adotarmos idênticos procedimentos para derivar o logaritmo natural dos salários para ambos os espaços geográficos no momento t (T-1=0), temos que:

$$\ln \bar{W}_{L,t} = \bar{X}_{L,t} \hat{\beta} + \bar{X}_{L,t} \hat{\beta}^L \\ \ln \bar{W}_{I,t} = \bar{X}_{I,t} \hat{\beta}$$

Assim, o logaritmo das diferenças salariais no momento t **será igual a:**

$$(11) \quad (\ln \bar{W}_L - \ln \bar{W}_I)_t = (\bar{X}_L - \bar{X}_I)_t \hat{\beta} + \bar{X}_{L,t} \hat{\beta}^L$$

Assim, as alterações na razão dos salários entre dois momentos distintos obtêm-se recorrendo às expressões (10) e (11), ou seja:

$$(12) \quad (\ln \bar{W}_L - \ln \bar{W}_I)_t - (\ln \bar{W}_L - \ln \bar{W}_I)_{t-1} = [(\bar{X}_L - \bar{X}_I)_t - (\bar{X}_L - \bar{X}_I)_{t-1}] \hat{\beta} \\ - \bar{X}_{L,t-1} \hat{\beta}^{L,T-1} - (\bar{X}_L - \bar{X}_I)_{t-1} \hat{\beta}^{T-1} + (\bar{X}_{L,t} - \bar{X}_{L,t-1}) \hat{\beta}^L$$

Se harmonizarmos esta notação com a do modelo de Oaxaca e substituírmos $\hat{\beta}^{T-1}$ por $(\hat{\beta}_{L,t} - \hat{\beta}_{L,t-1})$ e $\hat{\beta}^{L,T-1}$ por $(\hat{\beta}_L - \hat{\beta}_I)_t - (\hat{\beta}_L - \hat{\beta}_I)_{t-1}$ na expressão (12), então teremos a seguinte expressão genérica:

$$(13) \quad (\ln \bar{W}_L - \ln \bar{W}_I)_t - (\ln \bar{W}_L - \ln \bar{W}_I)_{t-1} = \hat{\beta} \left[(\bar{X}_L - \bar{X}_I)_t - (\bar{X}_L - \bar{X}_I)_{t-1} \right] \\ + \bar{X} \left[(\hat{\beta}_L - \hat{\beta}_I)_t - (\hat{\beta}_L - \hat{\beta}_I)_{t-1} \right] + (\hat{\beta}_{L,t} - \hat{\beta}_{L,t-1}) (\bar{X}_L - \bar{X}_I)_{t-1} \\ + (\hat{\beta}_L - \hat{\beta}_I)_t (\bar{X}_{L,t} - \bar{X}_{L,t-1}).$$

A decomposição da expressão (13) permite-nos identificar quatro componentes nas alterações dos salários relativos entre regiões ao longo do tempo.

O primeiro termo identifica a parcela da alteração que é atribuível às variações, ao longo do tempo, nas características médias dos trabalhadores e estabelecimentos que compõem o Litoral e o Interior, assumindo que a estrutura salarial $\hat{\beta}$ é a mesma entre esses espaços geográficos e constante no tempo. Se as características se mantiverem inalteradas, o primeiro termo da expressão (13) será igual a zero.

O segundo termo identifica a parcela da alteração nos salários relativos que é imputada às mudanças ao longo do tempo na estrutura salarial do Litoral em relação ao Interior, assumindo que não há alterações nas características dos trabalhadores/estabelecimentos. Assim, se não houver qualquer alteração na estrutura salarial do Litoral relativamente à do Interior, esta componente será igual a zero.

Mesmo não ocorrendo qualquer alteração no primeiro e segundo termos, ou seja, mesmo que não se registre mudanças nas características médias relativas a ambas as áreas geográficas, nem alterações nas taxas de rendibilidade ou prémios salariais, os salários relativos entre o Litoral e o Interior podem alterar-se. Assim, o terceiro termo permite avaliar o efeito das alterações que afectam indistintamente e na mesma proporção as estruturas salariais de ambas as regiões e que acabam por alterar a dispersão salarial total em virtude das diferenças existentes nas características dos trabalhadores/estabelecimentos.

Da mesma forma, se um aumento de igual amplitude ocorrer nas características de ambos os espaços geográficos, a diferença salarial inter-regional alterar-se-á se as estruturas salariais forem diferentes. Se, por exemplo, as taxas de rendibilidade são maiores no Litoral e se regista um aumento dos anos médios de escolaridade de igual montante para ambas as regiões, então as diferenças salariais irão aumentar. Esta parcela da alteração da diferença salarial é-nos indicada pelo quarto termo da expressão (13).

Por fim, repare-se que a decomposição da variação do logaritmo da rácio dos salários médios mensais entre o Litoral e o Interior ao longo do tempo pode ser sensível às hipóteses adoptadas, nomeadamente quanto às características ou à estrutura salarial no caso de estarmos num mercado onde não se registam diferenças salariais. A importância de cada uma das componentes revelou-se, no entanto, robusta em relação às diferentes formulações, tendo-se adoptado pela seguinte expressão final:

$$(14) \quad (\ln \bar{W}_L - \ln \bar{W}_I)_t - (\ln \bar{W}_L - \ln \bar{W}_I)_{t-1} = \hat{\beta}_{L,t} \left[(\bar{X}_L - \bar{X}_I)_t - (\bar{X}_L - \bar{X}_I)_{t-1} \right] \\ + \bar{X}_{I,t-1} \left[(\hat{\beta}_L - \hat{\beta}_I)_t - (\hat{\beta}_L - \hat{\beta}_I)_{t-1} \right] + (\hat{\beta}_{L,t} - \hat{\beta}_{L,t-1}) (\bar{X}_L - \bar{X}_I)_{t-1} \\ + (\hat{\beta}_L - \hat{\beta}_I)_t (\bar{X}_{I,t} - \bar{X}_{I,t-1}).$$

Com base nas expressões (8) e (14), podemos agora medir o contributo das características dos trabalhadores/estabelecimentos e das diferenças nas taxas de rendibilidade para a dispersão salarial total entre as regiões do Centro Litoral e do Centro Interior nos anos de 1988 e 1996, bem como as alterações registadas no logaritmo da rácio dos salários médios entre os anos de 1988 e 1991 e entre 1991 e 1994. (Repare-se que, para este exercício, a escolha dos anos é algo arbitrária, visando-se sobretudo a ilustração do método).

3.3 IDENTIFICAÇÃO DAS PRINCIPAIS COMPONENTES DA DISPERSÃO SALARIAL

Neste ponto, vamos identificar as componentes da dispersão salarial total entre o Centro Litoral e o Centro Interior nos anos de 1988 e 1996 (primeiro e último ano, respectivamente, do período em análise). Como já tivemos oportunidade de salientar na secção 3.1, assumimos que a dispersão salarial é composta por duas parcelas: uma primeira que reflecte as diferenças nas características dos trabalhadores, empregos e/ou estabelecimentos; e uma segunda que reflecte a existência de diferentes estruturas salariais. Para obtenção destes dois efeitos, recorreremos então à expressão (8), utilizando para média das características os valores constantes do Anexo II. Por sua vez, os coeficientes associados a essas características resultam da estimação da função salário da região Centro Litoral e Centro Interior constantes do Quadro 1.

De acordo com o Quadro 2, em 1988, as diferenças existentes entre características eram responsáveis por cerca de 66% da dispersão total dos salários entre o Centro Litoral e o Centro Interior, enquanto que as diferenças entre taxas de rendibilidade e prémios salariais representam 34%. Por outras palavras, o efeito *retorno* era responsável por um ganho adicional de cerca de 5% (pontos-base) por parte dos trabalhadores do Litoral, enquanto o efeito *médias* era responsável pelos restantes 9%, de um *prémio* total de aproximadamente 14%. Tendo em conta que, de acordo com os resultados da estimação da função salário apresentados no Quadro 1, se determinou que os trabalhadores do Interior auferiam, em 1988, menos 5.5% do que os seus homólogos do Litoral, verifica-se assim que a variável *dummy* de localização capta, aproximadamente, o efeito *retorno*.

De entre as variáveis utilizadas para explicar o comportamento dos ganhos mensais nominais, destaca-se claramente o sector de actividade e a dimensão do estabelecimento (ambos incluídos na variável *indústria*), responsáveis por 0.122/0.140, ou seja, 87% do total da diferença salarial entre o Litoral e o Interior.

Quanto às variáveis de capital humano, verifica-se que as diferenças no nível de escolaridade, antiguidade contribuem para que os salários mensais nominais sejam maiores no Centro Litoral (o efeito *médias* é positivo). No entanto, as respectivas taxas de rendibilidade são superiores no Interior, sendo, no global, este efeito (efeito *retorno*) superior ao efeito *médias*, o que origina um efeito total favorável aos trabalhadores do Interior. Neste contexto, a variável experiência (uma *proxy* da formação geral) parece ser a única excepção: pelo efeito *médias*, os salários médios do Interior seriam ligeiramente superiores aos do Litoral, mas o efeito *retorno* é favorável aos trabalhadores do Litoral (a taxa de rendibilidade entre trabalhadores com os mesmos anos de experiência é superior no Litoral). Como este último efeito é muito superior ao efeito *médias*, a dispersão salarial líquida a favor do Litoral é positiva (0.0119).

De mais difícil interpretação é variável *outras variáveis pessoais*, por resultar da agregação de vários termos de interacção entre os níveis de escolaridade, antiguidade e experiência. O mesmo se poderá dizer da variável *constante*, pois esta resulta da diferença nos coeficientes estimados dos termos constantes das funções salário restritas, representando os efeitos na dispersão total dos salários de variáveis não incluídas no modelo.

No ano de 1996, a dispersão total dos salários aumentou, comparativamente ao ano de 1988, em cerca de 3% (pontos-base), sendo que agora as diferenças entre as características pessoais dos trabalhadores, empregos e estabelecimentos são responsáveis por 63% da dispersão total dos salários entre o Centro Litoral e o Centro Interior, contra 66% em 1988. Pode concluir-se assim que a desigualdade salarial resultante do efeito *retorno* tem vindo a assumir um papel crescente na dispersão total dos salários inter-regional, sendo responsável, em 1996, pelo facto dos trabalhadores do Litoral auferirem, em média, mais 6.06% do que os trabalhadores do Interior.

Tal como em 1988, a fonte predominante de dispersão salarial em 1996 é a especialização sectorial, com a variável *indústria* a representar 81% da diferença observada. Quanto às variáveis de capital humano, verifica-se que o contributo das diferenças entre as dotações de capital humano aumentou em 1996, nomeadamente no que se refere à antiguidade e à experiência. Quanto ao efeito *retorno*, é de realçar o aumento deste efeito na variável antiguidade e no logaritmo das horas, que se deve sobretudo à variação nos valores da elasticidade dos ganhos em relação às horas de trabalho mensais estimados na especificação geral da função salário.

Em síntese, o exercício de decomposição da dispersão dos salários permitiu confirmar que os salários nominais são, em média, mais elevados no Centro Litoral, resultando esta disparidade salarial sobretudo do efeito *médias*, ou seja, de diferenças entre características de oferta e procura de trabalho. Deve salientar-se, no entanto, que as diferenças nas taxas de rendibilidade e prémios salariais – a verdadeira fonte de desigualdade segundo Oaxaca – têm vindo a acentuar-se em favor do Litoral. De particular importância e motivo de preocupação parece ser o papel extremamente decisivo do sector de actividade na explicação da dispersão observada.

3.4 EVOLUÇÃO DA DISPERSÃO TOTAL DOS SALÁRIOS

Vimos na secção anterior que os salários no Litoral não só são superiores aos do Interior como também têm vindo, comparativamente, a aumentar. Pretende-se agora, com base no modelo de Farber e Newman, apresentado na secção 3.2, identificar as fontes que contribuem para o agravamento da desigualdade inter-regional, seleccionando, para ilustração do método, os períodos 1988-1991 e 1991-1994. (Entre 1994 e 1996, a variação do salário médio relativo foi praticamente nula).

As quatro componentes da variação do logaritmo da rácio dos salários médios mensais entre o Litoral e o Interior, obtidas a partir da equação (14), estão indicadas no Quadro 3, sendo de destacar, desde logo, o 2º e 3º termos, isto é, o efeito das alterações nas estruturas salariais ocorridas no Litoral ou no Interior ao longo do tempo (2º termo) e o efeito das alterações nas estruturas salariais que afectaram indistintamente ambas as regiões (3º termo).

Repare-se que, com base nos resultados do Quadro 2, tínhamos concluído que a principal componente da dispersão total dos salários, em cada um dos anos analisados, eram as diferenças existentes entre características dos trabalhadores, empregos e estabelecimentos (efeito *médias*), mas que, paralelamente, o efeito *retorno* parecia estar a adquirir maior importância. Se considerarmos agora a evolução da dispersão salarial no tempo, é possível confirmar de facto, através do peso do 2º e 3º termos, que o agravamento da dispersão se deve sobretudo às alterações na estrutura salarial.

No período 1988-1991, verificou-se um aumento substancial nas diferenças das estruturas salariais (2º termo) a favor do Litoral. A decomposição de cada um dos termos

segundo as características observadas mostra, por outro lado, que as alterações nas taxas de rendibilidade afectaram sobretudo as variáveis de capital humano, nomeadamente o nível de escolaridade e a antiguidade. Em relação ao período 1991-1994, pelo contrário, as taxas de rendibilidade das variáveis de capital humano evoluíram favoravelmente para a região Centro Interior, pelo que o 2º termo, sendo novamente (a par do 3º termo) muito substancial, actua em sentido inverso ao do período anterior (-0.014). Sendo as características de procura e oferta de trabalho relativamente estáveis (o 1º termo é, em ambos os períodos, de reduzida expressão), encontra-se assim explicada a modesta variação registada na rácio do logaritmo dos salários médios entre o Centro Litoral e o Centro Interior.

Tomando em conjunto os resultados dos Quadros 2 e 3, parece pois poder concluir-se que os factores explicativos da dispersão salarial num dado momento e da sua evolução ao longo do tempo não são necessariamente os mesmos. Com efeito, enquanto as características dos trabalhadores, dos empregos e dos estabelecimentos são as componentes decisivas na explicação da dispersão salarial entre os trabalhadores do Centro Litoral e do Centro Interior, já o agravamento da dispersão salarial parece resultar mais das alterações ocorridas nas taxas de rendibilidade e prémios salariais e portanto da estrutura salarial das duas sub-regiões.

4. CONCLUSÃO

Neste estudo, procurou-se identificar as componentes da dispersão dos salários médios entre o Litoral e o Interior da região Centro. Tratou-se de uma abordagem preliminar, que visa esclarecer, a partir das funções salário típicas dos modelos de capital humano, se a dispersão total dos salários entre as aquelas duas sub-regiões se deve à existência de diferentes características dos trabalhadores e dos estabelecimentos/empregos ou à ocorrência de diferentes taxas de rendibilidade e à prática de diferentes prémios salariais para trabalhadores e estabelecimentos com idênticas características, residindo nesta última componente a noção de desigualdade salarial inter-regional.

Na primeira parte do trabalho, concluímos que a fonte principal de dispersão dos salários, em cada um dos anos amostrais compreendidos entre 1988 e 1996, seriam as características dos trabalhadores, dos empregos e dos estabelecimentos, com destaque para o contributo da variável *indústria*, o que deixa sugerir que sem alterar a matriz de especialização sectorial não será possível corrigir as actuais desigualdades salariais inter-regionais.

Na segunda parte, a análise da evolução da dispersão salarial mostrou que o agravamento da desigualdade salarial entre o Litoral e o Interior prende-se mais com as alterações ocorridas nas diferenças das taxas de rendibilidade e prémios salariais do que com a variação nas diferenças de características dos trabalhadores, dos empregos ou das indústrias, que, em larga medida, se mantiveram estáveis.

Certamente muitos outros aspectos ficaram por analisar. Em particular, parece-nos que seria importante proceder, em trabalhos futuros, a uma análise semelhante aplicada a outras regiões do país. Por outro lado, seria igualmente interessante avaliar a amplitude da dispersão salarial inter-indústrias em cada um dos níveis de qualificação, bem como determinar, com maior precisão, se as indústrias que pagam salários elevados para a execução de tarefas complexas pagam igualmente salários elevados para a execução de tarefas de menor complexidade.

QUADRO 1 – Resultados da estimação da especificação geral da função salário para a totalidade dos trabalhadores por conta de outrem (Região Centro)

	REGIÃO CENTRO			
	1988	1991	1994	1996
CONSTANTE	8.3782 (0.0455)	8.8351 (0.0517)	8.4827 (0.0526)	9.0037 (0.0462)
ILETRADOS	-0.0898 (0.0092)	-0.0702 (0.0113)	-0.0511 (0.0143)	-0.0953 (0.0138)
PRIMÁRIA	-0.1004 (0.0054)	-0.0899 (0.0056)	-0.0786 (0.0057)	-0.0730 (0.0051)
PREPARATÓRIO	-0.1333 (0.0058)	-0.0986 (0.0058)	-0.0752 (0.0056)	-0.0697 (0.0049)
SECUNDÁRIO	0.0915 (0.0064)	0.1043 (0.0065)	0.0396 (0.0065)	0.0469 (0.0054)
ENSINO MÉDIO	0.2043 (0.0211)	0.1699 (0.0218)	-	-
BACHARELATO	0.3468 (0.0248)	0.2918 (0.0209)	0.2907 (0.0147)	0.1898 (0.0107)
LICENCIATURA	0.3270 (0.0152)	0.4065 (0.0149)	0.4383 (0.0141)	0.4258 (0.0095)
ANTIGUIDADE	0.0200 (0.0003)	0.0199 (0.0003)	0.0202 (0.0004)	0.0215 (0.0004)
ANTIGUIDADE²	-0.0002 (0.00003)	-0.0002 (0.00002)	-0.0002 (0.00001)	-0.0002 (0.00004)
EXPERIÊNCIA	0.0132 (0.0004)	0.0146 (0.0004)	0.0132 (0.0003)	0.0140 (0.0003)
EXPERIÊNCIA²	-0.0002 (0.00003)	-0.0002+ (0.0004)	-0.0002 (0.00001)	-0.0002 (0.00002)
ILETRADOS*ANT	-0.0105 (0.0004)	-0.0121 (0.0004)	-0.0140 (0.0005)	-0.0125 (0.0005)
PRIMARIA*ANT	-0.0064 (0.0003)	-0.0071 (0.0003)	-0.0084 (0.0003)	-0.0087 (0.0002)
PREPARATÓRIO *ANT	0.0002+ (0.0003)	-0.0013 (0.0004)	-0.0031 (0.0003)	-0.0028 (0.0003)
SECUNDÁRIO *ANT	-0.0030 (0.0004)	-0.0033 (0.0004)	0.0011 (0.0004)	0.0022 (0.0004)
ENSINO MÉDIO*ANT	-0.0100 (0.0014)	0.0006* (0.0015)	-	-
BACHARELATO*ANT	-0.0057 (0.0015)	-0.0004+ (0.0013)	0.0002+ (0.0011)	0.0054 (0.0008)
LICENCIATURA*ANT	-0.0042 (0.0011)	-0.0032 (0.0011)	-0.0027* (0.0011)	0.0043 (0.0008)

ILETRADOS*EXP	-0.0022 (0.0004)	-0.0038 (0.0004)	-0.0039 (0.0005)	-0.0037 (0.0004)
PRIMARIA*EXP	-0.0013 (0.0003)	-0.0026 (0.0003)	-0.0028 (0.0003)	-0.0032 (0.0003)
PREPARATÓRIO *EXP	0.0030 (0.0004)	0.00002+ (0.0004)	-0.0009 (0.0003)	-0.0020 (0.0003)
SECUNDÁRIA*EXP	-0.0007** (0.0004)	-0.0014 (0.0004)	0.0030 (0.0005)	0.0013 (0.0004)
ENSINO MÉDIO*EXP	0.0000+ (0.0013)	0.0006+ (0.0014)	-	-
BACHARELATO*EXP	-0.0048* (0.0019)	0.0027** (0.0015)	-0.0008+ (0.0011)	0.0030 (0.0009)
LICENCIATURA*EXP	-0.0081 (0.0011)	0.0003+ (0.0011)	0.0012+ (0.0011)	-0.0010+ (0.0007)
LnHORAS	0.3261 (0.0086)	0.3179 (0.0099)	0.4231 (0.0100)	0.3470 (0.0088)
MULHER	-0.1440 (0.0015)	-0.1719 (0.0016)	-0.1482 (0.0017)	-0.1628 (0.0015)
OCUP1	0.0015 (0.0021)	0.0982 (0.0023)	0.0940 (0.0026)	0.0696 (0.0024)
OCUP2	0.1596 (0.0019)	0.1824 (0.0021)	0.1747 (0.0024)	0.1598 (0.0021)
OCUP3	0.3297 (0.0037)	0.3692 (0.0041)	0.3580 (0.0044)	0.3398 (0.0037)
OCUP4	0.3791 (0.0034)	0.4130 (0.0036)	0.4260 (0.0041)	0.4142 (0.0036)
OCUP5	0.4881 (0.0060)	0.5353 (0.0059)	0.5213 (0.0059)	0.5207 (0.0049)
OCUP6	0.5987 (0.0069)	0.6284 (0.0069)	0.6084 (0.0070)	0.5789 (0.0053)
FIRM5A9	0.0623 (0.0033)	0.0728 (0.0034)	0.0964 (0.0033)	0.0845 (0.0027)
FIRM10A49	0.1449 (0.0029)	0.1642 (0.0030)	0.1727 (0.0029)	0.1629 (0.0024)
FIRM50A99	0.1855 (0.0032)	0.2040 (0.0033)	0.2294 (0.0033)	0.2154 (0.0028)
FIRM>100	0.2955 (0.0031)	0.2733 (0.0032)	0.3242 (0.0032)	0.3002 (0.0027)
EXTMINERAIS	0.2053 (0.0093)	0.2341 (0.0088)	0.2876 (0.0097)	0.2878 (0.0084)
INDALBEBIDAS	0.1178	0.1259	0.1710	0.1645

	(0.0029)	(0.0034)	(0.0038)	(0.0034)
INDMADCORTIÇA	-0.0036+ (0.0032)	0.0480 (0.0036)	0.0924 (0.0041)	0.0814 (0.0042)
INDPAPEL	0.3340 (0.0040)	0.3855 (0.0046)	0.3882 (0.0056)	0.3580 (0.0050)
INDQUIMICA	0.2358 (0.0035)	0.2085 (0.0039)	0.2787 (0.0049)	0.2560 (0.0041)
INDPRODMINERAIS	0.1964 (0.0026)	0.2375 (0.0029)	0.3006 (0.0034)	0.2859 (0.0030)
INDMETAL	0.2200 (0.0048)	0.2280 (0.0061)	0.3324 (0.0077)	0.3574 (0.0106)
INDPRODMETÁLICOS	0.1599 (0.0026)	0.1972 (0.0027)	0.2924 (0.0030)	0.2762 (0.0027)
OUTINDTRANSFORM	0.1127 (0.0130)	0.1021 (0.0137)	0.1350 (0.0203)	0.1534 (0.0050)
ELEGASAGUA	0.7464 (0.0047)	0.8394 (0.0059)	0.8799 (0.0071)	0.8409 (0.0067)
OBRAPUBLICAS	0.1263 (0.0029)	0.1255 (0.0030)	0.1696 (0.0035)	0.1475 (0.0031)
COMERGROSSO	0.1947 (0.0032)	0.1823 (0.0035)	0.2699 (0.0039)	0.2281 (0.0035)
COMERARETALHO	0.1350 (0.0031)	0.1252 (0.0034)	0.2180 (0.0035)	0.2004 (0.0030)
RESTCAFE	0.0502 (0.0044)	0.0381 (0.0047)	0.0739 (0.0044)	0.0511 (0.0039)
TRANSARMAZENAGEM	0.4326 (0.0035)	0.3654 (0.0038)	0.4433 (0.0044)	0.4021 (0.0041)
COMUNICAÇÕES	0.6237 (0.0040)	0.5245 (0.0045)	0.6456 (0.0051)	0.6049 (0.0049)
BANCOS	0.7966 (0.0047)	0.7023 (0.0052)	0.9133 (0.0057)	0.9101 (0.0050)
SEGUROS	0.8889 (0.0089)	0.7484 (0.0106)	0.8229 (0.0104)	0.7859 (0.0097)
OPIMOVEIS	0.2055 (0.0080)	0.1199 (0.0065)	0.2710 (0.0068)	0.1779 (0.0123)
SERSOCIAIS	0.1362 (0.0043)	0.1259 (0.0043)	0.1835 (0.0041)	0.1942 (0.0037)
SERCULTURAIS	0.0194+ (0.0128)	0.0481 (0.0126)	0.1671 (0.0131)	0.1498 (0.0133)
SERDOMÉSTICOS	0.1053 (0.0044)	0.1318 (0.0050)	0.2035 (0.0055)	0.2119 (0.0050)

INTERIOR	-0.0550 (0.0013)	-0.0627 (0.0015)	-0.0519 (0.0016)	-0.0647 (0.0014)
R² ajustado	0.67	0.61	0.57	0.61
F- estatístico	5408	4777	4515	5967
Desvio-Padrão	0.23496	0.27724	0.31397	0.29766
N. de observações	158338	184248	193212	221690

Notas: Os desvios-padrão dos coeficientes estão entre parêntesis. Todos os coeficientes são estatisticamente significativos a um nível de 1%, excepto os coeficientes que apresentam as seguintes anotações: * estatisticamente significativo para um nível de 5%; ** significativo a 10%; + não significativo. Variáveis omissas: Secundário Unificado; OCUP0 (profissionais não qualificados); FIRM1A4 (estabelecimentos com menos de 4 trabalhadores); HOMENS; INDTEXVE (indústrias têxteis, do vestuário e do couro); e LITORAL. As estimações foram efectuadas a partir dos dados dos quadros de pessoal do DETEFP, Ministério do Trabalho e da Solidariedade. A definição das variáveis encontra-se no Anexo 1.

QUADRO 2 – A dispersão total dos salários no Centro Litoral e Centro Interior (1988 e 1996)

	1988			1996		
	Efeito médias	Efeito retorno	ln (W _L) – ln (W _I)	Efeito médias	Efeito retorno	Ln (W _L) – ln (W _I)
Escolaridade	0.0098	-0.0237	-0.0139	0.0088	-0.0223	-0.0135
ANTIGUIDADE	0.0030	-0.0416	-0.0386	0.0065	0.6926	0.6991
EXPERIÊNCIA	-0.0015	0.0134	0.0119	-0.0040	0.0102	0.0062
MULHER	0.0033	-0.0114	-0.0081	0.0035	-0.0147	-0.0112
OUT.VAR.PESSO AIS	-0.0024	-0.00114	-0.0138	0.0028	-0.0093	-0.0065
OCUPAÇÕES	0.0127	0.0170	0.0297	0.0115	0.0141	0.0256
INDÚSTRIA	0.0667	0.0508	0.1215	0.0758	0.0587	0.1345
CONSTANTE		0.0508	0.0508		-0.6687	-0.6687
TOTAL	0.0916	0.0479	0.1395	0.1049	0.0606	0.1655
%	66	34	100	63	37	100
DISPERSÃO OBSERVADA			0.138			0.164

Notas: Os cálculos baseiam-se na equação (8):

$$\ln(\bar{W}_L) - \ln(\bar{W}_I) = (\bar{X}_L - \bar{X}_I)\hat{\beta}_L + \bar{X}_I(\hat{\beta}_L - \hat{\beta}_I),$$

onde o primeiro termo representa o efeito médias e o segundo o efeito retorno. Os coeficientes $\hat{\beta}_L, \hat{\beta}_I$ foram obtidos a partir da função salário estimada no Quadro 1 e os valores médios \bar{X}_I, \bar{X}_L são os que constam do Anexo 2. As variáveis OUTRAS VARIÁVEIS PESSOAIS (OUT.VAR.PESSOAIS), INDÚSTRIA e ANTIGUIDADE incluem, respectivamente, os efeitos dos termos de interdependência entre os níveis de escolaridade, a antiguidade e a experiência, os efeitos relativos aos sectores de

actividade e à dimensão dos estabelecimentos e o efeito da variável logaritmo das horas (LnHORAS). Neste último caso, ao associarmos as variáveis logaritmo natural das horas trabalhadas e antiguidade, presume-se que a formação profissional específica num local de trabalho é tanto mais elevada quanto mais horas forem trabalhadas e quanto mais anos de antiguidade possuírem os trabalhadores. A definição das variáveis encontra-se no Anexo 1.

QUADRO 3 – Análise da dispersão total dos salários no Centro Litoral e Centro Interior (1988-1991)

	1988-1991				
	1º TERMO	2º TERMO	3º TERMO	4º TERMO	$[\ln (W_{it}) - \ln (W_{it-1})]_t$ - $[\ln (W_{it}) - \ln (W_{it-1})]_{t-1}$
Escolaridade	-0.0013	0.0310	0.0011	0.0001	0.0310
ANTIGUIDADE	0.0065	0.6104	0.0002	-0.0079	0.6091
EXPERIÊNCIA	0.0008	0.0030	-0.0002	0.0001	0.0037
MULHER	0.0034	-0.0035	0.0007	-0.0026	-0.0019
OUT.VAR.PESSO AIS	-0.0034	-0.0191	-0.0006	0.0045	-0.0186
OCUPAÇÕES	-0.0020	-0.0139	-0.0005	-	-0.0164
INDÚSTRIA	0.0019	-0.0008	0.0009	0.0014	0.0034
CONSTANTE	-	-0.5943	-	-	-0.5943
TOTAL	0.006	0.013	0.002	-0.004	0.017
DISPERSÃO OBSERVADA					0.018

Os coeficientes $\hat{\beta}, \hat{\beta}$ foram obtidos a partir da função salário estimada no Quadro 1 e os valores médios \bar{X}, \bar{X} são os que constam do Anexo 2. As variáveis OUTRAS VARIÁVEIS PESSOAIS, INDÚSTRIA e ANTIGUIDADE incluem, respectivamente, os efeitos dos termos de interdependência entre os níveis de escolaridade, a antiguidade e a experiência, os efeitos relativos aos sectores de actividade e à dimensão dos estabelecimentos e o efeito da variável logaritmo das horas (LnHORAS). Neste último caso, ao associarmos as variáveis logaritmo natural das horas trabalhadas e antiguidade, presume-se que a formação profissional específica num local de trabalho é tanto mais elevada quanto mais horas forem trabalhadas e quanto mais anos de antiguidade possuírem os trabalhadores. A diferença entre o valor total e a variação observada deve-se aos arredondamentos efectuados. A definição das variáveis encontra-se no Anexo 1.

REFERÊNCIAS BIBLIOGRÁFICAS

- BELLANTE, DON (1979). The North-South Differentials and the Migration of Heterogeneous Labor. *American Economic Review* 69, 166-175.
- BEESON, PATRICIA E. e GROSHEN, ERICA L. (1991). Components of City-Size Wage Differentials, 1973-1988. *Economic Review*, Federal Reserve Bank of Cleveland 27, 10-24.
- COELHO, PHILIP R. P. e GHALI, MOHEB A. (1971). The End of the North- South Wage. *American Economic Review* 61, 932- 937.
- DICKENS, WILLIAM T. e KATZ, LAWRENCE F. (1987). Inter-industry Wage Differences and Theories of Wage Determination. *NBER Working Paper* 2271.
- FARBER, STEPHEN C. e NEWMAN, ROBERT J. (1987). Accounting for South/Non-South Real Wage Differentials and for Changes in those Differentials Over Time. *Review of Economics and Statistics* 69, 215-223.
- GERKING, SCHELBY e WERICK, WILLIAM (1983). Compensating Differences and Interregional Wage Differentials. *Review of Economic and Statistics* 65, 483-487.
- HANUSHEK, ERIC A. (1973). Regional Differences in the Structure of Earnings. *Review of Economic and Statistics* 55, 204-213.
- JACKSON, LORIE D. (1986). The Changing Nature of Regional Wage Differentials From 1975 to 1983. *Economic Review*, Federal Reserve Bank of Cleveland 1, 12-23.
- KRUEGER, ALAN B. e SUMMERS, LAWRENCE H. (1988). Efficiency Wages and the Interindustry Wage Structure. *Econometrica* 56, 259-293.
- MARTINS, ANA PAULA (1991). Human Capital Earnings Functions: The Portuguese Case. *Working Paper*, Universidade Católica Portuguesa.
- OAXACA, RONALD (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review* 14, 693-709.
- SAHLING, LEONARD G. e SMITH, SHARON P. (1983). Regional Wage Differentials: Has the South Risen Again. *Review of Economic and Statistics* 65, 131-135.
- SANTOS, FILIPE (1999). Estrutura e Dispersão Salarial na Região Centro: Função-Salário, Decomposição Salarial e Análise Intertemporal. *Dissertação de Mestrado*, Faculdade de Economia da Universidade de Coimbra.
- SANTOS, MARIA CLEMENTINA (1995). Education and Earnings Differentials in Portugal. *Dissertação de Doutoramento*, Universidade do Porto.
- VIEIRA, JOSÉ CABRAL (1992). Diferenças Salariais e Afectação no Mercado de Trabalho: Uma Aplicação nos Açores, 1989. *Dissertação de Mestrado*, Universidade Nova de Lisboa

ANEXO 1 – Definição das variáveis

VARIÁVEIS	DESCRIÇÃO
<i>LnW</i>	Logaritmo natural dos ganhos mensais em escudos
Níveis de escolaridade	
<i>EDUCAÇÃO</i>	Número de anos completos de escolaridade.
<i>ILETRADOS</i>	1 se o trabalhador não souber ler nem escrever ou não tiver concluído o ensino básico primário ou primeiro ciclo, ou então 0.
<i>PRIMARIA</i>	1 se o nível de escolaridade completo for de 4 anos (conclusão do ensino básico primário ou primeiro ciclo), ou então 0.
<i>PREPARATÓRIO</i>	1 se o nível de escolaridade completo de 6 anos (conclusão do ensino básico preparatório ou segundo ciclo), ou então 0.
<i>SEC. UNIFICADO</i>	1 se o nível de escolaridade completo for de 9 anos (conclusão do curso geral dos liceus ou terceiro ciclo), ou então 0.
<i>SECUNDÁRIO</i>	1 se o nível de escolaridade completo for de 11 anos (conclusão do ensino secundário, ou curso complementar dos liceus), ou então 0.
<i>ENSINO MÉDIO</i>	1 se o nível de escolaridade completo for de 12 anos (conclusão dos cursos de enfermagem, ou dos cursos ministrados pelos Institutos Comercial e Industrial e Magistérios) ou então 0.
<i>BACHARELATO</i>	1 se o nível de escolaridade completo for de 14 anos (conclusão do bacharelato), ou então 0.
<i>LICENCIATURA</i>	1 se o nível de escolaridade completo for de 16 anos (conclusão da licenciatura), ou então 0.
Antiguidade	
<i>ANTIGUIDADE</i>	Número de anos ao serviço na mesma empresa.
<i>ANTIGUIDADE²</i>	ANTIGUIDADE ao quadrado
Experiência	
<i>EXPERIÊNCIA</i>	Experiência Potencial = idade - anos de escolaridade - antiguidade - 6
<i>EXPERIÊNCIA²</i>	EXPERIÊNCIA ao quadrado.
Termos de interdependência	
<i>ILETRADOS*ANT</i>	Termo interactivo (ILETRADOS x ANT)
<i>PRIMARIA*ANT</i>	Termo interactivo (PRIMARIA x ANT)
<i>PREPARATÓRIO*ANT</i>	Termo interactivo (PREPARATÓRIO x ANT)
<i>SEC. UNIFICADO*ANT</i>	Termo interactivo (SEC. UNIFICADO x ANT)
<i>SECUNDÁRIO*ANT</i>	Termo interactivo (SECUNDÁRIO x ANT)
<i>ENSINO MÉDIO*ANT</i>	Termo interactivo (ENSINO MEDIO x ANT)
<i>BACHARELATO*ANT</i>	Termo interactivo (BACHARELATO x ANT)
<i>LICENCIATURA*ANT</i>	Termo interactivo (LICENCIATURA. x ANT)
<i>ILETRADOS*EXP</i>	Termo interactivo (ILETRADOS x EXP)
<i>PRIMARIA*EXP</i>	Termo interactivo (PRIMARIA x EXP)
<i>PREPARATÓRIO*EXP</i>	Termo interactivo (PREPARATÓRIO x EXP)
<i>SEC. UNIFICADO*EXP</i>	Termo interactivo (SEC. UNIFICADO x EXP)
<i>SECUNDÁRIO*EXP</i>	Termo interactivo (SECUNDÁRIO x EXP)
<i>ENSINO MÉDIO*EXP</i>	Termo interactivo (ENSINO MEDIO x EXP)

VARIÁVEIS	DESCRIÇÃO
<i>BACHARELATO*EXP</i>	Termo interactivo (BACHARELATO x EXP)
<i>LICENCIATURA*EXP</i>	Termo interactivo (LICENCIATURAx EXP)
Horas trabalhadas	
<i>LnHORAS</i>	Logaritmo das horas mensais trabalhadas.
Sexo	
<i>MULHER</i>	1 se estivermos perante um trabalhador do sexo feminino, ou então 0.
Dimensão do estabelecimento	
<i>FIRM1A4</i>	1 se o número de pessoas ao serviço no estabelecimento for maior ou igual a 1 e menor ou igual a 4, ou então 0.
<i>FIRM5A 9</i>	1 se o número de pessoas ao serviço no estabelecimento for maior ou igual a 5 e menor ou igual a 9, ou então 0.
<i>FIRM10A49</i>	1 se o número de pessoas ao serviço no estabelecimento for maior ou igual a 10 e menor ou igual a 49, ou então 0.
<i>FIRM50A 99</i>	1 se o número de pessoas ao serviço no estabelecimento for maior ou igual a 50 e menor ou igual a 99, ou então 0.
<i>FIRM100</i>	1 se o número de pessoas ao serviço no estabelecimento for maior ou igual a 100, ou então 0.
Sectores de actividade	
	(CAE REV) – não estão representados as CAEs 21 e 22 (indústria extractiva do carvão e do petróleo bruto e gás natural), a CAE 42 (abastecimento de água) e a CAE 92 (serviços de saneamento e limpeza). Os sectores da agricultura, silvicultura, caça e pesca e da administração pública e defesa nacional foram também excluídos.
<i>EXMINERAIS</i>	1 se o trabalhador exercer a sua actividade nas indústrias da extracção de minerais não metálicos e rochas industriais (CAE 29), ou então 0. Para o ano de 1996, passa a incluir a divisão 14 da subsecção CB (CAE REV 2).
<i>INDALBEBIDAS</i>	1 se o trabalhador exercer a sua actividade nas industria de alimentação, bebidas e tabaco (CAE 31), ou então 0. Para o ano de 1996, passa a incluir a divisão 15 da subsecção DA (CAE REV 2).
<i>INDTEXVESTUARIO</i>	1 se o trabalhador exercer a sua actividade nas industrias texteis, do vestuário e do couro (CAE 32), ou então 0. Para o ano de 1996, passa a incluir as divisões 17, 18 e 19 da subsecção DB (CAE REV 2).
<i>INDMADCORTIÇA</i>	1 se o trabalhador exercer a sua actividade nas industrias da madeira e da cortiça (CAE 33), ou então 0. Para o ano de 1996, passa a incluir a divisão 20 da subsecção DD (CAE REV 2).
<i>INDPAPEL</i>	1 se o trabalhador exercer a sua actividade nas industrias do papel; artes gráficas e edição de publicações (CAE 34), ou então 0. Para o ano de 1996, passa a incluir as divisões 21 e 22 da subsecção DE (CAE REV 2).
<i>INDQUIMICA</i>	1 se o trabalhador exercer a sua actividade nas industrias químicas, derivados de petróleo, do carvão e dos produtos de borracha de plástico (CAE 35), ou então 0. Para o ano de 1996, passa a incluir as divisões 23, 24, 25 e 37 das subsecções DF, DG, DH e DN (CAE REV 2).
<i>INDPRODMINERAIS</i>	1 se o trabalhador exercer a sua actividade nas industrias de produtos minerais não metálicos, com excepção dos derivados do petróleo (CAE 36), ou então 0. Para o ano de 1996, passa a incluir a divisão 26 da subsecção DI (CAE REV 2).
<i>INDMETALICOS</i>	1 se o trabalhador exercer a sua actividade nas industrias metalúrgicas de

VARIÁVEIS	DESCRIÇÃO
	base (CAE 37), ou então 0. Para o ano de 1996, passa a incluir a divisão 27 da subsecção DJ (CAE REV 2).
<i>INPRODMETAIS</i>	1 se o trabalhador exercer a sua actividade na fabricação de produtos metálicos e de máquinas, equipamento e material de transporte (CAE 38), ou então 0. Para o ano de 1996, passa a incluir as divisões 28, 29, 30, 31, 32, 33, 34 e 35 das subsecções DJ, DK, DL e DM (CAE REV 2).
<i>OUTINDUSTRIAS</i>	1 se o trabalhador exercer a sua actividade noutras indústrias transformadoras (CAE 39), ou então 0. Para o ano de 1996, passa a incluir a divisão 36 da subsecção DN (CAE REV 2).
<i>ELEGASAGUA</i>	1 se o trabalhador exercer a sua actividade no sector da electricidade, gás e agua (CAE 41), ou então 0. Para o ano de 1996, passa a incluir a divisão 40 da secção E (CAE REV 2).
<i>OBRAPUBLICAS</i>	1 se o trabalhador exercer a sua actividade no sector da construção civil e obras públicas (CAE 50), ou então 0. Para o ano de 1996, passa a incluir a divisão 45 da secção F (CAE REV 2).
<i>COMERGROSSO</i>	1 se o trabalhador exercer a sua actividade no sector da comércio por grosso (CAE 61), ou então 0. Para o ano de 1996, passa a incluir a divisão 51 da secção G (CAE REV 2).
<i>COMEARETALHO</i>	1 se o trabalhador exercer a sua actividade no sector da comércio a retalho (CAE 62). Para o ano de 1996, passa a incluir as divisões 50 e 52 da secção G (CAE REV 2).
<i>RESTCAFE</i>	1 se o trabalhador exercer a sua actividade no sector dos restaurantes, hotéis, e cafés (CAE 63), ou então 0. Para o ano de 1996, passa a incluir a divisão 55 da secção H (CAE REV 2).
<i>TRANSARMAZENAGEM</i>	1 se o trabalhador exercer a sua actividade no sector dos transportes, armazenagem (CAE 71), ou então 0. Para o ano de 1996, passa a incluir as divisões 60, 61, 62, 63 da secção I (CAE REV 2).
<i>COMUNICAÇÕES</i>	1 se o trabalhador exercer a sua actividade no sector das comunicações (CAE 72), ou então 0. Para o ano de 1996, passa a incluir a divisão 64 da secção I (CAE REV 2).
<i>BANCOS</i>	1 se o trabalhador exercer a sua actividade no sector dos bancos e outras Instituições Monetárias e Financeiras (CAE 81), ou então 0. Para o ano de 1996, passa a incluir a divisão 65 da secção J (CAE REV 2).
<i>SEGUROS</i>	1 se o trabalhador exercer a sua actividade no sector dos seguros (CAE 82), ou então 0. Para o ano de 1996, passa a incluir as divisões 66 e 67 da secção J (CAE REV 2).
<i>OPIMOVEIS</i>	1 se o trabalhador exercer a sua actividade no sector das operações sobre imóveis e serviços prestados as empresas (CAE 83), ou então 0. Para o ano de 1996, passa a incluir as divisões 70, 71 e 72 da secção K (CAE REV 2).
<i>SERSOCIAIS</i>	1 se o trabalhador exercer a sua actividade no sector dos serviços sociais e similares prestados a colectividade (CAE 93), ou então 0. Para o ano de 1996, passa a incluir as divisões 80, 85 e 91 das secções M e N (CAE REV 2).
<i>SERCULTURAIS</i>	1 se o trabalhador exercer a sua actividade no sector dos serviços recreativos e culturais (CAE 94), ou então 0. Para o ano de 1996, passa a incluir a divisão 92 da secção O (CAE REV 2).
<i>SERDOMESTICOS</i>	1 se o trabalhador exercer a sua actividade no sector dos serviços pessoais

VARIÁVEIS	DESCRIÇÃO
	e domésticos (CAE 95), ou então 0. Para o ano de 1996, passa a incluir as divisões 74, 93 e 95 da secção O (CAE REV 2).
Níveis de Qualificação	Baseia-se nos níveis de qualificação, funções e formação regulamentados pelo DL 121/78 de 02 de Junho. Excluíram-se os Aprendizizes e Praticantes.
OCUP0	1 se o trabalhador for Profissional não Qualificado, ou então 0.
OCUP1	1 se o trabalhador for Profissional Semiqualficado, ou então 0.
OCUP2	1 se o trabalhador for Profissional Qualificado, ou então 0.
OCUP3	1 se o trabalhador for Profissional Altamente Qualificado, ou então 0.
OCUP4	1 se o trabalhador for Encarregado, Contramestre ou Chefes de Equipa, ou 0.
OCUP5	1 se o trabalhador for Quadro Médio, ou então 0.
OCUP6	1 se o trabalhador for Quadro Superior, ou então 0.
Localização geográfica	As NUTS III Beira Interior Norte, Beira Interior Sul, Cova da Beira, Dão Lafões, Pinhal Interior Norte, Pinhal Interior Sul e Serra da Estrela constituem o Centro Interior. As NUTS III Baixo Mondego, Baixo Vouga e Pinhal Litoral formam o Centro Litoral.
INTERIOR	1 se o trabalhador exerce a sua actividade no Centro Interior, ou então 0.

ANEXO 2 – Valores médios das varáveis

	LITORAL				INTERIOR			
	1991	1991	1994	1996	1988	1991	1994	1996
LnW	10.748	11.174	11.445	11.543	10.610	11.018	11.281	11.379
EDUCAÇÃO	5.324	5.653	6.134	6.419	4.968	5.369	5.767	6.101
ANTIGUIDADE	9.164	8.454	7.962	8.207	9.082	8.125	7.753	7.575
EXPERIÊNCIA	15.849	16.031	16.055	15.768	15.992	16.098	16.481	16.310
MULHER	0.278	0.312	0.369	0.373	0.299	0.351	0.396	0.393
ILETRADOS	0.084	0.06	0.034	0.029	0.062	0.041	0.029	0.02
PRIMARIA	0.568	0.525	0.454	0.418	0.662	0.597	0.525	0.484
PREPARATÓRIO	0.161	0.2	0.238	0.251	0.137	0.183	0.223	0.236
SEC. UNIFICADO	0.073	0.088	0.143	0.144	0.069	0.087	0.118	0.123
SECUNDARIO	0.099	0.107	0.105	0.12	0.061	0.079	0.088	0.109
ENSINO MEDIO	0.003	0.004	–	–	0.003	0.003	–	–
BACHARELATO	0.004	0.006	0.011	0.015	0.002	0.004	0.008	0.012
LICENCIATURA	0.007	0.011	0.015	0.023	0.004	0.006	0.008	0.015
ILETRADOS*ANT	1.131	0.814	0.433	0.388	0.893	0.556	0.365	0.241
ILETRADOS*EXP	2.130	1.525	0.895	0.737	1.622	1.084	0.745	0.537
PRIMARIA*ANT	5.518	4.930	4.211	4.184	6.213	5.253	4.713	4.397

	LITORAL				INTERIOR			
	1991	1991	1994	1996	1988	1991	1994	1996
PRIMARIA*EXP	10.105	10.009	9.324	8.677	11.508	11.181	10.681	10.090
PREPARATÓRIO*ANT	1.072	1.150	1.406	1.549	0.876	0.958	1.199	1.316
PREPARATÓRIO*EXP	1.751	2.310	2.996	3.280	1.475	2.079	2.770	3.029
SEC. UNIFICADO*ANT	0.540	0.599	1.173	1.176	0.525	0.628	0.905	0.911
SEC. UNIFICADO*EXP	0.767	0.951	1.671	1.671	0.738	0.936	1.371	1.418
SECUNDARIO*ANT	0.813	0.852	0.604	0.698	0.484	0.553	0.478	0.603
SECUNDARIO*EXP	0.955	1.039	0.938	1.090	0.563	0.734	0.775	0.971
ENSINO MEDIO*ANT	0.024	0.023	_	_	0.024	0.023	_	_
ENSINO MEDIO*EXP	0.034	0.039	_	_	0.035	0.038	_	_
BACHARELATO*ANT	0.028	0.044	0.069	0.088	0.019	0.026	0.052	0.062
BACHARELATO*EXP	0.039	0.058	0.091	0.127	0.018	0.031	0.076	0.105
LICENCIATURA*ANT	0.049	0.061	0.091	0.130	0.023	0.028	0.038	0.061
LICENCIATURA*EXP	0.069	0.089	0.115	0.185	0.033	0.051	0.064	0.123
LnHORAS	5.237	5.196	5.176	5.179	5.234	5.187	5.173	5.191
OCUP0	0.138	0.133	0.131	0.124	0.152	0.139	0.127	0.127
OCUP1	0.253	0.243	0.219	0.207	0.292	0.267	0.248	0.216
OCUP2	0.496	0.499	0.516	0.511	0.467	0.495	0.517	0.537
OCUP3	0.042	0.039	0.044	0.052	0.033	0.032	0.038	0.037
OCUP4	0.048	0.054	0.048	0.049	0.036	0.043	0.039	0.04
OCUP5	0.012	0.016	0.021	0.027	0.012	0.014	0.018	0.025
OCUP6	0.012	0.017	0.021	0.031	0.008	0.011	0.012	0.019
FIRM1A4	0.044	0.051	0.073	0.083	0.071	0.081	0.108	0.12
FIRM5A9	0.073	0.08	0.104	0.117	0.093	0.107	0.126	0.136
FIRM1049	0.303	0.303	0.329	0.32	0.341	0.324	0.337	0.334
FIRM5099	0.147	0.153	0.159	0.152	0.136	0.154	0.134	0.141
FIRM100	0.432	0.414	0.336	0.327	0.359	0.334	0.295	0.269
EXMINERAIS	0.006	0.007	0.007	0.007	0.002	0.004	0.005	0.005
INDALBEBIDAS	0.062	0.053	0.051	0.05	0.054	0.050	0.050	0.057
INDTEXVestuário	0.099	0.092	0.077	0.071	0.298	0.285	0.271	0.236
INMADCORTIÇA	0.045	0.041	0.043	0.026	0.068	0.060	0.050	0.039
INDPAPEL	0.03	0.028	0.024	0.024	0.023	0.019	0.012	0.012
INDQUÍMICAS	0.054	0.053	0.040	0.044	0.01	0.010	0.010	0.014
INPRODMINERAIS	0.14	0.139	0.121	0.123	0.026	0.023	0.024	0.023
INDMETAL	0.027	0.019	0.015	0.005	0.002	0.002	0.001	0.001
INPRODMETÁLICOS	0.134	0.145	0.151	0.159	0.051	0.063	0.071	0.071

	LITORAL				INTERIOR			
	1991	1991	1994	1996	1988	1991	1994	1996
OUTINDÚSTRIAS	0.003	0.004	0.001	0.02	0	0.000	0.001	0.018
ELEGASÁGUA	0.016	0.012	0.010	0.009	0.025	0.020	0.016	0.014
OBRAPUBLICAS	0.077	0.092	0.084	0.085	0.108	0.117	0.113	0.127
COMERGROSSO	0.065	0.062	0.066	0.063	0.052	0.051	0.050	0.053
COMEARETALHO	0.063	0.063	0.087	0.108	0.078	0.081	0.096	0.125
RESTCAFE	0.022	0.024	0.039	0.039	0.024	0.027	0.038	0.041
TRANSARM	0.046	0.050	0.042	0.038	0.042	0.034	0.031	0.028
COMUNICAÇÕES	0.027	0.030	0.030	0.024	0.034	0.030	0.020	0.017
BANCOS	0.027	0.027	0.029	0.026	0.029	0.034	0.033	0.03
SEGUROS	0.006	0.005	0.006	0.005	0.004	0.004	0.005	0.005
OPIMOVEIS	0.007	0.010	0.015	0.003	0.004	0.015	0.011	0.002
SERSOCIA IS	0.021	0.025	0.041	0.044	0.034	0.043	0.063	0.064
SERCULTURAIS	0.003	0.004	0.004	0.003	0.001	0.001	0.001	0.001
SERDOMÉSTICOS	0.02	0.017	0.019	0.024	0.029	0.028	0.026	0.017
N. de observações	100519	115697	120226	140180	57819	68551	72986	81510

MAJOR GUIDELINES TOWARDS THE DEVELOPMENT OF A PORTUGUESE SAM^(*)

PRINCIPAIS ORIENTAÇÕES PARA O DESENVOLVIMENTO DE UMA MATRIZ DE CONTABILIDADE SOCIAL

Autor: António Machado Lopes

- Director-Adjunto do Gabinete de Estudos e Conjuntura - Instituto Nacional de Estatística

ABSTRACT:

- This article presents the general framework adopted to build a SAM for the Portuguese economy within the architecture of the national statistical system. The paper is divided in four sections. The first is an assessment of the availability of data to develop a SAM in the near future. The second and the third sections discuss the main breakdown, links and sources, as well as some shortcomings of the available data. The last section concludes with some examples of links between different accounts.

KEY-WORDS:

- *National Accounts; Social Accounts; Household Budget Survey; Household Consumption Patterns; LEG-SAM.*

RESUMO:

- Este artigo apresenta a estrutura geral de desenvolvimento de uma Matriz de Contabilidade Social para Portugal, no quadro do sistema estatístico nacional. O artigo está dividido em quatro secções. Na primeira são avaliadas as condições para a criação de uma SAM num futuro próximo. Na segunda e na terceira secções são discutidas as principais desagregações a considerar, as articulações entre as diferentes contas e as fontes disponíveis, bem como algumas das suas principais limitações. Na última secção são dados alguns exemplos de ligações entre diferentes contas

PALAVRAS-CHAVE:

- *Contas Nacionais; Contabilidade Social; Inquérito aos Orçamentos Familiares; Padrões de Consumo das Famílias; LEG-SAM.*

^(*) Este texto resulta da participação do INE no LEG-SAM (Leader Group - Social Accounting Matrix), grupo de projecto liderado pelo CBS da Holanda (Central Bureau of Statistics) e apoiado pelo Eurostat. Neste grupo participam ainda os institutos de estatística do Reino Unido, Itália, Grécia, Bélgica, Finlândia, Noruega. Formalmente, o texto é principalmente uma "composição" de alguns materiais produzidos até ao momento, resultantes da colaboração interdepartamental entre o Gabinete de Estudos e Conjuntura (GEC) e o Departamento de Coordenação e Contas Nacionais (DCCN). O texto muito beneficiou das sugestões e comentários do Dr. Daniel Santos, bem como da reflexão feita pelo grupo interdepartamental, no qual cabe destacar a participação da Dra. Isabel Quintela do DCCN. Todas as imprecisões, erros e outros lapsos são, porém, da minha responsabilidade.

1. GENERAL REMARKS:

The Portuguese statistical system has not a set of reconciled data on labour supply and demand, expenditure and income of households, in the framework of a system of satellite accounts: Labour Accounts and Social-Economic Accounts. Only National Accounts (NA) are available, for the moment on ESA 79 basis. So, the statistical information system excludes interim integrated information systems, with the exception of NA. Had a global integrated system been available in a regular basis and a SAM would appear as the piece of information that would close the system. However, being the major objective to build a pilot SAM as soon as possible, this implies for the near future:

- A) A **top down approach**, in which the results of the National Accounts (NA) are taken as constraints (in general, this implies to take the values and structures of the NA as a *“fait accompli”*, adjusting the estimates to the NA aggregates).
- B) A **low level of assessment of the quality of the other sources**, namely those related to expenditure and income of households and supply of the different types of labour force. Note that this does not mean an uncritical view of existing data, but only the assumption of its use without any major changes.
- C) A **pilot SAM for the year 1995** on ESA 79 basis.
- D) A **“learning by doing”** process, in order to achieve the expertise in building a SAM on ESA 95 basis, as soon as data in this format is available from the national accounts information system.

In the medium term, one can imagine there will be some kind of network connecting data needs for compiling a SAM with the development of the Labour Accounts and the improvement of NA quality. In fact, this network should be one of the outputs of the pilot exercise. For instance, the needs for a SAM built in a regular basis could be reflected into Labour Accounts and National Accounts frameworks, adjusting the architecture of these accounts to the input needs of the SAM. The same applies to the questionnaires of Household Budget Survey (HBS) and Labour Force Survey (LFS) which could be improved too.

2. CLASSIFICATIONS, LINKS AND SOURCES (EXPANSION TOWARDS AN ORIENTED EMPLOYMENT/HOUSEHOLD SAM)

2.1. PRODUCTION:

The **level A6** for the activity classification would be a good choice, due to the weakness of reconciled data on labour force/wage income on one hand, and data on income/expenditure of households on the other. Nevertheless, a more detailed breakdown, such as A17, should be used in the working process of the pilot SAM (and this applies to other kind of classifications). This will allow assessing the dimension of discrepancies between data sources, improving the quality of the pilot. Besides, it approximates this exercise to the real situation of producing a SAM according to users needs. The source is 1995 data of National Accounts, on ESA 79 basis.

2.2. GENERATION OF INCOME (1):

It is possible to consider **six types of labour for each productive branch**, defined by gender and level of education (low, medium and high levels). **The objective** is the

distribution of total wage income generated in each branch, which was evaluated in the NA, to the six types of labour.

The **main sources** are the Labour Force Survey and Quadros de Pessoal (these are administrative data; enterprises employing at least one paid employee have the legal obligation to send their list of personnel to the Ministry of Employment). The former source would be used to estimate the total annual worked hours for each type of labour (eventually, the Earnings Survey can also be considered). The latter would allow estimating the average hour compensation per type of labour.

2.3. GENERATION OF INCOME (2):

The following **classification of Households** into sub-sectors is possible, according to:

1. The main source of income (wage income, self-employment income, income property, pension income, other sources).
2. The dimension of “wage-households” (one member, two without child, two or more with one or more children).
3. At least on a preliminary basis, the “self employment income households” can be divided by number of employees (one, more than one). In addition, and on the same basis, both the “self employment income households” and the “income property households” can be classified using the same categories of dimension as those of “wage-households”.
4. The “pension-households” related or not with age. At least on a preliminary basis, both can be classified by dimension, (one member, two or more members for the first kind of household; one member, two without child, two or more with one or more children, for the second one).
5. At least on a preliminary basis the “Other-households” can be classified by the same categories of dimension as those of “wage-households”.

The objective is to establish the link between these groups of households and the value of each kind of income they receive. Note that the wage income should be distributed to all types of household, and the same applies to the other kinds of income.

The main sources are the Labour Force Survey and the Quadros de Pessoal for wage-income and the Household Budget Survey for the allocation of other income. Concerning wage-income, the Labour Force Survey and the Quadros de Pessoal provide estimates that should be constrained by the global results of wage income of NA (see 2.2.Generation of income (1)). The allocation of wage income should be done according to the structure of wage income by types of households, estimated from the results of the Household Budget Survey. The final result should be consistent with branch generation of wage income (NA), labour factor distribution of wage income (LFS, QP) and household sub-sector distribution of wage income (HBS).

In addition, the HBS will be the main source for the distribution of gross operating surplus to the households.

2.4 ALLOCATION AND SECONDARY DISTRIBUTION OF INCOME

The **first objective** is the breakdown of in/outflows of income property to/from each types of household, obtaining as a balance the total primary income. **Secondly**, it is necessary

to estimate the outflow of social contributions, direct taxes and other income transfers and inflow of social benefits and other income transfers, in order to obtain the disposable income of each type of household.

Once again, the results of the Household Budget Survey could be used as the **main source**, restricted to National Accounts global results. Other sources, such as data from tax and social security files can be used as well.

The principal drawback of this approach, based on the Household Budget Survey, is the strong probability that property income (level and weight) is underestimated, owing to the reluctance of households to answer these questions.

Another drawback is the “asymmetry” of the sub-matrices. In fact, is the breakdown by recipient households and the breakdown by payer households is possible, but the breakdown of correspondent sectors cannot be done so easily. The subdivision of these ones will be no more detailed than General Government/Other sectors.

2.5. USE OF INCOME

The breakdown of final consumption by types of products and types of households (sub-sector) will be done taking the previous referred sources (NA and HBS). The consumption weight of each type of household in total consumption will be given by the HBS and should be taken as a constraint. The weights of each product on the total consumption, given by NA, should also be considered a constraint. The value of the total consumption (NA) will be distributed to the relevant matrix. An interactive process, based on an algorithm developed by Richard Stone will reconcile the level of each element of that matrix both with the figures of consumption of each product (NA) and the weight consumption of each type of household (HBS) on the total consumption (NA). See the example in 4.1.

3. SOME SHORTCOMINGS OF AVAILABLE DATA

The relevant sources show some fragility but they still have a sufficient degree of quality to allow its use in the exercise of building a pilot SAM, at least with the 1995 as a reference year.

Labour Force Survey: Regarding the year of 1995, the one that is considered as reference for the pilot SAM, available data is of acceptable quality. However, for 1996 and 1997, the results of preliminary data exploration do not advise the use of the Labour Force Survey without any changes of its raw data. This is a severe constraint for the building of a SAM referring to 1996 or 1997.

Household Budget Survey: There are two main drawbacks related to the use of this source. The first is the reluctance of households to answer to questions concerning their level of income, which implies additional problems to evaluate the weights of each source of income. This can produce some distortions in the breakdown by type of household. The other drawback is that national accounts underused this source to estimate the levels of consumption of each product. Again, this can produce significant discrepancies between the structures of consumption given by National Accounts estimates and by the results of Household Budget Survey (see the correspondent example of consumption breakdown).

Quadros de Pessoal: This is an administrative which is not representative of household small businesses (the very small enterprises), although it will be used to estimate the wage rates. As a consequence, it is probable that wage rates, especially in some activities, will be overestimated, implying overestimates of wage-incomes. In addition, there are some branches that are not well covered, namely those where public services have major weights, such as education and wealth care. This will distort some estimates resulting from the combination of this data source and the LFS (see the correspondent example of breakdown of compensation by branch, gender and level of education, in 4.2.).

Income tax and Social Security files: The unit household in the fiscal sources is not compatible with the institutional unit implicit in the NA and in the HBS. The fiscal source would only enable the breakdown of the households into three groups of income earners: Compensations of employees, Income from individual employers and own-account workers, Income from retirement pensions. All the other categories of income (such as property income and transfers) are excluded. Besides, all households with a level of income below a limit in which there is an exemption in taxation are also excluded. The fiscal evasion is also a difficult problem to deal with. As a consequence income tax files can be used only on a complementary way. Concerning the social security files, there is no way to establish the link between the individual files and the type of household.

National Accounts: National Accounts will be the fundamental source, restraining strongly the building of the pilot SAM. On the other hand, 1995 data of National Accounts, although produced in a well-defined framework, with a standard methodology and sources, has a data basis comparatively out of date. As a result, some deviation from the true economic structures may exist, although it may not be easily noticed (since the “true economic structures” are not observable at least on a coherent framework).

4. SOME EXAMPLES OF EXPANSION TOWARDS THE SAM

4.1. CONSUMPTION BREAKDOWN

The following exercise combines two sources, the NA estimates of consumption by products, classified at the level A6 of ESA 95 (the figures in the last columns of the table) and the Household Budget Survey (HBS). Take the next table: according to the main source of income, there are five types of households and the level of consumption of each one is broken down by product (there are six products, defined according to classification A6 of ESA 95). Summing the elements of each row, the consumption by product is obtained (this is the first column labelled with “Total”). Summing the elements of this column, the HBS estimates of total consumption is also obtained. Each element of the last but two rows represents the level of consumption by type of household (the row labelled “total”), according to the HBS data. The next row shows the correspondent structure of consumption.

HOUSEHOLD CONSUMPTION PATTERNS AND TOTAL CONSUMPTION (NON-RECONCILED DATA)

Unit: 10⁶ escudos

PRODUCTS	HBS-HOUSEHOLDS							Constraints	
	1	2	3	4	5	Total		NA total	
A	215081	64943	16994	95880	11902	404800	0.053	624319	0.064
B	1990397	609941	167101	659186	131297	3557922	0.467	5655241	0.580

C	14353	2104	726	13879	106	31168	0.004	12642	0.001
D	934568	261229	60508	190085	58977	1505367	0.198	1498510	0.154
E	663253	195646	137295	191554	42300	1230050	0.161	997579	0.102
F	525911	122325	55318	149748	38592	891893	0.117	956110	0.098
Total	4343564	1256187	437942	1300333	283174	7621200	1.000	9744401	1.000
Constraints HBS	0.570	0.165	0.057	0.171	0.037	1.000			
Constraints HBS/AN	5553643	1606150	559949	1662595	362064	9744401			

It can be seen that the HBS underestimate the NA total consumption, although some components of the former are higher than the latter. In other words, the structures are significantly different. Nevertheless, it is possible to reconcile these two sets of data. The underlying hypotheses are, of course, that both the NA levels of consumption and the HBS structure are sufficiently good estimates of the true values.

Firstly, note that the last row results from the breakdown of NA total consumption by type of household, according to their weight in the total consumption. It means that, by construction, each element of this row vector reconciles the NA total consumption and the HBS structure of consumption by household. Next, using the RAS procedure, the matrix is filled with new estimates of the breakdown of consumption by products and type of households:

HOUSEHOLD CONSUMPTION PATTERNS AND TOTAL CONSUMPTION (RECONCILED DATA)

Unit: 10⁶ escudos

PRODUCTS	HBS-HOUSEHOLDS							Constraints	
	1	2	3	4	5	Total		N Accounts	
A	333786	99530	28150	144420	18433	624319	0.064	624319	0.064
B	3177975	961741	284780	1021536	209209	5655241	0.580	5655241	0.580
C	5895	853	318	5532	44	12642	0.001	12642	0.001
D	933341	257638	64500	184251	58780	1498510	0.154	1498510	0.154
E	537425	156556	118743	150649	34206	997579	0.102	997579	0.102
F	565221	129832	63458	156207	41392	956110	0.098	956110	0.098

HBS/AN	4343564	1256187	437942	1300333	283174	9744401	1.000	9744401	1.000
HBS	0.570	0.165	0.057	0.171	0.037	1.000			

4.2 BREAKDOWN OF COMPENSATION BY BRANCH, GENDER AND EDUCATION LEVEL

The objective is to distribute the compensation by branch (roughly at level A17) by gender and level of education (four levels were considered: G1-level 1 of ISCED-97, G2-level 2 of ISCED-97, G3-level 3 of ISCED-97 and G4-level 5 of ISCED). The procedure has two stages: the aim of the first is to obtain preliminary estimates and the aim of the other is to reconcile them with the NA figures.

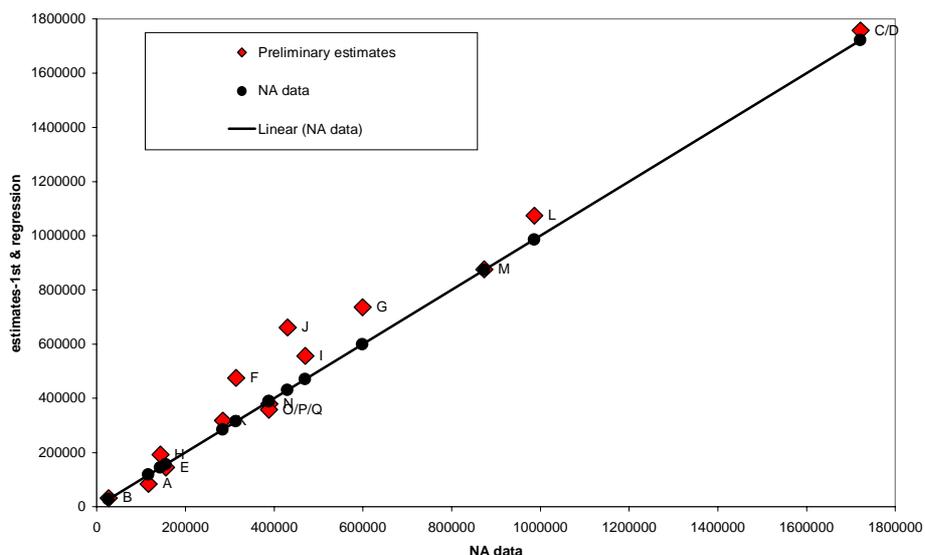
Stage one: preliminary estimates

Data was taken from Quadros de Pessoal, namely figures of earnings by gender, level of education and branch and organised in a matrix [C*] with a (8,14) dimension. This matrix was adjusted in order to incorporate the employer's actual and imputed social contributions. Since this data is available in the NA only for the total of each branch, it was assumed that the implicit tax of actual social contribution was the same in each branch. This means that the tax was considered the same, independently of the kind and amount of labour force used in the branch. The imputed social contribution was assumed to be proportional to the amount of wage and salaries and uniform in each branch. The first assumption is very close to reality. The second assumption must be seen as a simplification derived from the availability of data but with no important effects considering the value of the imputed contributions. Consequently, the corrected earnings were defined according to $W_{ij}^g = W_{ij}^1 * (1 + t_{isc}) / (1 - t_{asc})$, where i refers to the type of labour and j refers to the branch. Each estimate was also adjusted to an annual scale, $W_{ij} = 48 * W_{ij}^g$.

The matrix of hours worked in each branch, by gender and education level, was taken from the same source (Quadros de Pessoal). This matrix, with the general element H_{ij} , was also scaled to an annual basis. Combining the two set of data, the average hourly earning, $w_{ij} = (W_{ij}/H_{ij})$, was estimated.

It was admitted that data of labour Force Survey should be used to build the relevant matrix of hours worked. This matrix was constructed on a four-term average of actual hours by the same breakdown, scaled to an annual basis. Actually three matrices were calculated, one for wage-employment, another for total employment and the last one for the rest of (Other) employment, each one with the general element H_{ij}^w , H_{ij}^t and H_{ij}^o , respectively. Preliminary estimates were calculated, by multiplying each w_{ij} by the correspondent H_{ij}^w . The matrix of compensation [C], with the general term C_{ij} , was generated, as a result. Knowing that each row refers to a kind of labour, and each column to a branch, summing the elements of a given column of the matrix [C] the compensation paid by the branch is obtained. Summing the elements of a given row the compensations of a given kind of labour is obtained. See chart 1 to evaluate the "quality" of the first estimates.

Compensations



Stage two: Reconciliation with NA data

The next step has the objective of reconciling the first estimates with data from National Accounts. An interactive process is developed, taking as constraint the NA figures of compensations by branch. Another (weak) constraint is given by the weights of each type of labour in the total compensation, obtained from the matrix [C*] of the first estimates. See the next two table, the first showing the final estimates of compensation breakdown, the second presenting the correspondent estimates of monthly compensation per worker. See also chart 2, for a preliminary view of the distribution of monthly compensation per worker.

COMPENSATIONS

UNIT: 10⁹ escudos

		A	B	C/D	E	F	G	H	I	J	K	L	M	N	O/P/Q	Total
M	G1	77.8	21.6	680.8	52.2	239.3	203.8	52.2	182.1	29.4	32.8	157.0	20.6	25.5	65.2	1840.5
	G2	0.7	1.9	202.0	32.4	26.2	95.2	16.2	73.5	82.0	28.0	77.1	10.2	10.9	24.4	680.8
	G3	1.8	0.6	136.5	23.3	18.9	65.0	6.6	43.4	101.8	36.6	64.6	17.6	8.5	24.5	549.7
	G4	5.5	0.7	162.7	23.1	17.4	40.9	1.7	70.0	108.9	82.8	406.1	249.1	105.2	43.4	1317.5
F	G1	27.2	1.4	354.9	3.7	1.4	72.9	43.8	12.9	11.1	18.5	45.2	51.9	65.5	157.3	867.6
	G2	1.1	0.3	76.4	5.9	3.9	53.0	12.0	31.0	32.5	22.2	50.5	29.7	34.3	26.7	379.4
	G3	2.4	0.2	60.8	5.8	3.7	39.6	6.4	26.8	37.4	29.7	47.3	28.4	21.2	23.7	333.4
	G4	0.3	0.0	47.4	9.4	3.7	28.7	4.7	30.4	27.3	33.4	138.2	466.2	117.7	23.0	930.6
M	M	85.9	24.9	1182.0	131.0	301.8	404.8	76.7	368.9	322.2	180.3	704.8	297.5	150.1	157.6	4388.5
	F	31.0	2.0	539.4	24.7	12.7	194.2	67.0	101.1	108.4	103.8	281.2	576.1	238.7	230.7	2511.0
TOTAL		116.9	26.8	1721.4	155.8	314.5	599.1	143.7	469.9	430.6	284.1	986.1	873.6	388.8	388.3	6899.5

MONTHLY COMPENSATIONS PER EMPLOYEE

UNIT: 10³ escudos

		A	B	C/D	E	F	G	H	I	J	K	L	M	N	O/P/Q	Total
M	G1	88.9	100.3	123.0	320.8	100.5	117.9	110.2	198.9	353.6	168.1	116.3	154.2	189.9	114.3	124.7
	G2	109.3	120.9	181.8	324.3	121.1	148.2	135.3	265.9	397.3	175.6	112.3	177.3	241.4	161.1	179.4
	G3	145.9	104.4	201.5	352.7	141.4	173.7	124.4	267.8	362.8	183.5	132.4	184.9	257.6	201.1	203.4
	G4	189.5	102.7	411.5	528.7	283.7	343.5	355.0	545.6	554.4	398.7	889.8	414.2	449.1	377.9	506.8
F	G1	61.5	41.2	78.7	249.1	75.9	82.7	76.6	157.6	216.0	86.1	91.5	105.3	144.9	61.7	80.3
	G2	80.0	85.7	113.9	293.7	94.1	100.3	89.0	224.3	328.6	128.5	84.6	122.3	168.7	124.0	123.2
	G3	93.5	52.3	129.3	269.3	98.5	106.4	103.1	197.3	281.7	131.0	90.2	141.9	180.8	128.0	132.5
	G4	135.6	0.0	234.3	433.6	137.9	199.1	161.5	307.3	348.9	210.0	324.9	224.6	249.7	230.1	242.6
M	M	92.9	101.8	153.1	351.8	108.1	141.3	117.8	248.9	420.2	236.6	236.4	335.1	336.0	164.3	184.0
	F	64.1	45.8	92.2	317.6	102.3	100.8	83.9	222.3	299.7	134.2	137.7	191.3	191.9	75.6	124.1
TOTAL		83.0	93.4	126.9	345.9	107.8	125.0	99.1	242.7	381.6	185.0	196.3	224.1	230.0	96.8	156.5

These are just some illustrative estimates, with no other objective than to show the procedures that can be developed to build a SAM for Portugal. Much work has to be done to gather, reconcile data and test the results. One of the major issues is the consistence of estimates based on different sets of data. For instance, even if the compensation per worker estimates has some expectable characteristics, such as a positive correlation between the compensation per worker and the level of education, as well as relative higher values for male, the gap between male and female compensation should be checked with other sources.



To briefly conclude, the building of a SAM is obviously facilitated when NA is taken as a constraint. The task would be more difficult, although more interesting, if the objective was the building of a system of integrated information, linking from the bottom data on labour, expenditure and output in a meaningful breakdown. In this framework much of intermediate reconciliation would have to be done and probably there would have to be also some changes in the procedures of raw data collection.

ANNEX: Branch/products classification and level of education

- A:** Agriculture, hunting, forestry
- B:** Fishing
- C/D:** Industry
- E:** Electricity, Gas and Water
- F:** Construction
- G:** Retail and Repair
- H:** Hotels and restaurants

- I:** Transport, Storage and Communication
- J:** Financial Activities
- K:** Real Estate, Renting and Business activities
- L:** Public Administration and Defence, Compulsory Social Security
- M:** Education
- N:** Health, Social Work
- O/P/Q:** Other Community, Social and Personal Services, Private Households With Employed Persons, International Organisms and Other Institutions

- G 1:** Level 1 of ISCED-97
- G2:** Level 2 of ISCED-97
- G3:** Level 3 of ISCED-97
- G4:** Level 5 of ISCED-97

ESTIMAÇÃO DE ESTADO EM MODELOS LINEARES POR BOCADOS EM TEMPO DISCRETO: COMPARAÇÃO NUMÉRICA DE DOIS PROCEDIMENTOS

DISCRETE-TIME PIECEWISE LINEAR FILTERING: NUMERICAL COMPARATION BETWEEN TWO PROCEDURES

Autora: Rita Cristina Pinto de Sousa

- Técnica Superior de Informática - Direcção Regional do Norte do Instituto Nacional de Estatística

RESUMO:

- O problema de estimação de estado em modelos lineares por bocados parcialmente observados em tempo discreto e com pequeno ruído de observação é tratado neste artigo. Para resolver um problema deste tipo há primeiro que seleccionar os intervalos de linearidade, nos quais podemos aproximar o filtro óptimo pelo correspondente filtro de Kalman-Bucy. Na separação de intervalos de linearidade, o problema consiste em decidir qual dos filtros deve ser aplicado. Dois tipos de testes são propostos para este fim: teste da razão de verosimilhança e teste de variação quadrática. Neste trabalho apresentamos e comparamos o desempenho destes testes com base em simulações numéricas.

PALAVRAS-CHAVE:

- *Estimação de estado, filtro de Kalman-Bucy, teste da razão de verosimilhança, teste de variação quadrática.*

ABSTRACT:

- This article regards a piecewise linear discrete time filtering problem with small observation noise. It is possible to separate the time intervals of linearity of the model and we can then approximate, over such an interval, the optimal filter by the corresponding Kalman-Bucy filter. When separating the intervals of linearity, the problem is to decide wich filter must be applied. Likelihood ratio and quadratic variation tests are proposed for this purpose. We present and compare the performance of both tests numerically, based on simulation results.

KEY-WORDS:

- *State estimation, Kalman-Bucy filter, likelihood ratio test, quadratic variation test.*

1. INTRODUÇÃO

Neste trabalho propomo-nos efectuar o estudo da estimação das variáveis de estado para um modelo de Markov escondido não linear em tempo discreto, em que as funções que o descrevem são lineares por bocados. Tomamos, portanto, como ponto de partida um modelo de Markov escondido (HMM - *Hidden Markov Model*), isto é, um processo estocástico, neste caso em tempo discreto, constituído por uma sequência de estados, não observáveis, que representa o processo de Markov e por uma outra sequência de variáveis aleatórias condicionalmente independentes, que representa o processo das observações. Nos modelos lineares gaussianos podemos recorrer ao filtro de Kalman-Bucy para estimar o estado em cada instante e respectiva covariância de erro de estimação. Nos modelos não lineares utiliza-se geralmente uma solução aproximada para o problema de estimação (Andrews-Grewal, 1993, Harvey, 1989 e Jazwinski, 1970, entre outros). De entre as aproximações possíveis o chamado Filtro de Kalman Estendido é uma das mais usadas pelas suas propriedades (Picard, 1991). Exemplos da aplicação do Filtro de Kalman Estendido em modelos não lineares podem ser encontrados por exemplo em Sorenson (1990). Existem diversos tipos de modelos não lineares, como é o caso dos modelos de limiar (“threshold”), aqui em estudo ¹. Trataremos assim, o caso em que as funções que descrevem o modelo são lineares por bocados. Neste caso, para obter as estimativas é necessário decidir sobre a região onde se situam as variáveis de estado. Sem perda de generalidade suporemos que estas regiões são \mathfrak{R}^- ou \mathfrak{R}^+ , que estão portanto separadas pelo limiar zero. Então, a estimação deverá ser precedida por dois testes, um para a detecção de passagens em zero e outro para a decisão sobre o sinal do estado, nos intervalos sem cruzamentos em zero. No que respeita ao primeiro teste, temos à escolha entre dois tipos de teste: um teste sobre as observações ou um teste sobre o estimador de Kalman-Bucy. Na decisão sobre o sinal do estado podemos também optar entre dois testes, o da razão de verosimilhança e o de variação quadrática. Quer os testes de passagens em zero quer os testes de decisão sobre o sinal foram estudados em Fleming *et al.* (1991), Oliveira-Roubaud (1991) e Oliveira (2000) para modelos discretos e em Fleming-Pardoux (1989), Fleming *et al.* (1988), Pardoux-Roubaud (1991) e Roubaud (1992 e 1995) para os análogos modelos contínuos. Em qualquer dos casos, estes métodos só podem ser aplicados se se verificar uma hipótese, dita hipótese de detectabilidade, genericamente designada por (HD). Existem dois tipos diferentes de hipóteses de detectabilidade, (HD1) e (HD2), conforme veremos na secção seguinte.

O principal objectivo deste trabalho consiste em analisar e comparar o comportamento destes dois testes de decisão sobre o sinal, baseados nas observações, (razão de verosimilhança e variação quadrática) na resolução do problema de estimação, face à hipótese de detectabilidade (HD2). Os resultados numéricos apresentados baseiam-se na simulação de alguns modelos e permitem fazer uma análise comparativa do desempenho dos testes em estudo. Para um estudo idêntico sob a hipótese (HD1) consultar Fleming *et al.* (1991).

Este trabalho divide-se em 7 secções que se distribuem da seguinte forma. Após esta breve introdução formulamos o problema em estudo na secção 2. Os filtros de Kalman-Bucy em paralelo, usados na estimação, são apresentados na secção 3. As secções 4 e 5 contêm o teste de detecção de passagens em zero e os testes de decisão sobre o sinal, respectivamente. Nestes últimos, podemos separar o teste da razão de verosimilhança (secção 5.1) e o teste de variação quadrática (secção 5.2). Na secção 6 são apresentados e analisados os resultados numéricos obtidos em alguns modelos, em que o método exposto foi aplicado. Por fim são apresentadas conclusões sobre o desempenho dos testes em estudo (secção 7).

¹ Veja-se, em Tong-Yeung (1989), um exemplo de aplicação destes modelos na área financeira.

2. FORMULAÇÃO DO PROBLEMA

Consideremos o seguinte modelo contínuo não linear:

$$\begin{cases} d\tilde{X}_t = b(\tilde{X}_t)dt + \sigma(\tilde{X}_t)dU_t, & 0 \leq t \leq T \\ d\tilde{Y}_t = h(\tilde{X}_t)dt + \varepsilon dV_t, \end{cases} \quad (2.1)$$

no intervalo $[0, T]$, em que os processos $\{U_k, k \geq 0\}$ e $\{V_k, k \geq 0\}$ são processos de Wiener standard independentes, com valores em \mathfrak{R} e Δt é um parâmetro positivo próximo de zero. Considerando uma discretização, no intervalo de tempo contínuo $[0, T]$, com $\Delta t = \varepsilon$, obtém-se o seguinte modelo não linear em tempo discreto:

$$\begin{cases} X_{k+1} = X_k + \varepsilon b(X_k) + \sqrt{\varepsilon} \sigma(X_k) u_k, & k = 0, 1, 2, \dots, K \\ Y_k = h(X_k) + \sqrt{\varepsilon} v_k, \end{cases} \quad (2.2)$$

em que $\{X_k, k \geq 0\}$ e $\{Y_k, k \geq 0\}$ são processos estocásticos que tomam valores em \mathfrak{R} . Suponhamos que estes correspondem aos processos de estado e de observação, respectivamente. As sequências $\{u_k\}_k$ e $\{v_k\}_k$ são processos estocásticos que pretendem modelar perturbações, ruídos ou incertezas. No nosso caso, são considerados como ruídos brancos gaussianos normalizados e independentes de X_0 . Os processos $\{X_k\}_k$ e $\{Y_k\}_k$ são indexados da maneira habitual pelos $K+1$ instantes, com $K = \text{int}(T/\varepsilon)$, resultantes da discretização no intervalo contínuo $[0, T]$.

No problema que pretendemos tratar suporemos que as funções b, σ e h são funções reais de variável real lineares por bocados :

- $b(x) = b_- x|_{\{x < 0\}} + b_+ x|_{\{x \geq 0\}}$
- $\sigma(x) = \sigma_- |_{\{x < 0\}} + \sigma_+ |_{\{x \geq 0\}}$
- $h(x) = h_- x|_{\{x < 0\}} + h_+ x|_{\{x \geq 0\}}$

O processo estocástico $\{(X_k, Y_k)\}_k$, modelado pelo sistema (2.2), pode ser visto como um modelo de Markov escondido, em que $\{X_k\}_k$ é um processo de Markov em espaço de estados infinito e o processo $\{Y_k\}_k$ corresponde a uma sequência de variáveis aleatórias, dependentes de $\{X_k\}_k$. A designação de modelo de Markov escondido é motivada pelo facto do processo de estado $\{X_k\}_k$ não ser observado, mas existir um processo directamente correlacionado com o mesmo, $\{Y_k\}_k$, que representa as observações do sistema em estudo.

O problema consiste em estimar o processo de estado não observado $\{X_k\}_k$ em cada instante k , dadas as observações até esse mesmo instante. No caso em que o modelo é um sistema linear gaussiano, podemos recorrer ao filtro de Kalman-Bucy que é um estimador óptimo, no sentido do erro quadrático médio (Jazwinski, 1970). Este estimador permite-nos calcular recursivamente as estimativas e respectiva covariância de erro:

$$\hat{X}_{k/k} = E[X_k / \gamma_0^k]$$

$$e \quad Q_{k/k} = E[(X_k - \hat{X}_{k/k})(X_k - \hat{X}_{k/k})^T / \gamma_0^k],$$

assim definidas para os modelos lineares gaussianos multidimensionais, onde γ_0^k é a σ -álgebra das observações até ao instante k . No caso linear gaussiano a lei condicional também é gaussiana e a sua média e matriz de covariância são suficientes para caracterizar essa distribuição, como é bem conhecido, constituindo o filtro proposto por Kalman e Bucy nos anos 60 (Kalman-Bucy, 1961). A necessidade motivada pelas aplicações práticas levou a que a teoria da filtragem linear fosse estendida aos sistemas não lineares dando origem ao filtro de Kalman Estendido (Andrews-Grewal, 1993 e Jazwinski, 1970 entre muitos outros). Embora este filtro, seja em geral eficiente em muitas situações práticas, a sua eficiência não é garantida, dependendo das condições do modelo. Picard (1986) demonstrou que a eficiência se verifica se a função h for injectiva, para além de outras hipóteses de regularidade bastante gerais sobre as funções que intervêm no modelo. Como estamos a lidar com uma função de observação unidimensional linear por bocados, h não será necessariamente injectiva. Quando h não é injectiva (não monótona), ou seja, no nosso caso se $h_- h_+ < 0$, as estimativas para as variáveis de estado não podem ser directamente obtidas pela estimação de $h(X_k)$. Nesta situação, o Filtro de Kalman Estendido não é em geral assintoticamente eficiente, sendo necessária a determinação da região onde $\{X_k\}_k$ se situa ou, se quisermos, do sinal de $\{X_k\}_{k \geq 0}$ em cada instante. Neste caso, dado que as funções no modelo são lineares por bocados, a ideia é a de considerar um estimador para cada um dos sistemas correspondentes aos (dois) ramos de linearidade das funções. Assim, mantêm-se dois filtros em paralelo, optando-se em cada instante pela aplicação de um ou de outro, consoante se toma a decisão de considerar que o processo evolui numa das regiões \mathfrak{R} ou \mathfrak{R}^* . É precisamente neste caso que nos situaremos neste trabalho, isto é, admitimos a seguinte hipótese:

$$h_- h_+ < 0 \tag{H1}$$

Naturalmente a questão da determinação do sinal do estado passa primeiro pela garantia de se ter uma variância condicional reduzida. Para isso, foram introduzidas duas hipóteses de detectabilidade (Fleming *et al.*, 1991 e Oliveira-Roubaud, 1995). São elas :

$$h_-^2 \sigma_-^2 \neq h_+^2 \sigma_+^2 \tag{HD1}$$

$$h_-^2 \sigma_-^2 = h_+^2 \sigma_+^2 \text{ e } b_+ \neq b_- \tag{HD2}$$

Sob qualquer destas hipóteses é possível, à custa de dois estimadores de Kalman-Bucy em paralelo, obter um estimador convergente, sub-óptimo na classe de estimadores lineares. Para além disso, qualquer das hipóteses apresentadas permite obter intervalos de tempo

suficientemente longos, em que o processo $\{X_k\}_k$ toma valores distantes de zero. Nestes intervalos podemos aplicar os filtros de Kalman-Bucy, ora para um ramo ora para outro, de acordo com o sinal das variáveis de estado nesses intervalos.

No caso da hipótese (HD1), o sinal de X_k é essencialmente determinado atendendo à variação quadrática da função $h(X_k)$. No caso da hipótese (HD2), é a função $b(X_k)$ (observada indirectamente através de Y_k) que contém a informação essencial para a determinação do sinal de X_k , dificultando o processo de estimação. É sob esta hipótese que se situarão os estudos e resultados apresentados ao longo deste artigo.

A obtenção das estimativas para as variáveis de estado do modelo (2.2) passa por duas etapas, que se traduzem em dois testes estatísticos:

1ª etapa: Detectar os intervalos de tempo em que a trajectória de $\{X_k\}_k$ não cruza o nível zero.

2ª etapa: Decidir, para cada um desses intervalos, se se tem $X_k > 0$ ou $X_k < 0$.

Na primeira etapa, utiliza-se um teste de detecção de passagens em zero e na segunda etapa, utiliza-se um teste de decisão sobre o sinal do estado. Na decisão sobre o sinal consideraremos dois tipos de testes: testes de razão de verosimilhança e testes de variação quadrática.

No que diz respeito ao teste de detecção de passagens em zero, estudos anteriores mostram que é possível estender, sem dificuldade, o teste proposto por Fleming et al (1991) sob a hipótese (HD1) ao problema sob a hipótese (HD2) (ver Oliveira-Roubaud, 1995). No que diz respeito ao teste de decisão sobre o sinal Oliveira-Roubaud (1991) propõem, sob a hipótese (HD2), dois testes da razão de verosimilhança e Oliveira (2000) propõe um teste de variação quadrática. O principal objectivo deste estudo consiste em analisar o comportamento do teste de variação quadrática e comparar o seu desempenho com o do teste da razão de verosimilhança, ambos os testes baseados nas observações, sob a hipótese de detectabilidade (HD2).

3. DOIS ESTIMADORES DE KALMAN-BUCY EM PARALELO

A ideia subjacente à utilização do estimador de Kalman Estendido em sistemas não lineares consiste em considerar que, nas vizinhanças de uma estimativa, o sistema não linear se comporta localmente de forma aproximada à do sistema linearizado, de tal forma que, ao aplicar o filtro de Kalman-Bucy correspondente a esse sistema localmente linearizado produzimos estimativas próximas das óptimas no sentido do erro quadrático médio. No nosso caso temos um sistema unidimensional linear por bocados portanto, em cada um desses intervalos de linearidade, a ideia é aplicar o filtro correspondente ao ramo em causa consoante se decide que o processo de estado está num ramo ou no outro. Para tomar estas decisões, que antecedem a aplicação dos filtros, são usados os testes de decisão sobre o sinal, que iremos tratar na secção 5. Para detectar os intervalos de linearidade é usado o teste apresentado na secção 4.

Consideremos dois modelos lineares correspondentes aos ramos em que se separam as funções intervenientes no modelo (2.2). Referenciemo-los como modelo (+) e modelo (-):

$$(+) \quad \begin{cases} X_{k+1} = (1 + \varepsilon b_+) X_k + \sigma_+ \sqrt{\varepsilon} u_k \\ Y_k = h_+ X_k + \sqrt{\varepsilon} v_k \end{cases}$$

$$(-) \quad \begin{cases} X_{k+1} = (1 + \varepsilon b_-) X_k + \sigma_- \sqrt{\varepsilon} u_k \\ Y_k = h_- X_k + \sqrt{\varepsilon} v_k \end{cases}$$

Podemos dizer que estes modelos resultam da permanência do processo de estado em \mathfrak{R}^+ e \mathfrak{R}^- , respectivamente. A separação do modelo original nestes dois modelos dá origem ao aparecimento de dois filtros que terão de ser calculados em simultâneo, um para cada modelo. Daí a designação de filtros de Kalman-Bucy em paralelo.

Sejam (\hat{X}_k^+, Q^+) e (\hat{X}_k^-, Q^-) os filtros associados aos sistemas lineares (+) e (-) :

$$\begin{cases} \hat{X}_{k/k}^+ = \hat{X}_{k/k-1}^+ + K_k^+ (Y_k - h_+ \hat{X}_{k/k-1}^+) \\ Q_{k/k}^+ = Q_{k/k-1}^+ - Q_{k/k-1}^+ h_+ (h_+ Q_{k/k-1}^+ h_+ + \varepsilon)^{-1} h_+ Q_{k/k-1}^+ \end{cases}, \quad (\text{FKB})^+$$

com $\hat{X}_{k/k-1}^+ = (1 + \varepsilon b_+) \hat{X}_{k-1/k-1}^+$, $K_k^+ = Q_{k/k-1}^+ h_+ (h_+ Q_{k/k-1}^+ h_+ + \varepsilon)^{-1}$ e

$$Q_{k/k-1}^+ = (1 + \varepsilon b_+)^2 Q_{k-1/k-1}^+ + \sigma_+^2 \varepsilon$$

e

$$\begin{cases} \hat{X}_{k/k}^- = \hat{X}_{k/k-1}^- + K_k^- (Y_k - h_- \hat{X}_{k/k-1}^-) \\ Q_{k/k}^- = Q_{k/k-1}^- - Q_{k/k-1}^- h_- (h_- Q_{k/k-1}^- h_- + \varepsilon)^{-1} h_- Q_{k/k-1}^- \end{cases}, \quad (\text{FKB})^-$$

com $\hat{X}_{k/k-1}^- = (1 + \varepsilon b_-) \hat{X}_{k-1/k-1}^-$, $K_k^- = Q_{k/k-1}^- h_- (h_- Q_{k/k-1}^- h_- + \varepsilon)^{-1}$ e

$$Q_{k/k-1}^- = (1 + \varepsilon b_-)^2 Q_{k-1/k-1}^- + \sigma_-^2 \varepsilon$$

As condições iniciais gaussianas que se consideram, habitualmente, são as seguintes :

- $\hat{X}_0^+ = \hat{X}_0^- = E(X_0)$
- $Q_0^+ = Q^+$ e $Q_0^- = Q^-$, onde Q^+ e Q^- são as variâncias estacionárias das leis condicionais.

As condições iniciais são importantes apenas no sentido em que influenciam o comportamento inicial do modelo, o que poderá não ser acompanhado pelo estimador, mas o efeito de dependência da condição inicial tende a desaparecer rapidamente no tempo (Oliveira-Roubaud, 1991 e Picard, 1991).

Podemos resumir o processo de estimação em três passos essenciais: a selecção dos intervalos de tempo em que um dos filtros conduz a uma boa estimativa (teste de detecção de passagens em zero, secção 4), a decisão de qual dos dois filtros se deve aplicar (teste de decisão sobre o sinal do processo não observado, secção 5) e por último, a estimação propriamente dita usando o estimador \hat{X}_k^+ ou \hat{X}_k^- , consoante o resultado da etapa anterior.

4. TESTE DE DETECÇÃO DE PASSAGENS EM ZERO

Para seleccionar os intervalos de tempo, ditos de linearidade, em que, para uma dada probabilidade próxima de um, o processo de estado não cruza o zero recorremos ao teste de detecção de passagens em zero. Este teste pode ser aplicado às observações ou ao próprio estimador de $\{X_k\}_k$, $\{\hat{X}_k\}_k$, tendo-se neste caso optado por um teste baseado nas observações. No caso em que o teste de detecção é baseado no próprio estimador, o procedimento é basicamente o mesmo, excepto naturalmente no que diz respeito à estatística do teste. Para mais detalhes poderá ser consultado o trabalho desenvolvido por Fleming *et al.* (1991) e Oliveira-Roubaud (1991).

Começemos por fixar uma trajectória do processo $\{X_k\}_k$ para um modelo de Markov (2.2) que obedece à hipótese de detectabilidade (HD2). No intervalo de tempo contínuo $[0, T]$, consideremos um intervalo $[a, b]$ em que essa trajectória não cruza o zero. A não existência de cruzamentos em zero corresponde à ocorrência de um dos seguintes acontecimentos:

$$A_- = \{ X_k < 0; k = i_0, i_0 + 1, \dots, i_0 + m \},$$

ou

$$A_+ = \{ X_k > 0; k = i_0, i_0 + 1, \dots, i_0 + m \},$$

com $i_0 = \lceil a/\varepsilon \rceil$ e $m = \lfloor (b-a)/\varepsilon \rfloor$. Num abuso de linguagem, por uma questão de simplificação do texto, designaremos um intervalo de tempo correspondente aos índices i_0 a $i_0 + m$ por intervalo de tempo discreto $[i_0, i_0 + m]$. Detectar um destes intervalos corresponde a testar se algum dos dois acontecimentos ocorre ou não e para isso, delineamos um teste. O raciocínio que reside a este teste é o seguinte.

Se analisarmos o modelo em tempo discreto (2.2) constatamos que $h(X) = 0$ sse $X = 0$ e que $h(X_k)$ “pequeno” implica, por continuidade de h , que X_k esteja próximo de zero. No entanto $h(X_k)$ não é observável, mas para ε “pequeno”, aproxima-se de Y_k . Assim, Y_k é um candidato à partida como estatística do teste na detecção de passagens em zero para o processo $\{X_k\}_k$. Uma dificuldade na aplicação deste teste em relação a um teste de hipóteses clássico, diz respeito ao facto de ser aplicado numa situação de observação parcial, baseando-

se num processo observado $(\{Y_k\})$ quando se pretende decidir sobre um processo não observado $(\{X_k\})$.

A nossa preocupação, agora, na construção do teste será precisamente a determinação de uma fórmula para a constante $c_{obs} > 0$, designada por alguns autores como “*cutoff value*”, que corresponde ao valor crítico do teste.

Este teste, baseia-se portanto na análise da ocorrência do seguinte acontecimento:

$$C = \{ |Y_k| \geq c_{obs} ; k = i_0, \dots, i_0 + m \}$$

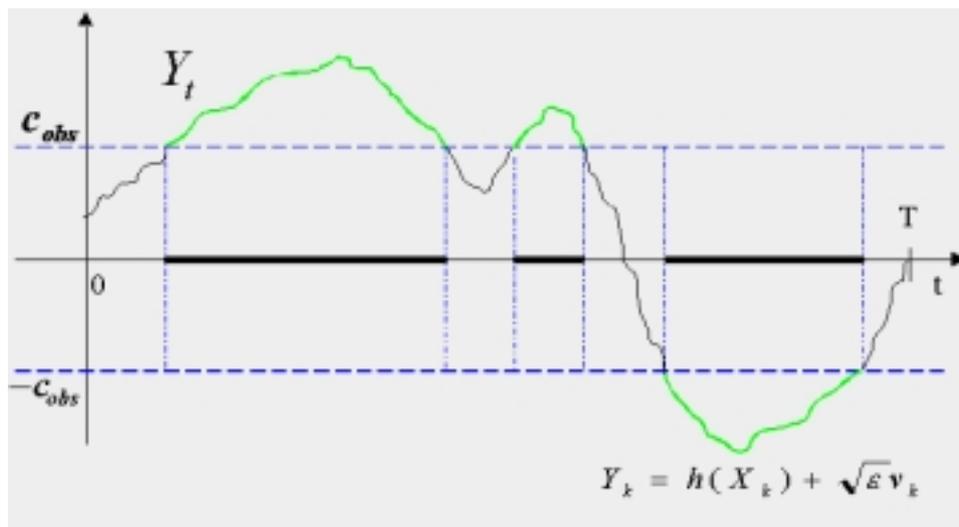


Figura 4.1 – Exemplo de aplicação do teste de detecção de passagens em zero.

Na figura 4.1 podemos observar uma trajetória do processo das observações, que obedece ao modelo (2.2), à qual se aplicou o teste. Os intervalos de tempo assinalados a negro correspondem aos intervalos em que ocorre o acontecimento C, ou seja, aos intervalos detectados como intervalos de linearidade. Se pretendessemos estimar o estado, para este exemplo, o procedimento a seguir seria a aplicação do teste de decisão sobre o sinal aos intervalos seleccionados e posteriormente a aplicação do respectivo filtro.

Na obtenção de uma fórmula para a constante c_{obs} pretendemos minimizar a probabilidade de erro de que um cruzamento não seja detectado, a qual deverá corresponder ao valor máximo para a probabilidade condicional

$$P \left\{ \{ X_k \cdot X_{k+1} < 0 \} \mid |Y_k| \geq c_{obs} , |Y_{k+1}| \geq c_{obs} \right\},$$

nos instantes $k = i_0, \dots, i_0 + m$. Minimizando a probabilidade de erro e fixando uma determinada probabilidade de risco α_d ², obtém-se (ver Sousa, 1999):

² Nas aplicações tomaremos $\alpha_d = 5\%$.

$$c_{obs} = \max(c_1, c_2), \quad (4.3)$$

em que

$$c_1 = \lambda \sqrt{\varepsilon} \frac{\sqrt{h_+^2 h_-^2 \sigma_-^2 + h_-^2 + h_+^2 (1 + \varepsilon b_-)^2}}{|h_- - h_+ (1 + \varepsilon b_-)|}$$

$$c_2 = \lambda \sqrt{\varepsilon} \frac{\sqrt{h_+^2 h_-^2 \sigma_+^2 + h_+^2 + h_-^2 (1 + \varepsilon b_+)^2}}{|h_+ - h_- (1 + \varepsilon b_+)|}$$

O parâmetro $\lambda > 0$ é dado por $\phi(\lambda) = 1 - \alpha_d$, com ϕ para representar a função de distribuição da lei normal reduzida.

Os intervalos detectados por este teste determinam os instantes para uma possível estimação por um dos filtros de Kalman-Bucy, sendo para isso necessário decidir agora sobre o sinal das variáveis de estado em cada um desses intervalos de linearidade, procedimento este que abordaremos a seguir.

5. TESTES DE DECISÃO SOBRE O SINAL DO PROCESSO NÃO OBSERVADO

Após a selecção dos intervalos em que o processo de estado, para uma dada probabilidade, não apresenta passagens em zero, é necessário determinar o sinal que este tomou ou, mais concretamente, é necessário decidir sobre qual dos filtros (\hat{X}_k^+ ou \hat{X}_k^- introduzidos na secção 3) deve ser usado em cada um desses mesmos intervalos. Tal como já adiantámos, na decisão sobre o sinal, podemos considerar dois tipos de testes, o da razão de verosimilhança e o de variação quadrática. Neste trabalho pretende-se analisar o teste de variação quadrática e compará-lo com o teste da razão de verosimilhança, para o caso da hipótese de detectabilidade (HD2), que relembramos a seguir.

$$(HD2): \quad h_-^2 \sigma_-^2 = h_+^2 \sigma_+^2 \quad \text{e} \quad b_+ \neq b_-$$

Os dois tipos de testes permitem decidir sobre o sinal do processo de estado optando entre duas alternativas, que correspondem à ocorrência do acontecimento A_- ou à ocorrência do acontecimento A_+ , onde

$$A_- = \{ X_k < 0; k = i_0, i_0 + 1, \dots, i_0 + m \},$$

$$A_+ = \{ X_k > 0; k = i_0, i_0 + 1, \dots, i_0 + m \}.$$

Os intervalos de tempo discreto $[i_0, i_0 + m]$ considerados provêm do primeiro teste, o teste de detecção, apresentado na secção 4.

5.1. TESTE DA RAZÃO DE VEROSIMILHANÇA

Consideremos o processo das observações correspondentes a um intervalo sem passagens em zero, $\{Y_k, i_0 + 1 \leq k \leq i_0 + m\}$. O índice i_0 corresponde ao primeiro instante a partir do qual se iniciará a contabilização da estatística.

Pretendemos estabelecer uma estatística de teste do tipo razão de verosimilhança para decidir sobre o sinal de $\{X_k\}_{k=i_0}$ com base nas observações. De acordo com o modelo (2.2) os valores observados para o processo $\{Y_k\}_k$ seguem a equação

$$Y_k = h(X_k) + \sqrt{\varepsilon} v_k,$$

que pode ser reescrita sob as hipóteses A_- e A_+ da seguinte maneira:

$$\begin{aligned} \text{Sob } A_+ : \quad Y_k &= (1 + \varepsilon b_+) X_k + \sqrt{\varepsilon} \dot{\Gamma} u_k^+ + \sqrt{\varepsilon} h_+ v_k^+, \\ \text{Sob } A_- : \quad Y_k &= (1 + \varepsilon b_-) X_k + \sqrt{\varepsilon} \dot{\Gamma} u_k^- + \sqrt{\varepsilon} h_- v_k^-, \end{aligned}$$

com $\dot{\Gamma}^2 = (h_- \sigma_-)^2 = (h_+ \sigma_+)^2$.

Seja L_n a estatística do teste, que se define à custa do logaritmo da razão de verosimilhança das medidas de probabilidade para os acontecimentos A_- e A_+ (Sousa, 1999). Para simplificar o cálculo da estatística, Oliveira-Roubaud (1995) partiram de um modelo mais simples, com passo de tempo Δt e sem ruído de observação, e deduziram-na com base numa versão discreta do Teorema de Girsanov (ver Oliveira, 1994):

$$L_n = \frac{1}{\dot{\Gamma}^2} \left[(b_+ - b_-) \sum_{k=i_0+1}^{i_0+n} Y_k (Y_{k+1} - Y_k) - \frac{1}{2} \Delta t (b_+^2 - b_-^2) \sum_{k=i_0+1}^{i_0+n} Y_k^2 \right].$$

Para o modelo (2.2), em estudo, usamos uma aproximação L_n^ε desta estatística com $\Delta t = \varepsilon$:

$$L_n^\varepsilon = \frac{b_+ - b_-}{2\dot{\Gamma}^2} (Y_{i_0+n+1}^2 - Y_{i_0+1}^2) - \frac{1}{2} n \varepsilon (b_+ - b_-) - \varepsilon \frac{b_+^2 - b_-^2}{2\dot{\Gamma}^2} \sum_{k=i_0+1}^{i_0+n} Y_k^2. \quad (5.4)$$

O afastamento entre as estatísticas L_n e L_n^ε é da ordem de grandeza ³ de $\sqrt{\varepsilon}$ (Oliveira-Roubaud, 1995). É esta estatística L_n^ε que tomaremos como estatística de teste. A regra de decisão do teste é a seguinte :

- Se $L_{N^*} \geq l_2$, decide-se pelo acontecimento A_+ ;
- Se $L_{N^*} \leq -l_1$, decide-se pelo acontecimento A_- ;
- Senão, não se decide.

Os valores $l_1 > 0$ e $l_2 > 0$ são duas constantes que forneceremos de seguida e N^* é um tempo de paragem definido por:

$$N^* = \inf \{n : L_n \geq l_2 \text{ ou } L_n \leq -l_1\} \wedge m$$

correspondente ao número mínimo de instantes necessário para decidir sobre o sinal do estado num dado intervalo detectado.

O passo seguinte será então a determinação dos valores dos parâmetros l_1 e l_2 que nos permitem aplicar a regra de decisão. Consideremos as probabilidades de erro do tipo I, p_- e p_+ , respeitantes à decisão sobre as hipóteses alternativas A_- e A_+ :

$$\begin{cases} p_+ = P(\text{'rejeitar } A_+ \text{'} | A_+) \\ p_- = P(\text{'rejeitar } A_- \text{'} | A_-) \end{cases} \quad (5.5)$$

Fixando as probabilidades de erro p_+ e p_- , os valores de l_1 e l_2 resultam da resolução do sistema (não linear) seguinte (ver Oliveira-Roubaud, 1995):

$$\begin{cases} p_+ = \frac{1 - e^{-l_2}}{e^{l_1} - e^{-l_2}} \\ p_- = \frac{1 - e^{-l_1}}{e^{l_2} - e^{-l_1}} \end{cases} \quad (5.6)$$

Como podemos constatar a solução deste sistema é independente dos parâmetros do modelo, isto é, os valores de l_1 e l_2 dependem apenas dos valores de erro admitidos para o teste.

O sistema de equações (5.6) por si só pode ter várias soluções e terá de ser resolvido numericamente. Do ponto de vista prático, para $p_+ = p_- = \alpha_s$, a dificuldade de obtenção de uma solução adequada do sistema (5.6) pode ser ultrapassada procurando uma solução em que $l_1 = l_2$. Nessa altura facilmente se verifica que:

³ **Notação 1** – Para um dado processo $\{G_k\}$ considera-se que $G_k = O(\varepsilon^q)$, onde $q \in \mathfrak{R}_+$, para significar que existem $c_1, c_2, c_3 > 0$ tais que

$$E[G_k^2] \leq c_1 \exp\{-c_2 k \varepsilon\} + c_3 \varepsilon^{2q}, \forall k \in N, \varepsilon > 0.$$

⁴ Em geral considera-se, na prática, $p_+ = p_- = 0.05$.

$$l_1 = l_2 = \ln \frac{1 - \alpha_s}{\alpha_s}$$

Esta solução foi adoptada em Oliveira-Roubaud (1995) com sucesso e será também adoptada neste estudo. Por exemplo, para $\alpha_s = 0.05$ obtém-se $l_1 = l_2 = 2.94$.

Veremos também que, do ponto de vista prático, é recomendável considerar como primeiro instante de cálculo da estatística de teste não o índice i_0 mas sim um índice ligeiramente posterior⁵. Esta pequena alteração em nada altera o espírito do teste. O seu efeito é apenas o de evitar pequenas instabilidades que possam ocorrer quando o sistema ainda esteja a “entrar num intervalo de linearidade”. Esta observação também é válida para o teste de variação quadrática, cujo procedimento analisamos na secção seguinte.

5.2. TESTE DE VARIAÇÃO QUADRÁTICA

Com o intuito de introduzir a estatística para este teste começemos por considerar as seguintes notações:

$$\begin{aligned}\Delta_k^+ &= Y_{k+1} - (1 + b_+ \varepsilon) Y_k, \\ \Delta_k^- &= Y_{k+1} - (1 + b_- \varepsilon) Y_k,\end{aligned}$$

e vejamos qual a ideia que preside ao teste de variação quadrática.

No caso de se ter $X_k > 0$ e $X_{k+1} > 0$ (acontecimento A_+), estes processos são dados pelas seguintes expressões :

$$\begin{aligned}\Delta_k^+ &= h_+ \sigma_+ \sqrt{\varepsilon} u_k + \sqrt{\varepsilon} v_{k+1} - (1 + b_+ \varepsilon) \sqrt{\varepsilon} v_k, \\ \Delta_k^- &= h_+ (b_+ - b_-) \varepsilon X_k + h_+ \sigma_+ \sqrt{\varepsilon} u_k + \sqrt{\varepsilon} v_{k+1} - (1 + b_- \varepsilon) \sqrt{\varepsilon} v_k,\end{aligned}$$

expressões essas que decorrem do próprio modelo (2.2). No caso de ocorrer $X_k < 0$ e $X_{k+1} < 0$ (acontecimento A_-), obtém-se:

$$\begin{aligned}\Delta_k^+ &= h_- (b_- - b_+) \varepsilon X_k + h_- \sigma_- \sqrt{\varepsilon} u_k + \sqrt{\varepsilon} v_{k+1} - (1 + b_+ \varepsilon) \sqrt{\varepsilon} v_k, \\ \Delta_k^- &= h_- \sigma_- \sqrt{\varepsilon} u_k + \sqrt{\varepsilon} v_{k+1} - (1 + b_- \varepsilon) \sqrt{\varepsilon} v_k.\end{aligned}$$

Analisando estas expressões facilmente concluímos que, se o acontecimento A_+ ocorre, a variável aleatória Δ_k^+ tem distribuição $N(0, \Gamma_+^2 \varepsilon)$ e se o acontecimento A_- ocorre, a variável aleatória Δ_k^- tem distribuição $N(0, \Gamma_-^2 \varepsilon)$, sendo

⁵ Nos exemplos de aplicação que apresentamos no capítulo 6 consideramos, de um modo geral, i_0 incrementado de 6 unidades.

$$\Gamma_+^2 = h_+^2 \sigma_+^2 + 1 + (1 + b_+ \varepsilon)^2 \quad \text{e} \quad \Gamma_-^2 = h_-^2 \sigma_-^2 + 1 + (1 + b_- \varepsilon)^2.$$

Podemos então dizer que a decisão entre as alternativas A_- e A_+ nos conduz naturalmente a um teste de hipóteses sobre as variâncias das variáveis aleatórias Δ_k^- e Δ_k^+ . Note-se que $\Gamma_+^2 - \Gamma_-^2 = O(\varepsilon)$ (ver nota de rodapé (3)), logo não será surpreendente se chegarmos a um teste que necessita de um tempo longo para tomar uma decisão.

Considere-se a razão das respectivas funções de densidade (gaussianas) :

$$Z_k = \ln \left(\frac{f_N(\Delta_k^+, 0, \Gamma_+^2 \varepsilon)}{f_N(\Delta_k^-, 0, \Gamma_-^2 \varepsilon)} \right).$$

Ela permite definir a seguinte estatística :

$$S_n = \sum_{k=i_0}^{i_0+n} Z_k,$$

com $0 \leq n \leq m$. Desenvolvendo a razão Z_k obtém-se :

$$Z_k = \ln \left(\frac{\Gamma_-}{\Gamma_+} \right) + \frac{1}{2\varepsilon} \left(\frac{\Delta_k^-^2}{\Gamma_-^2} - \frac{\Delta_k^+^2}{\Gamma_+^2} \right)$$

e portanto

$$S_n = (n+1) \ln \left(\frac{\Gamma_-}{\Gamma_+} \right) + \frac{1}{2\varepsilon} \sum_{k=i_0}^{i_0+n} \left(\frac{\Delta_k^-^2}{\Gamma_-^2} - \frac{\Delta_k^+^2}{\Gamma_+^2} \right). \quad (5.7)$$

Construída a estatística de teste, S_n , falta agora determinar a região crítica para a estatística, que é calculada atendendo às probabilidades de erro consideradas, p_- e p_+ ⁶, definidas como na secção anterior (ver (5.5)).

Seja $[-l_1, l_2]$, com $l_1 > 0$ e $l_2 > 0$, a região crítica para a estatística S_n , que é construída para um erro do tipo I, α_s . Os valores para l_1 e l_2 podem ser determinados aproximadamente se em (5.5) substituirmos os segundos membros por expressões aproximadas. Temos então o seguinte sistema de equações não linear:

⁶ Em geral consideraremos $p_+ = p_- = 0.05$.

$$\left\{ \begin{array}{l} p_+ = \frac{1 - e^{-(\mu_+ / \gamma_+^2) l_2}}{e^{(\mu_+ / \gamma_+^2) l_1} - e^{-(\mu_+ / \gamma_+^2) l_2}} \\ p_- = \frac{1 - e^{-(\mu_- / \gamma_-^2) l_1}}{e^{-(\mu_- / \gamma_-^2) l_2} - e^{-(\mu_- / \gamma_-^2) l_1}} \end{array} \right., \quad (5.8)$$

em que

$$\begin{aligned} \mu_+ &= \ln \left(\frac{\Gamma_-}{\Gamma_+} \right) + \frac{1}{2} \left(\frac{1}{\Gamma_-^2} - \frac{1}{\Gamma_+^2} \right) \Gamma_+^2, \\ \mu_- &= \ln \left(\frac{\Gamma_-}{\Gamma_+} \right) + \frac{1}{2} \left(\frac{1}{\Gamma_-^2} - \frac{1}{\Gamma_+^2} \right) \Gamma_-^2, \end{aligned}$$

$$\begin{aligned} \gamma_+ &= \frac{1}{2} \left(\frac{1}{\Gamma_-^2} - \frac{1}{\Gamma_+^2} \right) \sqrt{\Gamma_+^4 + 2(1 + b_+ \varepsilon)^2}, \\ \gamma_- &= \frac{1}{2} \left(\frac{1}{\Gamma_-^2} - \frac{1}{\Gamma_+^2} \right) \sqrt{\Gamma_-^4 + 2(1 + b_- \varepsilon)^2}. \end{aligned}$$

A maneira como estas expressões são deduzidas assim como detalhes sobre a construção do teste são apresentados em Oliveira (2000).

Podemos agora estabelecer a regra de decisão para o teste (de variação quadrática) :

- Se $S_{N^*} \leq -l_1$, decidir pelo acontecimento A_- ;
- Se $S_{N^*} \geq l_2$, decidir pelo acontecimento A_+ ;
- Caso contrário, não decidir,

onde N^* é um tempo de paragem definido da mesma forma que no teste anterior (secção 5.1), agora para a estatística S_n .

Dado que pretendemos comparar o comportamento dos testes, torna-se interessante caracterizar a performance do teste em termos do tempo necessário para atingir uma decisão, razão pela qual medimos aquilo a que chamaremos o tempo médio de espera para uma decisão. Considerando os tempos de paragem N^+ e N^- que representam a variável aleatória N^* sob as hipóteses A_+ e A_- respectivamente, definimos os correspondentes tempos de espera $T^+ = N^+ \varepsilon$ e $T^- = N^- \varepsilon$ contabilizados no intervalo de tempo contínuo $[0, T]$. Oliveira-Roubaud (1995) e Oliveira (2000) apresentam fórmulas aproximadas para os tempos de paragem T^+ e T^- .

Voltando ao sistema de equações não linear (5.8), convém notar que a determinação da sua solução, de um ponto de vista prático, não é imediata. De facto esta tem geralmente de ser obtida por métodos numéricos, de entre os métodos habitualmente usados para resolução de sistemas não lineares com os inconvenientes que tão bem conhecemos. Existe em geral um conjunto de soluções possíveis e aproximadas, ficando o sucesso do teste muito dependente

da escolha de uma solução inicial para $-l_1$ e l_2 no arranque do método numérico. Por outro lado, não temos garantia do sistema ter sempre solução e por vezes, quando existe mais do que uma solução, torna-se necessário optar pelo par de valores (l_1, l_2) mais “adequado”. Como seria de esperar os resultados do teste são muito sensíveis aos parâmetros l_1 e l_2 e por isso torna-se conveniente determinar um procedimento para a sua determinação na prática. No caso da hipótese (HD1) Fleming *et al.* (1991) contornaram esta dificuldade usando a seguinte aproximação:

$$l_1 = -\left|\frac{1}{\theta_+}\right| \log(\alpha_s) \quad \text{e} \quad l_2 = -\left|\frac{1}{\theta_-}\right| \log(\alpha_s),$$

com $\theta_+ = \frac{\mu_+}{\gamma_+}$ e $\theta_- = \frac{\mu_-}{\gamma_-}$.

(5.9)

Estas aproximações foram testadas em alguns exemplos de aplicação e revelaram-se eficazes no caso da hipótese (HD1). Infelizmente, como veremos mais à frente (secção 6), o mesmo não se pode dizer para a hipótese (HD2). Os resultados numéricos que o ilustram são apresentados na secção 6.

6. EXEMPLOS DE APLICAÇÃO

Esta secção é dedicada ao estudo numérico do comportamento dos testes de decisão sobre o sinal. O principal objectivo consiste em analisar o desempenho do teste de variação quadrática, na decisão sobre o sinal do estimador a aplicar, e compará-lo com o teste da razão de verosimilhança quando se verifica a hipótese (HD2). Para isso, recorreremos à simulação numérica de alguns exemplos de modelos.

A análise baseia-se em 100 trajectórias geradas, para cada exemplo, num intervalo de tempo contínuo $[0,15]$, o que, para $\mathcal{E}=0.01$, corresponde a considerar uma partição em 1500 instantes de tempo discreto ($K=1500$). A simulação dos processos de estado e observações foi feita partindo da condição inicial $X_0 \sim N(-5,0.1)$.

A performance dos testes é medida através das percentagens de instantes em que os testes actuam e das percentagens de erro na decisão sobre o sinal para o processo não observado. As probabilidades de erro são comparadas com as frequências de decisões erradas encontradas nas simulações, Pe_v e Pe_{vq} . Para o cálculo destas frequências consideram-se apenas os intervalos em que efectivamente não houve cruzamentos em zero, independentemente de estarem ou não bem detectados pelo primeiro teste. Os testes de decisão sobre o sinal são ainda analisados tendo em conta os tempos médios de paragem, que correspondem aos tempos (mínimos) de que, em média, um teste precisa para decidir. A análise dos tempos de paragem permite-nos comparar a rapidez de decisão dos testes ou até mesmo analisar o efeito dos parâmetros do modelo na rapidez de decisão. Nesta secção apresentamos alguns dos resultados mais importantes, sob a forma de tabela ou gráfico. Para valores mais detalhados poderão consultar Sousa (1999).

Nesta secção serão utilizadas as seguintes notações:

- α_d : nível de significância para o teste de detecção de passagens em zero.
- α_s : nível de significância para o teste de decisão sobre o sinal.

- $-l_1, l_2$: valores críticos para a estatística do teste de decisão sobre o sinal.
- pe_v, pe_{vq} : percentagens de erro nos testes da razão de verosimilhança e de variação quadrática, para decisão sobre o sinal.
- pid_v, pid_{vq} : percentagens de instantes de decisão para os testes da razão de verosimilhança e de variação quadrática.
- ET_v^-, ET_v^+ : médias teóricas do tempo necessário para decisão no teste de verosimilhança.
- \bar{T}_v^-, \bar{T}_v^+ : médias (empíricas) do tempo necessário para decisão no teste da razão de verosimilhança.
- ET_{vq}^-, ET_{vq}^+ : médias teóricas do tempo necessário para decisão no teste de variação quadrática.
- $\bar{T}_{vq}^-, \bar{T}_{vq}^+$: médias (empíricas) do tempo necessário para decisão no teste de variação quadrática.
- i_0 : instante a partir do qual se começa a contabilizar a estatística do teste de decisão ⁷.

No caso do teste da razão de verosimilhança, a solução $l_1 = l_2 = \ln\left(\frac{1-\alpha_s}{\alpha_s}\right)$ do sistema de equações não lineares apresentado no secção 5 (sistema 5.6) é adequada ao nosso problema (Oliveira-Roubaud, 1991). Para uma probabilidade de erro $\alpha_s = 0.05$ obtém-se $l_1 = l_2 = 2.94$. No caso do teste de variação quadrática esses valores dependerão dos parâmetros do modelo. Veremos a seguir que nos deparamos com algumas dificuldades quando pretendemos obter uma boa solução para o par (l_1, l_2) , tendo mesmo que recorrer a métodos de aproximação numérica (como por exemplo, a função 'fsolve' da linguagem de programação Matlab que permite a resolução numérica de sistemas não lineares). As duas figuras que se seguem ilustram bem a importância de uma boa escolha para l_1 e l_2 como soluções de (5.8). Nas Figuras 6.1 e 6.2 estão representados os resultados para o teste de variação quadrática com duas soluções diferentes para os parâmetros l_1 e l_2 , para uma mesma trajectória do processo de estado. Estes valores foram obtidos numericamente através da função 'fsolve' do Matlab.

⁷ Conforme foi mencionado na secção 5.1, dado que estes testes assentam sobre condições de estacionaridade considerou-se i_0 incrementado em 6 unidades. No entanto alguns ensaios foram realizados com i_0 incrementado de 11 unidades, de modo a estudar a sensibilidade do teste relativamente a este parâmetro.

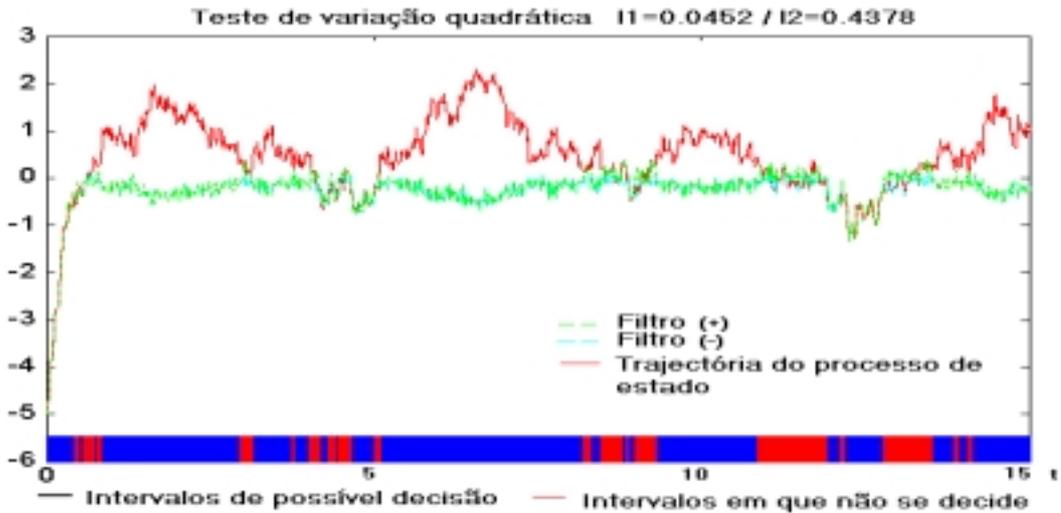


Figura 6.1 - Resultado da filtragem aplicando o teste de variação quadrática, em que $l_1 = 0.0452$, $l_2 = 0.4378$.

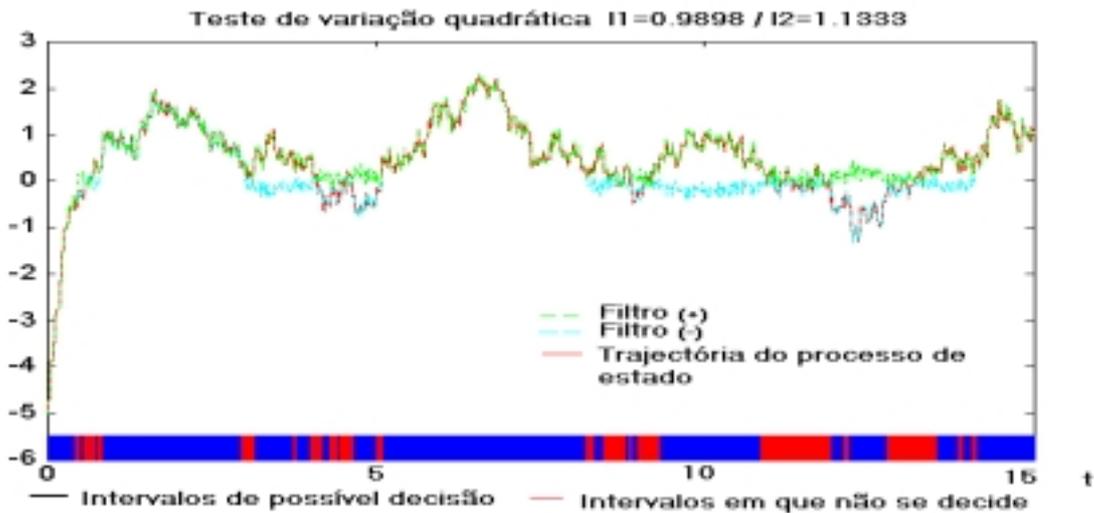


Figura 6.2 - Resultado da filtragem aplicando o teste de variação quadrática, em que $l_1 = 0.9898$, $l_2 = 1.1333$.

A zona identificada com barra preta corresponde aos intervalos seleccionados pelo teste de detecção de passagens em zero. A trajetória do processo de estado está representada a contínuo, sendo os filtros $(FKB)^+$ e $(FKB)^-$ representados a tracejado em tons mais claros. Nos instantes em que estes traços coincidem, significa que foi tomada uma decisão e, se eles seguirem a trajetória do processo de estado, a decisão foi correcta.

Na Figura 6.1, verificamos que o teste de variação quadrática decidiu em praticamente todos os intervalos de detecção e quase sempre de forma errada. O facto do valor l_1 estar muito próximo de zero faz com que o teste decida muitas vezes pelo sistema (-), sem que corresponda à realidade. Para este modelo, na simulação de 100 trajetórias, incluindo a da figura, obteve-se um erro na decisão muito superior a 5%. Podemos assim concluir que os valores $l_1 = 0.0452$ e $l_2 = 0.04378$ não conduzem a bons resultados de estimação. Vejamos agora o que se passa se escolhermos um outro par de valores (l_1, l_2) (ver Figura 6.2). Neste exemplo, uma abordagem do sistema de equações (5.8) análoga à sugerida em Fleming *et al.*

(1991) para os modelos sob a hipótese (HD1) (ver expressão (5.9)) permite encontrar valores de l_1 e l_2 mais equilibrados e que dão origem a probabilidades de erro do tipo I inferiores a 5% (a saber $p_+ = 0$ e $p_- = 0.0413$). A simples alteração dos parâmetros l_1 e l_2 proporcionou um funcionamento adequado do teste, muito semelhante ao do teste da razão de verosimilhança, decindo correctamente pelo filtro (+) em intervalos em que anteriormente não decidia.

Do ponto de vista prático, nem sempre é fácil encontrar soluções para os parâmetros l_1 e l_2 , que conduzem a probabilidades de erro adequadas. Há que ensaiar diferentes alternativas, partindo do que foi sugerido por Fleming *et al.* (1991).

Passamos agora ao estudo do comportamento dos testes em alguns exemplos (casos 1 a 4). Começemos por considerar o modelo (2.2) com os seguintes parâmetros:

$$\begin{cases} b_+ = -0.01, b_- = -0.5 \\ h_+ = 1, h_- = -1 \\ \sigma_+ = 1, \sigma_- = 1 \end{cases} \quad \text{(Caso 1)}$$

Os resultados de aplicação dos testes estão resumidos na Tabela 6.1. Para este exemplo de aplicação, obteve-se uma percentagem de erro na decisão do teste da razão de verosimilhança muito semelhante à do teste de variação quadrática (6.56% e 6.82%, respectivamente). As percentagens de instantes de tempo com decisões também foram muito semelhantes (44.46% e 41.79%, respectivamente). Vejamos o que se passa se neste exemplo se aumentar a diferença entre os parâmetros b_- e b_+ . Consideremos assim dois exemplos com as seguintes alterações no parâmetro b_- :

$$b_- = -1 \quad \text{(Caso 2)}$$

$$b_- = -4 \quad \text{(Caso 3)}$$

Da análise da Tabela 6.1 retiramos que, ao aumentar $|b_+ - b_-|$ se obteve uma maior percentagem de instantes de decisão (pid_v e pid_{vq}) e, por outro lado, uma diminuição no erro da decisão (pe_v e pe_{vq}), para ambos os testes. O facto da decisão ser mais rápida nestes últimos casos pode também ser constatado através dos valores médios empíricos \bar{T}_v^- , \bar{T}_v^+ e \bar{T}_{vq}^- , \bar{T}_{vq}^+ que diminuíram consideravelmente. Note-se que é natural esperar que o aumento da distância entre os parâmetros b_- e b_+ facilite a decisão sobre o sinal nesses intervalos, uma vez que quanto maior $|b_+ - b_-|$ mais rapidamente os valores das variáveis de estado se afastam do limiar zero.

Casos	b_-	b_+	pid_v	pid_{vq}	pe_v	pe_{vq}	\bar{T}_v^-	\bar{T}_v^+	\bar{T}_{vq}^-	\bar{T}_{vq}^+
1	-0.5	-0.01	41.46%	41.79%	6.56%	6.82%	1.2	2.68	1.46	6.66
2	-1	-0.05	47.21%	34.74%	3.95%	4.41%	0.42	1.44	0.67	3.86

3	-4	-0.05	61.36%	58.77%	4.77%	4.59%	0.09	0.37	0.17	0.5
---	----	-------	--------	--------	-------	-------	------	------	------	-----

Tabela 6.1 – Resultados de decisão nos testes da razão de verossimilhança e de variação quadrática para diferentes valores de b_- e b_+ .

O mesmo acontece, relativamente à detecção de intervalos sem passagens em zero.

Comparando os dois testes de decisão sobre o sinal, entre si, verificamos que, embora as percentagens de erro sejam semelhantes em ambos os testes, os tempos médios de decisão para o teste de variação quadrática são ligeiramente superiores.

Tentemos agora analisar o efeito da quantidade $\dot{\Gamma}^2 = (\sigma_- h_-)^2 = (\sigma_+ h_+)^2$ no comportamento dos testes. Se analisarmos a estatística utilizada no teste de razão de verossimilhança (expressão (5.4)) podemos constatar que as alterações em $\dot{\Gamma}^2$ deverão influenciar muito os valores da estatística. Retomando o caso 2, em que $(\sigma_- h_-)^2 = (\sigma_+ h_+)^2 = 1$, vejamos o que sucede se diminuirmos esta quantidade. Consideremos um novo exemplo em que apenas se diminui o valor $(\sigma_- h_-)^2 = (\sigma_+ h_+)^2$:

$$\begin{cases} \sigma_- = 1, h_- = -0.8 \\ \sigma_+ = 1, h_+ = 0.8 \end{cases}$$

(Caso 4)

Os resultados são apresentados na tabela seguinte:

Casos	h_-	h_+	σ_-	σ_+	pid_v	pid_{vq}	pe_v	pe_{vq}	\bar{T}_v^-	\bar{T}_v^+	\bar{T}_{vq}^-	\bar{T}_{vq}^+
2	-1	1	1	1	47.21%	34.74%	3.95%	4.41%	0.42	1.44	0.67	3.86
4	-0.8	0.8	1	1	45.06%	25.98%	8.56%	1.72%	0.32	1.37	1.23	5.3

Tabela 6.2 – Resultados de decisão nos testes da razão de verossimilhança e de variação quadrática, para diferentes funções

$$\sigma h(x_k) \quad (b_- = -1, b_+ = -0.05).$$

Enquanto que para $\dot{\Gamma}^2 = 1$ (caso 2) se obtiveram decisões em 47.21% dos instantes e erro em 3,95% das decisões, para $\dot{\Gamma}^2 = 0.64$ (caso 4) obtiveram-se decisões em $pid_v = 45.06\%$ dos instantes e $pe_v = 8.56\%$ de decisões erradas. Quanto ao teste de variação quadrática, verificou-se uma diminuição na frequência de decisões erradas (4.4% e 1.7% respectivamente).

A diminuição da quantidade $|h_-| = |h_+|$ provoca uma diminuição na grandeza dos valores observados reduzindo ligeiramente a percentagem de instantes que verificam o acontecimento C do teste de detecção. Por outro lado, a estatística do teste da razão de verossimilhança necessita de um menor número de instantes de tempo para atingir os valores críticos, mas vem acompanhada de um aumento do erro de decisão pe_v . Verificámos ainda que, se $\dot{\Gamma}^2 = (\sigma_- h_-)^2 = (\sigma_+ h_+)^2$ aumenta consideravelmente este teste toma decisões erradas

com maior frequência revelando um mau comportamento na prática. Uma razão para este comportamento poderá estar em que quando $\hat{\Gamma}^2$ aumenta, o termo principal da estatística deste teste (ver (5.7)) diminui de peso. No entanto, o primeiro termo, que nestes casos é sempre negativo, assume maior importância levando por vezes a decisões precipitadas. O facto da estatística atingir o valor $-l_1$, quando as variáveis estão do lado positivo, leva a concluir que, nestes casos, o teste poderá não ser tão eficaz. Como verificámos que a performance do teste está muito relacionada com a maneira como são determinados os valores de l_1 e l_2 , enveredámos esforços para obter outros valores possíveis para estas constantes o que se revelou infrutífero. Não quer isto dizer, evidentemente, que eles não existam mas sim que seria necessário uma investigação especificamente com este propósito do âmbito da análise numérica. Esta situação mostra que, na prática, é necessário ter em atenção as características dos modelos e verificar, por via da simulação, se o teste de facto permite obter bons resultados quando aplicado. Pelos exemplos que foram testados constata-se que, aparentemente, quando $|\sigma_- h_-| = |\sigma_+ h_+|$ é elevado (ultrapassa a unidade) torna-se problemática a aplicação do teste de variação quadrática, sendo nesses casos aconselhável dar preferência ao outro teste de decisão.

Se por um lado o teste de variação quadrática tem algumas limitações em termos de aplicação, por outro lado este teste revela-se mais eficaz nos intervalos próximos de zero que ainda correspondem a intervalos detectados. Em geral o teste de variação quadrática toma mais frequentemente decisões nessas zonas, comparativamente ao teste da razão de verosimilhança. Este comportamento é ilustrado nas Figuras 6.3 e 6.4 relativas ao caso 3. Destacaram-se duas regiões (Zoom1 e Zoom2, Figura 6.5) para, nesta escala de cinza, facilitar a visualização dos filtros quando não há decisão de sinal e, por consequência, não há estimação de estado.

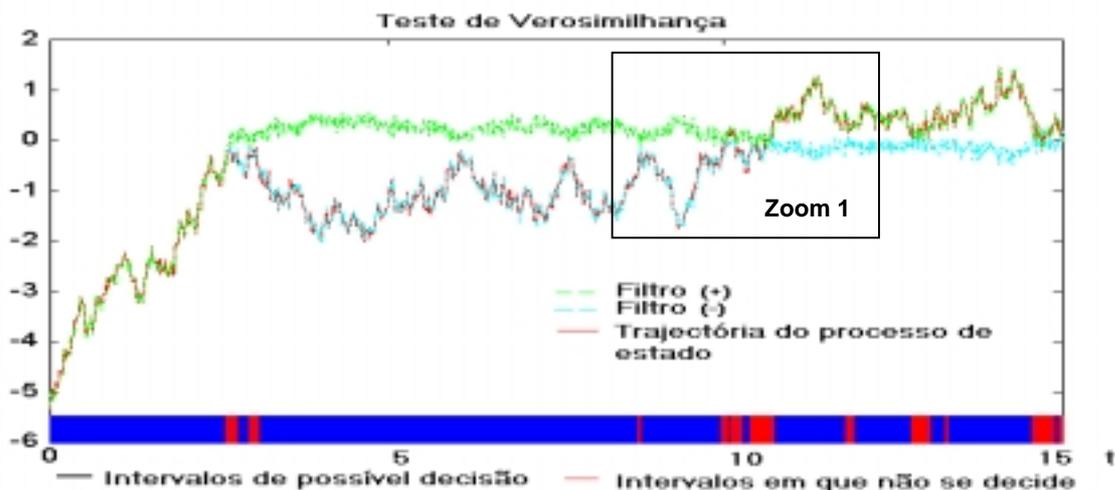


Figura 6.3 – Resultado da filtragem utilizando o teste de verosimilhança para uma trajetória do caso 3.



Figura 6.4 – Resultado da filtragem utilizando o teste de variação quadrática para uma trajectória do caso 3.

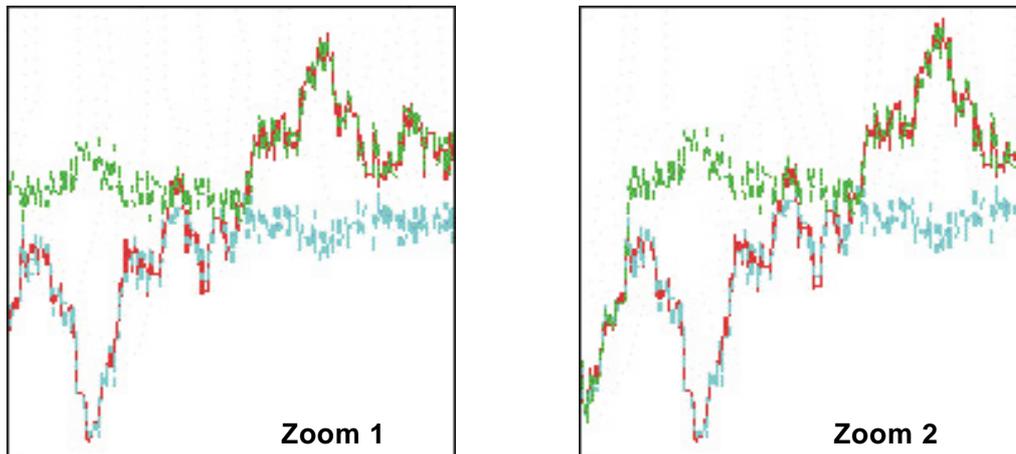


Figura 6.5 – Ampliações das regiões Zoom1 e Zoom2 das figuras 6.3 e 6.4.

Como podemos ver, em ambos os testes de decisão sobre o sinal, existem zonas no intervalo de tempo contínuo $[10,15]$ onde o teste é aplicado mas não permite tomar uma decisão (nesses intervalos a estatística não chega a atingir os limites $-l_1$ e l_2). Para os instantes anteriores a $t=10$, pode-se destacar na Figura 6.4 um intervalo longo, identificado com barra a preto, onde o teste de variação quadrática conduz a uma decisão o que não acontece com o teste da razão de verosimilhança (Figura 6.3). O teste de variação quadrática opta correctamente pelo filtro do lado negativo nesse intervalo. Este é apenas um exemplo dos muitos em que situações como esta ocorrem. De facto, em zonas de detecção mais próximas de zero o teste de variação quadrática, se encontrar um intervalo suficientemente longo para poder decidir, parece funcionar melhor do que o teste da razão de verosimilhança. Podemos dizer que este último tem geralmente melhor desempenho do que o teste de variação quadrática por estar menos sujeito a decisões erradas, mas perde performance no que respeita à percentagem de decisões tomadas.

7. CONCLUSÕES

Ao longo deste trabalho pretendeu-se analisar os modelos de Markov escondidos unidimensionais lineares por bocados em tempo discreto, numa perspectiva de estimação do processo de estado não observado. Mediante uma hipótese de detectabilidade sobre o modelo, os estimadores tradicionais (Filtro de Kalman-Bucy) permitem estimar o estado, desde que precedidos da aplicação de uma sequência de dois testes estatísticos. Foram estudados o teste de detecção de passagens em zero e dois testes de decisão sobre o sinal para o processo das observações, atendendo à hipótese de detectabilidade (HD2). Sob esta hipótese, compara-se o desempenho do teste de variação quadrática, na decisão sobre o sinal, com o teste da razão de verosimilhança.

Confirmámos os resultados obtidos por Oliveira-Roubaud (1995) segundo os quais o teste de razão de verosimilhança pode ser usado com sucesso no procedimento de estimação do estado. Verificámos que o teste de variação quadrática não é geralmente melhor do que o anterior mas permite-nos por vezes obter estimativas em intervalos de tempo em que o teste da razão de verosimilhança não toma uma decisão, como é o caso de intervalos com poucos instantes ou em que o processo de estado toma valores muito próximos de zero. Verificámos no entanto existirem exemplos em que o teste de variação quadrática não fornece bons resultados e dificilmente cumpre os níveis de erro exigidos. Não se aconselha portanto a aplicação do teste de variação quadrática sem prévia verificação do comportamento sob simulação. Aparentemente o seu mau desempenho acontece quando estão em causa valores elevados para $(\sigma h)^2$. Um estudo mais demorado, com um número muito maior de simulações e um passo de tempo inferior seria necessário para averiguar das razões do aparecimento de valores de erro mais elevados e verificar se, de facto, são significativamente diferentes de 5%. Concluimos também que o comportamento dos testes está muito dependente dos valores encontrados para os parâmetros l_1 e l_2 (valores críticos da estatística do teste) e seria necessário realizar um estudo numérico aprofundado para melhorar a determinação destes parâmetros (que resultam da resolução de um sistema de equações não linear). A aproximação sugerida por Fleming et al (1991) (ver expressão 5.9) nem sempre se revelou adequada para os modelos que verificam a hipótese de detectabilidade (HD2), embora se considere que a expressão é adequada para uso nos modelos em que se verifica a hipótese (HD1). A qualidade dos testes deveria melhorar consideravelmente se utilizássemos um passo de tempo inferior ao utilizado para a discretização do modelo (2.2), o que acarretaria no entanto uma grande sobrecarga em termos computacionais.

AGRADECIMENTOS

Ao Instituto Nacional de Estatística, em especial ao Professor Doutor Paulo Gomes, pelo interesse e apoio demonstrados para a publicação deste artigo. À Prof. Doutora Paula Milheiro Oliveira, docente na Faculdade de Engenharia da UP, pela sua disponibilidade para sugestões e comentários que se revelaram extremamente úteis.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDREWS, ANGUS P. e GREWAL, MOHINDER S. (1993). *Kalman Filtering, Theory and Practice*, Prentice Hall, New Jersey.

- FLEMING, W. H. e PARDOUX, E. (1989). *Piecewise monotone filtering with small observation noise*, SIAM J. Control and Optimization, vol. 27, No. 5, pp. 1156-1181.
- FLEMING, W. H, JI, D. e PARDOUX, E. (1988). *Piecewise linear filtering with small observation noise*, Proc. 8th Int. Conf. on Analysis and Optim. Of Systems, Antibes 1988, Lect, No. 111, pp. 725-739, Springer, 1988.
- FLEMING, W. H., JI, D., SALAME, P. e ZHANG, Q. (1991). *Piecewise monotone filtering in discrete-Time with small observation noise*, IEEE Transactions on Automatic Control, Vol. 10, No. 10, pp 1181-1186.
- HARVEY, ANDREW C. (1989). *Forecasting, Structural time series models and Kalman filter*, Department of Statistical and mathematical Sciences, London.
- JAZWINSKI, ANDREW H. (1970). *Stochastic processes and filtering theory*, Academic Press.
- KALMAN, R. E. e BUCY, R. S. (1961). *New results in linear filtering and prediction theory*, J. Basic Eng., vol. 83, pp. 95-108.
- OLIVEIRA, P. MILHEIRO (1994). *Uma Versão discreta do Teorema de Girsanov*, II Congresso Anual da Sociedade Portuguesa de Estatística, Luso, pp. 211-217.
- OLIVEIRA, P. MILHEIRO E ROUBAUD, MARIE-CHRISTINE (1991). *Filtrage lineaire par morceaux d'un systeme en temps discret avec petiti bruit d'observation*, Unité de Recherche, INRIA-Sophia Antipolis, No. 1451.
- OLIVEIRA, P. MILHEIRO e ROUBAUD, M. C. (1995). *Discrete-Time piecewise linear filtering with small observation noise*, IEEE Transactions on Automatic Control, Vol. 40, No. 12, pp 2149-2152.
- OLIVEIRA, P. MILHEIRO (2000). *A sequential method for the estimation of a partially observed piecewise AR(1) process*, pp 1-5 (preprint).
- PARDOUX, E. e ROUBAUD, M.C. (1991), *Finite dimensional approximate filters in the case of high signal-to-noise ratio*, Conference on the Memory of Zakai, Haifa.
- PICARD, J. (1986). *Nonlinear filtering of one-dimensional diffusions in the case of a high signal-to-noise ratio*, SIAM J. APPL. MATH., vol. 46, No. 6, pp. 1098-1125.
- PICARD, J. (1991). *Efficiency of the extended Kalman filter for non linear systems with small noise*, , SIAM J. APPL. MATH, (51), 3, pp. 843-855.
- ROUBAUD, M. C. (1992). *An approximate filter for a partially observed piecewise linear system with small observation noise*, Systems & Control Letters, North-Holland, No. 20, pp. 239-247.
- ROUBAUD, M. C. (1995). *Filtrage linéaire par morceaux avec petiti bruit d'observation*, Applied Mathematics Optimization, Springer-Verlag, New York, No. 32, pp. 163-194.
- SORENSEN, H. W. (1985). *Kalman Filtering: Theory and application*, Department of Applied Mechanics and Engineering Sciences, University of California, San Diego.
- SOUSA, RITA C. P. (1999). *Estimação de estado em modelos lineares por bocados parcialmente observados: comparação numérica de dois testes*, Relatório da Tese de Mestrado em Estatística, Dep. Matemática Aplicada, F.C.U.P.
- TONG, HOWELL e YEUNG, IRIS (1989). *On tests for self-exciting threshold autoregressive-type non-linearity in partially observed time series*, University of Kent at Canterbury, UK.

UM ESTUDO SOBRE A SATISFAÇÃO NO TRABALHO - UMA APLICAÇÃO DO MODELO LOGIT ORDENADO

ORDERED LOGIT MODELS: A STUDY ON JOB SATISFACTION

Autor: Paulo Costa Parente¹

- Técnico Superior de Estatística no Gabinete de Estudos e Conjuntura do Instituto Nacional de Estatística

RESUMO:

- Com este estudo pretende-se mostrar de que forma podemos utilizar modelos Logit ordenados para analisar a satisfação no trabalho. Assim, procurou-se estudar de que forma a satisfação dos trabalhadores da Direcção de Pessoal de uma grande empresa de transportes da Área Metropolitana de Lisboa, variou com as suas características. Os resultados obtidos permitem concluir que, salvo algumas excepções, o nível de satisfação dos trabalhadores não varia com o sexo, idade e antiguidade na empresa, mas o mesmo já não se verifica relativamente ao nível hierárquico, habilitações e local de trabalho.

PALAVRAS-CHAVE:

- *Satisfação no trabalho, Modelos Logit ordenados.*

ABSTRACT:

- With this study we intend to show how one can use ordered Logit models to analyse job satisfaction. Thus, we tried to study how the satisfaction of the workers of the Personnel Department of a large company of transports of the Metropolitan Area of Lisbon varies, with their own characteristics. Our results lead to the conclusions that, with some exceptions, the satisfaction of workers does not vary with their gender, age and experience in the company. On the other hand, it varies with their hierarchic level, their qualifications, and the place where they work.

KEY-WORDS:

- *Job satisfaction, Ordered Logit Models.*

¹ Este artigo foi produzido como parte integrante dos elementos de avaliação do curso de Microeconometria do Programa de Doutoramento e Mestrado da Faculdade de Economia da Universidade Nova de Lisboa no ano lectivo de 1999/2000, leccionado pelo Professor Doutor João Santos Silva, a quem agradeço toda a ajuda dispensada. Agradeço, ainda, à Dra. Regina Soares e Dra. Teresa Bago d'Uva pelos comentários à versão preliminar do artigo e, especialmente, ao meu irmão, Dr. José Nuno Parente, por ter cedido os dados para a realização deste trabalho.

INTRODUÇÃO

O objectivo deste estudo é mostrar de que forma podemos aplicar modelos Logit ordenados para estudar a satisfação no trabalho. Com este propósito procurou-se analisar a satisfação dos trabalhadores da Direcção de Pessoal de uma grande empresa de transportes da Area Metropolitana de Lisboa. Mais precisamente, pretendeu-se verificar de que forma o nível de satisfação variou com as características dos indivíduos.

Para tal solicitou-se a estes trabalhadores a resposta a um extenso inquérito, em que teriam de classificar alguns itens relacionados com a satisfação de 1 (concordo) a 5 (discordo), para além de fornecerem alguma informação sobre as suas próprias características.

Temos assim um modelo em que a variável dependente é qualitativa e ordenada (a classificação atribuída ao item pelos indivíduos) e em que as variáveis independentes são as características dos indivíduos. Contudo, quando a variável dependente é discreta ordenada não é correcto utilizar a regressão linear, pelo que é necessário aplicar uma metodologia diferente, que tenha em conta esta propriedade.

Este trabalho está organizado da seguinte forma. Na primeira secção é descrito com algum detalhe o modelo com variável dependente discreta ordenada bem como os testes de especificação apropriados.

É feita, na segunda secção, uma exposição do inquérito utilizado e é revelada qual a amostra estudada. Nesta secção são indicados quais os itens sobre os quais o estudo se debruçou e de que forma foram incorporadas as variáveis relativas às características nos modelos analisados neste trabalho. Na mesma secção são, ainda, apresentados e interpretados os resultados empíricos obtidos.

Por último na terceira secção é feita uma síntese com as principais conclusões.

1. O MODELO COM VARIÁVEL DEPENDENTE DISCRETA ORDENADA

1.1. DESCRIÇÃO DO MODELO

Foi afirmado anteriormente que quando a variável dependente é discreta ordenada a regressão linear não é apropriada. A razão de tal situação está relacionada com o facto da regressão linear não ter em conta a natureza discreta dos dados e de tratar a diferença entre 4 e 3 da mesma maneira a diferença entre 3 e 2 quando, de facto, não há nada que a isso obrigue. Assim, é necessário utilizar uma metodologia apropriada em que a variável dependente é discreta ordenada. É precisamente esta metodologia que será aplicada neste estudo.

O modelo é construído a partir de uma relação linear entre uma variável latente contínua Y_i^* e um vector de regressores \bar{x}_i ($k \times 1$):

$$(1.1.1) \quad Y_i^* = \beta^T \bar{x}_i + \varepsilon_i \text{ com } i = 1, \dots, N$$

em que β é um vector de parâmetros ($k \times 1$) e ε_i é o termo de erro.

Apesar de não se conseguir observar Y_i^* consegue-se observar:

$$Y_i=0 \text{ se } Y_i^* \leq e_0$$

$$Y_i=1 \text{ se } e_0 < Y_i^* \leq e_1$$

$$Y_i=2 \text{ se } e_1 < Y_i^* \leq e_2$$

· ·

· ·

· ·

$$Y_i=J \text{ se } e_{J-1} < Y_i^*$$

$$\text{com } e_0 < e_1 < \dots < e_{J-1}$$

em que os e_j ($j=0, \dots, J-1$) são limites desconhecidos dos intervalos que vão ser estimados juntamente com o vector β .

Note-se que neste caso $E(Y_i^* | \bar{x}_i) = \beta^T \bar{x}_i$ o que já não acontece com $E(Y_i | \bar{x}_i)$.

Seja $Var(\varepsilon_i | \bar{x}_i) = \sigma^2$. Note-se que se dividirmos a equação (1.1.1) por σ obtemos:

$$(1.1.2) \quad \frac{Y_i^*}{\sigma} = \frac{1}{\sigma} \beta^T \bar{x}_i + \frac{\varepsilon_i}{\sigma} \text{ com } i = 1, \dots, N$$

Mas a definição de Y_i não se altera. Este facto implica que o Y_i não depende da escala em que o Y_i^* é medido. Contudo, uma vez que só é observado o Y_i nunca conseguiremos estimar a variância condicional do ε_i .

Assim, uma vez que é impossível estimar a variância condicional do ε_i , ao especificarmos a distribuição condicional de ε_i em \bar{x}_i assumimos que esta variância é uma constante conhecida. Todavia ao se fazer isto tem que se reconhecer que de facto não se irá estimar o vector β mas sim $\frac{\beta}{\sigma}$.

Para estimarmos o modelo é necessário especificar uma distribuição condicional de ε_i em \bar{x}_i . Podemos assumir as mais diversas distribuições condicionais de ε_i em \bar{x}_i , mas na prática as que são mais utilizadas são a Normal standardizada e a Logística.

Se admitirmos que a distribuição condicional de ε_i é a Normal standardizada então o modelo tem o nome de Probit ordenado. No caso de admitirmos que essa distribuição é a Logística então teremos o modelo Logit ordenado. Neste caso a função distribuição dos erros é a seguinte:

$$F(\varepsilon_i) = \frac{\exp(\varepsilon_i)}{1 + \exp(\varepsilon_i)}$$

Será, de facto, este último o modelo utilizado neste trabalho. A opção por este modelo não tem grande suporte teórico. Deveu-se apenas a duas razões: Em primeiro lugar e tal como afirma Johnston & DiNardo (1997), aparentemente na prática, existe pouca diferença entre o modelo Probit e Logit ordenado pelo que será indiferente escolher entre estes dois modelos, em segundo lugar porque, tal como se poderá ver na subsecção 1.3., os testes de especificação nos modelos Logit ordenados são de aplicação prática mais rápida e fácil.

Para tornar o modelo o mais geral possível vai-se considerar que $F(\cdot)$ é a função de distribuição condicional do ε_i .

Assim para $j=0, \dots, J$ temos

$$\begin{aligned} P(Y_i = j | \bar{x}_i) &= P(e_{j-1} < Y_i^* \leq e_j | \bar{x}_i) = P(e_{j-1} < \beta^T \bar{x}_i + \varepsilon_i \leq e_j | \bar{x}_i) = P(e_{j-1} - \beta^T \bar{x}_i < \varepsilon_i \leq e_j - \beta^T \bar{x}_i | \bar{x}_i) = \\ &= F(e_j - \beta^T \bar{x}_i) - F(e_{j-1} - \beta^T \bar{x}_i) \end{aligned}$$

com $e_{-1} = -\infty$ e $e_J = +\infty$

Vai-se, agora, definir as seguintes variáveis *dummy*:

$$z_{ij} = \begin{cases} 1 & \text{se } y_i = j \\ 0 & \text{se } y_i \neq j \end{cases} \quad (i=1,2,\dots,N, j=1,2,\dots,J)$$

Logo pode-se escrever

$$P(Y_i = y_i | \bar{x}_i) = \prod_{j=0}^J [F(e_j - \beta^T \bar{x}_i) - F(e_{j-1} - \beta^T \bar{x}_i)]^{z_{ij}}$$

Considerando uma amostra aleatória (Y_i, \bar{x}_i) então a Função de Verosimilhança é dada por

$$L = \prod_{i=1}^N P(Y_i = y_i | \bar{x}_i) = \prod_{i=1}^N \prod_{j=0}^J [F(e_j - \beta^T \bar{x}_i) - F(e_{j-1} - \beta^T \bar{x}_i)]^{z_{ij}}$$

Assim o logaritmo da verosimilhança é dado por

$$(1.1.3) \quad \ln(L) = \sum_{i=1}^N \sum_{j=0}^J z_{ij} \times \ln [F(e_j - \beta^T \bar{x}_i) - F(e_{j-1} - \beta^T \bar{x}_i)]$$

O estimador de máxima verosimilhança é obtido maximizando (1.1.3) em relação aos parâmetros desconhecidos $e_j (j=1, \dots, J-1)$ e β .

É de notar ainda, que se o modelo tiver um termo independente β_0 então não existe forma de identificar separadamente este parâmetro e os limites e_j ($j=1, \dots, J-1$). Para se identificar completamente o modelo é necessário impor uma restrição adicional sobre qualquer um destes parâmetros. Tradicionalmente a restrição imposta é $e_0=0$ pelo que, também, foi introduzida neste trabalho.

1.2. INTERPRETAÇÃO DOS PARÂMETROS DO MODELO

Geralmente o objectivo do investigador é estimar o vector de parâmetro β de forma a analisar a relação entre os regressores e a variável latente não observável Y_i^* . Será este tipo de interpretação que será feito neste trabalho.

Tal como salienta Becker & Kennedy (1992), uma vez que não é possível estimar σ então não se pode atribuir nenhum significado à magnitude absoluta dos coeficientes estimados.

Alguns autores como Becker & Kennedy (1992), Greene (1997), Johnston & DiNardo (1997), e Crawford et al.(1998) procuraram mostrar de que forma se deve interpretar os resultados se o objectivo do investigador é estudar a relação entre os regressores e a variável observável Y_i . Estes autores mostraram que embora o impacto de uma alteração nas variáveis explicativas para as probabilidades de escolha de 0 e J tenham um efeito inequívoco, o impacto nas classificações intermédias não pode ser determinado a priori.

Este tipo de estudo não será feito neste trabalho, em primeiro lugar porque não é esse o seu objectivo e, também, porque a análise feita por estes autores só é válida se as variáveis exógenas são contínuas. Como se poderá verificar na secção 2, os regressores incluídos nos modelos estudados neste trabalho são variáveis *dummy* pelo que uma análise deste tipo não teria sentido.

Todavia, Crawford et al.(1998) demonstram algumas proposições com consequências importantes. Estes autores mostraram que ao considerarmos o modelo Logit ordenado estamos a impor certas restrições no modelo que um investigador, que admite esta hipótese não se apercebe. Estas restrições transparecem claramente quando admitimos que os regressores são contínuos e quando estudamos a relação entre os regressores e a variável observável Y_i . Embora, se possa afirmar que estas proposições não são aplicáveis nos casos estudados neste trabalho, dão alguma intuição das limitações que foram auto-impostas quando se admitiu que a distribuição dos erros é Logística.

Para não complicar a notação vai-se ignorar o índice referente à ordem da observação i , considerar $P_j = P(Y = j | \bar{x})$ e x_k o elemento de ordem k do vector \bar{x} então, as proposições deduzidas por estes autores foram as seguintes:

Proposição 1: No modelo Logit ordenado se β_k é positivo² então:

$$a) \quad \partial P_1 / \partial x_k \leq 0$$

$$b) \quad \partial P_j / \partial x_k \begin{matrix} > \\ < \end{matrix} 0 \text{ se e só se } (e_{j-1} - \beta^T \bar{x}) \begin{matrix} > \\ < \end{matrix} (e_j - \beta^T \bar{x})$$

² Se este parâmetro for negativo então as desigualdade seguintes são invertidas.

$$c) \partial P_j / \partial x_k \geq 0$$

Proposição 2: No modelo Logit ordenado se β_k é positivo³ então:

a) Se $\partial P_m / \partial x_k \geq 0$ então $\partial P_n / \partial x_k \geq 0$
para qualquer $n \geq m$, $m=2, \dots, J$;

b) Se $\partial P_m / \partial x_k \leq 0$ então $\partial P_n / \partial x_k \leq 0$
para qualquer $n \leq m$, $m=2, \dots, J$;

Proposição 3: No modelo Logit ordenado onde os elementos x_k são funcionalmente independentes, se o sinal de $\partial P_j / \partial x_r$ muda em \bar{x} , então o sinal de $\partial P_j / \partial x_s$ também muda em \bar{x} para todo o r e s .

Note-se que as proposições enunciadas são válidas em qualquer modelo com variável dependente discreta ordenada, com exceção das proposições 1 b) e 2, que só são válidas se os erros ε_i tem valor esperado zero e a sua função densidade é simétrica e unimodal.

Combinando as proposições 1 e 2 podemos ver que $\partial P_j / \partial x_k$ muda apenas *uma vez* de sinal (à medida que j cresce de 1 para J) de negativo para positivo. É exactamente para esta restrição que é necessário chamar a atenção pois poderá ser muito forte⁴. Os autores mostram um exemplo em que a variável dependente é o grau de instrução das jovens mulheres Australianas e um dos regressores é um índice de atitude que representa a visão sobre as mulheres que trabalham. O modelo Logit ordenado implica que uma alteração do índice de atitude que aumenta a probabilidade associada com um nível de grau de instrução, também aumenta as probabilidades associadas com todos os níveis altos de instrução. Tal como afirmam Crawford et al. (1998), muitos investigadores poderão considerar esta restrição demasiado forte.

A proposição 3 diz-nos que se a derivada muda de sinal para uma variável explicativa num ponto então no mesmo ponto as derivadas em relação às outras variáveis explicativas também mudam de sinal.

1.3. TESTES DE ESPECIFICAÇÃO

Os testes de especificação desempenham um importante papel na avaliação dos modelos econométricos. Antes de utilizarmos um modelo econométrico como base para um estudo é necessário testá-lo exaustivamente. Aliás, Hendry (1980, citado por Godfrey, 1988) afirmou que as três regras de ouro em Econometria são “*testar, testar e testar*”.

³ Se este parâmetro for negativo então as desigualdade seguintes são invertidas.

⁴ É sabido que o modelo Logit ordenado tem subjacente hipóteses diferentes do modelo Logit multinomial (também conhecido por modelo Logit condicional de McFadden) em que a variável dependente observada não é ordenada e o decisor faz a sua escolha de forma determinística maximizando a sua utilidade. Esta restrição é ultrapassada se considerarmos o modelo Logit multinomial. Contudo, na demonstração de Crawford et al. (1998) das proposições 1 b) e 2, foi admitido que os erros têm função densidade simétrica (em relação a zero) e unimodal. Assim uma questão se levanta: Será que estes resultados se mantêm num modelo com variável dependente ordenada mas cuja função densidade dos erros viola estas hipóteses?

Embora os modelos com variável dependente ordenada sejam conhecidos desde a década de 70⁵, só na década de 90 é que houve a preocupação de utilizar testes de especificação neste tipo de modelos.

Em relação ao modelo Probit ordenado já existem alguns artigos dedicados a este tema dos quais se destacam Butler & Chatterjee (1997) que utiliza testes baseados na estimação do modelo por GMM (generalized method of moments) para testar a hipótese de normalidade dos erros e a exogeneidade das variáveis explicativas, Johnson (1996) e Glewwe (1997) que utilizam o teste de score (também conhecido por teste do multiplicador de Lagrange) para aferir da veracidade da hipótese de normalidade dos erros.

No caso do modelo Logit ordenado Murphy (1996) mostrou que, tal como em muitos outros casos, o teste de score para variáveis omitidas, heterocedasticidade, forma funcional incorrecta e assimetria pode ser calculado através de uma regressão artificial⁶. Um dos motivos para a escolha do modelo Logit ordenado foi o facto do autor providenciar um método simples e rápido de calcular a estatística de teste. Em qualquer caso e como já foi afirmado anteriormente, na prática, não existe grande diferença entre o modelo Probit ordenado e Logit ordenado.

O teste consiste em definir

$$G_{ij} = P(\varepsilon_i > e_j - \beta^T \bar{x}_i) = [1 + \exp(e_j - \beta^T \bar{x}_i)]^{-1}$$

$$P_{ij} = P(Y_i = j | \bar{x}_i) = G_{ij-1} - G_{ij}$$

$$u_{ij} = \frac{y_{ij} - P_{ij}}{\sqrt{P_{ij}}} \quad \text{e} \quad v_{ij} = \frac{1}{\sqrt{P_{ij}}} \frac{\partial P_{ij}}{\partial \theta}$$

Em que $\theta^T = (\beta^T, e^T, \gamma^T, \alpha)$, em que e é um vector que inclui os e_j e γ é um vector de parâmetros que são incluídos na expressão da variância quando existe heterocedasticidade $\sigma^2 = \pi^2 f(\bar{w}_i^T \gamma) / 3$, com \bar{w}_i um vector função do vector dos regressores \bar{x}_i e $f(0)=1$ e α é um parâmetro que mede o grau de assimetria da distribuição quando a alternativa a G_{ij} é $H_{ij} = [1 + \exp(e_j - \beta^T \bar{x}_i)]^{-\alpha}$, onde $\alpha > 0$. Se $\alpha < 1$ a cauda da esquerda da distribuição é mais pesada e se $\alpha > 1$ a cauda da direita é mais pesada. Se $\alpha=1$ a distribuição é Logística. Obviamente que o objectivo do investigador é considerar $H_0: \gamma=0$ no caso que queremos testar a existência de homocedasticidade e $H_0: \alpha=1$ quando se pretende testar a simetria.

Neste trabalho estas duas hipóteses foram testadas apenas separadamente. As variáveis utilizadas no vector \bar{w}_i foram $\hat{\beta}^T \bar{x}_i$ e $(\hat{\beta}^T \bar{x}_i)^2$.

⁵ De facto, segundo Greene (1997) o modelo Probit ordenado foi introduzido por McKelvey R. e Zavoina W. em 1975 no *Journal of Mathematical Sociology*, 4, pp. 103-120 com o título "A statistical model of ordinal level dependent variables".

⁶ Os testes de especificação para o modelo Probit ordenado também podem ser calculados através de uma regressão artificial, contudo Murphy (1996) apenas apresentou as fórmulas a utilizar no caso do modelo Logit ordenado.

O teste consiste em estimar o modelo Logit ordenado, calcular u_{ij} e v_{ij} substituindo os parâmetros pelas estimativas e fazer a regressão dos u_{ij} nos v_{ij} ao longo das $J+1$ alternativas e N indivíduos. O estatística do teste score é dada por $\hat{u}^T \hat{u}$ (em que o \hat{u} corresponde ao vector dos valores ajustados desta regressão auxiliar)⁷. A distribuição assintótica desta estatística é a Qui-Quadrado com os graus de liberdade iguais ao número de restrições que se pretende testar. Assim, no teste de heterocedasticidade a distribuição é $\chi^2(2)$ e no teste de simetria a distribuição é $\chi^2(1)$.

Obviamente que para fazer este teste é necessário calcular $\frac{\partial P_{ij}}{\partial \theta}$. As fórmulas destas derivadas no caso do modelo Logit ordenado podem ser vistas em Murphy (1996).

Note-se que os testes de especificação utilizados no modelo Logit simples⁸ podem ser aplicados para aferir da qualidade destes modelos Logit ordenados. Isto é verdade porque os testes de especificação são feitos de modo a investigar as propriedades dos erros em (1.1.1). Note-se que podemos estimar os parâmetros do modelo através de um modelo Logit simples sem alterar a equação (1.1.1). Basta considerar um número inteiro qualquer z , com $0 < z < J$, definir uma variável auxiliar

$$Y_i' = \begin{cases} 0 & \text{se } Y_i \in \{0, 1, 2, \dots, z\} \\ 1 & \text{se } Y_i \in \{z+1, z+2, \dots, J\} \end{cases}$$

e tomar Y_i' como variável dependente no modelo Logit simples. Uma vez que o teste de score anteriormente descrito testa a heterocedasticidade e simetria no modelo o único teste que tem interesse neste contexto é um teste para aferir se a distribuição dos erros é Logística. Este teste é um caso particular do teste da matriz de informação. Na prática consiste em acrescentar no modelo $(1 - 2 \times \hat{p}_i)$ com

$$\hat{p}_i = \frac{\exp(\hat{\beta}^T \bar{x}_i)}{1 + \exp(\hat{\beta}^T \bar{x}_i)}$$

e analisar o rácio t do parâmetro referente a esta variável. Este teste será designado neste trabalho por TBMLS de “Teste Baseado no Modelo Logit Simples”.

Para verificar se a forma funcional na equação (1.1.1) era correcta, utilizou-se o teste RESET de Ramsey. Este teste assenta no pressuposto de que se a forma funcional de um modelo 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 introduzir no modelo para além de todos os regressores a variável $(\hat{\beta}^T \bar{x}_i)^2$, e em seguida realizar um teste t ao seu coeficiente.

⁷ Esta estatística é assintoticamente equivalente a $NJ \times R^2$ da regressão auxiliar.

⁸ Este resultado também é válido nos modelos Probit.

2. MODELOS DE AVALIAÇÃO DA SATISFAÇÃO DOS TRABALHADORES

2.1. O INSTRUMENTO DE MEDIDA

O questionário utilizado neste trabalho foi construído e aferido para Portugal por uma empresa de consultoria sediada em Chicago, a *International Survey Research (I.S.R)* fundada em 1974, que constrói e implementa *surveys* para companhias Multinacionais a nível mundial.

Este questionário, denominado por B.E.S.T (*Benchmarking employee satisfaction trends*), tem vindo a ser aplicado na Europa nomeadamente no Reino Unido a partir de 1975 expandindo-se posteriormente a outros países europeus. Recentemente este instrumento está a ser aplicado em filiais de Multinacionais em Portugal. O questionário têm como objectivo avaliar as percepções dos membros da organização numa escala tipo Likert em 5 pontos cujo eixo vai de “Concordo” a “Discordo”, sendo constituído por 90 itens definidos como frases descritivas da organização, em que o trabalhador se vai posicionar.

O questionário utilizado é constituído pelos seguintes domínios:

- | | |
|----------------------------------|---|
| - <i>Organização do trabalho</i> | - <i>Eficiência Operacional</i> |
| - <i>Gestão</i> | - <i>Formação profissional e informação</i> |
| - <i>Segurança no trabalho</i> | - <i>Supervisão</i> |
| - <i>Comunicações</i> | - <i>Performance e desenvolvimento</i> |
| - <i>Remuneração</i> | - <i>Ética da companhia</i> |
| - <i>Satisfação no trabalho</i> | - <i>Imagem da Companhia</i> |
| - <i>Condições de trabalho</i> | - <i>Relações laborais</i> |
| - <i>Qualidade</i> | - <i>Envolvimento</i> |
| - <i>Benefícios</i> | - <i>Liderança e direcção</i> |

2.2 A AMOSTRA

Considerou-se como população os trabalhadores da Direcção de Pessoal de uma grande empresa de transportes da area Metropolitana de Lisboa. A amostra é constituída pelos trabalhadores da Direcção de Pessoal da empresa em 22 de Fevereiro de 1999.

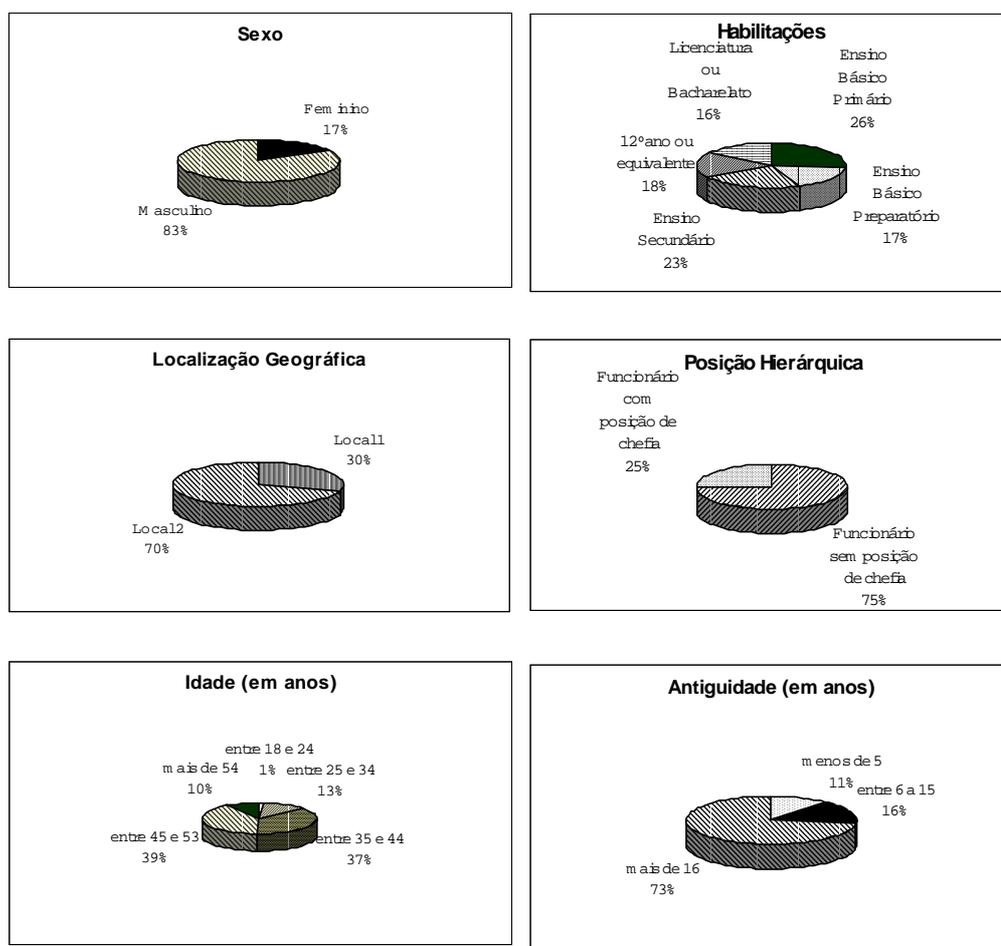
O período de entrega e recolha do questionário decorreu de 22 de Fevereiro de 1999 a 5 de Março de 1999. A Direcção de Pessoal da empresa em 22 de Fevereiro de 1999 era constituída por 130 trabalhadores. Uma vez que não foi possível contactar com todos os trabalhadores, foram entregues 124 questionários e destes obtiveram-se 101 respostas e 23 não respostas.

Os inquiridos foram questionados relativamente a algumas das suas características nomeadamente quanto ao sexo, idade, habilitações, antiguidade na empresa, posição hierárquica e uma variável específica relacionada com a localização geográfica do trabalho do inquirido. Esta variável foi introduzida uma vez que a Direcção de Pessoal da empresa está

localizada em dois locais geograficamente distintos no distrito de Lisboa que serão designados por Local 1 e Local 2.

São exactamente estas variáveis que vão ser utilizadas como regressores nos modelos que são apresentados neste trabalho.

De forma a ter uma maior percepção sobre quais são as características dos indivíduos que constituem a amostra são apresentados alguns gráficos.



2.3. APRESENTAÇÃO DAS VARIÁVEIS

As variáveis dependentes dos modelos que são apresentados correspondem às avaliações que foram obtidas para cada item. Dos 90 itens que constituem o inquérito foram estudados 12 e destes são apenas apresentados os resultados referentes a 3. A escolha dos 12 itens que se estudou inicialmente baseou-se apenas na opinião do autor sobre a relevância dos mesmos, e após a estimação dos vários modelos, escolheu-se os 3 itens a partir dos quais se podiam tirar o máximo de conclusões. Estes 3 itens são os seguintes:

- A. As prioridades ou objectivos de trabalho são mudados com tanta frequência que tenho dificuldade em fazer o meu trabalho.**
- B. O meu trabalho dá-me poucas oportunidades de pôr em prática os meus conhecimentos e capacidades.**
- C. O meu trabalho dá-me realização pessoal.**

Poderá parecer ao leitor que os itens B e C estão relacionados, pelo que pelo menos um destes é redundante. Contudo, como se irá ver na subsecção 2.4. as conclusões que se tira dos modelos em que estes itens correspondem a variáveis dependentes são ligeiramente diferentes.

Como já se afirmou anteriormente os regressores utilizados na estimação dos modelos correspondem às características dos indivíduos. Uma vez que a informação obtida sobre as características dos indivíduos é do tipo discreto (mesmo a informação relativa à idade e à antiguidade foi, apenas, solicitado aos inquiridos que indicassem em que intervalo a sua idade ou antiguidade se situava) os regressores dos modelos estudados são variáveis *dummy*.

Note-se que os coeficientes de uma variável *dummy* devem ser interpretados como um indicador da forma como comportamento de uma categoria se desvia de uma base. Essa base corresponde à categoria que é omitida (se esta categoria fosse incluída existiria multicolinearidade perfeita)

Assim em relação ao sexo foi definida a seguinte variável:

$$S_i = \begin{cases} 1 & \text{se } i \text{ é do sexo masculino} \\ 0 & \text{se } i \text{ é do sexo feminino} \end{cases}$$

Neste caso a base é o indivíduo do sexo feminino.

Quanto à idade foram construídas as seguintes variáveis artificiais:

$$I_{1i} = \begin{cases} 1 & \text{se } i \text{ tem entre 18 e 24 anos} \\ 0 & \text{caso contrário} \end{cases} \quad I_{2i} = \begin{cases} 1 & \text{se } i \text{ tem entre 25 e 34 anos} \\ 0 & \text{caso contrário} \end{cases}$$

$$I_{3i} = \begin{cases} 1 & \text{se } i \text{ tem entre 35 e 44 anos} \\ 0 & \text{caso contrário} \end{cases} \quad I_{4i} = \begin{cases} 1 & \text{se } i \text{ tem entre 45 e 54 anos} \\ 0 & \text{caso contrário} \end{cases}$$

A base corresponde ao indivíduo com 55 anos ou mais.

No que respeita à habilitação foram criadas as seguintes variáveis:

$$H_{1i} = \begin{cases} 1 & \text{se } i \text{ tem Ensino Básico Primário} \\ 0 & \text{caso contrário} \end{cases} \quad H_{2i} = \begin{cases} 1 & \text{se } i \text{ tem Ensino Básico Preparatório} \\ 0 & \text{caso contrário} \end{cases}$$

$$H_{3i} = \begin{cases} 1 & \text{se } i \text{ tem Ensino Secundário} \\ 0 & \text{caso contrário} \end{cases} \quad H_{4i} = \begin{cases} 1 & \text{se } i \text{ tem 12º ano ou equivalente} \\ 0 & \text{caso contrário} \end{cases}$$

A base é o indivíduo com Licenciatura ou Bacharelato.

Em relação à antiguidade foram definidas as seguintes variáveis:

$$A_{1i} = \begin{cases} 1 & \text{se } i \text{ tem menos de 5 anos} \\ 0 & \text{caso contrário} \end{cases} \quad A_{2i} = \begin{cases} 1 & \text{se } i \text{ tem entre 6 e 15 anos} \\ 0 & \text{caso contrário} \end{cases}$$

A base corresponde ao indivíduo com mais de 16 anos de antiguidade na empresa.

A posição hierárquica de cada indivíduo foi introduzida no modelo através da seguinte variável *dummy*:

$$HQ_i = \begin{cases} 1 & \text{se } i \text{ é um funcionário com posição de chefia} \\ 0 & \text{se } i \text{ não é um funcionário sem posição de chefia} \end{cases}$$

A base é o funcionário sem posição de chefia.

Por fim, foi criada uma variável que indica em que local trabalha um indivíduo:

$$L_i = \begin{cases} 1 & \text{se } i \text{ trabalha no Local 1} \\ 0 & \text{se } i \text{ trabalha no Local 2} \end{cases}$$

Neste caso a base é constituída pelo indivíduo que trabalha no Local 2.

Quando se trabalha com várias variáveis *dummy* que correspondem a uma só característica dos indivíduos (como por exemplo as variáveis relacionadas com idade, habilitações e antiguidade) por vezes é dito que não se deve fazer o teste de significância de uma só destas variáveis com o argumento que se mudássemos de base (a variável omitida) esse parâmetro passaria a ser significativo. Assim, neste caso só o teste F de significância conjunta das variáveis *dummy* referentes a essa característica é que teria sentido.

Contudo, os testes individuais de significância dos parâmetros poderão, também, ter sentido. Como já foi afirmado anteriormente cada parâmetro dá o desvio em relação a uma base. Se através do rácio *t* chegarmos à conclusão que, por exemplo o parâmetro relativamente à variável I_{1i} não é significativo então é porque não existe diferença de opinião entre os indivíduos com idade entre 18 e 24 anos e os indivíduos com mais de 54 anos (que correspondem à base). Obviamente se mudarmos a base e considerarmos que esta corresponde aos indivíduos com idade entre 45 e 54 anos, o coeficiente da variável I_{1i} poderá ser significativo mostrando que existe diferenças de opinião entre os indivíduos com idade entre 18 e 24 anos e os indivíduos com idade entre 45 e 54 anos. Note-se que quando mudamos a base o teste de significância do parâmetro não tem o mesmo sentido de que quando utilizamos a base antiga. Assim, é correcto fazer os testes de significância individual a este tipo de variáveis *dummy*.

Por último, é de referir que todos os modelos têm uma constante pelo que considerou-se $e_0=0$.

2.4. APRESENTAÇÃO E INTERPRETAÇÃO DOS RESULTADOS EMPÍRICOS

A estratégia utilizada na construção dos modelos consistiu na inclusão de todas as variáveis *dummy*, sendo o passo seguinte a identificação das que eram significativas em termos estatísticos.

O software para estimação do modelo foi o TSP (versão 4.5 - Hall & Cummins, 1999) e foi utilizado o método de Newton tanto para a maximização da Função de Verosimilhança como para a aproximação da matriz Hesseana.

Em relação aos testes de especificação, estes mostraram que os modelos se comportavam bem, não existindo problemas de heterocedasticidade, simetria e forma funcional pelo que o procedimento utilizado para eliminar as variáveis revelou-se sempre correcto.

Em alguns modelos, verificou-se no início alguns problemas de “Separação quase-completa”⁹ devido, principalmente, às variáveis *dummy* relacionadas com a idade. O conceito de “Separação quase-completa” foi introduzido por Albert & Anderson (1984) e consiste no facto das variáveis independentes preverem “quase” completamente a variável dependente. Neste caso a estimação por Máxima Verosimilhança não irá convergir e as estimativas dos parâmetros irão divergir para mais/menos infinito. Albert & Anderson (1984) referem, ainda, o seguinte:

“The difficulties associated with complete and quasicomplete separation are small sample problems. With increasing sample size the probability of observing a set of separated data points tends to zero, no matter the sampling scheme. Complete separation may occur with any type of data, but is unlikely that quasicomplete separation will occur with truly continuous data.”

Uma vez que a dimensão da amostra é de 101 observações e que os regressores dos modelos estudados são variáveis *dummy* então não é de estranhar que este problema tenha ocorrido. Sempre que surgiu esse problema as variáveis que o causavam foram retiradas dos modelos.

Os resultados dos modelos estimados foram os seguintes:

Modelo 1

Variável dependente: Item A

Nº de observações: 101

	Variáveis que correspondem Aos parâmetros estimados						Limites dos intervalos		
	Constante	L	H ₁	H ₃	I ₁	HQ	e ₃	e ₂	e ₁
Estimativas	3,4090	-0,9604	-1,2461	-1,3476	-2,7694	1,0874	3,2997	1,6392	1,5334
Desvio Padrão	0,5316	0,4504	0,4936	0,4805	1,6240	0,4812	0,3891	0,3093	0,3034
Estatística t	6,4129	-2,1322	-2,5243	-2,8047	-1,7053	2,2599	8,4797	5,3004	5,0538

Estatística do teste de Wald 23,2793 p-value 0,0003

⁹ Tradução do inglês “Quasicomplete separation”.

Modelo 2

Variável dependente: Item B

Nº de observações: 101

	Variáveis que correspondem aos parâmetros estimados			limites dos intervalos		
	Constante	HQ	A ₂	e ₃	e ₂	e ₁
<i>Estimativas</i>	1,7387	0,9200	0,9419	3,2884	2,1155	2,0295
<i>Desvio Padrão</i>	0,3213	0,4259	0,5273	0,3625	0,3107	0,3071
<i>Estatística t</i>	5,4118	2,1599	1,7862	9,0704	6,8086	6,6080

Estatística do teste de Wald 7,403955 *p-value* 0,0247

Modelo 3

Variável dependente: Item C

Nº de observações: 101

	Variáveis que correspondem aos parâmetros estimados			limites dos intervalos		
	Constante	L	HQ	e ₃	e ₂	e ₁
<i>Estimativas</i>	-0,5099	1,5967	-1,4577	3,5639	1,8609	1,7970
<i>Desvio Padrão</i>	0,4156	0,4730	0,5204	0,4897	0,2775	0,2723
<i>Estatística t</i>	-1,2270	3,3759	-2,8014	7,2779	6,7059	6,5995

Estatística do teste de Wald 18,2496 *p-value* 0,0001

Os resultados dos testes de especificação destes modelos são apresentados no seguinte quadro:

Testes de Especificação

<i>Variável dependente</i>	<i>Modelo 1</i>		<i>Modelo 2</i>		<i>Modelo 3</i>		<i>Distribuição da estatística</i>
	<i>Item A</i>		<i>item B</i>		<i>item C</i>		
	<i>Estatísticas de teste</i>	<i>Valor</i>	<i>p-value</i>	<i>Valor</i>	<i>p-value</i>	<i>valor</i>	<i>p-value</i>
<i>Heterocedasticidade</i>	1,594	0,451	2,551	0,279	1,255	0,534	$\chi^2(2)$
<i>Simetria</i>	1,286	0,257	0,732	0,694	0,244	0,621	$\chi^2(1)$
<i>TBMLS</i>	a)		-0,261	0,794	-1,006	0,315	$N(0,1)$
<i>RESET</i>	-0,411	0,681	-0,633	0,526	1,028	0,304	$N(0,1)$

a) Neste caso não foi possível fazer este teste devido ao problema de "Separação quase-completa" no modelo binário.

O teste de Wald indicado tem como hipótese nula a igualdade de todos os parâmetros do modelo a zero, com excepção do termo independente e dos limites dos intervalos. A distribuição da estatística de teste neste caso é a Qui-Quadrado com o número de graus de liberdade igual ao número restrições que se está a testar¹⁰.

Quando à interpretação dos parâmetros é necessário primeiro perceber claramente qual o significado das variáveis latentes não observáveis implícitas nestes modelos.

¹⁰ Nos Modelos 2 e 3 o número de graus de liberdade é 2 e no modelo 1 o número de graus de liberdade é 5.

Note-se que a variável dependente observável nestes modelos é uma variável ordenada que toma os seguintes valores consoante a resposta do inquirido:

- 1 - Concorda;
- 2 - Tende a concordar;
- 3 - Sem opinião;
- 4 - Tende a discordar;
- 5 - Discorda.

Por outro lado, existem perguntas que estão elaboradas de forma positiva (é o caso do item C) e outras elaboradas de forma negativa (é o caso do item A e B).

Tendo em atenção estes dois factos então podemos atribuir o seguintes significado às variáveis latentes não observáveis:

Modelo 1- <i>Índice de satisfação com o ritmo com que são mudadas as prioridades e os objectivos de trabalho.</i>
Modelo 2- <i>Índice de satisfação com as oportunidades de pôr em prática os conhecimentos do trabalhador.</i>
Modelo 3- <i>Índice de insatisfação com o nível de realização pessoal obtido pelo trabalhador.</i>

Um índice de insatisfação significa que quanto maior é essa variável maior é a insatisfação e um índice de satisfação indica que quanto maior é essa variável maior é a satisfação.

Uma primeira conclusão que se pode tirar ao analisar todos os modelos é que os índices de satisfação ou insatisfação dos indivíduos não se alteram com o sexo. Este facto poderá indicar que ambos os sexos são tratados na empresa de igual modo. Outra interpretação possível deste resultado é a de que o peso dos indivíduos do sexo masculino no Departamento de Pessoal é tão grande que poderá estar a influenciar a falta de discriminação entre sexos na empresa.

A análise dos vários modelos, também permite concluir que os funcionários com posição de chefia apresentam uma maior satisfação que os restantes. Este resultado pode ter duas interpretações. Em primeiro lugar, pode estar relacionado com o facto de existir uma grande proximidade orgânica entre a Administração da empresa e os funcionários com posição de chefia, pelo que é natural que estes não queiram manifestar algum desagrado face às tomadas de decisão que afectam o seu nível de satisfação. Por outro lado, os trabalhadores com posição de chefia, também tomam parte em decisões no que diz respeito à gestão na empresa pelo que apresentarão sempre respostas mais favoráveis que os outros trabalhadores.

Pode-se, também, ver que na maioria dos modelos não existem diferenças de opinião sobre satisfação entre indivíduos cuja idade ou antiguidade é diferente contudo existem algumas excepções.

No Modelo 1 o coeficiente de I_1 é negativo, notando que só existe um indivíduo com idade inferior a 24 anos na Direcção de Pessoal da empresa, então este facto poderá indicar que este indivíduo está a ter maior dificuldade que os restantes em suportar o ritmo com que

são mudadas as prioridades e os objectivos da Companhia. Este facto é natural devido à sua inexperiência.

A partir da análise do Modelo 2 conclui-se que os indivíduos com antiguidade entre 6 e 15 anos revelam uma maior satisfação que os indivíduos com maior ou menor antiguidade.

Ao analisar o Modelo 1 podemos ver que são os indivíduos com Ensino Básico Primário e os indivíduos com Ensino Secundário que revelam maior dificuldade em acompanhar o ritmo com que são alteradas as prioridades e os objectivos de trabalho comparativamente com os outros trabalhadores. A justificação deste facto estará no baixo nível de formação destes trabalhadores. Note-se que a razão deste resultado não ter surgido para os indivíduos com Ensino Básico Preparatório poderá estar relacionada com o tipo de tarefas desempenhadas pelos trabalhadores com este nível de habilitações.

Verifica-se que os trabalhadores do Local 1 revelaram uma maior insatisfação que os trabalhadores de Local 2 sobre o ritmo com que são mudadas as prioridades e os objectivos da empresa (Modelo 1). Este resultado deverá ficar a dever-se ao facto de se ter processado uma reestruturação interna no Departamento de Pessoal no Local 1 no período em que foi realizado o inquérito.

3. CONCLUSÕES

Procurou-se com este trabalho verificar de que forma o nível de satisfação dos trabalhadores varia com as suas próprias características.

Numa síntese global dos resultados obtidos pode-se concluir que existe no Departamento de Pessoal da empresa estudada algum consenso relativamente à satisfação dos trabalhadores nomeadamente no que diz respeito a características relacionadas com o sexo, idade e antiguidade. No que respeita à idade, o único exemplo que viola este consenso corresponde ao único indivíduo com idade inferior a 24 anos que mostra maior insatisfação relativamente ao ritmo com que a organização muda os seus objectivos e prioridades. Em relação à antiguidade existe também uma excepção ao consenso evidenciado, nomeadamente os indivíduos com antiguidade entre 6 e 15 anos mostram uma maior satisfação com as oportunidades para pôr em prática os seus conhecimentos.

Contudo, verificou-se a existência de discordância de opiniões relativamente à satisfação principalmente entre níveis hierárquicos diferentes, indivíduos com habilitações diferentes (embora este resultado só tenha surgido no modelo 1, em relação ao modelos expostos, verificou-se ser verdadeiro em outros modelos, que não foram apresentados) e trabalhadores que desempenham a sua actividade em locais diferentes.

Em relação à metodologia utilizada, é sabido que não existem grandes diferenças entre o modelo Probit ordenado e Logit ordenado pelo que a opção por este último não se revela controversa. Os resultados dos testes de especificação suportam esta ideia.

Note-se que com excepção do sexo e do local todas as variáveis independentes dos modelos podem ser consideradas ordenadas. Assim, se ignorarmos as duas variáveis referidas teríamos um modelo em que todas as variáveis seriam discretas e ordenadas (tanto a dependente como as independentes). Ronning & Kukuk (1996) mostram como estimar este tipo de modelos admitindo que subjacente a ele existe um modelo linear em que a tanto a variável endógena como os regressores, são não observáveis. Todavia, estes autores para estimar o modelo admitem uma hipótese muito forte, nomeadamente a Normalidade Multivariada entre todas as variáveis não observáveis. Por este motivo o método proposto por

estes autores não foi utilizado. Talvez, no futuro seja criado um método mais geral que permita deixar de lado uma hipótese tão restrita como a Normalidade Multivariada.

Estes modelos com variável dependente ordenada, apesar de serem conhecidos desde a década de 70, não têm tido grandes desenvolvimentos desde então, constituindo, assim, uma área onde, porventura, existirá muito por desvendar.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBERT, A. & ANDERSON J. (1984), On the existence of maximum likelihood estimates in logistic regression models, *Biometrika*, 71,1, pp.1-10;
- BECKER, W. & KENNEDY, P. (1992), A Graphical Exposition of the Ordered Probit, *Econometric Theory*, 8, pp.127-131;
- BUTLER, J. & CHATTERJEE, P. (1997), Tests of Specification of univariate and bivariate ordered probit, *The Review of Economics and Statistics*, 79 (May), pp. 343-347;
- CRAWFORD, D., POLLAK, R. & VELLA, F.(1998), Simple Inference in Multinomial and Ordered Logit, *Econometric Reviews*, 17(3), pp. 289-299;
- GLEWWE, P. (1997), A Test of the Normality assumption in the Ordered Probit, *Econometric Reviews*, 16(1), pp. 1-19;
- GODFREY, L.G. (1988), *Misspecification Tests in Econometrics*, Econometric Society Monographs;
- GREENE, W. (1997), *Econometric Analysis*, Third Edition, Prentice Hall International Editions;
- HALL, B.H. & CUMMINS C. (1999), Time Series Processor, Version 4.5, Reference Manual, TSP International;
- JOHNSON, P. (1996), A Test of the Normality Assumption in the Ordered Probit Model, *Vassar College - Working Paper No. 34*, (Documento WWW)
<http://noether.edu:80/econ/wpseries/wp34.pdf>;
- JOHNSTON, J. & DINARDO, J. (1997), *Econometric Methods*, Fourth Edition, McGraw-Hill International Editions;
- MURPHY, A. (1996), Simple LM tests of mis-specification for ordered logit models, *Economics Letters*, 52, pp. 137-141;
- RONNING, G. & KUKUK, M. (1996), Efficient Estimation of Ordered Probit, *Journal of the American Statistical Association*, Vol. 91, No. 435, pp. 1120-1129.