Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

Mestrado em ECONOMETRIA APLICADA E PREVISÃO

ECONOMETRIC ANALYSIS OF PRIVATE MEDICINES EXPENDITURE IN PORTUGAL

MÓNICA SOFIA INÁCIO DUARTE INÊS

Orientação

NICOLETTA ROSATI

Constituição do Júri

JOSÉ M. M. PASSOS **(Presidente)**

MIGUEL R. A. GOUVEIA **(Vogal)**

NICOLETTA ROSATI **(Vogal)**

Junho de 2007

Resumo

O Serviço Nacional de Saúde Português estabelece que o acesso a cuidados de saúde deve depender essencialmente das necessidades clínicas. Condicionado nas necessidades individuais, o acesso e utilização de medicamentos não deveria depender de factores económicos como rendimento, classe social, nível de educação ou o acesso a farmácias ou postos de vendas de medicamentos. Utilizando dados das últimas duas realizações do Inquérito Nacional de Saúde (1995/96 e 1998/1999), este estudo testa a existência de inequidades nas despesas com medicamentos, condicionadas na necessidade, relacionadas com o rendimento, com a densidade de farmácias e com possuir seguro de saúde privado ou relacionado com o local de trabalho. Foi aplicado um modelo em duas partes. Para a probabilidade individual de efectuar despesas com medicamentos, foi adoptado um estimador LOGIT modificado para acomodar a dupla natureza dos zeros da variável dependente e que permitisse assimetria. Para modelar as despesas positivas com medicamentos foram utilizadas as propriedades da pseudo verosimilhança através da utilização de um modelo de Poisson. Não se detectou má especificação do modelo em duas partes e concluiu-se que existem inequidades na despesa privada com medicamentos relacionadas com a existência de seguro de saúde privado ou relacionado com o local de trabalho, o rendimento e a densidade de farmácias.

Classificação JEL: C21, C5, I18

Palavras-chave: Inquérito Nacional de Saúde - Portugal, Medicamentos, Despesa, Modelo em duas partes; Poisson pseudo máxima verosimilhança.

Abstract

The Portuguese National Health System states that access to health care should depend mainly on need. Conditional on need, access to pharmaceuticals should not depend on socio-economic factors such as income, social class, education or geographical factors such as the access to pharmacies. This study uses data from the last two waves of National Health Survey (1995/1996 and 1998/1999) and focuses on equity issues testing for the existence of insurance inequalities, income-related and pharmacies density related inequalities. A two-part model was adopted. To model the probability of occurrence of medicines private expenditure, a modified LOGIT model was specified accounting for the double nature of the zeros of the dependent variable and asymmetry. In the second part a Poisson pseudo maximum likelihood estimator was adopted. No misspecification was detected in the two-part model. The main results showed inequity in Portuguese private medicines expenditures with respect to supplementary health insurance (private and job related), income and pharmacies density.

JEL classification: C21, C5, I18

Keywords: Portuguese National Health Survey, Medicines, Expenditure, Two-part model, Poisson pseudo maximum likelihood estimator.

INDEX

TABLE INDEX

FIGURE INDEX

ABREVIATIONS AND DEFINITIONS

Abbreviation	Definition
2PM	Two-part model
ADSE	Insurance plan that covers Portuguese civil servants
APE	Average partial effect
BMI	Body mass index
CAR	Coarseness at random
EU	European Union
GPML	Gamma pseudo-maximum-likelihood
ICD-9-CM	International classification of diseases, 9th Revision, clinical modification
INE	Portuguese national statistical institute
INFARMED	Portuguese regulatory agency for medicines
JRHS	Job-related health system
LAU	Local administrative units (for statistics)
LPM	Linear probability model
MAR	Missing at random
NHS	National health system
NLS	Non-linear least squares
NUTS	Territorial units for statistics
OECD	Organisation for economic co-operation and development
OLS	Ordinary least squares
OTC	‘Over the counter’ medicine
PHI	Private Health Insurance
PNHS	Portuguese national health survey
PPML	Poisson pseudo-maximum-likelihood
SAMS	Insurance plan that covers Portuguese bank employees
sd	Standard deviation
WHO	World Health Organization
ZIM	Zero inflated models
ZIP	Zero inflated Poisson model

ACKNOWLEDGEMENTS

Access to Portuguese National Health Survey is gratefully acknowledged. Portuguese National Health Survey explanations were kindly provided by Carlos Dias (ONSA - National Observatory of Health).

Special thanks are due to Prof. Nicoletta Rosati for the dedicated supervision and to Prof. Montezuma Dumangane for the clarification of some specification issues. Thanks are also due to Félix Bernardo, Filipa Duarte-Ramos, Kathrin Schneider, Marta Soares, Pedro Saramago, Teresa Bago d'Uva and Tiago Lopes for their constructive comments and suggestions. Finally, I am grateful to my family for their unconditional support.

This research has also benefited from participation in the “Applied Health Economics” course under the Marie Curie Training Program - June 2006, hosted by the Health, Econometrics and Data Group/University of York. Attendance, travel and accommodation expenses were funded by the European Union.

Previous analysis were presented under the title “Pseudo-Panel Data Framework: Evidence from the National Health Survey – Portugal” at the 22nd International Conference on Pharmacoepidemiology & Therapeutic Risk Management held in Lisbon, Portugal, August 2006.

All errors, omissions and opinions are sole responsibility of the author.

Comme un fou se croit Dieu,
nous nous croyons mortels.

- Delalande, Discours sur les ombres

1. INTRODUCTION

The Portuguese pharmaceutical sector is currently undergoing a legislative reform in the supply and financing of medicines. The sustainability of the Portuguese pharmaceutical sector is endangered by increased access to health care and supply of services, and by the increasing costs of the health resources. The political imperatives underneath these changes are wide ranging and include the ultimate desire to sustain public expenditure on medicines while benefiting from higher efficacy/safety for more recent and expensive medicines. It is also believed that the cost-sharing formula between patients and the National Health System (NHS) should be improved in order to deliver optimal utilisation of medicines, more according to individual health needs. Recent government decisions point at the pharmaceutical market as a primary target of policy control/regulation aiming at a reduction in health care expenditure. Facing these difficulties, prices have been reduced and NHS cost-sharing formula with patients has been modified recently.

The Portuguese National Health System states that access to health care should depend on need. Conditional on need, the utilization of medicines should not depend on socio-economic factors such as income, social class, education, or geographical factors such as the access to pharmacies and medical doctors. There is horizontal inequity when health care utilisation is shaped by *non-need variables* and individuals with the same needs consume different amounts of health care. Despite the current Portuguese political focus on medicines utilisation and expenditure, the issue of equity has received little attention in the literature. Recent studies, under the Ecuity Project, have indicated Portugal as being one of the countries with the highest level of inequalities in health care utilisation (Van Doorslaer, 2004; Van Doorslaer, 2004a). Lourenço and Ferreira (2004) and Lourenço *et al.* (2006) have also recently devoted research to the issue of equity in Portuguese healthcare utilization. However, there is no empirical evidence about the factors that establish Portuguese individual private expenditure on medicines and if there is still scope for inequality. Hence, studies focused on the expenditure on medicines are of great importance to identify the determinants of private expenditure and detect the existence of inequity.

This study has two main motivations.

1. The first motivation is to contribute to current knowledge, within the current Portuguese Health System reform process, through the study of private medicines expenditure in Portugal, using the most recent data available from Portuguese National Health Survey (PNHS). Within this evaluation it is also possible to account for the supply of health services, in particular, density of pharmacies, at area of residence level, a factor that is often excluded from analyses mainly due to the lack of data, and has increased importance under current Portuguese pharmaceutical market reforms. Alongside income-related inequity and geographical-related inequity are estimated. To our knowledge, it is the first study to model Portuguese individual private expenditure with medicines.

2. The second motivation is clearly analytical, revisiting an economist's old and increasingly frequent problem: how to obtain consistent estimates, under unknown forms of heterogeneity. Nonlinear models are common in health economics since measures of outcomes are often based on limited dependent variables. This research faces the challenge of adopting the most appropriate econometric specification in order to avoid inconsistent estimators of the model parameters, aware that the extent of the inconsistencies will depend both on the specific characteristics of the sample and the model adopted.

Summarizing, it is a major research purpose to make contributions to Portuguese literature on equity and enlighten the use of adequate econometric specifications, valid and useful to the analysis of health care expense determinants.

The remainder of the document is organized as follows: the next section describes the main characteristics of the Portuguese Health Care System giving focus to the pharmaceutical market. Section 3 describes the Portuguese National Health Survey, the type of information it contains and the variable definitions. Section 4 contains the empirical analysis: econometric model specification and estimation results. Finally, section 5 contains some concluding remarks.

2. PORTUGUESE HEALTH CARE SYSTEM

The Portuguese National Health System (NHS), established in 1979 (Law 56/1979), constituted a crucial step towards the pursuit of equity, introducing the universal coverage for health care, nearly free[1], financed by general taxation. The Portuguese government statement of equity is compatible with the principle of ''equal opportunity for those in equal need'' (Pereira, 1990).

Most of Portuguese NHS patients are registered with a general practitioner, who can diagnose diseases, prescribe medicines and manage chronic conditions. General practitioners acts like 'gatekeeper of the system' and only prescribed medicines are reimbursed by the NHS. Since 1994, the NHS also reimburses prescriptions from private doctors. Medicines private expenditure is deductible up to 30% under the Portuguese Taxation System.

The pharmaceutical market is complex and has some country-specific features. This chapter gives some insights on Portuguese regulation, expenditure and financing of medicines in order to enlighten the underlying process that generates individual private expenditure on medicines.

2.1. REGULATION OF THE PHARMACEUTICAL MARKET

The economic regulation of the pharmaceutical market in Portugal focuses mainly in distribution, prices and reimbursement.

Distribution

[1] The Portuguese Constitution was modified since 1979 and the expression "free" was replaced by "nearly free".

Drugs restricted to medical prescription are only sold in pharmacies but in 2005 the government has approved the sales of ‘over the counter’ (OTC[2]) medicines in other licensed establishments such as supermarkets special areas and health related commercial areas (Law 134/2005).

Ownership of pharmacies (for restricted medical prescription drugs) is limited to pharmacists, and licensing of new establishments is issued by INFARMED (Portuguese regulatory agency for medicines) based on a maximum number of pharmacies per inhabitants, being the geographic distance to existing pharmacies also a criteria.

However, the Portuguese government is at the moment undergoing reforms to liberalize ownership of pharmacies and increase the current number of pharmacies by 300 new units (Portuguese Government, 2006) and extend the pharmacies opening hours from 40 to 55 hours *per* week (Law 53/2007).

Prices

The revenue of pharmacies and wholesalers in the sale of prescription medicines (as proportion of the market price) is limited and regulated by the Ministry of Economy, as well as the maximum market price authorized, bounded by the minimum of the prices established in Spain, Italy and France (Law 29/90).

Recently, the medicines price mechanism suffered changes with the inclusion of Greece in the group of reference price countries and the use of the average price to set the upper bound of Portuguese price (Law 65/2007).

Nevertheless, the government regulated a compulsory global down pricing of 6% in September 2005 (Law 618-A/2005) and a second global down pricing of 6% by

[2] OTC: medicines available without prescription. OTC medicines reimbursed by NHS are only distributed in pharmacies.

January 2007 (Law 30-B/2007 and Law 2496/2007), which was justified with the sustainability of global Portuguese expenditure with medicines.

Reimbursement

Once the Ministry of Economy has established the maximum market price, NHS reimbursement is discussed at Ministry of Health/INFARMED who can require an economic evaluation study of the medicine (Law 19064/1999) in order to allow a more informed decision.

In some circumstances, NHS reimbursement can be conditioned by the total number of individuals to be treated or reimbursement only occurs if the medicine Portuguese market price is reduced. A positive decision from INFARMED will make the medicine drug reimbursed by the NHS.

National Health System

The medicines reimbursement scheme of the Portuguese National Health System is essentially structured on the therapeutic indications of each drug medicine (Law 1474/2004), yielding five basic levels[3] of reimbursement: 100, 95, 70, 40 and 20 percent.

The 95% reimbursement level was introduced in 2005, when the 100% reimbursement level was shortened in order to keep only life-saving medicines. The 20% reimbursement level is defined as transitory (Law 205/2000) and it was introduced in 2000 with the aim of accommodate drug medicines with less efficacy/effectiveness evidence.

[3] The Reference Price System (Law 270/2002), implemented since March 2003, has some additional special features that are not reported here.

As established in the 2007 Portuguese Budget (Portuguese Government, 2006), three of the five reimbursement levels of NHS were reduced recently from 70, 40 and 20 percent to 69, 37 and 15 percent respectively.

Since 1992, and to overcome some recognized income-related inequality in medicines utilization, the Portuguese government introduced a ***special reimbursement status*** for the elderly with retirement funds below minimum national wage, adding up to 15% to the basic medicines reimbursement rate (Law 118/1992). Alongside to this special group, Portuguese National Health System introduced over the years an ***exemption reimbursement status*** characterized with full reimbursement (100%) to all primary care medicines needed by patients with major particular clinical diseases such as lupus, paramyloidosis, thalassaemia, sickle cell disease or haemophilia.

Job-Related Health Systems

A non-trivial part of the Portuguese population benefits from double health care expenditures coverage. This pattern of coverage is due to the maintenance of Job-Related Health Systems (JRHS) of health coverage established in the previous social insurance system, prior to 1979. Many of these systems are financed by public funds.

The JRHS provide medicines co-payments on the following format **i)** medicines direct co-payment[4] instead of NHS reimbursement, **ii)** medicines co-payment following the NHS reimbursement or **iii)** co-payment of medicines that are not currently reimbursed by the NHS and that would be hardly affordable as private expenditure.

[4] Specific JRHS medicines reimbursement scheme, although similar to the NHS scheme.

Private Health Insurance

The private health insurance (PHI) is often understood as a possible approach of addressing some health system challenges given its position as an alternative source of health care financing.

Colombo and Tapay (2004) refer that "private health insurance presents both opportunities and risks for the attainment of health system objectives and it has also given rise to considerable equity challenges in many countries".

Indeed, increased access to private insurance can promote greater health care utilisation (Jones, 2007) mainly due to:

i) *Moral hazard effect*: the level of utilization is greater when insurance decreases the private out-of-pocket cost of health care.

ii) *Risk reduction effect*: the chosen level of utilization is greater under the financial certainty created by insurance.

iii) *Access effect*: insurance may provide access to expensive medical technologies that would not be affordable otherwise.

These three factors will be jointly analyzed as the ***insurance effect*** on utilisation.

The study by Jones *et al.* (2007) on the impact of supplementary private health insurance on the number of specialist visits reported a positive insurance effect, concluding that private insurance contributes to 'pro-rich' inequality. The research used data from the European Community Household Panel for Ireland, Italy, Portugal, Spain and the UK.

The next section focuses on Portuguese medicines expenditure and the pattern exhibited through the last 20 years. The literature suggested some factors responsible for such pattern, which are also discussed here.

2.2. MEDICINES EXPENDITURE

Over the last years, the Portuguese government applied several policies intended to sustain public expenditure on medicines.

According to the INFARMED statistical information on medicines (INFARMED, 2001), the *per capita* annual expenditure has greatly increased: between 1995 and 1999 it grew from 165.27€ to 235.99€, about 70.7€ *per capita*. After reforms implemented in recent years, the pharmaceutical market reached a total of 3105 million euros in 2005, equivalent to 294.29€ *per capita*. In ten years, the total annual medicines expenditure *per capita* increased 78%.

The Portuguese National Health Service is financed mainly through general taxation and as depicted in Figure 1 is facing increasing expenditure on medicines.

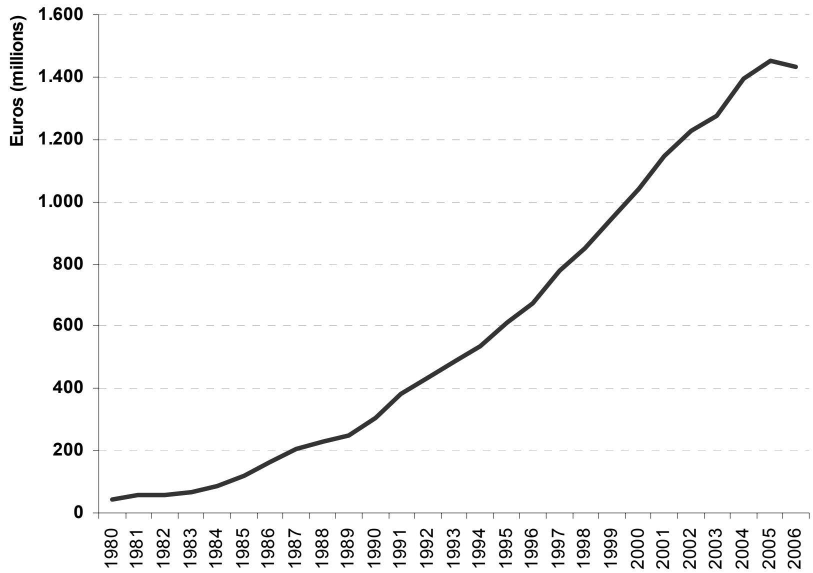

Figure 1. Portuguese NHS medicines expenditure. Primary care setting.
Source: General Health Department (DGS 1980-1994) and Infarmed (1995-2006).

Data from year 1999 (INFARMED, 2001) illustrate that 62.5% of the pharmaceutical market beneficiated from co-sharing payments with NHS and 13% with JRHS. The remaining 24.5% were financed entirely by individual private expenditure and a proportion of this expenditure was covered *a posteriori* by private health insurance schemes. However, in 1997, private insurance accounted for only 1.7% of total private financing health sources (Gouveia Pinto and Aragão, 2004). Recent data (Vaz *et al.* 2006) pointed out that nearly 20% of the total expenditure on medicines occurs without any direct cost-sharing scheme (NHS, JRHS our PHI) and as out-of-pocket expenditure. Portugal is identified as having one of the highest ratios of private - public financing of health care in the EU (Pereira, 1999).

Barros (1998) investigated the 'black box of health care expenditure growth determinants' in 24 OECD countries and concluded that health system characteristics often reported as determinants of health expenditures such as aging population, type of health system and the existence of gatekeepers are found to be non-significant. Gouveia Pinto and Teixeira (2002) points out that the Portuguese increasing pharmaceutical expenditure is mostly due to the prices of new pharmaceuticals. Oliveira and Gouveia Pinto (2005) analysed the content and impact of policies designed to reform the Portuguese health system between 1979 and 2002 and concluded that, despite the great improvement in health outcomes, the Portuguese health system is barely achieving its goals, mainly in terms of the equity of access and utilisation. Therefore, two important research questions arise:

What are the determinants of medicines expenditure? Can we disentangle between different types of inequity?

These are major research questions and the answers will be based on the Portuguese Health Survey data. In the next chapter, the data and the variable definitions are described. This set of variables will be used to construct the econometric model of the private medicines expenditure determinants.

3. DATA AND VARIABLES

The research results depend upon the econometric model specification chosen and the list of covariates used. In order to investigate horizontal inequity it is essential to differentiate between ***need variables***, which shape the use of health care and ***non-need variables*** that in equitable systems, should not affect it. Underlying the research is a set of value judgements about the factors that reflect *need* and the factors that reflect *non-need*. It is recognised that different value judgements may influence results about the existence of inequity (Morris, 2005).

In the first section of this chapter the Portuguese National Health Survey is described. In the second section the definition of the variables used in the econometric model specification is explained.

3.1. PORTUGUESE NATIONAL HEALTH SURVEY

The Portuguese National Health Survey (PNHS) is jointly carried out by the Health Ministry, National Institute of Health 'Dr. Ricardo Jorge', National Observatory of Health and the National Statistical Institute. Four national rounds have been completed until present: 1986/1987, 1995/1996, 1998/1999 and 2005/2006[5].

Unfortunately micro data from the most recent survey (2005/2006) is not yet available for research due to the ongoing data validation process.

Table 1 presents the major research areas covered by PNHS, over its waves. The second column reports some of the main variables included in the models. It can be perceived that this survey has great richness of information on health status, health care use and lifestyles. Dental Health related questions were only available on the 1998/1999 wave.

[5] Detailed information available at National Observatory of Health Website: www.onsa.pt.

Table 1. PNHS areas and relevant research variables

Area	Relevant Research Variables	PNHS wave 1986/ 1987	PNHS wave 1995/ 1996	PNHS wave 1998/ 1999
Socio-demographic	Gender, age, area of residence, education, employment status	x	x	x
General health status	Weight and height[2], self-assessed health[3], insurance coverage		x	x
Acute illness or injury	Number of days with daily activities limited due to illness or injury	x	x	x
Longstanding illness	Type of daily limitations and health condition associated	x	x	x
Chronic conditions	Specific chronic conditions such as diabetes, asthma, bronchitis, allergy, high blood pressure, back pain with medical/nursing diagnose confirmation	x	x	x
Health care use	Medical visits, use of sleeping pills	x	x	x
Dental health				x
Health care expenditure and income	Expenditure on medicines and household total net monthly income	x	x	x
Tobacco Use	Smoker behaviour	x	x	x
Food intake and drink behaviour	Drinking behaviour	x	x	x
Children Health[1]		x	x	x
Physical Activity[2]	Practise of exercise		x	x

1 – Less than 5 years of age

2 – Respondent with 18 years of age and older

3 – Self-respondent with 15 years of age and older

Some of the variables of the first national survey were measured in different scales from those used in the subsequent waves and information on other variables was not even collected (e.g. private insurance). Therefore, 1986/1987 data was not included in this research. The 1995/1996 and 1998/1999 waves of the Portuguese National Health Survey were collected during May 1995 - April 1996 and October 1998 - September 1999, respectively.

This research is based on pooled data, which include a total of 98324 individual records (49718 from 1995/1996 and 48606 from 1998/1999).

The sample was representative of the five regions[6] of continental Portugal: Norte, Centro, Lisboa and Vale do Tejo, Alentejo and Algarve. The island regions of Açores and Madeira were excluded from both surveys. Households were selected randomly from the 1991 population census, with local interviewers asked to follow pre-determined instructions (Ministério da Saúde, 1998).

Sampling rates were adjusted in each part of the year, in order to incorporate seasonal differences (Ministério da Saúde, 2001). The samples included participants who lived in private households (collective housing, e.g., hospitals and retirement houses were excluded).

Face to face structured interviews were held with all individuals who lived in the selected household, including non-family members. An appointed member of the household was asked to provide information on behalf of absent household members at the time of the interview. Little is known about non-respondents and external sources of data are not available to assess the representativeness of the sample. The overall response rate was 88% in 1995/1996 and 82% in 1998/1999. The Health Ministry validated the answers by repeating the interview in 10% of the initial sample (Ministério da Saúde, 2001).

The PNHS data set has been subject to a number of empirical studies by researchers (Giraldes, 2003; Marques-Vidal and Dias, 2005a; Marques-Vidal and Dias, 2005b; Marques-Vidal and Dias, 2006; Marques-Vidal, 2006a; Marques-Vidal, 2006b) although the majority of these projects are analyses focused on lifestyles, routines or epidemiologic measures such as disease prevalence estimation. Given the richness of information, health economics research making use of PNHS data is very interesting. Recent research conducted by Barros *et al.* (2005) use the 1998/1999 PNHS data in order to estimate the impact of health

[6] According to the Nomenclature of Territorial Units for Statistics (NUTS 2) as defined by the Eurostat (http://ec.europa.eu/eurostat/ramon/nuts/home_regions_en.html).

insurance coverage beyond National Health Insurance on the demand for several health services. Lourenço and Ferreira (2004) have analysed the income effect on the demand for dental health care and Lourenço *et al.* (2006) have focused on health care utilization (number of visits).

3.1.1. VARIABLE DEFINITIONS

Empirical analyses of demand for health care have mostly used reduced form equations as specification models for health care utilisation, depending on need, socio-economic variables, and demographic characteristics (Jones, 2007a). Van Doorslaer *et al.* (2004) and Jones *et al.* (2007) specify models for medical care, which include insurance plan, health status, socio-economic and demographic variables. Empirical studies of drug utilisation, such as Grootendorst (1995) and Street *et al.* (1999), control for age, eligibility for reimbursement, insurance status, health status, employment and demographic variables. Following these examples explanatory variables related to morbidity (clinical need), socio-economic status, and demographic characteristics were selected in order to characterize Portuguese individual medicines expenditure.

Dependent variable

The dependent variable (Y) results from the question on the expenditure on medicines in the 2 weeks before the survey. The data was originally measured in *escudos* (former Portuguese currency, until the introduction of the European currency, in 2002: 1 euro = 200,482 *escudos*) and then converted into euros. The PNHS 1995/1996 data on Y was updated to 1999 global health prices, using the medicines and pharmaceutical products price index (INE, 2006): between 1996 and September 1999, prices of medicines increased 4.2%.

According to the question formulation, the dependent variable Y includes all out-of-pocket costs with medicines, which include the private expenses shared with health care payers such as NHS. Even so, very few individuals reported expenditure on medicines in this short period: 76% had no expenses and only 10% report positive expenditure above 19.96 euros. Table 2 shows detailed statistics on private expenditure on medicines.

Table 2. Private expenditure on medicines – euro. (N= 97767)

Y summary statistics	
Mean	5.62 €
Minimum	0 €
Maximum	1647.60 €
Standard Deviation	16.70
Variance	278.73
Skewness	17.45

On average, respondents reported 5.62 euros (sd 16.70 euros) of expenditure on medicines in the two weeks preceding the survey. Figure 2 shows a positively skewed empirical distribution of the nonzero realizations of Y, with 99% of PHNS respondents reporting a positive expenditure inferior to 72.76euros.

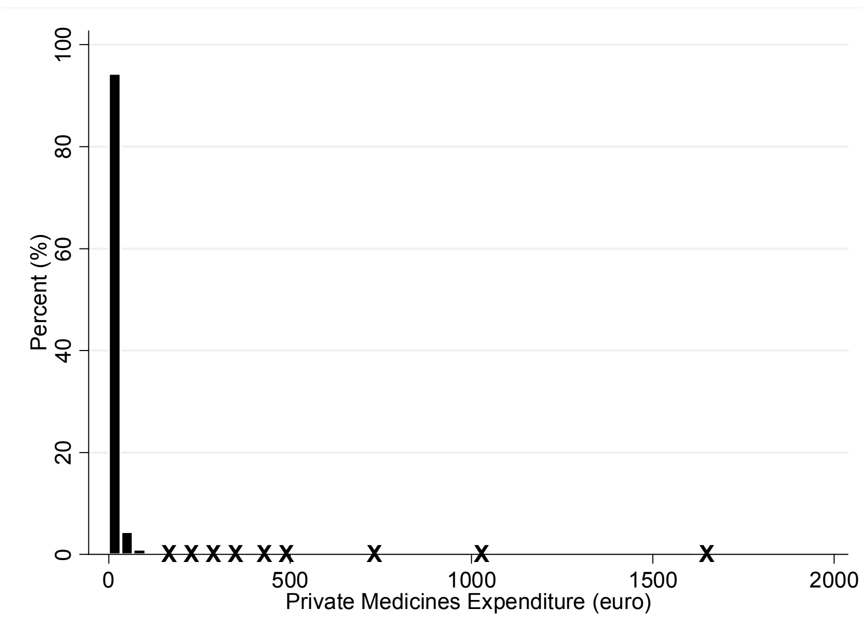

Figure 2. Histogram[7] of non-zero private medicines expenditure (N=23475). Individuals with 18 years of age and older.

[7] Non visible histogram bars are marked with the X symbol in the figure.

The most important determinant of expenditure on medicines is like to be the need for those medicines, which was accounted for with a wide range of health indicators available in the PNHS pooled data set.

Proxy respondents

The PNHS information about children or absent household members at the time of the interview was obtained from other appointed household member. Within the respondents over 15 years of age, the self-response rate was 69.89% in the PNHS pooled data. Other surveys that allowed proxy respondents were exploited in empirical models such as in Grootendorst (1995), Street *et al.* (1999) and Morris *et al.* (2005). Here a dummy variable was coded for control of this large number of proxy response: PROXRESP equals 1 if the individual do not give the answer, 0 if self-response.

The PNHS sample strategy presents several problems. The major identified issue is the existence of measurement error in the dependent variable. Although it may be possible that a proxy respondent is able to characterize socio-economic-demographic characteristics of absent household members, it is rather unlikely that the proxy respondent provides accurate information on private medicines expenditure or in crucial covariates such as need-variables.

For simplicity, the group of individuals whose responses were given by other household member will be here onwards named “Proxy-respondents”.

A preliminary analysis comparing the subgroups aged 18 years and older of self-respondents (N= 45999) with responses given by other household member (N= 33029) showed that 80.63 % reported no expenditure on medicines in the proxy-respondents sample, while 68.78% reported no expenditure on medicines in the self-respondents sample.

Table 3. Private expenditure on medicines (euro) by sub-sample. Individuals aged 18 years and older.

Y summary statistics	Self respondents	Proxy respondents
Mean	7.65€	4.78€
Minimum	0€	0€
Maximum	1647.60€	748.20€
Standard Deviation	19.64	15.41
Variance	385.63	237.39
Skewness	18.74	10.15

As reported in Table 3, the unconditional mean of private expenditure on medicines is 37.5% lower in the proxy-respondents sub-sample. Moreover, the histogram of the logarithm of non-negative values of Y depicted in Figure 3, indicates that the proxy-respondents sub-sample suffers from heaping. The logarithm scale was chosen to show the histograms due to the large skewness of the dependent variable (Figure 2).

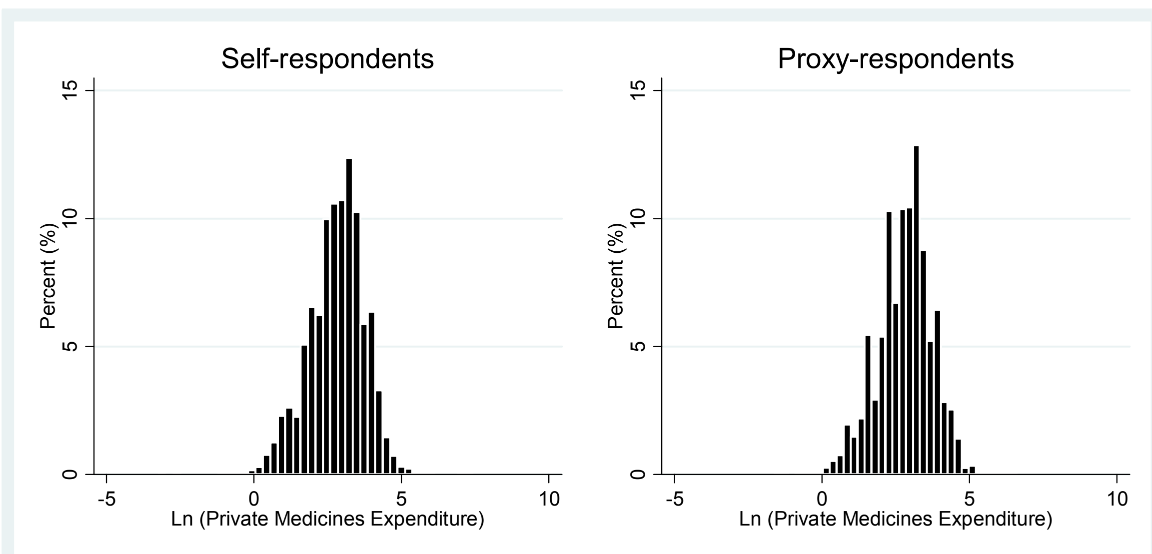

Figure 3. Histogram of the logarithm of non-zero private medicines expenditure. Self-respondents *versus* Proxy-respondents. Individuals with 18 years of age and older.

A naive estimation ignoring dependent variable measurement errors, due to heaping, conducts to inconsistent estimators of the parameters of interest (Wolff and Augustin, 2003). Wolff and Augustin (2003) and Bernardo (2007) have analyzed the impact of dependent heaped duration data in non-linear models concluding that disregarding these measurement errors will result in asymptotically biased estimators.

Therefore, in this research, the econometric models will concentrate on the sub-sample of PNHS self-respondents as approached by Barros *et al.* (2005) and Lourenço and Ferreira (2004).

The PNHS sampling scheme uses households as primary sampling units; once a household is included in the sample, all individuals in that family are interviewed. Since there is at least one self-respondent individual in each family, none of the initially chosen families is excluded from the analysis of self-respondents, maintaining the validity of the sampling process of the PNHS.

Need Variables

Acute Illness

A set of dummy variables is used to capture whether the individual is hampered by any health condition. This variable results from the question about the number of days in the last 2 weeks with limitations in daily activities due to illness or injury. The answer was coded as:

- NOTHAMP if the individual reported zero days;
- MILDHAMP if the individual reported less than 4 days;
- MODHAMP if the individual reported 4 to 13 days;
- SEVHAMP if the individual reported 14 days (maximum coded).

The MILDHAMP category was coded alongside with the Portuguese National Social Security health protection scheme that in the majority of the situations only provides social insurance for illness longer than 3 days (Law 146/2005).

SEVHAMP category was coded to capture health conditions coded in the upper bound of the number of days with limitations in daily activities due to illness or injury that can be more severe than temporary illness.

Specific Illnesses and Chronic Conditions

Information on specific cause of illness in the last two weeks, the cause of having at least one medical visit in the three months preceding the survey and the cause of having ear/eye complaints or verbal communication limitations was included. The underlying cause of these limitations was coded in both PNHS waves under the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM[8]).

Other survey information on specific chronic diseases such as asthma, diabetes, bronchitis, allergies, high blood pressure, back pain, urinary tract complaints and sleeping disturbances (proxy question was taking medication to sleep) was also included according to their classification under ICD-9-CM.

The variable ACCIDENT equals 1 if the observed health limitation was due to accident, 0 otherwise. If it is not the accident case, a set of dummies variables was created to capture the individual specific co-morbidities. Each dummy equals 1 if observed health limitation is related to a specific ICD-9-CM category, 0 otherwise. This set of dummies is not mutually exclusive and the individuals can report different number of co-morbidities. Given some less frequent morbidities reported such as “Infectious and Parasitic Diseases”, “Neoplasm”, “Blood and Blood-Forming Organs”, "Complications of Pregnancy and Childbirth ", "Skin and Subcutaneous Tissue", "Congenital Anomalies" and "Certain Conditions of the Perinatal Period", the variable *Other Diseases* is an aggregate dummy that equals

[8] International Classification of Diseases, Clinical Modification. 9th version. (http://www.who.int/classifications/icd/en/)

1 if observed health limitation is related to one of the above health conditions, 0 otherwise.

Medical Visits

Reported morbidities measures are used to capture the individual need from medicines but it can be reasonable to admit that some unobserved component on morbidity is positively correlated with individuals past visits to medical doctors. Moreover, general practitioners acts like gatekeepers in the Portuguese health system and medicines expenditure is often the consequence of a medical prescription. Alongside these arguments, an indicator of whether the individual had at least one visit to a doctor in the previous 3 months (VISITMD) was included in the set of *need variables*.

Self-Reported General Health

Self-assessed health (SAH) was measured on a five-point scale ranging from ‘‘very good’’ to ‘‘very poor”. SAH is a simple subjective measure of health that gives an ordinal ranking of self-perceived health status. Health economics literature refers self-assessed health as a major health outcome (Jones, 2004). This variable has also been shown to be a powerful predictor of mortality (Idler and Kasl 1995; Idler and Benyamini 1997). SAH has been extensively used in previous researches of the relationship between health and use of medical care (Grootendorst, 1995; Van Doorslaer, 2004; Morris, 2005; Barros, 2005) and between health and socioeconomic status and lifestyles (Contoyannis and Jones, 2004; Balia and Jones, 2007).

Healthy Lifestyles

As well as other factors, the adoption of potential risky health-related behaviours could be also considered in models for health care use (Balia and Jones, 2007). Hence, to better explain inequities in private medicines expenditure, it is appropriate to account for individual lifestyles. Lifestyles were considered as a set of behaviours, which can influence health and the need for medicines. Four indicators of whether the individual has healthy lifestyles were included in the covariate set:

- NOALCOHOL equals 1 if the individual has not consumed alcohol[9] in the last year or at utmost does it one time *per* week, 0 otherwise;
- NOOBESE equals 1 if the individual has a body mass index (BMI[10]) lower than 25 kg/m^2, 0 otherwise;
- EXERCISE equals 1 if the individual currently practises physical exercise or has daily activities with physical effort such as walking, climbing/descending stairs and carrying heavy objects, 0 otherwise;
- NOSMOKER equals 1 if the individual currently does not smoke tobacco, 0 otherwise.

Non Need Variables

Gender

The sex of the individual is considered using the dummy variable FEMALE, in order to account for sex-specific health care requirements and preferences. A common result in studies of health care utilisation is that women use it more frequently, even after controlled for need. The determinants of health care, or its effects, can also vary by gender (Grootendorst, 1995; Bago d'Uva, 2005; Morris, 2005), which could also motivate splitting the analysis.

Age

If the morbidity measures are not as accurate as desired to capture all of the individual need from health care, some authors argue that the age factor (AGYRS)

[9] Alcohol includes wine, beer, brandy, liquors and whisky or gin.
[10] BMI is a measure calculated by the ratio between weight and height. The variables weight and height were collected only in individuals with 18 years age or older. According to the WHO criteria, a body mass index superior to 25kg/m^2 is indicator of obesity.

can account for this ‘imperfect measurement of health status’ (Bago d’Uva, 2005). However, as depicted in Figure 4, the PNHS variable age has abnormal[11] concentrations at some values: the distribution of the reported age has strong spikes which certainly result from measurement error.

This phenomenon is common in survey data and is referred in the literature as age heaping: e.g. respondents often approximate their true age using some digit preference (Heitjan and Rubin, 1990). Hence, the age variable can suffer from several types of coarseness. In a study concerning children from Tanzania, Heitjan and Rubin (1990) were confronted with age peeks at six month multiples and used multiply imputed data in order to accomplish their research without having to confront the interfering issues of heaped age.

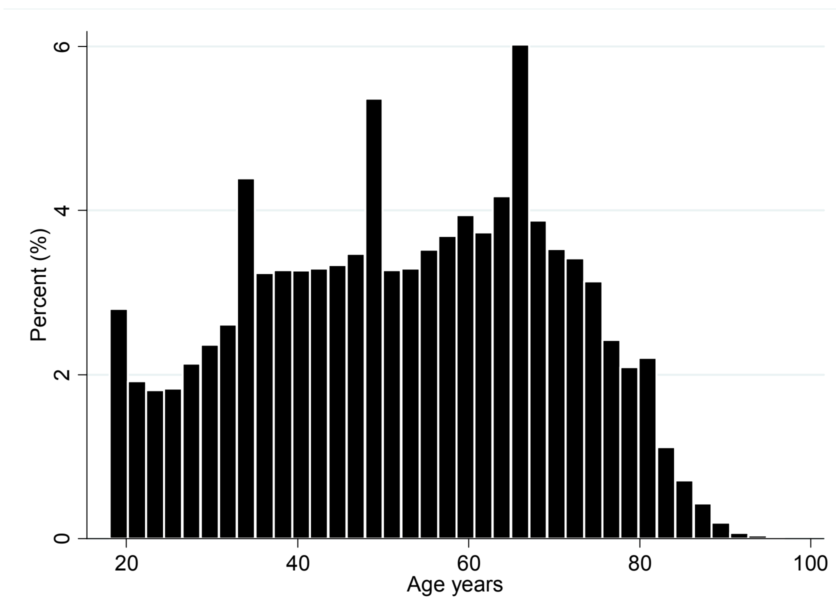

Figure 4. Histogram of individual’s age years. Self-respondents with 18 years of age and older.

[11] Abnormal: different pattern from the Portuguese population census.

Another issue concerning age years and medicines expenditure is that children's pathways to care are distinct from those of adults (Deb and Holmes, 2000): usually they do not take their own decisions regarding health care utilisation and hence it was decided to include in the model sample only individuals aged 18 years and older.

Therefore, in order to attenuate the age heaping effect, age is here represented by seven categories:

- AGE1824 equals 1 if 18 ≤ AGYRS ≤ 24, 0 otherwise;
- AGE2534 equals 1 if 25 ≤ AGYRS ≤ 34, 0 otherwise;
- AGE3544 equals 1 if 35 ≤ AGYRS ≤ 44, 0 otherwise;
- AGE4554 equals 1 if 45 ≤ AGYRS ≤ 54, 0 otherwise;
- AGE5564 equals 1 if 55 ≤ AGYRS ≤ 64, 0 otherwise;
- AGE6574 equals 1 if 65 ≤ AGYRS ≤ 74, 0 otherwise;
- AGE75 equals 1 if AGYRS ≥ 75, 0 otherwise.

This approach allows for a flexible relationship between age and medicines expenditure, without impose a linear or polynomial function, while it attenuates the major age spikes observed in the histogram.

Simply ignoring the heaping occurrence is the most frequent decision reported in the literature (Jürges, 2005). The aggregation of age coarse data in age bands relies in the assumption of coarseness at random (CAR), a concept similar to missing at random (MAR) for item missing data. To our knowledge, none of the study based on PNHS data has mentioned this age heaping occurrence.

Portuguese Region (area of residence)

Individual drug expenditures, even conditional on need and other factors, can differ between geographical regions due to differences such as medical prescription patterns or attitudes towards health care expenses. Continental Portugal was divided into five main geographical regions with a mean population of 1943469 residents (range 375841 to 3621210) in the year 1999. The variable region of residence[12] was represented by five categories LISBOAVT, CENTRO, NORTE, ALENTEJO and ALGARVE using a set of mutually exclusive set of dummies.

Marital status

Marital status has been incorporated in previous studies of the use of medical health care (Manning and Mullahy, 2001; Bago d'Uva, 2005). The individual marital status is here considered with a set of dummies: SINGLE, WIDOW, DIVSEP and MARRIED. DIVSEP equals 1 if the respondent reports divorced or separated, 0 otherwise.

Education

The total number of completed school years (SCHLYRS) is frequently used to control for education in models for health care use. Figure 5 shows the empirical distribution of the number of completed school years reported in the PNHS sample. As depicted, the PNHS variable number of completed school years has high concentrations at particular values: the percentage distribution showed spikes at 3, 6 and 9 completed school years and a very strong spike at 4 completed school years. The Portuguese mandatory number of school years changed through the last decades from four to nine years. Hence, these spikes can account for the behavioural pattern related to the Portuguese mandatory education.

[12] According to the European Union NUTS 2 level definition.

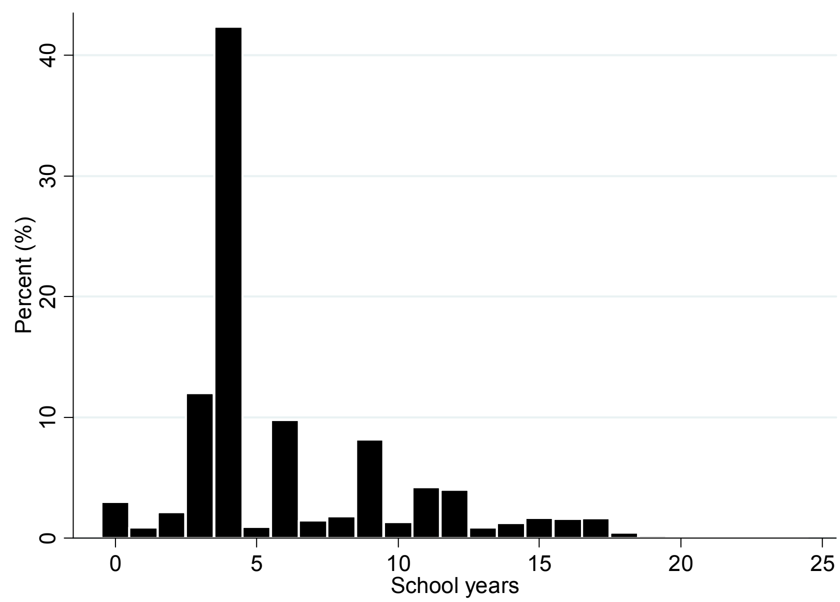

Figure 5. Histogram of number of school years. Self-respondents with 18 years of age and older.

Therefore, formal education was included in the model using the PNHS question on graduate level achieved where education is measured by a set of dummies: NOSCHL for children not yet in school, individuals unable to read or write and individuals with basic skills in reading and writing, BASICSCHL for basic school (until 9th grade), SECSCHL for secondary school (10th to 12th grade) and GRADSCHL for university degree or more.

Employment status

PNHS respondents were asked whether they have worked in the two weeks before the survey and the non-working cause was recorded. Employment status is here measured by a set of dummies: EMPLOYED, UNEMPLOYED, RETIRED, DISABLED and OTHERUNEMP.

EMPLOYED equals 1 if the respondent has worked or if the respondent has not worked due to vacations or atmospheric conditions. The respondents that were not working due to sickness were classified here as EMPLOYED. The UNEMPLOYED dummy equals 1 if the respondent was not working and was currently unemployed or searching for the first employment.

RETIRED equals 1 if the respondent was not working and is retired, waiting for retirement or has a retirement fund. DISABLED equals 1 if the respondent was not working and is disabled. DISABLED also incorporate individuals who reported long standing severe physical limitations such as being confined to bed, to a wheel chair or to their own house.

OTHERUNEMP equals 1 if the respondent was not working and his 'employment status' was keeping house, student, undertaking military service, living from rents or interests or was not working due to other unspecified reasons.

Time Variable

A time dummy was also incorporated in the covariates set: T1 equals 1 if the individual participated in the 1998/1999 survey, 0 otherwise.

T1 can account for the changes between 1996 and 1999 in the set of medicines prescribed for each health condition that embrace e.g. the on-going replacement of old-less-expensive medicines by new medicines, with more safety and efficacy but usually at higher prices. The time variable can also capture the effect of multiple events that affect the prescription, use and expenditure with medicines. Some pharmacotherapeutic guideline's changes such as the reduction in critical limits of glycaemia to undertake pharmacologic treatment with antidiabetics or the reduction of cholesterol levels in prescription guidelines to initiate the treatment with statins.

These factors are impossible to detect or disentangle from each other within this data set. Therefore, they will be jointly analyzed as the ***time effect*** on expenditure.

Major research variables

The effects of factors such as income, pharmacies density and additional health insurance (private or job-related) on conditional expected value of Portuguese private medicines expenditure are a major research interest.

Health Insurance

The potential effect of health insurance on demand for health care was taken into account using information about additional voluntary insurance coverage: PHI=1 if the individual is covered by a private health insurance plan, 0 otherwise. It should be noticed that the PNHS information does not make it possible to differentiate between private insurance purchased voluntarily by the individual or a private health insurance plan provided by the employer. The dummy variable JRHI=1 measures whether the individual is covered by a specific Portuguese job-related health insurance plan such as ADSE[13], SAMS[13] or others.

Income

The income variable is derived from the disposable income of the whole household (i.e. after deductions for income tax and mandatory national insurance). Total household income includes all net monetary income received by the household members during the previous month. It includes income from work, private non-work income such as interests or rents, pensions and other direct social transfers received (e.g. children social support).

The information was collected according ten household income bands of different widths. The last income category is an open-end top band. The approach of Morris *et al.* (2005) was followed in order to compute individual income from the PNHS household income bands.

[13] ADSE is the insurance plan that covers Portuguese civil servants; SAMS covers bank employees. ADSE is the largest JRHI plan.

The median level of household income within each band was the chosen measure of household income for all individuals within each band. The mean and standard deviation parameters of the log normal distribution were determined by minimising the sum of squared differences between actual and generated numbers of a log normal distribution in each band. The average income was calculated as the value of half the cumulative density within each band.

The estimate of household income for each individual whether adult or child was then equivalised to allow for differences in household size and composition using the Organization for Economic Co-operation and Development (OECD) modified scale. This scale gives a weight of 1 to the first adult, 0.5 to any other household member aged 14 and over, and 0.3 to each child under 14 in the household. The modified OECD scale replaced the original OECD scale (first adult person weight of 1.0, additional adult 0.7, children 0.5), which had been extensively used and was known previously as the Oxford scale.

The PNHS1995/1996 equivalised income was updated to 1999 prices making use of the consumer price index (INE, 2006) that registered a 10,40% increase between years 1996 and 1999. The natural logarithm of the equivalised income (LNINC) was chosen to be the income measure, like in other health care use econometric studies (Van Doorslaer, 2004; Morris, 2005).

3.2. EXTERNAL DATA

Supply of health services can still be determinant in medicines utilisation but it is often excluded from empirical analysis, mainly due to lack of data. To account for access to health services, the data set was augmented with external information on pharmacies and pharmacies extensions[14] density (PHD) at LAU 1[15] level.

[14] Pharmacy extension: establishment dependent from a pharmacy (headquarters).

Currently and as regulated by INFARMED a pharmacy is open to cover at least 4000 inhabitants. In 1999, Continental Portugal was divided into 278 districts, with a mean population of 34954 residents (range 1904–580436) and a mean number of 10 pharmacies (range 1–340 pharmacies) *per* district.

Pharmacies density

Pharmacies density was incorporated in the models by the number of pharmacies *per* 4000 inhabitants. The data on the number of pharmacies, pharmacies extensions and population statistics at district level was obtained from Portuguese National Statistical Office. The lowest density (0.4 pharmacies *per* 4000 inhabitants) was obtained in the Fafe district, with 5 pharmacies that supply drug medicines to 51597 inhabitants. However, the Crato district has 5 pharmacies and a lower current number of inhabitants: 4356.

3.3. SUMMARY STATISTICS

Non-responses are coded as "missing values" in the PNHS dataset. Non-responses occur when an individual responds to the survey as a whole but does not answer to a particular question. If observations with missing values are systematically different from those who respond to the question then we have potential sample selection bias. However, if the missing values occur at random (MAR), the observations with missing values for at least one of the variables used in the model could be dropped without major inference repercussions.

The final data set includes 41601 individual self-respondents records observed in the pooled PNHS sample. For clarity of presentation, the summary statistics for the variables included in the econometric models are presented by three subsets of

[15] The Local Administrative Units (LAU) Eurostat definition contains districts (LAU level 1) and municipalities (LAU level 2) (http://ec.europa.eu/eurostat/ramon/nuts/introannex_regions_en.html).

variables: non-need variables (Table 4), need variables (Table 5) and dependent variables plus major research variables (Table 6).

Table 4. Summary statistics of non-need variables (N=41601)

	Group Variable	Variable	Proportion
Non Need Variables	Gender	Female	0.613
	Age	Age1824	0.065
		Age2534	0.119
		Age3544	0.160
		Age4554	0.171
		Age5564	0.194
		Age6574	0.187
		Age ≥ 75	0.103
	Portuguese Region	Lisboa and Vale do Tejo	0.259
		Centro	0.211
		Norte	0.280
		Alentejo	0.143
		Algarve	0.107
	Marital Status	Married	0.732
		Single	0.122
		Widow	0.120
		Divorced or separated	0.026
	Education	No school	0.222
		Basic school (until 9th grade)	0.641
		Secondary school (10th to 12th grade)	0.065
		Graduate school (university degree or more)	0.072
	Employment	Employed	0.484
		Otherunemp (includes keeping house, student, undertaking military service, living from rents or interests and other)	0.194
		Unemployed	0.043
		Retired	0.278
		Disabled (also includes individuals being confined to bed, to a wheel chair or to their own house)	0.001
Other	Time (1999 wave)	T1	0.505

Each of the set of exclusive dummies has an omitted category to avoid ‘dummy variables trap’ that would generate perfect collinearity in the regression models if a dummy variable were included for each category. Jointly, the omitted categories highlighted in the tables by grey shading, define the ‘reference individual’.

The reference individual is self-respondent, male, between 18 and 24 years old, live in the Lisboa and Vale do Tejo region, is employed, married, has only basic skills in reading and writing, has not been hampered by his health condition in the last two weeks, has not visited a medical doctor in the past three months, has very good SAH, practises exercise, does not smoke tobacco, is not obese, does not consume or consumes alcohol prudently, has been surveyed in the 1995/1996 wave and does not have private insurance or any job-related insurance.

Table 4 shows that 61.3 per cent of the self-respondents sample are women and that the individuals between 55 and 64 age years are the largest group, with 19.4 per cent of the sample. Most of the respondents live in the Norte region (28 per cent) and the majority are married (omitted category) followed by those who are single (12.2 per cent) at the time of the survey. A large proportion of the sample has education qualifications until the 9th grade (64.1 per cent) and is employed (omitted category, 48.4 per cent) followed by those who are retired (27.8 per cent). From the self-respondents, 50.5 per cent has responded to the survey in the 1998/1999 wave.

Table 5 reports descriptive statistics for the full list of need variables included in the econometric models. After individuals not hampered by health in the last two weeks (the omitted category), the individuals mildly hampered are the next largest group, with 5.6 per cent of the sample. Musculoskeletal and circulatory system diseases are the largest groups reported, with 55.3 and 27.5 per cent. Most of the respondents have visited a medical doctor in the past three months (59.9 per cent) and the majority reported fair health (43.3 per cent) at the time of the survey. The majority of the respondents reported healthy lifestyles: 47.2 per cent are not obese, 69.7 practises exercise, 83.3 per cent does not smoke and 61.1 per cent does not consume alcohol or does it prudently.

Table 5. Summary statistics of need variables (N=41601)

	Group Variable	Variable	Proportion
Need Variables	Acute Illness (number of days in the last 2 weeks with limitations in daily activities due to illness or injury)	Not hampered (0 days)	0.875
		Mildly hampered (0 < days ≤ 3)	0.056
		Moderately hampered (3 < days ≤ 13)	0.031
		Severely hampered (14 days)	0.038
	Illnesses and Chronic Conditions (ICD-9M classification)	Accident	0.017
		Endocrine and Metabolic	0.087
		Mental Disorders	0.019
		Nervous System	0.034
		Circulatory System	0.275
		Respiratory System	0.166
		Digestive System	0.064
		Genitourinary System	0.097
		Musculoskeletal System	0.553
		Symptoms and Signs	0.249
		Injury and Poisoning	0.148
		Other Diseases	0.057
	Visit medical doctor	Visitmd	0.599
	Self-Assessed Health (SAH)	Very Good	0.025
		Good	0.269
		Fair	0.433
		Poor	0.220
		Very Poor	0.054
	Healthy Lifestyles	Exercise	0.697
		Non obese	0.472
		Non smoker	0.833
		Non alcohol heavy behaviour	0.611

Table 6 shows descriptive statistics for the dependent variable and the major research variables included in the econometric models.

Table 6. Summary statistics of dependent variable and major research variables (N=41601)

Group Variable		Variable	Mean (Std. Dev.) or proportion	Min	Max
Dependent Variable		**Y** - private medicines expenditure in the last 2 weeks (euro).	7.607 (19.763)	0.00	1647.60
Dependent Variable		**DY**, equals 1 if Y ≥ 0, 0 otherwise.	0.314 (0.464)	0	1
Major Research Variables	Job-related health insurance	**JRHI**	0.164	0	1
	Private health insurance	**PHI**	0.044	0	1
	Equivalised income	**Income** (euro)	374.107 (253.035)	40.30	1966.33
	Logarithm equivalised income	**LNINC**	5.739 (0.599)	3.70	7.58
	Pharmacies density	**PHD** (number pharmacies/ 4000 inhabitants)	1.301 (0.597)	0.39	4.09

Table 6 shows that 68.64 per cent of the model sample reported zero private medicines expenditure and the average expenditure is 7.61€, with a range from 0€ to 1647.60€. Supplementary health insurance is not frequently observed in the model sample: 16.4 per cent of the respondents have job-related health insurance and only 4.4 per cent have private health insurance. The mean of the equivalised monthly net income is 374.11€ (1999 prices) and the average pharmacies density at the time of the interview was 1.3 pharmacies/4000 inhabitants. The equivalised monthly net income is incorporated in the regression models by its natural logarithm (LNINC) although for completeness, summary statistics of the original variable (Income) are also reported in Table 6.

4. EMPIRICAL MODEL

Econometric analysis of the determinants of expenditure on private expenditures on medicines is relevant and crucial given the current reforms the Portuguese NHS is undergoing. Conditional on individual need and other socio-economic and demographic factors, this research has three main objectives:

- **i)** To test for income-related inequity;
- **ii)** To test for pharmacies density inequity;
- **iii)** To identify the existence of insurance effects.

In order to achieve these purposes through an econometric framework it is necessary to use a consistent estimator, to be able to make valid inference about the parameters of interest.

In the first section of this chapter, the model specifications are discussed and in the last section the estimation results are presented.

4.1. ECONOMETRIC SPECIFICATION

Health economists know *a priori* that many outcomes of interest require non-standard empirical methods. Examples are health care expenditures, length-of-stay, health care use and others. Manning and Mullahy (2001) characterize such data by:

- i) Non-negative measurements of the outcomes;
- ii) Positively skewed empirical distribution of the nonzero realizations;
- iii) Nontrivial fraction of zero outcomes.

Econometric strategies for the analysis of such data have been discussed extensively in the literature (Manning, 2005), but none of the articles on this topic has provided evidence on how well their estimators would behave under a range of data conditions, nor have they provided an algorithm for choosing among the alternatives (Santos Silva and Tenreyro, 2006).

Non-negative measurements and positively skewed empirical distribution

The pattern exhibited by private medicines expenditures is non-trivial to analyse given that we deal with a continuous non-negative variable with a skewed positives distribution, as showed in Figure 2 which reveals the long and heavy right tail of the empirical distribution. Classical econometric methods, such as OLS are clearly inappropriate. The literature on count data modelling deals frequently with patterns similar to this and solutions are encountered. However, given the non-integer characteristic of the dependent variable, a suitable specification is a challenging question.

Nontrivial fraction of zero outcomes

It is worth mentioning the large proportion of zero observations in the data sample: 68.78% of the self-respondents report zero expenditures with medicines in the previous 2 weeks. This high proportion cannot be ignored in the analysis and it represents an additional study challenge.

The large fraction of zero outcomes in the sample can be due to the time frame in PNHS instrument, which can be too short to capture all private expenditure on medicines taken in that period. Part of the zero values can actually correspond to healthy individuals and be captured by the model, but other zeros can refer, for instance, to individuals that suffer from chronic conditions. For e.g. individuals with heart disease (undertaking daily medications) can purchase their drug medicines for three months or more (three monthly packages) and hence can report zero expenditure in the last two weeks.

Although determinant (Street, 1999), individual eligibility to specific NHS pharmaceutical reimbursement levels (between 1995 and 1999: 100, 70 and 40 per cent reimbursement) was not included as a covariate due to lack of information. Hence, individuals that purchased medicines with 100 per cent reimbursement or individuals suffering from diseases with exemption status (100 per cent reimbursement to all medicines) can also report zero expenditure, for all the time frame considered. INFARMED statistics report 6.4% and 7.8% as the proportion of drug medicines reimbursed at 100% by the NHS in the years 1996 and 1999.

Summarising, zero expenditure can either be due to a healthy condition, or to a illness that implies purchase of medications for periods longer than 2 weeks or entitlement to full reimbursement. Therefore, zero outcomes can have different natures:

i) Zero medicines expenditures due to healthy conditions and remaining socio-economic-demographic characteristics;
ii) Zero outcomes due to the inability of the PNHS question to capture existent positive expenses;
iii) Zero outcomes due to the Portuguese reimbursement schemes (not controlled for).

Therefore, the chosen econometric specification cannot ignore the different nature of the zero private medicines expenditure and should, in fact, explicitly account for these data characteristics. However, given the available data, the zeros from nature ii) and iii) cannot be identified separately and these zeros will be jointly mentioned as "extra zeros" and the "double nature of the zero outcomes" will justify the models used.

Zero Inflated Models

Zero inflated models (ZIM) are popular to deal with the excess of zeros in count data econometric studies (Grootendorst, 1995; Mullahy, 1997) and will be the natural starting point since it allows us to incorporate a fraction of extra zeros.

However, this research focuses on private medicines expenditure (Y), a continuous dependent variable and hence a methodological challenging and crucial question arises:

Given the need to incorporate a fraction of extra zeros, which econometric specification should be modified in order to accommodate this data specificity?

Consider health care demand measured in terms of private medicines expenditure (Y), which is assumed to depend on a set of observable exogenous[16] explanatory variables, denoted by the vector x. Let us denote π the fraction of extra zeros.

The central research questions involve the conditional expected value, denoted by $E[Y|x]$, and its associated partial effects $\delta(x) = \frac{\partial E[Y|x]}{\partial x}$. The effects of factors such as income (LNINC), pharmacies density (PHD) and private health insurance (PHI) on conditional expected value of Portuguese private medicines expenditure are of major interest in this study.

Density functions for non-negative continuous random variables such as Lognormal, Gamma, and Weibull exclude zero from their support. Moreover, continuous densities do not allow a positive mass at any peculiar value in their supports such as the zero value (Deb and Trivedi, 2002). Thus, in the continuous case, there exists no density that can account specifically for the fraction π of extra zeros. Consequently, it is not possible to devise a continuous version of the zero inflated Poisson (ZIP), popular econometric specification for data with large fraction of zero counts. For this same reason, in a mental disorders expenditures determinants study, Deb and Holmes (2000) extend a finite mixture model, in the

[16] This is an econometric simplification that is often a study limitation.

continuous case, only to the sample of individuals with strictly positive expenditures.

A recent research of Santos Silva and Tenreyro (2006) implemented a set of experiments to model a continuous variable in order to study alternative estimators under different forms of heteroskedasticity with the presence of rounding errors in the dependent variable. They proposed a Poisson pseudo-maximum-likelihood estimator (PPML) to model bilateral trade. The idea of using the Poisson framework to obtain consistent estimates is very attractive and will be assessed to incorporate the degenerate fraction π of zeros.

Another econometric strategy to deal with excess zeros problem have been two-part models (2PM) or hurdle models extensively applied in the past (Grootendorst, 1995; Street, 1999; Lourenço and Ferreira, 2004) to model skewed outcomes data, usually with large amounts of zero observations. Here, the process that determines the binary decision is different from the process that determines the amount of expenditure once the hurdle is crossed (Deb and Trivedi, 1997). In the case of health care use, the motivation for two-part models comes from the 'principal-agent' theory of demand, which suggests that the individual initially chooses whether to seek treatment, but once in treatment it is the physician who determines the medicines to be taken.

Hence, two major and competing econometric strategies will be here considered to account for the π extra zeros in the data sample:

- One-part model: PPML;
- Two-part model (2PM).

4.2. ONE-PART MODEL

The fundamental propriety of the private medicines expenditure conditional mean is that $E[Y|x]$ must clearly be positive: $E[Y|x] > 0$, since all observations of y are non-negative. Therefore, the popular Ordinary Least Squares (OLS) estimator will not be considered in the one-part specification discussion.

However, the log OLS specification, widely used in empirical studies, is the benchmark for strictly positive expenditure data and for the second part of 2PM specifications (Street, 1999; Lourenço and Ferreira, 2004). In this empirical application, estimating the parameters of interest by OLS using log-linear model is actually unfeasible given the zero values of pharmaceutical expenditure for more than 60% of the individuals.

Given the basic propriety of conditional mean, $E[Y|x] > 0$, a natural starting (Mullahy, 1998) stochastic model can be formulated as,

$$y_i = \exp(x_i \beta) + \varepsilon_i \qquad (1)$$

with $y_i \geq 0$ for *i=1,2,…, N*, where $E[\varepsilon_i|x] = 0$. For notation simplicity, the index *i* is dropped from here onwards.

The generalized linear model framework relies on the correct specification of a mean and variance function for the observed variable y, conditional on x (Manning and Mullahy, 2001). Although the central structure of the model is the exponential conditional mean $E[Y|x] = \exp(x\beta)$, the relationship involving the conditional mean and the conditional variance $Var[Y|x]$ is a key attribute to consider. This relationship can be described using the general structure,

$$Var[Y|x] = \sigma^2 \upsilon(x) \qquad (2)$$

where $\upsilon(x)$ is the variance function. Manning and Mullahy (2001) enunciated some particular forms of $\upsilon(x)$ that originates special cases,

- **i)** $\upsilon(x) = 1$, the homoskedastic nonlinear regression model (variance of y conditional on x is given only by σ^2 and it is unrelated to x);
- **ii)** $\upsilon(x) = k_1 \mu(x;\beta)$, the 'Poisson-like' structure where $k_1 > 0$ and the variance is proportional to the mean, which is itself a function of x;
- **iii)** $\upsilon(x) = k_2 (\mu(x;\beta))^2$, the 'Gamma-like' structure where $k_2 > 0$ and the standard deviation is proportional to the mean.

Poisson Pseudo-Maximum-Likelihood

Santos Silva and Tenreyro (2006) analyzed several estimators to model the continuous variable bilateral trade, including log-linear OLS, non-linear least squares (NLS), gamma pseudo-maximum-likelihood (GPML) and the new proposed Poisson pseudo-maximum-likelihood (PPML). These authors pointed out that, under the assumption that $Var[Y|x] \propto E[Y|x]$, the parameters β could be estimated solving the first-order conditions,

$$\sum_{i=1}^{n} [y_i - \exp(x_i \beta)] x_i = 0 \qquad (3)$$

which is numerically equivalent to the Poisson pseudo-maximum-likelihood estimator, extensively used in count data empirical modelling since the summation on the left-hand side of equation (3) has expectation zero if $E[Y|x] = \exp(x\beta)$.

Hence the Poisson PML is consistent under the weaker assumption of correct specification of the conditional mean, independently of considerations on the distribution or type of positive support of the variable. Therefore, Y does not need to be Poisson distributed or even integer. This is a very interesting econometric strategy to model non-integer data such as expenditures, since it is consistent, reasonably efficient under a wide range of heteroskedasticity patterns as tested by Santos Silva and Tenreyro (2006) and straightforward to implement. Windmeijer *et al.* (2006) applied this estimator to model schizophrenia expenditures in a cost-efectiveness study.

The PPML model should be modified to add the inflate parameter π accounting for the degenerated distribution on $Y = 0$, making use of the following set of mixture probabilities denoted as P^M,

$$\begin{aligned} P^M &= \Pr(Y = 0|x, \beta, \pi) = \pi + (1-\pi)P(Y = 0|x, \beta) \quad y = 0 \\ P^M &= \Pr(Y = y|x, \beta, \pi) = (1-\pi)P(Y = y|x, \beta) \quad y = 1,2,3,... \end{aligned} \tag{4}$$

In the case of count data, maximum-likelihood estimation allows identification of both parameters π and β. However, to use pseudo-maximum-likelihood results we need to specify the conditional expected value, which in this case is given by

$$E[Y|x, \beta, \pi] = (1-\pi)\exp(x\beta) = \exp(\ln(1-\pi) + x\beta) \tag{5}$$

It is clear that if a constant term is present in the linear combination $x\beta$, this cannot be identified separately from the term $\ln(1-\pi)$ therefore, in the case of non-count data, the zero-inflated Poisson cannot be used; otherwise **the π proportion cannot be identified**.

This one-part model clearly presents attractive features such as consistency in the presence of heteroskedasticity and supporting dependent variable zero values;

however, it cannot be used due to the double nature of the zero values present in the data, since we are interested in estimating π .

Another option is to use Non Linear Least Squares (NLS), since it is an asymptotically valid estimator for the equation (5). It is defined by

$$\left[\hat{\beta}, \hat{\pi}\right] = \arg\min \sum_{i=1}^{n} \left[y_i - (1-\pi)\exp(x\beta)\right]^2 \qquad (6)$$

However, this estimator can be very inefficient as it ignores the heteroskedasticity that is present in this type of data. The NLS estimator is optimal when the variance of y conditional on x is given by a constant, hence unrelated to x. In practise, this is an unrealistic hypothesis and the form of $Var[Y|x]$ is usually unknown[17]. This estimator it is not frequently used in empirical applications due to its considerable inefficiency.

A one-part specification will not satisfy the purposes of this research namely the estimation of the extra fraction of zeros by an estimator consistent in the presence of heteroskedasticity and efficient. Hence, the option for a two-part specification is considered.

4.3. TWO-PART MODEL

Another common choice for data with large amounts of zero outcomes has been two-part models (Mullahy, 1998) where a two-part regression is performed, first modelling the binary outcome and then the non-zero values in a standard linear regression.

[17] If the form of $Var[Y/x]$ were known, we could weight the NLS estimator. The appropriate weights can be also obtained non-parametrically, although this estimator is rather cumbersome to implement as noted by Santos Silva and Tenreyro (2006).

Given $y \geq 0$, it is interesting and especially useful to decompose $E[Y|x]$ in two factors,

$$E[Y|x] = \Pr(Y > 0|x).E[Y|y > 0, x] \quad (7)$$

Two-part models are constructed permitting that that the mechanism that explains the zero outcomes or $\Pr(Y > 0|x)$, differs from the one that explains the positive values: $E[Y|y > 0, x]$. Given the binary variable $DY = \begin{cases} 0 \text{ when } Y = 0 \\ 1 \text{ when } Y > 0 \end{cases}$ the relationship $\Pr(Y > 0|x) = \Pr(DY = 1|x)$ holds and standard binary index models can be applied (Wooldridge, 2002). However, these standard models assume that there is only one type of zeros in the sample.

Hence, they are not appropriate when the zeros in the sample do not all have the same nature. Therefore, once more we need to explicitly include the fraction of excess zeros in the first part of this specification.

4.4.1 PART ONE

In the first part we focus on the estimation of $\Pr(Y = 0)$. To do so we use *DY* as the binary dependent variable and model $\Pr(Y = 0) = \Pr(DY = 0)$ allowing for a percentage π of extra zeros. We obtain $\Pr(DY = 0) = 1 - \Pr(DY = 1)$ where conditional on a set of exogenous variables x, the probability $\Pr(DY = 1)$ can be defined as,

$$\Pr(DY = 1|x, \alpha, \pi) = (1 - \pi)F(x\alpha) \quad (8)$$

where π is the parameter to accommodate the extra zeros in the data set, α is a vector of parameters and $F(x\alpha)$ the link function. This is a model where a proportion of π individuals in the sample are characterized by a degenerate distribution on *DY=0*. A simple way to model binary data is to use the linear function $F(x\alpha) = x\alpha$. However, the linear probability model (LPM) has heteroskedastic error terms and a major drawback: the predicted values can lie outside the range [0,1]. The usual way to avoid this is to use a non-linear function $F(x\alpha)$ for which popular choices are 'S' curves, that are bounded to the interval [0,1] whatever the values of $x\alpha$. Figure 6 intuitively shows the restriction imposed on $\Pr(DY = 1 | x, \alpha, \pi)$ modelled by an 'S' curve, in the case $\pi \geq 0$.

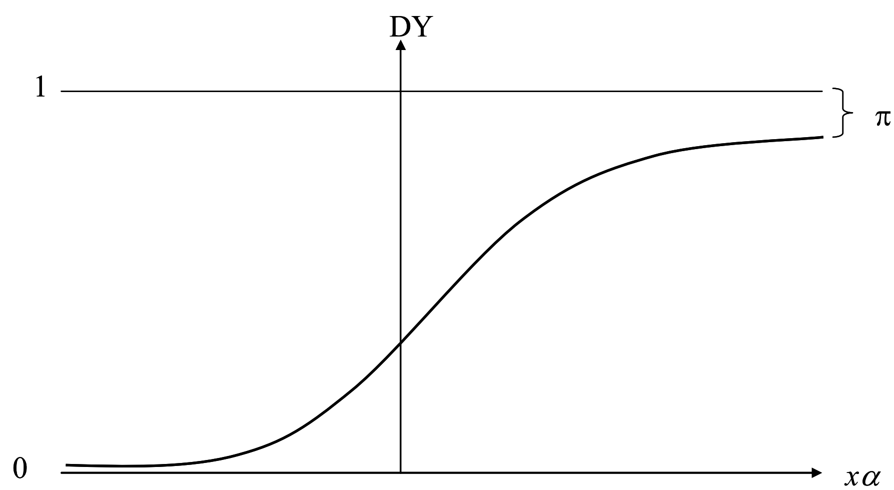

Figure 6. 'S' curve, model specification with asymptote.

In the case π =0, the specification (8) is simply the usual binary choice model, and the relation $0 \leq \Pr(DY = 1 | x, \alpha, \pi) \leq 1$ holds. The inclusion of the π parameter restricts this assumption to: $0 \leq \Pr(DY = 1 | x, \alpha, \pi) \leq (1 - \pi)$. When π >0, the model identifies a proportion π of individuals that do not report expenditure on medicines, independently of need or other characteristics, such as the ill individuals that do not report expenditure over the last two weeks or those entitled to full exemption.

For notation simplicity let F represent $F(x\alpha)$ and *f* the corresponding derivative. From equation in (8), the log-likelihood function individual contribution can be written as follows:

$$
\begin{aligned}
LL(\alpha,\pi) &= (1-DY)\log[1-(1-\pi)F] + DY\log[(1-\pi)F] \\
&= (1-DY)\log[\pi+(1-\pi)(1-F)] + DY\log[F] + DY\log(1-\pi)
\end{aligned} \quad (9)
$$

The first order conditions are,

$$
\begin{aligned}
\frac{\partial LL(\alpha,\pi)}{\partial\alpha} &= (1-DY)\frac{(1-\pi)(-f)}{[\pi+(1-\pi)(1-F)]} + DY\frac{(1-\pi)(f)}{(1-\pi)(F)} \\
\frac{\partial LL(\alpha,\pi)}{\partial\pi} &= (DY-(1-\pi)F)\frac{1}{\pi(1-\pi)-(1-\pi)^2(1-F)}
\end{aligned} \quad (10)
$$

This research starts with the obvious candidates for $F(x\alpha)$: logit, probit and complementary log-log probability models that have been widely employed in empirical applications of standard binary index models. In order to account for asymmetry, power functions of the probability models were also considered. Summarizing, the following link functions $F(x\alpha)$ were assessed,

i) Logit function, $F(x\alpha) = \frac{\exp(x\alpha)}{1+\exp(x\alpha)}$

ii) Probit function: $F(x\alpha) = \Phi(x\alpha)$

iii) Complementary log-log function: $F(x\alpha) = 1-\exp(-\exp(x\alpha))$

iv) Asymmetric logistic function: $F(x\alpha) = \left(\frac{\exp(x\alpha)}{1+\exp(x\alpha)}\right)^{\theta}$, $\theta > 0$

Once discussed the specification of Part One, it is necessary to specify the model for the strictly positive observations under part two of the 2PM.

4.4.2 PART TWO

The conditional mean $E\left[Y|x, y>0\right]$, part two of the 2PM, is modelled based only on the positive outcomes. Therefore, the presence of excess zeros does not have any implication on the modelling of the positive expenditure, which implies that for this part a PPML estimator can be applied. Mullahy (1998) also suggested a modified version of the 2PM, where the conditional expected value can be specified directly as,

$$E\left[Y|x,\beta, y>0\right]=\exp(x\beta) \tag{11}$$

Although the medicines expenditure is clearly non-discrete, the parameters can be consistently estimated using standard Poisson regression models as suggested by Santos Silva and Tenreyro (2006). A modified Park test used by these authors indicated that this specification is adequate relative to other types of distributions, such as the Gamma or inverse-Gaussian. Also, unlike the latter, the Poisson PML specification gives equal weight to extreme values, which is appropriate as data from respondents that are high medicines users should be given equal weight to those from low medicines users (Windmeijer, 2006). The 2PM is finally fully specified as,

$$E\left[Y|x\right]=(1-\pi)F(x\alpha)\exp(x\beta) \tag{12}$$

Due to the two independent processes considered, two vectors of parameters will be estimated: α and β. Hence, it is not straightforward the interpretation of the 2PM results and some issues connected with inference should be also considered.

4.4.3 INFERENCE WITH 2PM

In the majority of the studies, the researchers are satisfied conducting inference directly on the parameters α and β (Mullahy, 1998) without regard to how such inference relate to inferences about quantities such $\delta(x) = \frac{\partial E\left[Y|x\right]}{\partial x}$. However, the central inferential questions of this research involve the conditional expected value, $E[Y|x]$, and its associated partial effects $\delta(x)$ which should be computed since the parameters α and β relate to the linear index. Therefore, unlike with linear models, the partial effects are not given by the coefficients directly, but can be computed from them. In the case of a continuous variable *x,* such as income or pharmacies density, from equation (12) we obtain the marginal effects,

$$\begin{aligned} \delta(x) &= (1-\pi)\left[f(x\alpha)\exp(x\beta) + \beta F(x\alpha)\exp(x\beta)\right] \\ &= (1-\pi)\exp(x\beta)\left[\alpha f(x\alpha) + \beta F(x\alpha)\right] \end{aligned} \qquad (13)$$

where $f(x\alpha) = \frac{\partial F(x\alpha)}{\partial x}$ is the corresponding derivative.

However, in the case of a binary variable, it does not make sense to use marginal effects that measure the impact of small changes, since the variable x_k only takes two values: zero or one. In this case, the partial effect of changing from zero to one can be captured by the difference of the two conditional mean,

$$\begin{aligned} \delta(x_k) &= E\left[Y|x_k = 1\right] - E\left[Y|x_k = 0\right] \\ &= (1-\pi)\left[\left[F(x\alpha)\exp(x\beta)\right]_{x_k=1} - \left[F(x\alpha)\exp(x\beta)\right]_{x_k=0}\right] \end{aligned} \qquad (14)$$

However, both the marginal and partial effect depend on the values of the remaining explanatory variables. A way to deal with this problem is to use average partial effects (APE): compute the effect for each observation, using their specific *x* values, and then report the sample mean of the effects (Jones, 2007).

A note of caution is that the average partial effects give information about the impact on $E[Y|x]$ of changing each variable but holding all the others constant. Hence, should be noted that the APE interpretation is a classical *ceteris paribus* analysis.

In the next section the most appropriate $F(x\alpha)$ function is chosen and, making use of the average partial effects calculated from the 2PM estimated model, inference is made about the questions on horizontal equity, which originally motivated this research.

4.4. ESTIMATION RESULTS

It is possible that, due to the sampling strategy used in the PNHS, observations are independent across households, but not within households. The implication is that if we use estimators that assume independence within these households, the standard errors on the regression coefficients may be too small and we will overestimate the statistical significance of the covariates in the models.

Robust covariance matrix estimator

The robust variance matrix estimator is consistent, and using it in place of the usual estimator means that the econometric model accounts for possible incorrect specification. Notice that:

- In a binary response model it is not possible to correctly specify $E\left[DY|x\right]$and to misspecify $Var\left[DY|x\right]$. Once we have specified $\Pr\left[DY=1|x\right]$, we have specified all conditional moments of *DY* given *x* (Wooldridge, 2002);
- In the second part of 2PM, the assumption $Var\left[Y|x\right] \propto E\left[Y|x\right]$ might not hold and in this case the PPML estimator does not take full account of the heteroskedasticity pattern.

Therefore, all inference was based on the Eicker-White (White, 1980) robust covariance matrix estimator.

Specification test

To check the adequacy of the second part of estimated 2PM, the RESET heteroskedasticity-robust test (Ramsey, 1969) was performed. This is a test for the correct specification of the conditional expectation, performed by checking the significance of an additional variable constructed as $(x'b)^2$, where *b* denotes the vector of estimated parameters. Technical complications arise in the maximization of the likelihood in binary specification of the 2PM with the additional variable $(x'b)^2$, possibly due to the large number of covariates and so it was not possible to perform the RESET test. In this case, due to computational simplicity, the more parsimonious LINK test (Pregibon, 1980; Pregibon, 1981) is attractive and was here used to choose the most appropriate specification for the binary outcome. The linear prediction (*x'b*) and its square $(x'b)^2$ are computed. In the auxiliary regression, the dependent variable DY is regressed using the same specification against the (*x'b*), $(x'b)^2$ and a constant term. If the model is correctly specified then $(x'b)^2$ should have no statistical significance. The parsimonious LINK test has been used frequently to test the specification of non-linear empirical models (Manning, 2005; Morris, 2005).

Estimated Model

All results were obtained using STATA version 9.2 from StataCorp (2006). We used the conventional significance level of 5 per cent and the statistically significant coefficients were highlighted in the tables in bold.

Recalling the especially useful two factors decomposition of $E[Y|x]$ in equation (7), it should be noted that these factors where modelled with two separate processes (2PM): part one $\Pr(Y > 0|x)$ and part two $E[Y|y > 0, x]$.

Unlike the second part of the 2PM, it was complex to obtain a good specification to model $\Pr(Y > 0|x)$. The popular link functions probit, logit, complementary log-log and negative binomial were assessed using the LINK test and all of them were non-appropriate specifications. For completeness, Appendix A contains estimation results using these link functions. The logistic function, accounting for asymmetry, $F(x\alpha) = \left(\frac{\exp(x\alpha)}{1+\exp(x\alpha)}\right)^{\theta}$, $\theta > 0$, was the chosen specification since no evidence of misspecification was found, as reported in Table 8. Bago d'Uva (2002) have also used a modified logit specification to model Portuguese data on the probability of visiting a medical doctor.

The sign of the estimated coefficient tells us about the qualitative effect of the covariates. The size of the coefficient tells us about the quantitative effect. However, since we are using non-linear models, in order to interpret the quantitative effect we need to compute average partial effects, using marginal effects to continuous explanatory variables and partial effects for binary explanatory variables. To compute the marginal effects, we need to incorporate $f(x\alpha)=\frac{\partial F(x\alpha)}{\partial x}$,

$$f(x\alpha)=\theta\left(\frac{\exp(x\alpha)}{1+\exp(x\alpha)}\right)^{\theta-1}\frac{\alpha\exp(x\alpha)(1+\exp(x\alpha))-\exp(x\alpha)\alpha\exp(x\alpha)}{(1+\exp(x\alpha))^{2}}$$

$$=\theta\alpha\frac{\exp(x\alpha)^{\theta}}{(1+\exp(x\alpha))^{\theta+3}}$$

into equation (13), obtaining the basis for the APE calculation for continuous variables,

$$\delta(x)=(1-\pi)\exp(x\beta)\left[\theta\alpha\frac{\exp(x\alpha)^{\theta}}{(1+\exp(x\alpha))^{3}}+\beta\left(\frac{\exp(x\alpha)}{1+\exp(x\alpha)}\right)^{\theta}\right]$$

$$=(1-\pi)\exp(x\beta)(\theta\alpha+\beta)\frac{\exp(x\alpha)^{\theta}}{(1+\exp(x\alpha))^{\theta+3}} \quad (15)$$

For clarity of presentation, the 2PM estimation results are presented here by subsets of variables: non-need variables (Table 7) and need plus major research variables (Table 8). The estimates were computed using the method of maximum likelihood estimation.

The first two columns of Table 7 and Table 8 contain variables group and abbreviation. For clarity purposes, the variables related to ICD-9 are presented by a summary description instead of abbreviation. The third column reports the 2PM part-one results and the last column report the 2PM part-two results.

The results reported include the estimated coefficients and standard error estimates are reported in parenthesis.

Table 7. 2PM, estimation results. Non-need variables.

	Group Variable	Variable	Part One	Part Two
			Coef. (Std. Err.)	Coef. (Std. Err.)
Non Need Variables	Gender	Female	**0.059 (0.023)**	-0.058 (0.03)
	Age	Age2534	**0.169 (0.052)**	0.004 (0.07)
		Age3544	**0.162 (0.054)**	0.048 (0.069)
		Age4554	**0.19 (0.057)**	0.119 (0.072)
		Age5564	**0.28 (0.065)**	0.123 (0.068)
		Age6574	**0.387 (0.077)**	0.099 (0.071)
		Age ≥ 75	**0.415 (0.082)**	0.08 (0.072)
	Portuguese Region	Centro	**0.133 (0.031)**	**-0.104 (0.032)**
		Norte	**0.28 (0.045)**	0.02 (0.029)
		Alentejo	**0.187 (0.041)**	-0.064 (0.034)
		Algarve	**0.258 (0.043)**	0.003 (0.036)
	Marital Status	Single	0.039 (0.032)	**-0.115 (0.034)**
		Widow	0.039 (0.032)	-0.036 (0.027)
		Divorced or separated	0.057 (0.055)	-0.014 (0.049)
	Education	Basic school (until 9th grade)	**0.187 (0.034)**	**0.106 (0.026)**
		Secondary school (10th to 12th grade)	**0.282 (0.057)**	**0.127 (0.059)**
		Graduate school (university degree or more)	**0.376 (0.064)**	**0.201 (0.061)**
	Employment	Otherunemp (includes keeping house, student, undertaking military service, living from rents or interests and other)	-0.018 (0.025)	**0.084 (0.035)**
		Unemployed	0.006 (0.042)	0.042 (0.048)
		Retired	**0.077 (0.03)**	0.025 (0.028)
		Disabled (also includes individuals being confined to bed, to a wheel chair or to their own house)	0.071 (0.264)	-0.002 (0.196)
Other	Time (1999 wave)	T1	**0.151 (0.024)**	**0.15 (0.021)**

The *p-value* for the Wald test is omitted and instead the coefficients statistically significant at 5% level were highlighted in bold. In the last three rows of Table 8 are reported the number of observations, the log-pseudo-likelihood and the *p-value* for the LINK and RESET test.

Table 7 shows that the determinants of the binary decision are different from those of the amount of expenditure once the hurdle is crossed. In fact, the same variable can have different sign or different significance in each of the two parts. For example, the gender effect, conditional on need, has statistical significance in both processes, has a positive sign in the probability of having medicines expenditures but once in expenditure has a negative sign. As frequently reported in other health care utilisation empirical studies, we find increased probability of individual medicines expenditure for females. However, conditional on particular need and other covariates, once in pharmacotherapy, females have lower medicines expenditure mean. The results presented in Table 7 also show a significant, positive and increasing gradient in the probability of having medicines expenditure as the population ages as depicted in Figure 7.

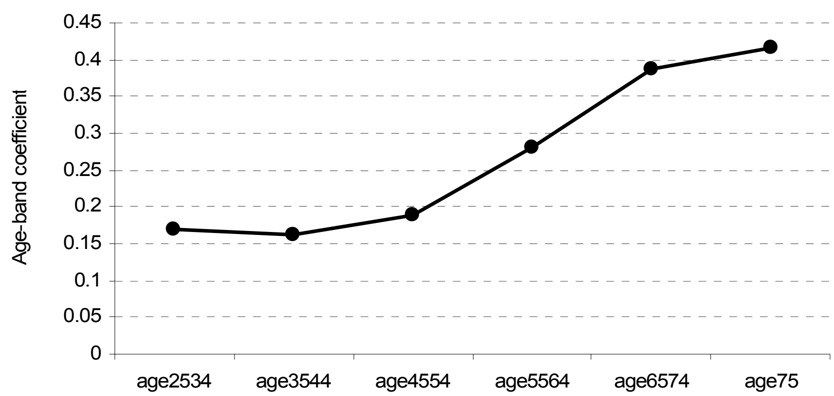

Figure 7. 2PM Part One: age-band estimated coefficients.

However, the age factor, conditional on need, is not statistical significant to determine the amount of medicines expenditure, once the hurdle is crossed. In fact, using a comprehensively set of covariates, it is reasonable to accept no differences in medicines expenditure conditional expectation across the age groups.

The Portuguese region also plays a role in the probability of medicines expenditures. Lisboa and Vale do Tejo is the reference class and the estimated coefficients for the remaining regions are positive and statistically significant. Once again, we find different results once the hurdle is crossed. Conditional on need, individuals that live in the Centro region have a lower medicines expenditure expectation. In the remaining regions (Norte, Alentejo and Algarve) the conditional expectation it is not statistically different from those individuals that live in the Lisboa and Vale do Tejo region.

Marital status was found without any relevance in the specification of probability of medicines expenditure, conditional on need. However, conditional on need, being single determines a lower conditional expectation. However, the results showed a significant positive and increasing gradient in medicines expenditure as the population education levels grows. This occurs in both of the two parts of the 2PM specification.

The occupational status is also a determinant and being retired is a positive and significant factor to have a higher probability of medicines expenditure, conditional on need, relative to the reference individual (who is employed). However, being retired is not significant in the second part and was found positive significance for been unemployed for other reasons. This means that an individual that is not working and whose ‘employment status’ was keeping house, student, undertaking military service, living from rents or interests, or was not working due to other unspecified reasons, has a higher medicines expenditure conditional expectation, relative to the reference individual who is employed.

The time dummy shows a positive and highly statistically significant effect. This sign was expected given the ongoing medicines technological advances, which penetrate the pharmaceutical market usually at higher prices.

Table 8 reports the 2PM coefficient estimates and standard errors for need and major research variables.

Table 8. 2PM, estimation results. Need and major research variables.

	Group Variable	Variable	Part One	Part Two
			Coef. (Std. Err.)	Coef. (Std. Err.)
Need Variables	Acute Illness (number of days in the last 2 weeks with limitations in daily activities due to illness or injury)	Mildly hampered (0 < days ≤ 3)	**0.822 (0.107)**	-0.024 (0.029)
		Moderately hampered (3 < days ≤ 13)	**1.086 (0.164)**	**0.119 (0.031)**
		Severely hampered (14 days)	**0.592 (0.082)**	**0.215 (0.064)**
	Illnesses and Chronic Conditions (ICD-9M classification)	Accident	0.116 (0.076)	**-0.133 (0.049)**
		Endocrine and Metabolic	**0.072 (0.034)**	**0.078 (0.03)**
		Mental Disorders	**0.235 (0.075)**	**0.191 (0.045)**
		Nervous System	**0.14 (0.052)**	-0.029 (0.041)
		Circulatory System	**0.317 (0.041)**	**0.095 (0.02)**
		Respiratory System	**0.335 (0.045)**	-0.015 (0.019)
		Digestive System	**0.207 (0.044)**	**0.081 (0.032)**
		Genitourinary System	**0.068 (0.033)**	0.001 (0.025)
		Musculoskeletal System	**0.154 (0.025)**	-0.002 (0.029)
		Symptoms and Signs	**0.395 (0.048)**	**0.12 (0.024)**
		Injury and Poisoning	**0.185 (0.034)**	0.028 (0.022)
		Other Diseases	**0.124 (0.043)**	0.096 (0.063)
	Visit medical doctor	Visitmd	**0.721 (0.087)**	**0.333 (0.028)**
	Self-Assessed Health (SAH)	Good	-0.067 (0.057)	0.102 (0.086)
		Fair	**0.133 (0.062)**	**0.276 (0.087)**
		Poor	**0.267 (0.073)**	**0.431 (0.088)**
		Very Poor	**0.433 (0.094)**	**0.506 (0.09)**
	Healthy Lifestyles	Exercise	**-0.082 (0.022)**	-0.012 (0.021)
		Non obese	-0.018 (0.018)	-0.039 (0.022)
		Non smoker	-0.004 (0.025)	-0.011 (0.047)
		Non alcohol heavy behaviour	**0.152 (0.03)**	**0.121 (0.025)**
Major Research Variables	**Suplementary health insurance**	JRHI - Job-related health insurance	-0.013 (0.027)	**0.161 (0.043)**
		PHI - Private health insurance	0.063 (0.043)	**0.136 (0.055)**
	Income	LNINC - Logarithm of net montly equivalised income	**0.072 (0.02)**	**0.084 (0.02)**
	Pharmacies density	PHD - Number pharmacies/ 4000 inhabitants, district level	**0.056 (0.019)**	-0.005 (0.018)
		_cons	-0.155 (1.467)	**1.652 (0.163)**
		π	**0.137 (0.038)**	
		θ	**9.231 (3.159)**	
Number of observations			41601	13048
Log PseudoLikelihood			-21196.646	-136415.490
LINK/RESET test p-value			0.902	0.836

As expected, the factors included in the models to account for the need for medicines are all positive and highly statistically significant. Exceptions occur in those who suffered an accident, those who reported good health and in some lifestyle indicators.

Those who suffered an accident have no different probability of medicines expenditure but once crossed the hurdle, have a lower medicines expenditure conditional expectation, relative to those that suffered no accident. In fact, the pharmacological need induced by an accident can be mostly accounted for in the hospital setting and it can be reasonable to accept this result to the primary care setting.

The self-assessed health (SAH) is also a determinant and those reporting fair, poor or very poor health have a significant higher probability of medicines expenditure, relative to the reference individual that reports very good health. However, reporting good health is not statistically different from those reporting very good health as showed in Table 8 and depicted in Figure 8. Identical interpretation can be given to the estimated coefficient of reporting good health in second part of the 2PM.

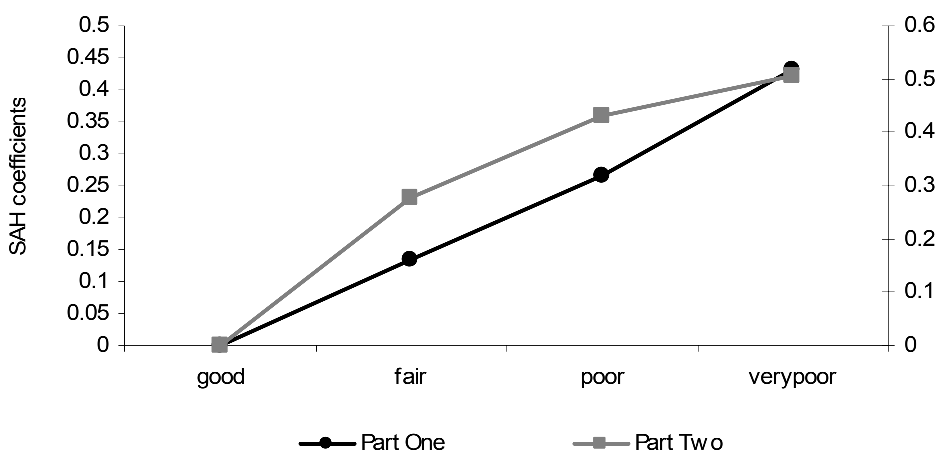

Figure 8. 2PM: SAH categories estimated coefficients.

However, as showed in Figure 8, the effect of moving from the very good health category to another one has different effects according to the two processes of 2PM.

The indicators of obesity and smoking behaviour are not determinants in none of the two processes of the 2PM model. However, the practise of exercise induces a lower probability of medicines expenditure.

In general, the interpretation of the impact of alcohol consumption on health is not simple due to the evidence that moderate consumption gives some positive effects on health (Balia and Jones, 2007). The results show that prudent alcohol consumption behaviour induces a higher probability of medicines expenditure and a higher positive medicines expenditure conditional expectation, relative to heavy drinkers. The reasons to heavy drinker behaviour lie in the grounds of psychology and it is complex to interpret this result, given the research context. Heavy drinkers can have lower positive medicines expenditure conditional expectation, conditional on need, due to several reasons with different basis such as severe non-compliance with medical prescription. Unlike this relation, the relationship between alcohol consumption and illicit drugs has been widely studied in the literature.

Finally, the parameters of major interest in this analysis are the coefficients of the variables related to supplementary health insurance (JRHI and PHI), income (LNINC) and access to pharmacies (PHD).

The supplementary health insurance effect is not statistically significant in the probability of medicines expenditure, conditional on need and other factors. However, having supplementary health insurance increases the positive medicines expenditure conditional expectation, relative to the reference individual that is covered only by the Portuguese NHS. The effect of JRHI and PHI is not evident in both parts of the 2PM but the clear, positive and statistically significant effect on $E\left[Y|y>0,x\right]$ will lead to a positive effect on $E\left[Y|x\right]$.

Table 8 shows that increasing income results in greater probability of medicines expenditure, conditional on need, and that the income effects on $E\left[Y|y>0,x\right]$ are also positive and statistically significant at the 5% level. Thus there is evidence of 'pro-rich' inequality in medicines expenditure.

It is realistic to admit that individuals that live in areas with greater number of pharmacies have greater utilisation of medicines. However, if we believe that the use of medicines should not be affected by access, then a PHD positive and statistically significant coefficient is evidence of horizontal inequity. This would imply that two individuals with the same morbidity should have the same medicines expenditure irrespective of whether they live very close to a pharmacy or whether they live in an area with lower pharmacies density. Table 8 shows that increasing the number of pharmacies density results in greater probability of medicines expenditure, conditional on need, although the PHD effect on $E\left[Y|y>0,x\right]$ are not statistically significant at the 5% level. However, the positive effect of PHD on $\Pr\left(Y>0|x\right)$ will lead to a positive effect on $E\left[Y|x\right]$.

The estimate of π is 13.7%, highly significant, corroborating the existence of excess of zeros in private medicines expenditure for the self-respondents sample.

Unlike other tested specifications for the binary model $\Pr\left(Y>0|x\right)$, with the modified asymmetric LOGIT there is no evidence of non-linearity in the index $x\alpha$ since the LINK test statistic has a p-value of 0.902. The second part was also well modelled using the Poisson pseudo maximum likelihood aproach since there is no evidence of misspecification and the chi-squared statistic for the RESET test is 0.04 with a p-value of 0.836.

Hence, the 2PM modified specification is valid for inference and the estimation results allow us to conclude about evidence for income-related inequity, pharmacies-density inequity and supplementary insurance inequity. In order to interpret quantitative implications of these findings we need to compute partial effects, using marginal effects for continuous explanatory variables (LNINC and PHD) and partial effects for binary variables (JRHI and PHI). The marginal/partial effects will be computed making use of equations (14) and (15).

Table 9 reports the estimated average partial effects (APE) for the binary variables JRHI and PHI and the marginal effects for the continuous explanatory variables LNINC and PHD.

Table 9. Average partial effects, major research variables.

	Group Variable	Variable	Average Partial Effect (sd)
Major Research Variables	Suplementary health insurance	JRHI - Job-related health insurance	1.295 € (1.17)
		PHI - Private health insurance	1.101 € (1.00)
	Income	LNINC - Logarithm of net monthly equivalised income	0.646 € (0.58)
	Pharmacies density	PHD - Number pharmacies/ 4000 inhabitants, district level	0.004 € (0.00)

As expected from the coefficients signal, all the average partial effects are positive. Having job-related health insurance will lead to an average positive effect on $E\left[Y|x\right]$ of 1.295€. The existence of private health insurance is also associated with an increase of the private medicines expenditure conditional expectation in 1.101€. Statistically significant, there is evidence of income semi-elasticity related to medicines expenditure. A small change of one per cent on equivalised income results in an effect on $E\left[Y|x\right]$ of 0.65€, equivalent to 8.5% of average medicines expenditure. However, a reformulation of PNHS in order to measure individual income with more accuracy or to increase the number of household-income-bands is desirable.

Table 9 also reports that one additional pharmacy *per* 4000 inhabitants in the residence district leads to a rather small positive effect of 0.004€ on the medicines expenditure conditional expectation.

5. CONCLUSION

The Portuguese National Health System, states that the access to health care should depend mainly on need. Conditional on need, access to medicines should not depend on socio-economic factors such as income or geographical factors such as the access to pharmacies. The PNHS data set was coded to account comprehensively for a set of morbidity measures using ICD-9M, the net household income was equivalised and external information on pharmacies density was added. This rich data set offers a number of advantages over those used in earlier studies using PNHS data. The detailed information on morbidity can avoid that the estimated effects of the variables are due to their correlation with omitted morbidity variables. The external information on pharmacies density also reduces the risk of omitted variable bias from this source.

An important remark is that 13.7% of the zero values reported on medicines expenditure is independent of the need reported in the morbidity measures. This excess of zeros in the dependent variable has clearly increased the model complexity and induced the adoption of a 2PM. The Poisson pseudo maximum estimator was an appropriate choice to the second part of 2PM and presents a clearly and powerful alternative to the 'misused' log OLS widely adopted in previous studies.

There is evidence of income inequity related to medicines private expenditure. A small change of 1% cent on equivalised income results in an positive and statistically significant average effect of 8.5% on medicines expenditure. This research demonstrated inequity in Portuguese private medicines expenditures with respect to supplementary health insurance, income and pharmacies density. Nonetheless, there are a number of limitations in the research that should be noticed.

First, due to measurement error in the dependent variable, we restricted the econometric analysis to the self-respondents sub-sample. Econometric models accounting for measurement error should be used in further research in order to use the PNHS full sample. A reformulation of PNHS in order to measure medicines expenditure more accurately is also highly recommended.

Second, it was not possible to include NHS medicines reimbursement levels although it can be expected that the extra zeros captured by the modified binary model can account partially for this omitted variable. The medicines prices are also not accounted for, due to lack of data.

Third, the possible endogeneity of the decision to purchase insurance (classical example of endogeneity in health economics) was not taken into account since it would require the use of instruments related to risk aversion of the individuals that is not observable. It should be also noticed that the PNHS data set does not allow identifying the purchaser of voluntary private health insurance (the individual or the employer).

Fourth, the measures of morbidity are predominantly based on self-reported health, which may be measured with errors that can be correlated with medicines expenditure.

Fifth, there may be reverse causality between medicines expenditure and morbidity. The last three study limitations also affect the majority of health care use empirical studies published in the literature.

Nevertheless, the econometric results provide first and new evidence on inequity in private medicines expenditure in Portugal. Since the extent of inequity can vary by subgroup, it may be difficult for the Health Ministry to devise policies to correct for this inequity. Whether the effects of these variables increase or decrease following the recent reforms in the Portuguese pharmaceutical market is an important question for future research, which can be studied using the most recent PNHS data, collected in 2005/ 2006.

APPENDIX A

This appendix contains estimation results with alternative $F(x\alpha)$ (Part One):

Logit function: $F(x\alpha)=\dfrac{\exp(x\alpha)}{1+\exp(x\alpha)}$;

Probit function: $F(x\alpha)=\Phi(x\alpha)$

Complementary log-log function: $F(x\alpha)=1-\exp(-\exp(x\alpha))$.

Table A.1 Part-one, estimation results, alternative models. Non-need variables.

	Group Variable	Variable	LOGIT	PROBIT	CLOG-LOG
			Coef. (Std. Err.)	Coef. (Std. Err.)	Coef. (Std. Err.)
Non Need Variables	Gender	Female	**0.1 (0.041)**	**0.06 (0.024)**	**0.072 (0.033)**
	Age	Age2534	**0.337 (0.09)**	**0.19 (0.051)**	**0.281 (0.073)**
		Age3544	**0.343 (0.092)**	**0.188 (0.052)**	**0.291 (0.075)**
		Age4554	**0.4 (0.093)**	**0.22 (0.053)**	**0.343 (0.076)**
		Age5564	**0.574 (0.095)**	**0.32 (0.054)**	**0.482 (0.077)**
		Age6574	**0.776 (0.101)**	**0.438 (0.058)**	**0.645 (0.082)**
		Age ≥ 75	**0.808 (0.112)**	**0.461 (0.064)**	**0.662 (0.091)**
	Portuguese Region	Centro	**0.246 (0.05)**	**0.143 (0.029)**	**0.191 (0.04)**
		Norte	**0.531 (0.048)**	**0.307 (0.028)**	**0.423 (0.039)**
		Alentejo	**0.351 (0.061)**	**0.205 (0.036)**	**0.284 (0.049)**
		Algarve	**0.449 (0.063)**	**0.27 (0.036)**	**0.344 (0.051)**
	Marital Status	Single	0.073 (0.061)	0.043 (0.035)	0.061 (0.05)
		Widow	0.082 (0.057)	0.045 (0.033)	0.07 (0.047)
		Divorced or separated	0.11 (0.1)	0.063 (0.058)	0.095 (0.08)
	Education	Basic school (until 9th grade)	**0.338 (0.048)**	**0.198 (0.028)**	**0.259 (0.039)**
		Secondary school (10th to 12th grade)	**0.521 (0.085)**	**0.305 (0.049)**	**0.413 (0.069)**
		Graduate school (university degree or more)	**0.677 (0.091)**	**0.399 (0.052)**	**0.523 (0.074)**
	Employment	Otherunemp (includes keeping house, student, undertaking military service, living from rents or interests and other)	-0.027 (0.047)	-0.017 (0.027)	-0.016 (0.038)
		Unemployed	0.028 (0.08)	0.01 (0.046)	0.018 (0.066)
		Retired	**0.134 (0.053)**	**0.078 (0.031)**	**0.094 (0.042)**
		Disabled (also includes individuals being confined to bed, to a wheel chair or to their own house)	0.104 (0.513)	0.063 (0.289)	0.032 (0.424)
	Time (1999 wave)	T1	**0.272 (0.033)**	**0.162 (0.019)**	**0.216 (0.026)**

Table A.2 Part-one, estimation results, alternative models. Need and major research variables.

	Group Variable	Variable	LOGIT	PROBIT	CLOG-LOG
			Coef. (Std. Err.)	Coef. (Std. Err.)	Coef. (Std. Err.)
Need Variables	Acute Illness (number of days in the last 2 weeks with limitations in daily activities due to illness or injury)	Mildly hampered (0 < days ≤ 3)	**1.447 (0.094)**	**0.857 (0.056)**	**1.116 (0.07)**
		Moderately hampered (3 < days ≤ 13)	**1.995 (0.177)**	**1.163 (0.103)**	**1.535 (0.132)**
		Severely hampered (14 days)	**0.977 (0.102)**	**0.589 (0.06)**	**0.708 (0.072)**
	Illnesses and Chronic Conditions (ICD-9M classification)	Accident	0.187 (0.135)	0.115 (0.079)	0.141 (0.106)
		Endocrine and Metabolic	0.114 (0.058)	**0.071 (0.035)**	0.085 (0.045)
		Mental Disorders	**0.426 (0.124)**	**0.247 (0.073)**	**0.332 (0.098)**
		Nervous System	**0.249 (0.087)**	**0.149 (0.052)**	**0.189 (0.067)**
		Circulatory System	**0.55 (0.043)**	**0.329 (0.025)**	**0.416 (0.034)**
		Respiratory System	**0.584 (0.052)**	**0.347 (0.03)**	**0.438 (0.041)**
		Digestive System	**0.354 (0.069)**	**0.215 (0.041)**	**0.273 (0.054)**
		Genitourinary System	**0.115 (0.057)**	**0.07 (0.034)**	0.081 (0.045)
		Musculoskeletal System	**0.272 (0.036)**	**0.163 (0.021)**	**0.217 (0.029)**
		Symptoms and Signs	**0.673 (0.047)**	**0.403 (0.027)**	**0.497 (0.036)**
		Injury and Poisoning	**0.333 (0.048)**	**0.196 (0.028)**	**0.255 (0.038)**
		Other Diseases	**0.229 (0.072)**	**0.134 (0.043)**	**0.175 (0.057)**
	Visit medical doctor	Visitmd	**1.345 (0.038)**	**0.787 (0.022)**	**1.133 (0.031)**
	Self-Assessed Health (SAH)	Good	-0.097 (0.114)	-0.065 (0.064)	-0.06 (0.096)
		Fair	**0.31 (0.115)**	**0.164 (0.064)**	**0.286 (0.097)**
		Poor	**0.545 (0.122)**	**0.305 (0.069)**	**0.471 (0.102)**
		Very Poor	**0.816 (0.146)**	**0.466 (0.082)**	**0.648 (0.118)**
	Healthy Lifestyles	Exercise	**-0.151 (0.037)**	**-0.088 (0.021)**	**-0.122 (0.03)**
		Non obese	-0.03 (0.034)	-0.018 (0.02)	-0.019 (0.027)
		Non smoker	0 (0.048)	-0.001 (0.027)	0.009 (0.039)
		Non alcohol heavy behaviour	**0.298 (0.039)**	**0.169 (0.022)**	**0.244 (0.031)**
Major Research Variables	**Suplementary health insurance**	JRHI - Job-related health insurance	-0.025 (0.05)	-0.014 (0.029)	-0.021 (0.04)
		PHI - Private health insurance	0.123 (0.079)	0.07 (0.046)	0.1 (0.063)
	Income	LNINC - Logarithm of net montly equivalised income	**0.125 (0.034)**	**0.075 (0.02)**	**0.098 (0.028)**
	Pharmacies density	PHD - Number pharmacies/ 4000 inhabitants, district level	**0.111 (0.031)**	**0.063 (0.018)**	**0.09 (0.025)**
		_cons	**-4.683 (0.252)**	**-2.737 (0.144)**	**-4.1 (0.207)**
		π	**0.252 (0.015)**	**0.248 (0.015)**	**0.307 (0.012)**
Number of observations				41601	
Log PseudoLikelihood			-21215.100	-21203.170	-21251.690
LINK test p-value			0.000	0.007	0.000

REFERENCES

Bago d'Uva T (2002). Informação assimétrica no mercado de seguros Português. Dissertação de mestrado em Ciências Actuariais. Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa.

Bago d'Uva T (2005). Latent class models for use of primary care: evidence from a British panel. *Health Economics.* Sep; 14(9):873-92.

Balia S, Jones AM (2007). Mortality, Lifestyle and Socio-Economic Status. *Journal of Health Economics, in press.*

Barros PP (1998). The black box of health care expenditure growth determinants. *Health Economics*. Sep;7(6):533-44.

Barros PP, Machado MP, Galdeano AS (2005). Moral Hazard and the demand for Health services: a matching estimator approach, *Working paper Economics Series 28*, pp.05-59.

Bernardo FF (2007). Impacto do heaping em modelos para dados de duração. Dissertação de mestrado em Econometria Aplicada e Previsão. Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa.

Colombo F, Tapay N (2004). Private Health Insurance in OECD Countries. The Benefits and Costs for Individuals and Health Systems. *OECD Health Working Paper* No. 15. (www.oecd.org/dataoecd/34/56/33698043.pdf)

Contoyannis P, Jones AM (2004). Socio-economic status, health and lifestyle. *Journal of Health Economics*. Sep; 23(5):965-95.

Deb P, Holmes AM, (2000). Estimates of use and costs of behavioural health care: a comparison of standard and finite mixture models. *Health Economics*, 9, 6.

Deb P, Trivedi PK (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal Of Applied Econometrics*, vol. 12, 313-336.

Deb P, Trivedi PK (2002). The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics* 21 (2002) 601–625.

Diário da República (1979). Law 56, Sep. 15.

Diário da República (1989). Law 46, 1a , Feb. 15.

Diário da República (1992). Law 118, 1a , Jun. 25.

Diário da República (1999). Law 19064, 2a , Oct. 06.

Diário da República (2004). Law 28/2004, Feb. 04.

Diário da República (2005). Law 129, 1a A, Aug. 11.

Diário da República (2005). Law 146/2005, Aug 26.

Diário da República (2006). Law 127, 1a 167, Jul. 04.

Diário da República (2006). Law 176, 1a 167, Aug. 30.

Diário da República (2006). Law 53-A, 2a , Dez. 29.

Diário da República (2007). Law 30-B, 1a, Jan. 05.

Giraldes, MR (2003). Private health insurance in Portugal: a comparative analysis of the National Health Surveys, 1995-1996 and 1998-1999, *Cadernos Saúde Pública*, Jan-Feb;19(1):263-8.

Gouveia Pinto C, Aragão F (2004). Rationing Health Care in Europe – Portugal in Rationing Medical Services in Europe: An Empirical Study. Amsterdam: IOS Press, pp. 149-196.

Gouveia Pinto C, Teixeira I (2002). Pricing and reimbursement of pharmaceuticals in Portugal. *The European Journal of Health Economics*; 3(4):267-70.

Grootendorst PV (1995). A comparison of alternative models of prescription drug utilisation. *Health Economics* 4, 183–198.

Heitjan DF, Rubin DB (1990). Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association*, 85 (410), 304-314.

Idler E,. Benyamini Y (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior* 38(1): 21-37

Idler EL, Kasl SV (1995). Self-ratings of health: do they also predict change in functional ability? *Journal of Gerontology* 50B: 344-353

INE (2006), INE Infoline Service. (www.ine.pt)

INFARMED (2001) Statistical Information. Ministry of Health. Lisboa.

Jones A, Koolman X, Rice N (2004). Health-related attrition in the BHPS and ECHP: using inverse probability weighted estimators in nonlinear models. *ECuity III Working Paper* 18

Jones A (2007a). Applied Econometrics for Health Economists. Radcliffe Medical Publishing.

Jones A, Van Doorslaer E, Koolman X (2007). The Impact of Supplementary Private Health Insurance on the Use of Specialists in European Countries, *Annales d'Economie et de Statistiques,* Vol. 83, April.

Jürges, H (2005). Unemployment, retrospective error, and life satisfaction. MEA discussion paper series 05089, Mannheim Research Institute for the Economics of Aging, University of Mannheim.

Lourenço OD, Ferreira PL (2004). Effects of income on the demand for dental health care. 5th European Conference on Health Economics. London, UK.

Lourenço OD, Ferreira PL, Barros PP (2006) Econometric Analysis of Health Care Utilization: An Alternative Hurdle Model Using Latent Class Models. 15th European Workshop on Econometrics and Health Economics. Thessaloniki, Greece.

Manning WG, Mullahy J (2001). Estimating log models: to transform or not to transform? *Journal of Health Economics*. Jul;20(4):461-94.

Manning GW, Basu A, Mullahy J (2005). Generalized modelling approaches to risk adjustment of skewed outcomes data, *Journal of Health Economics* 24, pp. 465–488

Marques-Vidal P, Dias CM (2005). Trends in overweight and obesity in Portugal: the National Health Surveys 1995-6 and 1998-9, *Obesity Research*, Jul;13(7):1141-5.

Marques-Vidal P, Dias CM (2005a). Trends and Determinants of Alcohol Consumption in Portugal: Results from the National Health Surveys 1995 to 1996 and 1998 to 1999, *Alcoholism: Clinical and Experimental Research*, Vol. 29, No 1, pp 89–97.

Marques-Vidal P, Dias CM (2006). Hypnotic consumption in the Portuguese population: data from the National Health Survey 1998-1999, *Pharmacoepidemiology Drug Safety.* Jan 15(1):63-9.

Marques-Vidal P., Goncalves A. and Dias C.M. (2006b). Milk intake is inversely related to obesity in men and in young women: data from the Portuguese Health Interview Survey 1998-1999, *International Journal of Obesity*. Jan;30(1): 88-93.

Marques-Vidal P, Ravasco P, Dias CM, Camilo ME (2006a). Trends of food intake in Portugal, 1987-1999: results from the National Health Surveys, *European Journal of Clinical Nutrition*, Jun 21.

Mullahy J (1997). Heterogeneity, excess zeros, and the structure of count data models. *Journal Of Applied Econometrics*, vol. 12, 337-350.

Mullahy J (1998). Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 17, 247–281.

Vaz JP, Mendes Z, Félix J, Inês M, Saramago M, Soares M (2006). Observatory of Pharmacies - Newsletter. (http://www.anf.pt/site/index.php?page=data/anf/cefar.php)

Ministério da Saúde (1998), Inquérito Nacional de Saúde - Manual do Entrevistador, Portugal.

Ministério da Saúde (2001) Inquérito Nacional de Saúde - Nota Metodológica.

Oliveira MD, Gouveia Pinto C (2005). Health care reform in Portugal: an evaluation of the NHS experience*. Health Economics*. Sep;14(Suppl 1):S203-20.

Pereira JA (1990). Equity objectives in Portuguese health policy, *Social Science and Medicine*, 31(1): 91/4.

Pereira JA, Campos AC, Ramos F, Simões J, Reis V (1999). *Health care reform and cost containment in Portugal*. In: E. Mossialos and J. Le Grand (*eds*.) Health Care and Cost Containment in the European Union, Ashgate, Aldershot, 1999.

Portuguese Government (2006). Orçamento de Estado para 2007.

Pregibon D (1980). Goodness of link tests for generalized linear models. *Applied Statistics*, 29, 15-24.

Pregibon D (1981). Logistic regression diagnostics. *Annals of Statistics* 9, 705–724.

Ramsey JB, (1969). Tests for Specification Errors in Classical Linear Least Squares Regression Analysis, *Journal of the Royal Statistical Society*, B, 31, 350-371.

Santos Silva JMC, Tenreyro S (2006). The log of gravity. *The Review of Economics and Statistics*. November 2006, 88 (4): 641-658.

Street A, Jones A, Furuta A (1999). Cost-sharing and pharmaceutical utilisation and expenditure in Russia, *Journal of Health Economics* 18, pp. 459–472.

Van Doorslaer E, Koolman X, Jones AM (2004). Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics*. 13(7): 629-47.

Van Doorslaer E, Koolman X (2004a). Explaining the differences in income-related health inequalities across European countries. *Health Economics*. 13(7):609-28.

White H (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* 48 (1980), 817–838.

Windmeijer F, Kontodimas S, Knapp M, Brown J, Haro JM (2006). Methodological approach for assessing the cost-effectiveness of treatments using longitudinal observational data: the SOHO study. *International Journal of Technology Assessment in Health Care*. 22(4):460-8.

Winkelmann R (2004). Health care reform and the number of doctor visits—an econometric analysis. *Journal of Applied Econometrics*. 19: 455–472.

Wolff J, Augustin T (2003): Heaping and its consequences for duration analysis – a simulation study. *Allgemeines Statistisches Archiv – Journal of the German Statistical Association* 87, 1-28.

Wooldridge JM (2002). *Econometric analysis of cross-section and panel data*. Cambridge, MA: Massachusetts Institute of Technology Press.