Unveiling health care consumption groups

a latent class approach in the Portuguese health data context

Óscar Manuel Domingos Lourenço

A thesis presented to the University of Coimbra in fulfilment of the requirement for the degree of Doctor in Philosophy in Mathematical Economics and Econometric Models, under the supervision of Professors Pedro Pita Barros and Pedro Lopes Ferreira.

Coimbra 2007

Acknowledgements

After all these years working in this dissertation, I have a long list of people who contributed in some way to it, to whom I would like to express my thanks. Surely, I can never show enough appreciation to all of them, however, I hope that these few words can express some of my gratitude. The first ones in my list are my supervisors, Pedro Pita Barros and Pedro Lopes Ferreira, for their inestimable professional and academic guidance, as well as for the support, encouragement, and friendship they always have showed.

My gratitude extends to João Santos Silva, José Murteira, Mattew Sutton, Pravin Trivedi, Tom van Ourti and Teresa Bago d'Uva for helpful discussions and suggestions about some aspects of the empirical applications.

A word of appreciation also goes to Teresa Lello and Carlota Quintal who proofread parts of previous versions of the text.

Faculdade de Economia da Universidade de Coimbra (FEUC), Centro de Estudos e Investigação em Saúde da Universidade de Coimbra (CEIS-UC), Fundação Calouste Gulbenkian (FCG), Fundação para a Ciência e Tecnologia (FCT) provided financial support to attend international conferences. The Observatório Nacional de Saúde (ONSA) provided the Portuguese National Health Survey, and the Instituto da Qualidade em Saúde (IQS) provided the Europep dataset.

I also thank my friends António Rodrigues, Isabel Clímaco, Joana Dias, Pedro Godinho, Paulo Melo and Vitor Raposo for the continuous encouragement and friendship demonstrated over the years. I cannot forget Mrs. Guiomar Fortunato for the excellent atmosphere provided in the office where I wrote most of this dissertation.

And last but not least, a special word to my wife Lurdes, for the incredible amount of patience she had had to me, especially in the last year and a half. Many projects were postponed in the name of this PhD. In this context, my son David needs a special mention, too. I owe him hours of play and bicycle riding.

Abstract

This thesis aims at understanding the processes underlying the consumption of medical care. It begins by exploring the possibility of unveiling (latent) health care consumption groups, specific to the population of actual users of health care, followed by generating insightful knowledge regarding a number of health economics issues relevant to the Portuguese health system.

Thus far, all empirical applications developed to identify latent health care consumption groups, have adopted specifications of the latent class family that split the overall population into two latent groups. Surprisingly enough no attention has been given to latent class models whose specification splits the population of health care users into latent groups. Such analysis can be relevant under two perspectives. Either on its own because it provides findings that reflects the patterns of utilization of actual users by latent group of actual users, or as a part of extensions of econometric specifications, for instance the hurdle model.

In the Portuguese health system a number of health economics related issues are of interest, e.g. the effect of the supplementary health insurance, the analysis of the patterns of health care utilization across urban and rural areas, the effect of time-costs, etc. We use two datasets with data at an individual level, combined with appropriate methodological tools producing knowledge about such issues.

One main finding regards the importance of identifying actual health care users with different levels of utilization, thus, different health care consumption groups. Moreover, our findings, based on actual data, revealed that the latent class model specified to split the users population into latent classes, when combined with a binary model aiming at explaining the probability of some medical care utilization, proved to outperform some of the most common econometric specifications usually adopted to analyse medical care utilization.

From a health policy perspective, our results provided various insights potentially useful for the Portuguese health system, from which we highlight four of them. First, supplementary health insurance does not affect medical care utilization among the actual users; second, actual health care users present a similar frequency of doctor visits regardless of residing in urban or rural areas; third users of primary medical care provided by public health centres are not reactive to the total time required to visit the doctor; and forth the occurrence of a delay in the appointment for the latent group of low users of public primary renders a decrease in the utilization of public primary care.

Contents

CHAPTER 1

GENERAL INTRODUCTION 13

1.1 — Motivation and Purpose of the Study 13

1.2 — Outline of the Thesis 15

CHAPTER 2

COUNT DATA MODELS FOR THE ANALYSIS OF HEALTH CARE UTILIZATION

GENERAL OVERVIEW 19

2.1 — Introduction 19

2.2 — The Poisson Regression Model 22

2.3 — Mixture Models to Incorporate Unobserved Heterogeneity 24

2.4 — The Negative Binomial Regression Model 26

2.5 — Latent Class Models 29

2.5.1 — LCM and posterior class analysis 34

2.6 — The Hurdle Regression Model 35

2.6.1 — On the appropriateness of the hurdle specification to explain health care utilization: some comments 42

2.7 — Models for Truncated Counts 45

2.8 — Specification of Latent Class Models for Truncated Samples 47

2.9 — Options Concerning LCM Specification and Estimation 51

2.10 — Model Comparison and Selection Procedures 55

2.10.1 — Likelihood Ratio Tests (Lr Tests) 56

2.10.2 — Voung Tests (Vut) 57

2.10.3 — Information Criteria (Ic) 59

2.11 — Interpretation of the Parameters in Nonlinear Regression Models 60

CHAPTER 3

SOME REMARKS ABOUT THE PORTUGUESE HEALTH CARE SYSTEM 67

CHAPTER 4

ECONOMETRIC ANALYSIS OF HEALTH CARE UTILIZATION: AN ALTERNATIVE HURDLE SPECIFICATION USING LATENT CLASS 83

4.1 — Introduction 83

4.2 — The Relevant Literature 85

4.3 — An Alternative Hurdle Formulation Using LCMs 87

4.3.1 — LCM for the health care users and posterior class analysis 97

4.4 — Data and Variables 98

4.5 — Analysis and Discussion of the Results 110

4.5.1 — Evaluation of fitted models: Does LCH-Cross hurdle outperform the competing alternatives? 111

4.5.2 — Do different models provide different evidence? 126

4.6 — General Conclusions 136

CHAPTER 5

THE UTILIZATION OF MEDICAL CARE IN PORTUGAL: EFFECT OF SUPPLEMENTAL INSURANCE AND PLACE OF RESIDENCE 141

5.1 — Introduction 141

5.2 — A personal Roadmap to the Most Relevant Literature 147

5.3 — Dependent Variables, Covariates and Preliminary Analysis of the Data 155

5.3.1 — Health care utilization measure and covariates 155

5.3.2 — Preliminary and exploratory analysis of the data 165

5.4 — Analysis of the Regression Results and Discussion 181

5.4.1 — Analysis of the latent classes of users 183

5.4.2 — Analysis of the effect of SHI 187

5.4.3 — The effect of place of residence on the utilization of medical care visits 198

5.4.4 — Analysis of the effect of the remaining covariates on the utilization of health care visits 205

5.4.5 — Analysis of class membership (model LCH-Cross) 213

5.5 — Conclusion 216

CHAPTER 6

ESTIMATION OF THE EFFECTS OF TIME-COSTS ON THE UTILIZATION OF PUBLIC PRIMARY CARE CENTRES IN PORTUGAL 221

6.1 — Introduction 221

6.2 — Effect of Time Costs on Utilization: Brief Review of The Literature 226

6.3 — Dataset, Dependent Variable and Covariates 241

6.3.1 — The questionnaire, dataset and sampling scheme 241

6.3.2 — The indicator of health care centre utilization 244

6.3.3 — The determinants of health centre utilization 246

6.3.4 — Summary statistics and sample characterization 263

6.4 — Specification of a Model to Estimate the Effect of Time-Costs on the Utilization of Health Centres 267

6.5 — Empirical Results and Discussion 273

6.5.1 — Evaluation of fitted models and model comparison 274

6.5.2 — Analysis of the latent classes of Health Centre users 281

6.5.3 — Analysis of the effect of the time-cost covariates 284

6.5.4 — Analysis of the impact of diverse covariates on the Health Centre utilization 294

6.5.5 — Analysis of Class membership of the Health Centre users 302

6.6 — Summary of the Conclusions, Policy Implications and Future Research 304

CHAPTER 7

FINAL REMARKS 313

Appendix A

STATA PROGRAMS.....317

BIBLIOGRAPHIC REFERENCES 331

List of Tables

Table 1 — Population covered by public health insurance arrangements before 1975

Table 2 —Empirical distribution of the dependent variable 'visits'

Table 3 — Definition of covariates

Table 4 — Summary statistics

Table 5 — Description of the competing models estimated

Table 6 — Likelihood ratio tests results

Table 7 — Results of various Vuong tests implemented to discriminate among non-nested alternatives

Table 8 — Information Criteria results

Table 9 — Parameter estimates of the Poisson and NB1 regression models

Table 10 — Parameter estimates of the Hurdle based on NB2

Table 11 — Parameter estimates of the LCM_NB1 model

Table 12 — Parameter estimates of LCH-Cross(NB2) model

Table 13 — Effect of selected covariates on the conditional mean for a number of models

Table 14 — Estimate of the impact, using AME, on the conditional mean function for the health care users population

Table 15 — Kolmogorov-Smirnov tests for equality of the distribution of 'visits' by health insurance status and by place of residence

Table 16 — Total number of doctor visits by health insurance status

Table 17 — Summary statistics of positive 'visits' by health insurance status

Table 18 — Proportion of individuals with at least one visit by health insurance status

Table 19 — Summary statistics of 'visits' by place of residence

Table 20 — Summary statistics of positive 'visits' by place of residence

Table 21 — Proportion of individuals with at least one visit by place of residence

Table 22 — Summary statistics of covariates by health insurance status

Table 23 — Summary statistics of the covariates by place of residence

Table 24 — Summary statistics of the distribution of fitted means by latent class on the users population

Table 25 — AME of health insurance status; LCH-Cross model

Table 26 — Estimates of the impact of the place of residence: LCH-Cross model

Table 27 — Estimates of the impact of place of residence on various descriptors taken from the LCH-Cross model

Table 28 — Parameter estimates, standard error and z-values of a Probit model to find the determinants of class membership to the low users class implied by the LCH-cross model

Table 29 — Summary of the empirical evidence regarding the effect of time costs on individual health care decisions

Table 30 — Empirical distribution of the dependent variable 'hc_visits'

Table 31 — Definition of the time-cost variables

Table 32 — Interactions between 'time-cost' and some other covariates

Table 33 — Interactions between 'app_delay' and some other covariates

Table 34 — Socioeconomic and demographic covariates

Table 35 — Health Status covariates

Table 36 — Health Centre/Family Doctor covariates included in the time-costs regression model

Table 37 — Definition of complete set of explanatory variables selected to enter into the time-costs model

Table 38 — Summary statistics of the explanatory variables

Table 39 — Description of the competing models estimated

Table 40 — Likelihood ratio tests

Table 41 — Information criteria statistics (BIC and CAIC)

Table 42 — Parameter estimates, robust standard error and the z-statistic for the two-component LCM based on the truncated negative binomial (type 2) model

Table 43 — Summary statistics of the distribution of fitted means by latent class and population

Table 44 — Estimates of the overall effect of 'App_delay' in the utilization of health care centres by latent class and overall population

Table 45 — AME of the socioeconomic and demographic variables

Table 46 — AME of the health status covariates

Table 47 — AME of the Characteristics of the Health Centre vs Family Doctor covariates.

Table 48 — Parameter estimates, standard errors and z-values of a probit model to find the determinants of class membership to the low users class

List of Figures

Figure 1 — Outline of the purposes and contribution of the study

Figure 2 — Box plots of the distribution of the average elasticity effects of income across various specifications

Figure 3 — Box plots of the distribution of the marginal effects of education across various specifications

Figure 4 — Box-plots of the distribution of the marginal effects of rural_area across various specifications

Figure 5 — Box plots of the distribution of the marginal effects of NHS-only across various specifications

Figure 6 — Box plots of the distribution of the individual income elasticities for the health care users population of models HNB2 and LCH-Cross(NB2)

Figure 7 — Two box plots of the distribution of the marginal effect of 'NHS-only' regressor for the models HNB2 and LCH-Cross(NB2)

Figure 8 — Empirical distribution of the dependent variable 'visits'

Figure 9 — Distribution of 'visits' by health insurance status

Figure 10 — Distribution 'visits' by place of residence

Figure 11 — Box-plots of the distribution of the fitted means by latent class for the model LCH-Cross

Figure 12 — Box-plots of the distribution of the effect of health insurance status (NHS and ADSE) on the probability of a doctor visit

Figure 13 — Box-plot of the distribution of the partial effects of health insurance status (NHS-only and ADSE) on the conditional mean of the overall users population

Figure 14 — Box-plots of the distribution of the effect of health insurance status (NHS-only and ADSE) on the conditional mean of the overall population.

Figure 15 — Box-plot of the distribution of the effect of place of residence on the probability of at least one visit

Figure 16 — Box-plots of the distribution of the effect of place of residence on the conditional mean of the users population, and by latent class.

Figure 17 — Box-plot of the distribution of the effect of place of residence on the conditional mean of the overall population

Figure 18 — Empirical distribution 'hc_visits'

Figure 19 — Empirical distribution of Age of the Health Centre users

Figure 20 — Box-plots of the distribution of fitted means, by latent class and population (model LCM_TNB2)

Figure 21 — Box-Plot of the distribution of the marginal effects of 'App_delay' by latent class

Figure 22 —Distribution of the impact of activity status, gender, health status and age on the effect of 'app_delay' for the low users population

Figure 23 — Box-Plot of health centre utilization elasticities relatively to 'total_time', by latent class

Chapter 1

General Introduction

1.1 — Motivation and purpose of the study

The health economics literature has seen, over the last two decades or so, a considerable increase in the number of studies endeavouring at understanding the processes underlying the consumption of health care by individuals. Better datasets, containing detailed information at individual level, have allowed for increased sophistication of analysis and more robust inferences, leading, naturally, to the derivation of better health policy implications. Good examples of the referred tendency to adopt sophisticated methodologies, applied to individual level datasets, aiming at explaining health care utilization patterns are found in Jones and O'Donnell (2002), which collects a number of studies clearly illustrating our point.

Despite the improvements verified in the quality of the datasets necessary to conduct such studies, given the inherent variation in the individual behaviour regarding medical care utilization, the persistence of unobserved heterogeneity in such datasets is a well known and acknowledge fact. Therefore, the econometric specifications adopted to study medical care utilization of the individuals require the introduction of components capturing and accounting for individual unmeasured factors, otherwise the model can produce biased results.

One major approach to control for non-observed individual heterogeneity is through the adoption of regression models based on the latent class framework, which assumes unobserved heterogeneity of discrete nature. Such discrete approach presents the advantage of making possible to unveil (latent) health care consumption groups of individuals within a given population. Excellent examples of this line of research is found in Deb and Trivedi (1997, 2002), Deb and Holmes (2000), Deb (2002), Atella *et*

al. (2004), Bago d'Uva (2005), Gerdtham and Trivedi (2001) among others. In all these empirical applications, the specification adopted accounted for individual unobserved factors of the members of the overall population (please, see box 'chapter 4' in Figure 1, lower right panel), generating thus, the split of that population into two latent groups of individuals.

Surprisingly enough no attention has been given to the literature of latent class models whose specification accounted for the unobserved factors specific to the population of health care users, generating, in this case, the division of the population of health care users into two latent groups of actual medical care users (please, see box 'chapter 4' in Figure 1, lower left panel). Such analysis can be relevant under two perspectives. Either on its own because it provides findings that reflect the behaviour of actual medical care users by latent class, or as a part of extensions of econometric models, for instance the hurdle model, designed to understand utilization patterns of the overall population, when two different stochastic processes govern the health care utilization.

As it is well accepted, the ultimate purpose of empirical studies, as the one presented here, is to generate empirical information allowing one to go from mere conjectures to facts about some relevant health economics problem. In this context, there are a number of issues on the Portuguese health system that demanded generating empirical information using rich datasets analysed with appropriate methodological tools. Namely, to learn something about the effect of the supplementary health insurance, to analyse the patterns of health care utilization across urban and rural areas in order to detect differences, in addition to the effect of income, as well as the effect of other covariates. Moreover, another relevant issue on the economics of medical care is the reaction of users to the time-price of medical care. In the Portuguese National Service, this constitutes a relevant issue, which has been almost completely ignored, as it may provide hints about how to alter that time-cost to influence the behaviour of people. Therefore, we estimate the effect of various

covariates, from which we highlight (exogenous) supplementary health insurance, place of residence (urban *vs* rural), income, time-costs, health status, etc., on the utilization of medical care services. The findings unveiled by such empirical analysis surely contribute to inform policy debate on the efficient organization and delivery of medical care services to achieve the ultimate goal of every health care system, which is to obtain gains in health. Since the early 70's, a number of empirical studies have been exploring similar questions, although in different health system contexts, and using data with different characteristics (Acton 1975; Coffey 1983; Janssen 1992; Pohlmeier and Ulrich 1995; Vera-Hernandez 1999; Gerdtham and Trivedi 2001; Buchmueller, Couffinhal *et al.* 2004; Jones, Koolman *et al.* 2004; Bago d'Uva 2005; Barros, Machado *et al.* 2005; Sarma and Simpson 2006)

In sum, the purpose of this work is to look at utilization patterns of health care in the Portuguese population exploring the possibility of unveiling (latent) health care consumption groups, specific to the users of medical care services, with the ultimate purpose of producing knowledge about health economics issues relevant to the Portuguese policy makers.

As a final remark, we claim that we consider this thesis as providing an investigation of issues relevant to the analysis of health care utilization, when utilization is measured as a count variable. We see it as contributing to the area of health econometrics, applied to the analysis of health care data, in addition to its contribution to the health economics of the Portuguese health system, by increasing our understanding of the determinants of utilization of medical care.

1.2 — Outline of the thesis

In addition to this introductory chapter, this dissertation contains six chapters, whose interplays are shown in Figure 1 that outlines the purposes and contribution of this thesis.

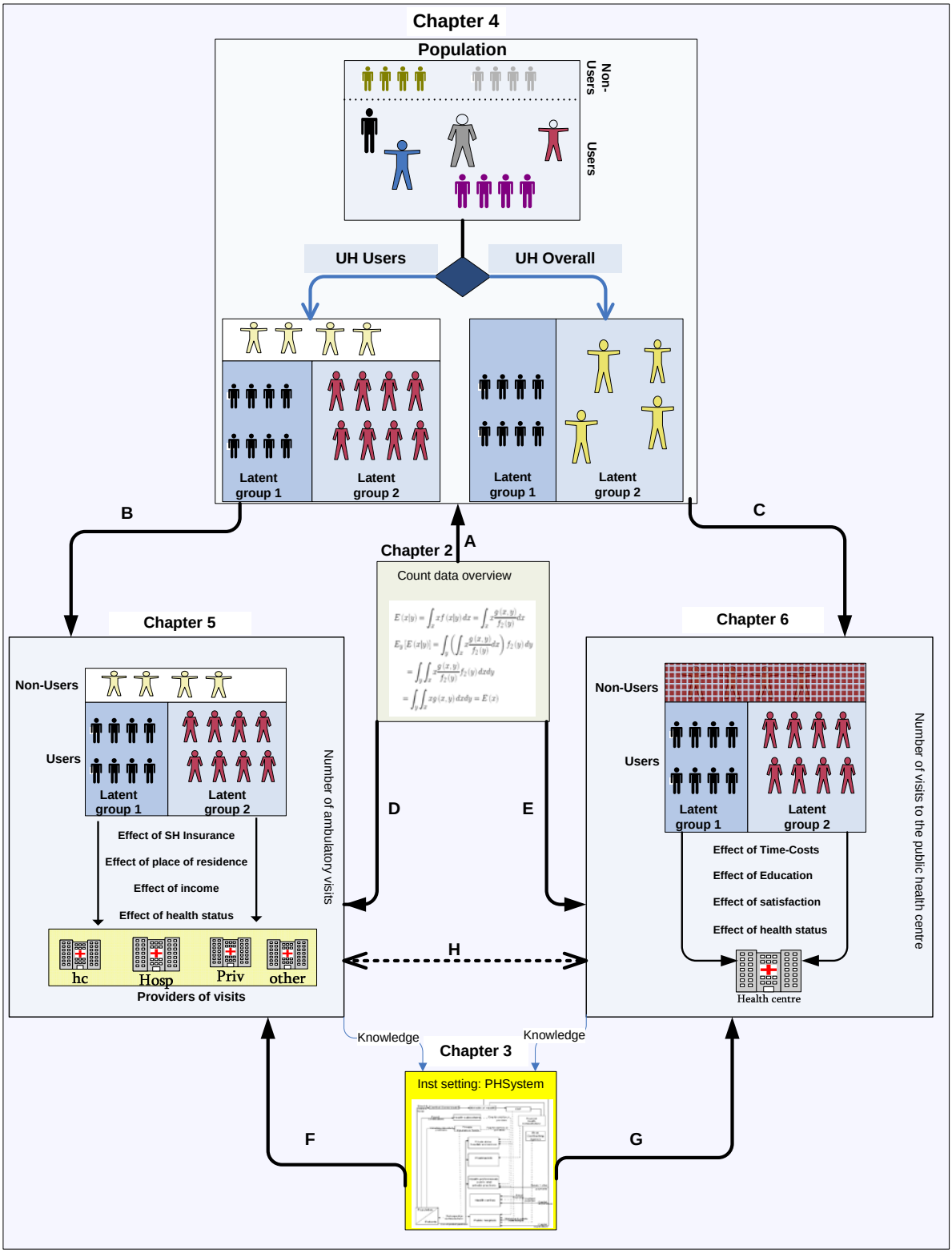

Figure 1 — Outline of the purposes and contribution of the study

To begin with, Chapter 2 provides an integrated methodological framework to support the applications developed in Chapters 4 to 6, as it is shown by the arrows A, D and E. The chapter can be thought to be divided in two parts. At its outset, the first part presents an overview of the count data framework that has tended to dominate the applied literature of the analysis of health care data. The second part of Chapter 2 focuses on other topics that the empirical research must address, namely the estimation of latent class models, the selection of the model with better fit and the interpretation of the parameters of non-linear regression models. Chapter 3, provides an overview of the Portuguese health care system, emphasizing its description from the delivery perspective. It is linked to Chapters 5 and 6, as it is shown in Figure 1, in the perspective of providing information that helps to understand the results of the empirical applications developed in those two chapters.

This first two chapters close what we see as the first part of the thesis, functioning as an umbrella to support the next three chapters.

In Figure 1, the arrows B and C convey the idea that Chapter 4 provides some methodological insights that are useful for the empirical applications presented in Chapters 5 and 6. The essay has the main purpose of re-evaluating the issue that discusses the better econometric framework to model count data. We contribute to this discussion by presenting a hurdle formulation that departs from the standard specifications in the way individuals with positive utilization are modelled, based on latent class models, explicitly assuming that the unobservable characteristics specific to the population of health care users. A second purpose of the chapter is to evaluate the robustness of the effect of various covariates, e.g. income, health insurance status, place of residence (rural *vs* urban) and education, on the conditional mean of various competing count data models under evaluation.

The essay presented in Chapter 5, which receive inputs from Chapters 2, 3 and 4, transmits some insights to the Portuguese health system, and aims at using the estimates that come out from Chapter's 4 preferred model to address some research

questions related to health economics. The first question evaluates the effect of health insurance status and of the place of residence (urban *vs* rural) on the utilization of medical care services in Portugal. The second question explores the possibility of unveiling latent classes of health care users, and analyses them both in terms of expected health care utilization and in terms of the characteristics of the individuals that each latent class encompasses.

Finally, Chapter 6, which receives inputs from Chapters 2, 3 and 4, and offers knowledge relevant to the Portuguese health system, reports the results of an essay developed to provide evidence on the effect of time-costs on the utilization of the public health care centres. The main goal of the study is to estimate the elasticity of health care centre utilization relative to the total time spent in visiting the health centre and to provide evidence about the effect of an appointment delay on the utilization of the health centre. The empirical application estimates a latent class model, specialized in handling truncated-at-zero data.

At last, the purpose of Chapter 7 is to present the overall conclusions that the thesis has brought out, suggesting as well some hints for possible future research.

Chapter 2
Count data models for the analysis of health care utilization: General overview

2.1 — Introduction

One first necessary precondition to study the utilization of medical services from an empirical viewpoint is that the economic variables representing individual consumption of medical care have an empirical counterpart. Such empirical counterpart should be feasible to implement and to measure with accuracy, actually reflecting the individual utilization of medical care services. Surely, the ideal empirical variable would be one single measure that incorporates all the diversity of types of medical care services present in the array of choices (e.g. consultations, hospitalizations, auxiliary tests, surgeries, medicines, etc) alongside with the quality inherent in all medical acts. Unfortunately, such a high quality indicator, possessing all those desirable characteristics is not measured in health surveys, therefore, the empirical investigator must content with the readily available indicators, which can be framed into two types.

On the one hand, some empirical studies use continuous variables as health care utilization indicators. Usually, in such cases, utilization is measured by the total expenditure on health care services or by the total expenditure on a particular health care item, e.g., dental health care, outpatient care, inpatient medical care, medicines, etc. (Duan, Manning *et al.* 1983; Manning, Newhouse *et al.* 1987; Keeler, Manning *et al.* 1988; Manning and Marquis 1996; Mullahy 1998; Ruiz, Amaya *et al.* 2007). On the other, an alternative stream of empirical models adopted discrete variables to measure utilization, for example, the total number of physician visits, the number of contacts to different types doctors, etc. (Cameron and Trivedi 1986; Pohlmeier and Ulrich 1995; Deb and Trivedi 1997; Geil, Million *et al.* 1997; Barros 1999; Vera-Hernandez 1999;

Gerdtham and Trivedi 2001; Yen, Tang *et al.* 2001; Deb and Trivedi 2002; Mocan, Tekin *et al.* 2004; van Doorslaer, Koolman *et al.* 2004; Winkelmann 2004; Bago d'Uva 2005; Lourenço and Ferreira 2005; Bago d'Uva 2006; Sarma and Simpson 2006). As one can easily conclude after a quick scan through the empirical literature on medical care demand, the studies based on discrete indicators show a clear tendency to dominate. In our view, the availability and accuracy of discrete data, at an individual level, relative to continuous data, is the main factor that contributes for this tendency.

The specification of models to analyse indicators of both families require modelling strategies especially adjusted to handle the features of the variable, which in the case of continuous variables is the high prevalence of individuals without expenditures and the non-negativity of the variable. In the case of discrete measures, the special features that the researcher must account for are their non-negativity, the large incidence of zeroes and the fact that the variable is integer. Thus, in sum, both types of medical care indicators present particular characteristics that rule out the utilization of simple linear regression models[1].

Given that the empirical applications of this dissertation adopt a discrete and non-negative variable to reflect health care utilization, we are particularly interested in presenting the methodologies and models most often adopted to treat this type of dependent variable.

The utilization of regression models to analyse count data has been widely used for the analysis medical care utilization. Several excellent textbooks and papers provide extensive overviews about the analysis of count data. Moreover, some of them even present applications for the health economics field (Hausman, Hall *et al.* 1981; Cameron and Trivedi 1986; Cameron and Trivedi 1998; Winkelmann 2003; Cameron

---

[1] For details, consult select parts of Jones (2000) and Jones and O'Donnell (2002).

and Trivedi 2005). Therefore, the purpose of this chapter is just to introduce the methods of count data directly related to the empirical applications developed in this dissertation. Consequently, we take the viewpoint of an applied econometrician who is primarily concerned with practical aspects of the utilization of such a methodology. Due to the objective of this dissertation, in this chapter, the specification of count data models that incorporates unobserved heterogeneity are in the forefront of our analysis. Moreover, we highlight the unobserved heterogeneity of discrete nature.

The first sections of the current chapter present some standard topics in the count data framework. Section 2.2 presents the Poisson regression model, usually taken as the natural starting point to analyse count data, however, as has been usually pointed out by several authors, the model ignores that the data does not fully account for all the variability in individual's choices. Therefore, section 2.3 goes a step further and provides a general definition of mixture models, which are a natural way to incorporate unobserved heterogeneity in the explicative models. Sections 2.4 and 2.5 present two examples of mixture models, respectively the negative binomial regression model, which, in some contexts, is a continuous mixture model, and a latent class model, which is a finite mixture model. Section 2.6 introduces the model that has been considered as the cornerstone specification in the analysis of health care utilization data — the hurdle regression model. In addition, yet in this section, we enumerate some of the critical remarks made to that specification. Next, Section 2.7 illustrates the adaptations that standard count data models must undergo to explain adequately truncated-at-zero count data variables. On the other hand, in Section 2.8 we present a less known issue of count data modelling, that arises when the researcher's goal is to specify a truncated-at-zero count model incorporating unobserved heterogeneity using mixture models, emphasizing the utilization of finite mixture models. The remaining sections of this chapter move away from specification topics, concentrating instead on other important methodological aspects the empirical researcher must address.

On the one hand, section 2.9 begins by discussing some practical issues that arise when the researcher aims at specifying and estimate latent class models. Being more specific, we explore the options regarding the number of latent classes and the baseline of the latent classes. On the other, Section 2.10 deals with aspects related to model selection techniques available to pick out the count data model that better fits the data. Closing the current chapter, Section 2.11 goes through the important issue of interpreting the output of (non-linear) count data models.

2.2 — The Poisson Regression Model

In the empirical analysis of data involving the number of events per time interval, the benchmark model for analyzing the effects of covariates in the dependent variable is the Poisson regression model, henceforth referred to as PRM (Cameron and Trivedi 1986; Cameron and Trivedi 1996; Winkelmann 2004).

Let us assume that y_i, $i = 1, 2, ..., N$ represent the dependent variable and let $\mathbf{x}_i^{'} = \left[x_{i1}, x_{i2}, \cdots, x_{ik} \right]$ denote a $(1 \times \mathrm{k})$ vector of independent variables determinants of y_i. The basic PRM is motivated by assuming that the probability function of y_i, conditional on the vector of covariates $\mathbf{x}_i$, has a Poisson distribution. The density of y_i given $\mathbf{x}_i$ under the Poisson model is completely determined after the specification of the conditional mean, $E\left(y_i \mid \mathbf{x}_i\right) = \lambda_i$. Hence, the probability function of $\left(y_i \mid \mathbf{x}_i\right)$ is given by

$$f\left(y_i \mid \mathbf{x}_i\right) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \qquad y_i = 0, 1, 2, ... \qquad [1]$$

In the above specification, λ_i represents the mean parameter of the Poisson probability function. The mean is usually parameterized as the exponential of a linear function, dependent on the vector of regressors, $\mathbf{x}_i$, and the unknown parameters collected in the $(k \times 1)$ vector β. Accordingly, the mean of the PRM in is given by

$$E(y_i \mid \mathbf{x}_i) = \lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}) \qquad [2]$$

This particular parameterization has been widely used as it ensures that the conditional mean of the model is always nonnegative.

From the above we have learned that the Poisson distribution is characterized by a single parameter, implying that the first moment of y_i, conditional on the covariates $\mathbf{x}_i$, equals the conditional variance, implying that the model imposes the equality of the conditional variance and mean, that is,

$$E(y_i \mid \mathbf{x}_i) = \lambda_i = V(y_i \mid \mathbf{x}_i) \qquad [3]$$

The mean-variance equality is a property known as '*equidispersion'* and there are empirical evidence showing that the property is seldom, or never, verified in applications (Cameron and Trivedi 1986; Pohlmeier and Ulrich 1995; Gurmu 1997; Cameron and Trivedi 1998; Grootendorst 2002; Winkelmann 2003). The most common result relating the conditional mean and variance is that the second generally exceeds the later. When this is the case, it is referred to as that the data presents '*overdispersion*' relative to the Poisson model. '*Overdispersion'* corresponds to a situation where the ratio $\frac{V(y_i \mid \mathbf{x}_i)}{E(y_i \mid \mathbf{x}_i)}$ exceeds that implied for the hypothesized data generating mechanism for y_i (Cameron and Trivedi 1986; Mullahy 1986; Cameron and Trivedi 1998). Another empirical finding that makes the PRM unattractive in empirical applications is the '*excess zeros'* problem. This occurs when the observed data shows a higher relative frequency of zeros than is consistent with the Poisson regression model (Gurmu and Trivedi 1996; Mullahy 1997).

'Overdispersion' and 'excess zeros', relative to the Poisson specification, are empirical properties that arise because the Poisson model does not incorporate the individual unobserved heterogeneity, present in the data whenever the covariates do not represent the full variability of individual behaviour regarding medical care decisions (Gurmu and Trivedi 1992; Gurmu and Trivedi 1996; Gourieroux and Visser

1997; Mullahy 1997). Failure to include unmeasured factors in the specification of the model leads to loss of efficiency and induces biases on variances and, consequently, on testing procedures (Cameron and Trivedi 1996; Gourieroux and Visser 1997).

Therefore, failure to include unobserved heterogeneity in the specification process makes the assumptions of the Poisson regression model untenable in empirical applications, motivating, therefore, the utilization of more general count data models to analyse health care utilization data. Some of these alternative count data models are no more than simple modifications of the Poisson regression model, tailored to include explicitly unobserved heterogeneity terms in the model specification. For example, one option is to include a random perturbation term in the mean parameter of the PRM and assume that that term follows a certain probability distribution.

However, despite the stringent conditions imposed by the Poisson regression model, Cameron and Trivedi (1998) point out that in order to make a valid inference about the conditional mean parameters, then the estimation of a PRM may be sufficient, as long as the estimates of the standard errors be adjusted.

In the applications developed in this project we are also interested in other respects of the conditional distribution of the count variable other than the mean, therefore we must go beyond the Poisson regression model, specifying and estimating more general count data models.

2.3 — Mixture models to incorporate unobserved heterogeneity

The specification of mixtures of distributions is a well-known and extensively used methodology to analyse a wide variety of actual situations, including the area of health services research. This approach is usually seen as an appropriate strategy for accounting the presence of individual unobserved factors.

Two general types of mixture models can be considered: on one hand, continuous mixture models[2], on the other, finite mixture models. The main differences between the two types of mixture specifications reside in the assumptions made about the distribution of individual unobserved heterogeneity. While in continuous mixtures the unobserved heterogeneity term is taken as a continuous random variable, in finite mixture models the unobserved heterogeneity is assumed a discrete random variable with a given, but generally unknown, number of support points.

Cameron and Trivedi (1998) and Lindsay and Lesperance (1995) define a continuous mixture model as a probabilistic model where the conditional density of the random variable y_i, is defined as

$$f\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}\right) = \int_{\upsilon_i} f\left(y_i \mid \mathbf{x}_i; \upsilon_i, \boldsymbol{\beta}\right) * h\left(\upsilon_i \mid \mathbf{x}_i, \boldsymbol{\gamma}\right) d\upsilon_i \qquad [4]$$

where $h\left(\upsilon_i \mid \mathbf{x}_i, \boldsymbol{\gamma}\right)$ is referred to as the mixing distribution of the mixture. Under this specification, the unobserved heterogeneity, represented by the random variable υ_i, is assumed to be continuous, with density function $h\left(\upsilon_i \mid \mathbf{x}_i, \boldsymbol{\gamma}\right)$. The parameter, or possibly, the vector of parameters represented by $\boldsymbol{\gamma}$, can be seen as the parameters of the mixing distribution to be estimated along with the other interest parameters, collected in the vector $\boldsymbol{\beta}$.

When the unobserved heterogeneity is assumed to be represented by a discrete random variable with P points of support, then the probability function of y_i, conditional on $\mathbf{x}_i$, is given by

$$f\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right) = \sum_{j=1}^{P} \pi_j f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right) \qquad [5]$$

---

[2] The continuous mixture models are also known as convolutions.

where the mixing probabilities of the mixture, $\pi_1, \pi_1, ..., \pi_P$ are such that $\sum_{j=1}^{P} \pi_j = 1$ and $\pi_j \geq 0, \quad j = 1, 2, ..., P$. In the discrete context, the density $f\left(y_i \mid \boldsymbol{x}_i\right)$ is a semi-parametric mixture for y_i because there is no need to assume any parametric assumptions about the mixing distribution to specify the mixture model (Wedel, Desarbo *et al.* 1993; Lindsay 1995; Cameron and Trivedi 2005).

Particular instances of these two different approaches dealing with the unobserved heterogeneity have been usually used to model medical care utilization. Despite the existence of several examples of mixture models, the most well known are the negative binomial model and the latent class model.

2.4 — The Negative Binomial regression model

The negative binomial model, thereafter referred to as NB, is a well-known, and popular statistical distribution for applied work in count data contexts. Besides its relevance as an end of line model to study some health economics related substantive issues, the NB model is also frequently used as a baseline model to specify more generalized count data models. For example, hurdle models, zero inflated specifications as well as latent class models. Some authors label the NB distribution as the most general and flexible probability model among all the densities available in the statistical literature for econometric models of count data (Deb and Holmes 2000).

Despite the characterization of the NB distribution could be done in a number of alternative ways (Cameron and Trivedi 1998; Winkelmann 2003), therefore can also be used in different contexts, in the medical care utilization analysis the model is widely adopted as the single strategy to account for the presence of unobserved heterogeneity in the data. When this is the case, the NB probability function can be characterized as a continuous Poisson-gamma mixture model, obtained by allowing the average parameter of the Poisson to vary randomly across the population according to a gamma

distribution. Under this assumption about the distribution of the random component included in the Poisson conditional mean parameter, it can be proved that the marginal distribution of $(y_i \mid \mathbf{x}_i)$ follows a negative binomial distribution (Cameron and Trivedi 1986; Pohlmeier and Ulrich 1995; Cameron and Trivedi 1996; Gurmu 1997; Cameron and Trivedi 1998; Winkelmann 2003; Cameron and Trivedi 2005).

Following, for instance, Deb and Trivedi (2002), the negative binomial probability function of the count variable y_i, conditional on the vector of covariates $\mathbf{x}_i$, can be written as,

$$f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}) = \frac{\Gamma(y_i + \eta_i)}{y_i! \Gamma(\eta_i)} \eta_i^{\eta_i} (\lambda_i + \eta_i)^{-(\eta_i + y_i)} \lambda_i^{y_i} \quad y_i = 0,1,2... \qquad [6]$$

where $\Gamma(.)$ represents the gamma function, $\lambda_i = \exp(\mathbf{x}_i' \boldsymbol{\beta})$ and $\eta_i = \left(\frac{1}{\alpha}\right) \lambda_i^k$. The parameter $\alpha(>0)$ that appears in η_i is the dispersion parameter. In this specification, the constant *k* is arbitrary and determines the functional relationship between the conditional mean and the conditional variance.

For the NB model, the conditional mean and variance of the negative binomial model are given, respectively, by

$$E(y_i \mid \mathbf{x}_i) = \lambda_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}) \qquad [7]$$

and

$$V(y_i \mid \mathbf{x}_i) = \lambda_i + \alpha \lambda_i^{2-k} \qquad [8]$$

In empirical applications the term *k* is usually held fixed. Setting *k = 0* or *k=1* gives rise to the most frequent models based on the negative binomial specification. Setting *k = 1* gives rise to the negative binomial type 1 model, henceforth referred to as NB1 on the other, setting *k* = 0, one obtain the negative binomial type 2, henceforth referred to as NB2.

Although the NB family of densities is considered as the most general and flexible discrete distribution to fitting count data (Deb and Holmes 2000), its utilization as end of line model, thus as a model used to incorporate the unobservables, may present some weaknesses. In the following, we provide a list of some weaknesses of the negative binomial regression model in applied work:

- Both types of the negative binomial model, when used as a mean to account for the individual unobserved heterogeneity, are fully parametric. Therefore, they are based on explicit assumptions about the distribution of the unobserved heterogeneity. However, economic theory that gives insight about the unknown functional form for the distribution of the unobserved factors is often lacking (Wedel, Desarbo *et al.* 1993), making the utilization of the NB distribution somewhat arbitrary. That is, the assumption that the unobserved heterogeneity follows a gamma distribution is only due to statistical convenience,
- Gurmu and Trivedi (1996) show that making the unobserved heterogeneity flexible, but without accounting for excess zeros, did not result in models with a good fit to the data. The consequence is that the observed health care data may exhibit a relative frequency of zeros ('*excess zeros*') inconsistent with the NB regression model (Pohlmeier and Ulrich 1995; Gerdtham 1997; Bago d'Uva 2005),
- The NB regression model belongs to the class of one-part models. Consequently, it relies on the hypothesis that the zeroes and positives share the same data generating mechanism. In the case of health care choices, this assumption may not correspond to the actual decision making process, because different agents decide about to contact or not to contact a doctor and about the intensity of care, after the first contact. This requires the specifications of different statistical processes to govern the choices in each part of the decision

process (Pohlmeier and Ulrich 1995; Gurmu and Trivedi 1996; Deb and Trivedi 1997; Deb and Trivedi 2002),

- Finally, in the negative binomial regression model the effect of a covariate in the conditional mean — $E\left(y_i \mid \mathbf{x}_i\right) = \lambda_i = \exp\left(\mathbf{x}_i^{'}\boldsymbol{\beta}\right)$ — is constrained to be equal across the support of the dependent variable. In consequence, if the impact of a covariate, say x_{ik}, in the mean parameter differs across the support of the dependent variable distribution, the NB regression model does not capture that different impact (Deb and Trivedi 2002).

Therefore, empirical health economists have pursued to the formulation of more general count data models, which, in some way, can overcome these issues raised in practical applications of the NB regression model.

An alternative approach to account for the presence of unobserved heterogeneity involves the formulation of finite mixture models. Finite mixtures and latent class models, henceforth referred to as LCM, are two families of models closely related (Aitkin and Rubin 1985 ; Wedel, Desarbo *et al.* 1993; Deb and Trivedi 1997).

2.5 — Latent Class models

In the LCM for integer and non-negative variables, the random variable, υ_i, that represents the individual unobserved factors is assumed discrete with *P,* generally unknown, points of support. In Latent Class models, it is assumed that a number of subpopulations compose the population under study. Expressed alternatively, the observed data is a mixture of the data of a finite, and generally unknown, number of sub-populations, however, the researcher does not observe which subgroup a given individual belongs to. Furthermore, the parameters that characterize the behaviour of the individuals of each latent class (sub-population) are, possibly, different, being possible to test this hypothesis.

Latent class models have been applied to the analysis of actual data coming from several different fields (Gritz 1993; Nagin and Land 1993; Wedel, Desarbo *et al.* 1993; Brannas and Rosenqvist 1994 ; Wang, Cockburn *et al.* 1998). Deb and Trivedi (1997) pioneered the application of this methodology to understand the factors that drive individual choices concerning medical care utilization. Since then a number of other researchers have used the same methodology to study health care utilization in diversity of health care contexts (Deb and Trivedi 1997; Deb and Holmes 2000; Deb 2001; Gerdtham and Trivedi 2001; Deb and Trivedi 2002; Jimenez-Martin, Labeaga *et al.* 2002; Atella, Brindisi *et al.* 2004; Bago d'Uva 2005; Bago d'Uva 2005; Lourenço and Ferreira 2005).

To specify a LCM, suppose that υ_i is a random variable representing the unmeasured factors. In addition let us assume that υ_i follows a discrete probability function with P points of support, that is, υ_i^j, $(j = 1, 2,..., P)$, with probability masses $0 \le \pi_j \le 1$, $(j = 1, 2,..., P)$, respectively, with $\sum_{j=1}^{P} \pi_j = 1$. In this specification, $\pi_1, \pi_2, ...\pi_P$ are referred to us as the mixing probabilities of the mixture. Moreover, let us assume that the distribution of the dependent variable y_i conditional on the vector of covariates $\mathbf{x}_i$ and on the unobserved heterogeneity $\upsilon_i^j, j = 1, 2, ..., P$, is defined by $f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)$, $(j = 1, 2,..., P)$, where $\boldsymbol{\beta}_j$ is a vector of parameters to estimate. Each probability function $f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)$, $(j = 1, 2,..., P)$ is designated as the component distribution of the mixture.

Under these assumptions, the distribution of $(y_i \mid \mathbf{x}_i)$, is given by the model presented by equation [5], which we reproduce here,

$$f\left(y_i \mid \mathbf{x}_i; \boldsymbol{\beta}\right) = \sum_{j=1}^{P} \pi_j f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right) \qquad [9]$$

where $\boldsymbol{\beta}' = \left[\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_p'\right]$ represents the vector of parameters that characterize the overall population. The mixing probabilities, π_j, $(j = 1, ..., P)$, are estimated together with all other model parameters, therefore the estimation procedure also provides an estimate of the parameters of the mixing distribution.

Under the general LCM framework, the conditional mean function and variance of the count in the overall population is given by, respectively,

$$E\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \sum_{j=1}^{P} \pi_j E_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right) = \sum_{j=1}^{P} \pi_j \lambda_{ij} \qquad [10]$$

$$V\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \sum_{j=1}^{P} \pi_j \left[V_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right) + \lambda_{ij}^2\right] - \left(\sum_{j=1}^{P} \pi_j \lambda_{ij}\right)^2 \qquad [11]$$

In the moments shown above,

$$\lambda_{ij} = E_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right) = \exp\left(\mathbf{x}_i' \beta_j\right) \qquad [12]$$

refers to the conditional mean of $f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)$, while $V_j\left(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}_j\right)$ refers to conditional variance. Note that the conditional mean of the j^{th} latent population is λ_{ij}.

The utilization of this specification for applied work requires two decisions from the researcher. On one hand the number of points of support (P), on the other, a specific distribution must be chosen for each component distribution of the mixture, that is, a precise density to represent each $f_j\left(y_i \mid \mathbf{x}_i, \beta_j, \upsilon_i^j\right)$, $j = 1, 2, ..., P$. Our choices concerning the value of P (that coincides with the assumed number of latent classes) and the specific form of probability functions used in the specification of the LCM estimated in this dissertation will be presented further on, in section 2.9 — Options concerning LCM specification and estimation.

Latent class models offer a number of advantages over the use of continuous mixture models. Deb and Trivedi (2002), Cameron and Trivedi (1998), Wedel *et al.* (1993) and Heckman and Singer (1984) provide a rather complete list of such advantages;

- Heckman and Singer (1984) show that estimates of such a model might provide good numerical approximations even when the distribution of the count is continuous,
- It is a specification that provides good results without being dependent on strong distributional assumptions regarding the mixing distribution,
- It provides a natural and intuitively attractive representation of unobserved heterogeneity in a finite, usually small, number of latent classes, each of which may be regarded as a 'type' or 'group' (Deb and Trivedi 2002),
- It is a conceptually simple specification. It provides a good alternative to the specification of continuous mixing densities for some parametric count models for which the marginal density may not have an analytical solution,
- The potential for identification of latent classes of individuals, based on non-observable characteristics, may be important for interpretation reasons because the impact of some variable may differ across latent classes. Moreover, it is equally possible to discern the impact of this same covariate on the population. This enhances the interpretation power of the models from this family, relative to more simpler, usually, single index models,
- Deb and Trivedi (2002) claim that the unobservable latent characteristics that partition the population in two latent classes are based on individual health status which is not captured by surveys. Therefore, in the case of two latent sub-populations it is possible to distinguish between a '*healthy'* population, characterized by low level of medical care utilization, and an '*ill'* population, characterized by an intense use of medical care resources.

Despite the wide recognition of the advantages of using the LCM specification for fitting health care count data, there are authors referring some disadvantages on its utilization. For example, note that, while the hurdle specification can be framed as a natural extension of an economic model the principal-agent model, only statistical reasoning drives the LCM specification (Jimenez-Martin, Labeaga *et al.* 2002).

In a latent class model with P latent classes, with the negative binomial density as baseline model, each component's distribution is given by

$$f_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta_j}\right) = \frac{\Gamma\left(y_i + \eta_{ij}\right)}{y_i!\Gamma\left(\eta_{ij}\right)} \eta_{ij}^{\eta_{ij}} \left(\lambda_{ij} + \eta_{ij}\right)^{-\left(\eta_{ij} + y_i\right)} \lambda_{ij}^{y_i} \quad y_i = 0,1,2... \qquad [13]$$

where $j = 1,2,...,P$ are the latent classes, $\eta_{ij} = \left(\frac{1}{\alpha_j}\right)\lambda_{ij}^k$, α_j is the dispersion parameter of the latent class j and $\lambda_{ij} = \exp\left(\mathbf{x}_i'\boldsymbol{\beta}_j\right)$ is the respective conditional mean, (j=1,2,…,P).

The assumptions regarding the impact of the unobserved heterogeneity on the parameters of the component distributions of the LCM influence the specification of each latent class's conditional mean. For example, one can assume that the unobserved heterogeneity is such that it only influences the intercepts of the conditional mean of each class and the dispersion parameter, therefore assuming equal slopes parameters across the latent classes. In this case, the unmeasured effects only cause the intercepts and dispersion parameters to be different across the latent classes. This constrained model is poor and misleading whenever the unobserved heterogeneity influences both the intercept and the slopes of the conditional mean, amounts to saying that the effect of covariates, given by the slopes, is different across the latent classes (Wedel, Desarbo *et al.* 1993). Thus, in this thesis, we privilege the more general LCM specification and allow for full heterogeneity by permitting all parameters in the P components to differ. Hence, we do not impose any restrictions regarding the vector of parameters β_j as well as the dispersion parameters α_j, $j = 1,2,...,P$.

2.5.1 — LCM and posterior class analysis

As explained above, every LCM assumes that the overall population consists of *P* latent classes of individuals. Under this framework, each individual may be regarded as belonging to one and only one latent class. Before the estimation of the model, the mixing probabilities $\pi_1, \pi_2, ..., \pi_P$ are unknown precluding any probabilistic classification of the individuals. However, once the latent class model is estimated, the mixing probability of the latent class model is disclosed, being then possible to estimate the posterior probability that individual *i (i=1,…,n)* belongs to latent class *j (j=1,2,…,P)* for every (i, j) pair.

For a known distribution of the mixing probabilities $\pi_j \ (j = 1, 2..., P)$ of the LCM specification, the posterior probability that individual *i (i=1,…,n)* belongs to the latent class *j* (j = 1, 2,…,P) is given by

$$P\left(I_i \in C_j \mid Y_i = y_i, \mathbf{x}_i, \boldsymbol{\beta}\right) = \frac{\pi_j * f_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)}{\sum_{k=1}^{P} \pi_k * f_k\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_k, \upsilon_i^k\right)} \quad j = 1, 2, ...P \qquad [14]$$

After computing the posterior probability for every individual in the sample using equation [14], the assignment of each individual to a given latent class should follow the following rule: individual *i (i=1,…,n)* is assigned to latent class $j \ (j = 1, 2..., P)$ if the posterior probability of belonging to that class is higher than the posterior probability of belonging to any other class k = 1,2,…,P, for all $j \neq k$. Mathematically, individual *i* is assigned to latent class *j* whenever the following inequality holds:

$$P\left(I_i \in C_j \mid Y_i = y_i, \mathbf{x}_i, \boldsymbol{\beta}\right) \geq P\left(I_i \in C_k \mid Y_i = y_i, \mathbf{x}_i, \boldsymbol{\beta}\right), \quad \forall_{k=1,2,...,P, \text{ and } j \neq k} \qquad [15]$$

This allocation of of a patient to a latent class may be valuable when the researcher aims at analysing the types of individuals who might belong to one or the other latent class. Unveiling the type of patients comprised in each class can be important from a health policy point of view, because it will allow health authorities to design policies

specific to the latent sub-populations of individuals (whose characteristics were make known by the posterior class analysis) and not to the overall population.

After model estimation, the introduction of the additional information conveyed by the individual's observed health care utilization (y_i) makes possible to infer something about the characteristics of the individuals comprised in each latent class (Deb 2001; Atella, Brindisi *et al.* 2004).

Atella *et al.* (2004) suggest the following procedure: estimate a binary regression model[3] where the dependent variable indicates class membership and the independent variables are, as is obvious, observed covariates. Notice that this kind of study unveils whether the class membership is influenced by observed individual characteristics.

2.6 — The Hurdle Regression model

In general, one can claim that the evolution from the Poisson regression model to the negative binomial and, from the later to the LCM specification is motivated only by statistical reasons, in an attempt to discern a general count data model that maximizes the data fitting. Jimenez-Martin *et al.* (2002) criticize this strategy of modelling based only on statistical criteria relegating economic reasoning to a second level of discussion. All of these specifications are underpinned on the assumption that a common data generating mechanism generates the zeroes and the positives observations. As it has been widely acknowledged, in the context of individual health care choices the decision process may involve two stages, possibly with different key decision makers in each stage (Zweifel 1981). Accordingly, the specification of empirical models to explain health care utilization should recognize this two-stage decisional structure.

---

[3] The specification of Probit and Logit regression models is common in similar contexts. In this dissertation, we choose the Probit specification

The hurdle specification belongs to the family of two-part regression models. This is one of the most important models and widely used in the study of medical care utilization data. Deb and Trivedi (2002) considered it as the methodological cornerstone of empirical analysis in modelling the utilization of medical care services. The relevance of this class of models in the empirical modelling of medical care count data is demonstrated by the large number of empirical studies using this type of framework (Grootendorst 1995; Pohlmeier and Ulrich 1995; Hakkinen, Rosenqvist *et al.* 1996; Geil, Million *et al.* 1997; Gerdtham 1997; Santos-Silva and Windmeijer 2001; Van Ourti 2004; Winkelmann 2004).

Cragg (1971) was the first author to propose an econometric specification suited to analyse consumption decisions that can be assumed as being made in two steps. In the specific case of the model suggested by Cragg, he argued that individuals, when making decisions about the purchase of durable goods, have to make two sequential decisions. While in the first step of the decision process, the individual decides whether to buy the durable item, in the second stage the individual decides about the quantity to buy. Therefore, two different stochastic models are in play and are needed to explain consumer behaviour. Later on, Mullahy (1986) proposed a model with a similar decisional structure, however tailored to model a count variable, that, in the specific application developed by Mullahy, represents the daily consumption of various beverages, measured as the number of cups, glasses, etc.

In the case of decisions regarding medical care utilization, this two-stage decision process is still relevant, perhaps even more, because in medical care decisions the key decision makers in each decision stage are potentially different. While in the first stage it is only the individual who decides whether to seek medical care, in the second stage it is the individual, along with the physician, who decides about the intensity of care to meet the health needs (Pohlmeier and Ulrich 1995; Gerdtham 1997; Santos-Silva and Windmeijer 2001). This decision structure regarding medical care utilization is in accordance with the principal-agent framework. The individual (the principal) decides

to initiate the caring process, with a visit to a doctor, but once the contact with the physician (agent) is done, it is he/she, possible jointly with the patient, who decides about the total number of visits needed to finish the treatment. The hurdle specification has been popular in health economics literature partly because the parameters can have a structural interpretation compatible with the intuition of a dual decision process in health care.

The hurdle specification is also interesting from a purely statistical point of view. It is a model particularly suited to handle an important characteristic of the data that represent medical care utilization, which is the large proportion of zeros and a long right tail (Pohlmeier and Ulrich 1995; Santos-Silva 2001; Deb and Trivedi 2002; Winkelmann 2003).

From a statistical point of view, the hurdle model consists of two parts. The first part consists in a binary model that identifies the factors that distinguish between users (positive observations) and non-users (zeroes) of medical care, while the second part determines the factors that influence the intensity of medical care chosen by those who have positive medical care utilization.

For a general formulation of the hurdle model, let us assume that $y_i,\ i = 1,2,...,N$, denotes the count variable and $\mathbf{x}_i^{'} = \left[x_{i1}, x_{i2}, \cdots, x_{ik}\right]$ denotes a $(1xk)$ vector of covariates. In our specification, we constrained the covariates to be equal in both parts of the hurdle model, although some authors insert different covariates in each part of the hurdle model (Van Ourti 2004). Furthermore, let us also assume that the functions $f_0(.)$ and $f_1(.)$ are discrete probability functions suited for counts. Assume that $f_0(.)$ governs the first part of the model and $f_1(.)$ governs the second stage of the model.

Under these conditions, the probability of being a non-user of health care is given by $P\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta_0}\right) = f_0\left(0 \mid \mathbf{x}_i\right)$, while the probability of being a health care user is defined by $P\left(y_i > 0 \mid \mathbf{x}_i, \boldsymbol{\beta_0}\right) = 1 - f_0\left(0 \mid \mathbf{x}_i\right)$. The second part of the hurdle is specified as

a truncated distribution, generically given by $f_1(y_i \mid \mathbf{x}_i;\boldsymbol{\beta}_1, y_i > 0)$. Therefore, the probability function of the hurdle model can be written as,

$$f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_0, \boldsymbol{\beta}_0) = \begin{cases} f_0(0 \mid \mathbf{x}_i;\boldsymbol{\beta}_0) & \text{if} \quad y_i = 0 \\ \left[1 - f_0(0 \mid \mathbf{x}_i;\boldsymbol{\beta}_0)\right] f_1(y_i \mid \mathbf{x}_i;\boldsymbol{\beta}_1, y_i > 0) & \text{if} \quad y_i = 1,2,..., \end{cases} \qquad [16]$$

However, the most usual form of writing the probability function of the full model is obtained replacing $f_1(y_i \mid \mathbf{x}_i;\boldsymbol{\beta_1}, y_i > 0)$ by $f_1(y_i \mid \mathbf{x}_i;\boldsymbol{\beta}_1) = \dfrac{f_1(y_i \mid \mathbf{x}_i;\boldsymbol{\beta}_1)}{1 - f_1(0 \mid \mathbf{x}_i;\boldsymbol{\beta}_1)}$, giving rise to the most common specification of the model, defined as;

$$f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1) = \begin{cases} f_0(0 \mid \mathbf{x}_i;\boldsymbol{\beta}_0) & \text{if} \quad y_i = 0 \\ \dfrac{1 - f_0(0 \mid \mathbf{x}_i;\boldsymbol{\beta}_0)}{1 - f_1(0 \mid \mathbf{x}_i;\boldsymbol{\beta}_1)} f_1(y_i \mid \mathbf{x}_i;\boldsymbol{\beta}_1) & \text{if} \quad y_i = 1,2,..., \end{cases} \qquad [17]$$

Different probability functions for $f_0(.)$ and $f_1(.)$ leads to different specifications of the hurdle model. Note that $\left[1 - f_0(0 \mid \mathbf{x}_i;\boldsymbol{\beta}_0)\right]$ is the probability of crossing the hurdle (contact decision was made), and $\left[1 - f_1(0 \mid \mathbf{x}_i;\boldsymbol{\beta}_1)\right]$ is the truncation normalization for $f_1(.)$.

Although various alternatives for $f_0(.)$ and $f_1(.)$ are available to use in applications, the binary part of Poisson or NB are usually adopted. Probit ane Logit models are two other widely used alternatives. On the other hand, for $f_1(.)$ the alternatives more often used are the Poisson and negative binomial densities (Winkelmann 2004). For example, Mullahy (1986) adopted the Poisson distribution in both stages of the model, while Pohlmeier and Ulrich (1995) and Deb and Trivedi (2002) selected the NB probability function.

The main reason pointed out to use the NB density in the second stage of the model, instead of using, for instance the Poisson, is that in that stage it is necessary to account for unmeasured heterogeneity. This is important because, besides the usual sources of unobserved heterogeneity like, for instance, unmeasured health status, other

individual characteristics, patient preferences, etc., household micro-data hardly measures the doctor related heterogeneity regarding medical care decisions about the duration of the treatment (Pohlmeier and Ulrich 1995).

To give a practical example of a specification of the hurdle model, consider the case when the negative binomial probability function serves as baseline for both stages of the hurdle model. Therefore, the first part of the model is defined by the two equations that follows,

$$P\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_0\right) = f_0\left(0 \mid \mathbf{x}_i, \boldsymbol{\beta}_0\right) = \left(\frac{\eta_{i0}}{\lambda_{i0} + \eta_{i0}}\right)^{\eta_{i0}} \qquad [18]$$

and

$$P\left(y_i > 0 \mid \mathbf{x}_i, \beta_0\right) = 1 - P\left(y_i = 0 \mid \mathbf{x}_i, \beta_0\right) =$$
$$= 1 - \left(\frac{\eta_{i0}}{\lambda_{i0} + \eta_{i0}}\right)^{\eta_{i0}} \qquad [19]$$

As pointed out previously, the hurdle model consists in a second equation to determine how much medical care is used conditional on utilization being positive. This stage of the hurdle is modelled assuming that the data follow the density for a truncated negative binomial distribution, defined by

$$f_1\left(y_i \mid \mathbf{x}_i, y_i > 0\right) = \frac{\Gamma\left(y_i + \eta_{i1}\right) {\eta_{i1}}^{\eta_{i1}} \left(\lambda_{i1} + \eta_{i1}\right)^{-\left(\eta_{i1} + y_i\right)} {\lambda_{i1}}^{y_i}}{y_i!\,\Gamma\left(\eta_{i1}\right)\left(1 - \left(\frac{\eta_{i1}}{\lambda_{i1} + \eta_{i1}}\right)^{\eta_{i1}}\right)} \quad y_i = 1, 2... \qquad [20]$$

Therefore, the probability distribution of the overall hurdle, based on the negative binomial probability distribution is given by

$$f\left(y_i \mid \mathbf{x}_i;\boldsymbol{\beta}_0,\boldsymbol{\beta}_1\right)=$$

$$=\begin{cases}\left(\dfrac{\eta_{i0}}{\eta_{i0}+\lambda_{i0}}\right)^{\eta_{i0}} & \text{if} \quad y_i=0 \\ \dfrac{\left[1-\left(\dfrac{\eta_{i0}}{\eta_{i0}+\lambda_{i0}}\right)^{\eta_{i0}}\right]}{\left(1-\left(\dfrac{\eta_{i1}}{\lambda_{i1}+\eta_{i1}}\right)^{\eta_{i1}}\right)}\dfrac{\Gamma\left(y_i+\eta_{i1}\right)}{y_i!\Gamma\left(\eta_{i1}\right)}\eta_{i1}{}^{\eta_{i1}}\left(\lambda_{i1}+\eta_{i1}\right)^{-\left(\eta_{i1}+y_i\right)}\lambda_{i1}{}^{y_i} & \text{if} \quad y_i=1,2,..\end{cases} \qquad [21]$$

In the specification shown above, $\lambda_{i0}=\exp\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_0\right)$ refers to the conditional mean of the model that describes the first part of the model, while $\lambda_{i1}=\exp\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_1\right)$ designates the conditional mean function of the probability model $f_1(.)$. Moreover, in line to what was presented in Section 2.4 — The Negative Binomial regression model — the parameters η_{i0} and η_{i1} are defined as follows $\eta_{i0}=\left(\dfrac{1}{\alpha_0}\right)\lambda_{i0}^{k}$ and $\eta_{i1}=\left(\dfrac{1}{\alpha_1}\right)\lambda_{i1}^{k}$. Since the first stage process only contains binary information about the dependent variable, the parameters of the first part of the model — the parameters $\left(\boldsymbol{\beta}_0,\alpha_0\right)$— are not separately identifiable, thus, to identify the model it is common to set $\alpha_0=1$ (Pohlmeier and Ulrich 1995; Gurmu and Trivedi 1996; Deb and Holmes 2000).

The conditional mean and variance of the dependent variable in the overall population is, respectively, given by

$$E\left(y_i \mid \mathbf{x}_i,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\right)=\frac{1-f_0\left(0 \mid \mathbf{x}_i,\boldsymbol{\beta}_0\right)}{1-f_1\left(0 \mid \mathbf{x}_i,\boldsymbol{\beta}_1\right)}*\exp\left(\mathbf{x}_i^{'}\beta_1\right) \qquad [22]$$

and

$$Var\left(y_i \mid \mathbf{x}_i,\boldsymbol{\beta}_0,\boldsymbol{\beta}_1\right)=\frac{1-f_0\left(0 \mid \mathbf{x}_i,\boldsymbol{\beta}_0\right)}{1-f_1\left(0 \mid \mathbf{x}_i,\boldsymbol{\beta}_1\right)}\times$$
$$\times\left[\lambda_{i1}+\alpha_1\lambda_{i1}^{2-k}+\lambda_{i1}^{2}\left(1-\frac{1-f_0\left(0 \mid \mathbf{x}_i,\boldsymbol{\beta}_0\right)}{1-f_1\left(0 \mid \mathbf{x}_i,\boldsymbol{\beta}_1\right)}\right)\right] \qquad [23]$$

where $f_0\left(0 \mid \mathbf{x}_i, \boldsymbol{\beta}_0\right)$ is defined by [18] and $f_1\left(0 \mid \mathbf{x}_i, \boldsymbol{\beta}_1\right) = \left(\frac{\eta_{i1}}{\eta_{i1} + \lambda_{i1}}\right)^{\eta_{i1}}$.

Considering the users population, its conditional mean is given by

$$E\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_1, y_i > 0\right) = \frac{\exp\left(\mathbf{x}_i' \beta_1\right)}{1 - f_1\left(0 \mid \mathbf{x}_i, \boldsymbol{\beta}_1\right)} = \frac{\exp\left(\mathbf{x}_i' \beta_1\right)}{1 - \left(\frac{\eta_{i1}}{\eta_{i1} + \lambda_{i1}}\right)^{\eta_{i1}}} \qquad [24]$$

Turning now to the estimation of the model, the maximum likelihood estimator (MLE) can be found by maximizing the following log-likelihood function,

$$LogL = \sum_{i=1}^{N} \left[(1 - d_i) * \left[\log\left(f_0\left(0 \mid \boldsymbol{x}_i; \boldsymbol{\beta}_0\right)\right)\right] + d_i * \log\left(\frac{1 - f_0\left(0 \mid \boldsymbol{x}_i; \boldsymbol{\beta}_0\right)}{1 - f_1\left(0 \mid \boldsymbol{x}_i; \boldsymbol{\beta}_1\right)} f_1\left(y_i \mid \mathbf{x}_i; \boldsymbol{\beta}_1\right) \right) \right] \qquad [25]$$

where $d_i = \begin{cases} 1 & if \quad y_i > 0 \\ 0 & if \quad y_i = 0 \end{cases}$

As is well known, assuming independence in both parts of the model the log-likelihood function is given by [25] which is the sum of the log-likelihoods of the first and the second part of the model. This implies that estimation can be performed in two steps. A first step estimates, using the full sample, a binary choice model that analysis, from a regression viewpoint, the probability of observing a zero or a non-zero medical care utilization. On the other hand, in the second stage of the model a truncated-at-zero count data model is estimated using only the individuals with positive medical care utilization (Pohlmeier and Ulrich 1995; Gurmu and Trivedi 1996; Gurmu 1997; Deb and Trivedi 2002; Grootendorst 2002; Winkelmann 2003; Bago d'Uva 2005).

From the above we learned that the hurdle regression model has a two-fold appeal to explain medical care data. On one hand it can be considered as the empirical counterpart of the principal-agent model, on the other it has been reported a good statistical behaviour to fit data with a high percentage of zeros and long right tails (Pohlmeier and Ulrich 1995; Cameron and Trivedi 1998). Despite this double advantage of the hurdle specification it has been subjected to some critical remarks.

Next, we provide a brief excursion through the most usual criticisms.

2.6.1 — On the appropriateness of the hurdle specification to explain health care utilization: some comments

First, Deb and Trivedi (1997) argue the specification imposes a sharp dichotomy between the population of users and non-users of health care services, which, according to the authors, is difficult to defend. This is the case because in the case of demand for health care, even a healthy individual will not typically be a non-user due to the always-open possibility of using some form of precautionary medical care. We believe that underpinning this criticism is the assumption that the hurdle specification models the process of health care utilization over the individual's life cycle, or, at least, over a long period of time. This may probably the case when the hurdle framework is adopted to model medical care data in panel-data contexts, thus, aiming at describe long-term individual behaviour. Surely, this is not the case for hurdle models applied to cross-sectional data. In this last context, what the applied econometrician is modelling is, certainly, only the process of individual health care utilization over a relatively short time-period. Under these circumstances the division of the population into users and non-users is a reasonable one, rendering the hurdle framework as appropriate to explain medical care utilization.

Second, the difficulties regarding the hurdle specification as a model to handle data coming from health surveys that collect information on medical care utilization over a fixed time period and not over an episode of illness (Deb and Trivedi 2002). In regards to this, Pohlmeier and Ulrich (1995) identified two potential sources of complications when the data are collected in such circumstances. To begin with, the hurdle model assumes that the first recorded visit is the response of the patient to a new illness spell. This assumption may not be verified when the measurement period is fixed. This occurrence leads to a possible misclassification of the first count, because with no

additional information, this count may belong to the caring process of an illness episode of the preceding period (Pohlmeier and Ulrich 1995), hence not corresponding to an individual decision as is assumed by the hurdle framework. The easy conclusion is that the probability of this type of misclassification decreases as the observation period increases. However, if the duration of the observation period increases, this leads to the second and more serious problem, which is due to the possibility of occurrence of multiple illness spells during the observation period. A longer observation period increases the probability of observing multiple illnesses spells (Pohlmeier and Ulrich 1995; Gerdtham 1997) making impossible to estimate the first and second stage parameters of the hurdle because, under this circumstances, only the first stage parameters are identified (Santos-Silva and Windmeijer 2001). Santos-Silva and Windmeijer (2001) actually proposed a two-part model that allows for multiple sickness spells, however, regarding this difficulty, we follow Pohlmeier and Ulrich (1995) and Gerdtham (1997). We note that in our application the observation period used to capture the data is a period of three months, a rather short period. Moreover, the empirical distribution of the dependent variable shows that roughly 70% of the individuals in the sample had at most one visit during the three-month period. Hence, in our applications of the hurdle framework we follow Pohlmeier and Ulrich (1995) and Gerdtham (1997) and assume that the occurrence of multiple illnesses spell is a rare event, meaning that we do not have identification problems when estimating hurdle models.

Third, another issue related to the hurdle specification has been pointed out by Deb and Trivedi (2002) who criticize the specification of the conditional mean of the users population presented by [24] for the case of a hurdle based on a single linear index. They clearly mentioned that in the hurdle model based either on the Poisson or on the NB probability function, the specification of the conditional mean for the users population is not likely to capture different responses to changes in x_{ik} in the right tail

of the distribution because the response of $E(y_i \mid \mathbf{x}_i, y_i > 0)$ to a regressor x_{ik} is fixed by $\exp(\mathbf{x}_i'\beta)$. This is similar to state that as all health care users are modelled by only one equation in the second stage, it does not allow for heterogeneity in behaviours among sub-groups of users. If it urns out that users are drawn from distinct behavioural groups, then, the hurdle will not be adequately represent the utilization for health care services.

Fourth, other criticisms made to the specification of the hurdle are concerned with issues regarding the specification of the model for the positive observations whenever the researcher wants to incorporate explicitly the unobserved heterogeneity supposedly present in the second part of the model. In applied work, it has been common to specify the second stage of the hurdle as a truncated-at-zero Poisson or a truncated-at-zero NB model, with the preference going to the utilization of the NB model due to its ability to account for the unobserved factors. However, Gurmu (1997) criticized the utilization of the truncated-at-zero negative binomial model to explain the positives because this can lead to poor results, therefore, to misleading conclusions. The author highlighted that the hurdle based on the NB distribution rests on the explicit assumption that the unobserved heterogeneity is gamma distributed. In addition Gurmu (1997) also notes that without any prior information about the true distribution of the unobserved heterogeneity, the gamma assumption cannot be fully justified, hence arbitrary. Therefore, in the case of a misspecification of the unobserved heterogeneity the estimation process leads to inconsistent estimates due to the consequent misspecification of the assumed distribution for the counts. From the applied point of view of using such estimates for policymaking, this possibly inconsistent estimates lead to biased policy advice. The inconsistent estimates arise in the case of a misspecification because hurdle models are nothing more than refinements of truncation and censoring, thus pseudo-maximum likelihood methods

are not applicable (Gurmu 1997; Gurmu, Rilstone *et al.* 1999), hence making this class of models non-robust to misspecifications.

In an attempt to respond to these criticisms, Gurmu (1997) suggested a semi-parametric mixture hurdle model not requiring prior knowledge of the distribution of the unobserved factors. However, even Gurmu's alternative semi-parametric mixture hurdle model may misspecify the part for strictly positive counts (Santos-Silva 2003). Santos-Silva (2003) identified an extra, much more subtle possibility of misspecification in hurdle models. The potential misspecification identified by Santos-Silva arises only in the second part of the hurdle when the researcher intends to account for the unobserved heterogeneity in the model specification. The specification and estimation of models for truncated counts with unobserved heterogeneity is a closely related matter. Thus, in what follows, we present some technical issues regarding the specification of models for truncated counts, and in Chapter 4, will return to the issues connected with the specification of the hurdle.

2.7 — Models for truncated counts

Truncated count data models are required to explain data collected by using sampling schemes that disregard observations above and/or below some value of the endogenous variable. The consequence of such endogenous sampling is that the researcher does not observe the complete distribution of the count. Although samples can be truncated from left or from right, with any integer as the truncation threshold, in empirical applications truncation from left, with one as the truncation threshold, is the most common form. This results in a truncated-at-zero dataset where zeros are not observed, being this type of truncation the one that interest us throughout this dissertation.

Models allowing for truncation have been the focus of much research, and several papers have analysed data with this characteristic, though pure applications to health

care data are somewhat uncommon (Creel and Loomis 1990; Grogger and Carson 1991; Gurmu 1991; Gurmu and Trivedi 1992; Brannas and Rosenqvist 1994). As it is well known, the straightforward application of standard count data models to truncated samples leads to inconsistent estimates, thus, suitable modification of standard count data models have to be made to make valid inference (Grogger and Carson 1991; Gurmu and Trivedi 1992; Santos-Silva 2003; Cameron and Trivedi 2005).

Let us assume that the discrete probability function $f(y_i \mid \mathbf{x}_i, \beta)$ governs the health care utilization of individual *i* in the overall population. Then, the adequate probability model in the sample, generically referred to as $f_s(y_i \mid \mathbf{x}_i, \boldsymbol{\beta})$, is defined by the following probability model

$$f_s(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}) = f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}, y_i > 0) = \frac{f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta})}{P(y_i > 0 \mid \mathbf{x}_i, \boldsymbol{\beta})} \quad y_i = 1, 2, \ldots \qquad [26]$$

The common specification for the conditional mean parameter of $f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta})$ continues to be $\lambda_i = \exp(\mathbf{x}_i' \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of unknown parameters.

As an example, let us assume that the data generating mechanism in the actual population is governed by a Poisson probability function with exponetial mean parameter λ_i. Therefore, the corresponding probability model in the sample is given by

$$f_s(y_i \mid \mathbf{x}_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!(1 - e^{-\lambda_i})} \quad y_i = 1, 2, \ldots \qquad [27]$$

This model is known in the econometric literature as the positive Poisson regression model (Grogger and Carson 1991; Gurmu and Trivedi 1992; Cameron and Trivedi 1998). For this specific model, the conditional mean and variance of the density in the truncated sample are given, respectively, by

$$E(y_i \mid \mathbf{x}_i, y_i > 0) = \lambda_{is} = \frac{\lambda_i}{1 - e^{-\lambda_i}} \qquad [28]$$

and

$$V\left(y_i \mid \mathbf{x}_i, \beta, y_i > 0\right) = \lambda_{is}\left[1 - P\left(y_i = 0 \mid \mathbf{x}_i, \beta\right)\lambda_{is}\right] \quad [29]$$

In the standard Poisson regression model, the parameters can be consistently estimated even in the presence of '*overdispersion*' caused by the existence of unobserved heterogeneity, because pseudo-maximum likelihood applies, making the parameter estimates robust to the presence of unobserved heterogeneity (Gourieroux, Monfort *et al.* 1984; Wooldridge 1997). Conversely, in the positive Poisson regression model, consistent parameter estimates require the complete and proper specification of all aspects of the distribution of the count. That is to say that disregarding unobserved heterogeneity in truncated models leads to inconsistent parameters estimates.

Therefore, in the modelling of truncated data, the econometrician shall use models adequate to account explicitly for unobserved heterogeneity.

As presented in Section 2.3, one common and well-known approach of accounting for the presence of unobserved heterogeneity in count data modelling is through the adoption of specifications based on latent class models. Brannas and Rosenqvist (1994) defend that the approach based on mixture models, that moves the model away from strong distributional assumptions, is even more relevant in the truncated context than in the non-truncated one. Because we are in a truncated data context, mixture models for truncated counts should be specified and estimated (Grogger and Carson 1991; Cameron and Trivedi 1998; Santos-Silva 2003). However, as pointed out clearly by Santos-Silva (2003), the specification of mixture models in a truncated context is not as linear as in the standard case, therefore, the next section shows some of the caveats that may arise when the researcher specifies mixture models to explain truncated data.

2.8 — Specification of latent class models for truncated samples

Santos-Silva (2003) studied the impact of endogenous sampling, of which truncation is a particular case (Santos-Silva 2003; Cameron and Trivedi 2005), in the distribution of the unobserved heterogeneity. The author highlighted that in truncated samples the

investigator may account for the unobservables into two different ways, and the choice of the correct one can be an issue open for debate.

As already shown above, the specification presented in [26] is the appropriate probability model to analyse truncated-at-zero count data. Let us now consider that the researcher's goal is to account for the unobserved heterogeneity, supposedly present in the data. Naturally, one natural approach to deal with the unobserved heterogeneity is through the specification of LCMs.

Accordingly, let us assume that the probabilistic model in the actual population is $f\left(y_i \mid \mathbf{x}_i\right)$ and is specified as a latent class model. Under this hypothesis, the data generating mechanism in the actual population is given by the specification presented through Equation [9], which we repeat this point,

$$f\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \sum_{j=1}^{P} \pi_j f_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right) \qquad [30]$$

where $\boldsymbol{\beta}' = \left[\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_p'\right]$ is a vector of parameters that characterizes the overall population and π_j, $\boldsymbol{j} = 1, 2, ..., \boldsymbol{P}$ are the mixing probabilities of the mixture[4].

In this context, it is straightforward to conclude that the denominator of the density function presented in [26] is given by

$$\left[1 - P\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}\right)\right] = \left[1 - \left(\sum_{j=1}^{P} \pi_j P_j\left(y_i = 0 \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)\right)\right] \qquad [31]$$

Plugging equations [30] and [31] into equation [26], one can conclude that the LCM based density in the truncated sample can be written as

$$f_s\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \frac{\sum_{j=1}^{P} \pi_j f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)}{1 - \left(\sum_{j=1}^{P} \pi_j P_j\left(y_i = 0 \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)\right)} \qquad [32]$$

---

[4] Please, see section 2.5 for details about this family of specifications.

The expression presented above can be expressed differently, as is presented below,

$$f_s\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \sum_{j=1}^{P} \left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)}{1 - \left(\sum_{j=1}^{P} \pi_j P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right) \right)} \right) \pi_j \qquad [33]$$

It is easy to show that in the previous expression a redefinition on the mixing probabilities of the mixture [30], that is, a redefinition of the $\pi's$, will allow one to re-express the density in the sample.

Simply, multiply and divide each term of the summation presented in [33] by $\left[1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)\right]$, to obtain the following,

$$f_s\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) =$$

$$= \sum_{j=1}^{P} \left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right) * \left[1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)\right]}{\left[1 - \left(\sum_{j=1}^{P} \pi_j P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right) \right)\right] * \left[1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)\right]} \right) \pi_j \qquad [34]$$

$$= \sum_{j=1}^{P} \left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)}{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)} \right) \hat{\pi}_j \qquad [35]$$

$$= \sum_{j=1}^{P} f_j^s\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right) \hat{\pi}_j \qquad [36]$$

In the density presented in [36], $f_j^s\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \beta_j\right)$ $j = 1, 2, ..., P$ is the probability function of the *j*th latent class of individuals in the truncated sample, that is,

$$f_j^s\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right) = \frac{f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \beta_j\right)}{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \upsilon_i^j, \beta_j\right)} \qquad [37]$$

In this case, the mixing probabilities in the truncated population were redefined, and presently are given by

$$\hat{\pi}_j = \frac{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \upsilon_i^j; \boldsymbol{\beta}_j\right)}{1 - \left(\sum_{j=1}^{P} \pi_j P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)\right)} \pi_j \qquad [38]$$

Summing up, to account for the presence of the unobservables in truncated samples, the density of the count in the truncated sample, $f_s\left(y_i \mid \mathbf{x}_i\right)$, can be written in two different ways

$$f_s\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \frac{\sum_{j=1}^{P} \pi_j f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)}{1 - \left(\sum_{j=1}^{P} \pi_j P_j\left(y_i = 0 \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)\right)} \qquad [39]$$

or

$$f_s\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \sum_{j=1}^{P} \left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)}{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)} \right) \hat{\pi}_j \qquad [40]$$

where $\hat{\pi}_j$ is defined above.

Santos-Silva (2003) and Bohning and Kuhnert (2006) claim that the truncated mixture of densities given by [39], is different from the mixture of truncated densities defined by [40], therefore care must be taken about which specification to use in applied work.

The specification given by the density model presented in [39] assumes that the unobserved factors correspond to the individuals actually present in the actual population implicitly making assumptions about the distribution of the unobserved heterogeneity in the overall population. On the other hand, the specification given by the density model presented in [40] assumes that the unobserved factors incorporated in the model belong to the individuals present in the truncated population, hence making assumptions about the distribution of the unobserved heterogeneity in this smaller population.

These two alternative, and different, specifications of the unobservables distribution lead the analyst to a crossroads regarding the choice of the proper specification to model the truncated data. Santos-Silva (2003) pointed out that it is different to choose between [39] and [40]. In addition, the author mentions that care is needed in deciding which model is the most suitable to fulfil the aims of the analysis. The specification given by [39] is the proper formulation when the population of interest is the actual population, while the researcher should use [40] if the empirical analysis aims at analysing the population induced by the sampling scheme (Santos-Silva 2003).

Therefore, when the study aims at analysing the actual population, the researcher assumes that the unobserved factors are in the overall population, and so, these unobserved factors contribute to generate the latent classes. Hence it makes sense first to specify a mixture model in the actual population and, subsequently, to truncate the mixture, resulting in specification [39]. On the contrary, when the target of the study is the truncated population, it is assumed that the unobserved factors are in this population and so, these unobserved factors aggregate in the population of the positives to form latent classes. Hence, in this case, one should at first specify a truncated distribution to represent each latent class of users, and, after that, mixture the truncated distributions that represent each latent class of users, resulting the density [40].

2.9 — Options concerning LCM specification and estimation

This section presents the choices and procedures regarding model specification and estimation that we have taken throughout this thesis. We begin by showing our choices relative to the econometric specifications based the latent class framework, which are prevalent in this dissertation.

In section 2.5 — Latent Class models —, we introduced the basics about LCMs, however from a generic point of view. To make such specifications functional from an

applied point of view, the analyst must make two decisions: first, the number of latent classes (P), and second, the probability function that governs the utilization within each one of the latent classes.

Beginning with the discussion concerning the number of latent classes, we first note that the consistent estimation of the value of the number of latent classes is a complex and still unresolved research question (McLachlan and Peel 2000; James, Preibe *et al.* 2001). This dissertation does not aim at contributing to this interesting, however, challenging research issue. In this thesis all specifications that encompass the specification of a latent class model assumes models with two latent classes, setting, therefore P = 2. A number of reasons support this choice:

- First, Wedel *et al.* (1993) mention that the empirical evidence has shown that a small number of latent classes provide enough flexibility to reproduce the data accurately,
- Second, the vast majority of studies that applied the LCM methodology to the analysis of medical care utilization have reported that two latent classes provide sufficient flexibility to explain medical care counts quite well (Deb and Trivedi 1997; Deb and Holmes 2000; Deb and Trivedi 2002; Jimenez-Martin, Labeaga *et al.* 2002; Atella, Brindisi *et al.* 2004). We should mention, however, that, to the extent of our knowledge, at least two empirical studies have reported three latent classes as the adequate number of classes to fit the data (Deb 2001; Bago d'Uva 2005),
- Third, LCMs when used in a regression contexts, are not parsimonious in the use of parameters, thus, it is likely that models with more than two latent classes demand too much for the data, becoming rapidly overparametrized. The result is that the LCMs with more than two latent classes are usually very difficult to estimate. In fact, we conducted some experiments with models with more than two latent classes and, as expected, experienced some difficulties in the estimation process.

In conclusion, in this dissertation we decided to disregard LCMs with more than two classes, thus restriction all our LCMs to two points of support.

Regarding the choice of the probability function for that governs medical care utilization within each latent class, that is, the choice of each $f_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right),\ j = 1, 2, ..., P$, (please see equation [9]). Despite the, always open, possibility of choosing densities from different families for different components, for instance, a combination of Poisson and NB probability functions, it has been usual in applications to assume densities of the same parametric family[5]. Therefore, in all specifications built upon the LCM framework, we choose the negative binomial density, or its truncated version for the case of some models specified in Chapter 4, to represent each component distribution of the LCM. In the context of the specification of LCMs, our preference for densities of this family is because this class of probability models offer general and flexible densities to model count data in a regression setting, and not from their well-known ability to account for the unobserved heterogeneity. Throughout this dissertation, our preferred approach to incorporate explicitly unobserved heterogeneity in the specifications is using latent class models, thus, in the finite mixture spirit. Moreover, in Chapter 4 we aim at comparing the statistical performance of an alternative specification of the hurdle model with the models previously used in the literature, therefore, the same underlying distribution is used here (Bago d'Uva 2006). In addition, the evidence favouring latent class models with NB as baseline distribution is numerous (Deb and Trivedi 1997; Deb and Holmes 2000;

---

[5] In some experiments, whose results are not reported in this dissertation, we estimated latent class models that combined densities from different families, for example, the Poisson combined with densities of the NB family. However, the results did not show any substantial improvements relative to the results presented in this study.

Deb and Trivedi 2002; Jimenez-Martin, Labeaga *et al.* 2002; Bago d'Uva 2005; Lourenço and Ferreira 2005).

Regarding the estimation techniques, all models presented in this dissertation were estimated by maximum likelihood (ML). Moreover, the reported standard errors are robust standard errors, estimated using the robust sandwich estimator. This estimator, although not assuring consistent estimates of the model parameters in the case of a misspecification, it provides such consistent estimates for the corresponding standard errors, even in the presence of misspecification of some part of the distribution (White 1982; Cameron and Trivedi 1998).

Contrarily to the one-component negative binomial regression and the popular hurdle regression models, the estimation of models involving latent class specifications can be challenging. Two methods have been proposed to estimate models involving LCMs: the Expected-Maximization algorithm (EM) to maximize the likelihood function and the direct optimization of a likelihood function.

The EM algorithm was proposed by Dempster *et al.* (1977) and is described in detail in Wedel *et al.* (1993) and McLachlan and Peel (2000) among others. Cameron and Trivedi (1998) point out that the EM algorithm may be slow to converge, especially if bad starting values are used to initialize the process and thus other approaches to maximize the likelihood function may also be worth considering. Therefore, all models involving LCM specifications were estimated by direct maximization of the likelihood function. However, estimating this class of models using direct optimization routines can be difficult as the likelihood function of such models may have multiple local maximum. Hence one cannot exclude the possibility of convergence to local optimums, (McLachlan and Peel 2000). Therefore, it is important to ensure that the numerical algorithm reaches the global maximum likelihood estimator. In the latent class models estimated in this thesis, we have guarded against the possibility of reaching local solutions by estimating repeatedly each LCM using a number of different initial solutions. To obtain the initial solutions to the optimization routines,

we used simpler models. Using this procedure, we estimated repeatedly every model (about 20 times) and did not observe convergence problems or convergence to local solutions.

We used Stata 9.0 to estimate all models. Moreover, we also used the statistical package Stata to perform all the numerical computations needed in the post-estimation analysis, namely, the implementation of hypothesis tests and the calculation of average marginal values and their respective standard errors. For each competing model considered, we wrote a Stata program to evaluate the log-likelihood function (Gould and Sribney 2003), which was afterwards maximized using the Stata ML command, with the unconstrained Broyden-Fletcher-Goldfard-Shanno (BFGS) algorithm as the optimization option. Appendixes A and B present the programs, written in Stata, to obtain the estimates presented throughout this dissertation.

2.10 — Model Comparison and Selection Procedures

In all applications developed in this dissertation, we have fitted the data using a wide range of alternative count data models, therefore, one has to use some criteria to assess the relative performance of the models in order to make the appropriate choice among the competing specifications. This section presents the techniques that we have used to compare and select the econometric specification that better fits the data. Cameron and Trivedi (1998), Winkelmann (2003) and Allvarez (2002) among others, present useful surveys regarding the theme of model evaluation and testing techniques in the count data framework.

One context where discriminating among competing econometric specifications has a well-defined framework is when the models are estimated by maximum likelihood and the competing models are nested. A more difficult situation arises when the competing models are non-nested, that is, when neither specification can be obtained from the other by imposing constraints on the model parameters. Non-nested models

arise often as competing alternatives when the application involves the analysis of count dependent variable.

Winkelmann (2003) suggests that there are two different ways to discriminate among non-nested specifications: hypothesis testing and model selection. From the alternatives available in the literature to discriminate among this type of non-nested econometric models, in this dissertation we compare them using two methods: Voung tests (VuT) and Information Criteria (IC) selection methods. In Winkelmann's (2003) classification, the Voung tests fit in the hypotheses testing category, while the information criteria belong to the model selection category.

2.10.1 — Likelihood ratio tests (LR tests)

There are three classical techniques for testing hypotheses when models are estimated by maximum likelihood: the likelihood ratio (LR) test, the Wald test and the Lagrange multiplier (LM) test (Davidson and Mackinnon 1993; Winkelmann 2003; Cameron and Trivedi 2005). Because they are simpler to use, throughout this dissertation we privileged the utilization of LR tests.

The likelihood-ratio test statistic is given by

$$LR = -2\left[\ln L(\beta_r) - \ln L(\beta_u)\right] \sim \chi^2(h) \qquad [41]$$

where $\ln L(\beta_r)$ is the maximized log-likelihood of the constrained model, $\ln L(\beta_u)$ is the maximized log-likelihood of the unconstrained model and h is the number of constraints. Under H_0, the statistics is asymptotically chi-square distributed, with h degrees of freedom. The implementation of this statistical test is straightforward being therefore unnecessary to present further details about it.

One case where it is usual to apply this class of tests is in the comparison of standard models with latent class formulations, being more accurate, models having inherent one population against the alternative of latent class models of two sub-populations (Deb and Trivedi 1997; Deb and Holmes 2000; Deb 2002). In the last situation,

however, care must be taken when analysing the results of the test of the hypothesis of one population against the alternative of two populations because the test involves a parameter restriction that is on the boundary of the parameter space, thus violating the standard regularity conditions for maximum likelihood (Davidson and Mackinnon 1993; Wooldridge 2002). The consequences of that violation are that, under the null, the LR test statistic does not follow the standard $\chi^2(q)$ distribution (Chernoff 1954; Cameron and Trivedi 1998). Bohning (1995) and Deb and Trivedi (1997) claim that, under this circumstances resorting on the usual $\chi^2(q)$ distribution is likely to under-reject the null hypothesis. Therefore, it is easy to argue that the output of this kind of tests is inconclusive only when it results in non-rejection of the null and not when it concludes for the rejection of the null hypothesis.

2.10.2 — Voung Tests (VuT)

The Vuong test is a statistical test based on the likelihood ratio principle, which is appropriate to discriminate between a pair of two competing non-nested econometric specifications (Vuong 1989).

Consider the problem of choosing, from a statistical point of view, between two non-nested probability models. Let us assume that the first model is referred to as F_{β_1}, with density function $f(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_1)$ and the second model, referred to as G_{β_2}, with probability function $g(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_2)$.

Under this framework, the null hypothesis amounts of considering that the competing models are equivalent, whereas the alternative hypothesis is that either model F_{β_1} is better than model G_{β_2} or model G_{β_2} is better than model F_{β_1}. Statistically, the hypothesis can be formulated as follows,

$$H_0 : F_{\beta_1} \text{ and } G_{\beta_2} \text{ are equivalent models}$$

$$vs$$

$$H_1 : \left(F_{\beta_1} \text{ is better than } G_{\beta_2}\right) \text{ or } \left(G_{\beta_2} \text{ is better than } F_{\beta_1}\right)$$

If the probability models under evaluation (F_{β_1} and G_{β_2}) are strictly non-nested, that is, neither can be obtained as a specialization of the other, Voung proposed the use the following test statistic,

$$V\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right) = \frac{\sum_{i=1}^{n} \ln\left(\frac{f\left(y_i \mid \mathbf{x}_i, \hat{\boldsymbol{\beta}}_1\right)}{g\left(y_i \mid \mathbf{x}_i, \hat{\boldsymbol{\beta}}_2\right)}\right)}{\hat{w}\sqrt{n}} = \frac{L_f\left(\hat{\boldsymbol{\beta}}_1\right) - L_g\left(\hat{\boldsymbol{\beta}}_2\right)}{\hat{w}\sqrt{n}} = \frac{LR\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right)}{\hat{w}\sqrt{n}} \qquad [42]$$

where $L_f\left(\hat{\boldsymbol{\beta}}_1\right)$ is the log-likelihood value of the model $f\left(y_i \mid \mathbf{x}_i, \hat{\boldsymbol{\beta}}_1\right)$ and, on the other hand, $L_g\left(\hat{\boldsymbol{\beta}}_2\right)$ is the log-likelihood value of the model $g\left(y_i \mid \mathbf{x}_i, \hat{\boldsymbol{\beta}}_2\right)$. Note that both these values are available after the estimation of the two competing alternatives.

In the test statistics $V\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right)$, presented in [42], the term $\hat{w}^2$ is a consistent estimate of its variance which can be easily computed, after the estimation of both models, using the following statistic,

$$\hat{w}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\ln\left(\frac{f\left(y_i \mid \mathbf{x}_i, \hat{\boldsymbol{\beta}}_1\right)}{g\left(y_i \mid \mathbf{x}_i, \hat{\boldsymbol{\beta}}_2\right)}\right)\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{f\left(y_i \mid \mathbf{x}_i, \hat{\boldsymbol{\beta}}_1\right)}{g\left(y_i \mid \mathbf{x}_i, \hat{\boldsymbol{\beta}}_2\right)}\right)\right]^2 \qquad [43]$$

Voung showed that $V\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right)$ follows a standard normal distribution. Accordingly, the critical value c that determines the decision rule of the test is taken from the normal distribution. Let us assume that the critical value is represented to as Z_α. Accordingly, assuming a test at the significant level α, the decision rule suggested by Vuong's (1989) is defined as:

- If $V\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right) > Z_\alpha$ then one rejects the null hypothesis of equivalent models in favour of F_{β_1} being better than G_{β_2}
- If $V\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right) < -Z_\alpha$ then one rejects the null hypothesis in favour of G_{β_2} being better than F_{β_1}.
- If $-Z_\alpha \leq V\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right) \leq Z_\alpha$ then, the data available do not permit to reject the null hypothesis, being impossible to discriminate the two competing models.

2.10.3 — Information Criteria (IC)

The Information criteria approach to model selection is based on information theory. It is a method of penalized likelihood based on the fitted log-likelihood function and a leading procedure for selecting one of several competing models (Sin and White 1996). The aim is not to detect the true model but the model that provides the most information about the real world. Despite the various selection statistics that may be used under this model selection methodology, two statistics have received the attention of researchers in count data contexts, the Akaike Information Criteria (AIC), or its consistent version, the Consistent Akaike Information Criteria (CAIC), and the Bayesian Information Criterion (BIC). As is referred by Sin (1996) this kind of model selection techniques can be soundly applied to a variety of models, including nested and non-nested models, linear and non-linear, correctly specified and misspecified models.

The Bayesian Information Criterion and the Consistent Akaike Information Criteria statistics are defined, respectively by [44] and [45], presented below,

$$BIC = -2\ln(L) + k\ln(n) \qquad [44]$$

$$CAIC = -2\ln(L) + k\left[1 + \ln(n)\right] \qquad [45]$$

In the expressions above, $\ln(L)$ represents the logarithm of the likelihood function in the maximum likelihood estimator, k is the number of parameters of the model and n the sample size.

Models presenting lower values for both statistics are preferred (Deb and Trivedi 2002).

2.11 — Interpretation of the parameters in nonlinear regression models

Conversely to linear regression models, where the interpretation of the regression coefficients is generally straightforward, for non-linear regression models the interpretation of the parameters is not so simple, requiring, most of the times, a thorough analysis. The objective of the current section is to indicate some possibilities about the interpretation of estimates in non-linear models, namely, in exponential regression models, which are the family of models that we use in the applications developed in this dissertation, as it has been motioned so far throughout this chapter.

The more elaborated interpretation of non-linear regression models are due to, essentially, for two reasons:

- In exponential regression models, the conditional mean function, which usually is the descriptor of the distribution that the investigator is interested in, is nonlinear. In our specifications, the simplest case is when the functional form of the conditional mean is the exponential of a linear index, that is, $E(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}) = \exp(\mathbf{x}_i'\boldsymbol{\beta})$. In the case of specifications that are more complex, like the models of the hurdle or LCM families, the conditional mean function for the overall population is much more complex. For example, in the case of hurdle models, the conditional mean function is given by [22] , that is, $\left\{E(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \frac{1 - f_0(0 \mid \mathbf{x}_i, \boldsymbol{\beta}_0)}{1 - f_1(0 \mid \mathbf{x}_i, \boldsymbol{\beta}_1)} * \exp(\mathbf{x}_i'\beta_1)\right\}$, and in the case of LCMs is

given by equation [10], that is $E\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_\mathbf{j}\right) = \sum_{j=1}^{P} \pi_j \exp\left(\mathbf{x}_i^{'} \beta_j\right)$. As is readily acknowledged, in all cases, the parameters β_k no longer measures the impact of a one-unit change of the regressor x_k in the mean function, requiring, therefore, more elaborated interpretation;

- The second difficulty in the interpretation of the results in nonlinear models, more prominently in count data models, occurs because, besides the conditional mean of the distribution, the researcher is usually interested in analysing some other descriptors of the conditional distribution of the count. Notice that discrete data allow estimating the probabilities of single outcomes after the model has been estimated. For example, it may be relevant to assess the effect of some selected covariates on single probabilities. Because these probabilities generally depend on the parameters and on the covariates, the evaluation of such effect may require more or less elaborated algebra.

Let us consider a generic nonlinear model. Assume, for now, that the researcher aims at studying the influence of a covariate, say x_k, on a generic descriptor of the conditional distribution that is used to model the dependent variable. Throughout the current section we call $G\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_1, \mathbf{x}_i^{'}\boldsymbol{\beta}_2, ...,\right)$ to such descriptor. In the case of the models estimated in this dissertation, the function $G\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_1, \mathbf{x}_i^{'}\boldsymbol{\beta}_2, ...,\right)$ may take several forms: in some cases, it depends on only one linear index and in other cases is function of two, or even three, linear indexes. For example, in the standard latent class models, if the interest lies in the analysis of the conditional mean function for the latent class *j* then the function $G(.)$ is given by $G\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_j\right) = E\left(y_i \mid \mathbf{x}_i, \beta_j\right) = \exp\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_\boldsymbol{j}\right)$. Conversely, if the interest lies in the analysis of the conditional mean function of the overall population, then $G\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_1, \mathbf{x}_i^{'}\boldsymbol{\beta}_2, ..., \mathbf{x}_i^{'}\boldsymbol{\beta}_j\right) = \sum_{j=1}^{P} \pi_j \exp\left(\mathbf{x}_i^{'}\boldsymbol{\beta}_\boldsymbol{j}\right)$. Yet another example, the conditional

mean function of the model from the hurdle family that consists of a binary model for the first stage and a latent class model for the second stage, which will be presented in Chapter 4 is a non-linear function dependent on three linear indexes.

In order to make the presentation easier, let us assume that the covariate x_{ik} is continuous and that the generic function $G(.)$ depends on one linear index only. The analysis for multiple index functions is straightforward, involving only more elaborated calculus.

In general, in applied work, the research is usually interested in the evaluation of the marginal effects, that is, the alteration that $G\left(\mathbf{x}_i'\boldsymbol{\beta}\right)$ suffers when x_k is changed by a small amount, for example, one unit change. As it is well known, such marginal effects are given by the partial derivative of the $G\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)$ in order to x_k.

Let ME_i denote the marginal effect associated to the covariate x_k, then,

$$ME_i = \frac{\partial G\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial x_{ik}} = \frac{\partial G\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial \mathbf{x}_i'\boldsymbol{\beta}} \frac{\partial\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial x_{ik}} \qquad [46]$$

Note that $\mathbf{x}_i'\boldsymbol{\beta}$ is a linear function in the vector $\mathbf{x}_i$, thus $\frac{\partial\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial x_{ik}} = \beta_k$. Therefore, it is easy to conclude the marginal effects are defined by

$$ME_i = \frac{\partial G\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial x_{ik}} = \frac{\partial G\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial \mathbf{x}_i'\boldsymbol{\beta}} \beta_k \qquad [47]$$

In order to abridge the notation in the subsequent analysis that we going to conduct, let us assume that the partial derivative of $G(.)$ in order to $\mathbf{x}_i'\boldsymbol{\beta}$ is given by the function $g(.)$, that is

$$\frac{\partial G\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial \mathbf{x}_i'\boldsymbol{\beta}} = g\left(\mathbf{x}_i'\boldsymbol{\beta}\right) \qquad [48]$$

hence, the marginal effect for the covariate x_k can be written as

$$ME_i = \frac{\partial G\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial x_{ik}} = g\left(\mathbf{x}_i'\boldsymbol{\beta}\right)\beta_k \qquad [49]$$

From expression [49], shown above, one can readily conclude that the marginal effect, unlike in linear models, is a function of both regressors and parameters, therefore, its magnitude depends on the vector of independent variables $\mathbf{x}_i$ together with the vector of parameters $\boldsymbol{\beta}$. We note that in the notation presented above, the marginal effect is identified by ME_i, where the subscript *i* is inserted to enhance the view that the marginal effects varies from individual to individual.

The direct interpretation of the individual marginal effects is difficult because, as it was already shown, they vary across individuals. In general, the researcher has two alternatives: it may analyse the effect of the covariate on the individuals, studying particular cases, or, on the other hand, can try to find a global indicator that rapidly convey information about the effect of interest. Cameron and Trivedi (1998; 2005) indicates three alternatives to compute such unified, easy to read, statistic.

To begin with, one may compute the marginal effect for each individual $\left(ME_i\right)$ and then average across all individuals, resulting in an average response for the population. In this situation, the overall indicator of the marginal effect, is an average marginal effect, henceforth referred to as AME, and is given by

$$AME = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial G\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{\partial x_{ik}} = \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{x}_i'\boldsymbol{\beta}\right)\beta_k \qquad [50]$$

As an alternative to the procedure presented above, consider the evaluation of expression [49], not for each individual *i*, but, alternatively, at the vector of sample means of the independent variables, referred to as $\bar{\mathbf{x}}$. This result on the following estimator

$$ME_{\bar{x}} = \frac{\partial G\left(\bar{\boldsymbol{x}}'\boldsymbol{\beta}\right)}{\partial x_{ik}} = g\left(\bar{\boldsymbol{x}}'\boldsymbol{\beta}\right)\beta_k \qquad [51]$$

A third possibility suggested by Cameron and Trivedi is to evaluate the estimate of the marginal effect for a typical individual with given characteristics, say $\mathbf{x} = \mathbf{x}^{\bullet}$. In this case, the global indicator for the marginal effect is given by

$$ME_{\mathbf{x}^{\bullet}} = \frac{\partial G\left(\mathbf{x}^{\prime\bullet}\boldsymbol{\beta}\right)}{\partial x_{ik}} = g\left(\mathbf{x}^{\bullet\prime}\boldsymbol{\beta}\right)\beta_k \qquad [52]$$

From the three alternative approaches to compute an overall indicator to assess the effect of a covariate on a generic function G(.) — AME, $ME_{\bar{\mathbf{x}}}$ and $ME_{\mathbf{x}^{\bullet}}$ — Cameron and Trivedi (2005) point out that the conceptually better and more relevant indicator is the average marginal effect (AME), when interpretation is to be used for policy making reasons.

The estimate of the AME is computed by plugging the consistent estimator of $\boldsymbol{\beta}$, which is $\hat{\boldsymbol{\beta}}$, into expression [50], resulting in

$$\hat{AME} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial G\left(\boldsymbol{x}_i^{\prime}\hat{\boldsymbol{\beta}}\right)}{\partial x_{ik}} = \frac{1}{n}\sum_{i=1}^{n} g\left(\boldsymbol{x}_i^{\prime}\hat{\boldsymbol{\beta}}\right)\hat{\beta}_k \qquad [53]$$

An issue that still needs to be addressed concerns the statistical distribution of the estimate of the average marginal effect. The knowledge of this distribution is relevant in order to test statistical hypothesis about this estimator, e.g. to test whether the estimated average marginal effect is statistically different from zero.

Following Ai and Norton (2003) and Greene (2003), the estimate of the average marginal effect follows a normal distribution with mean $AME = \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{x}_i^{\prime}\boldsymbol{\beta}\right)\beta_k$ and variance, found by applying the Delta method (Oehlert 1992; Wooldridge 2002), given by

$$\sigma^2_{AME} = \frac{\partial}{\partial \boldsymbol{\beta}^{\prime}}\left[\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{x}_i^{\prime}\boldsymbol{\beta}\right)\beta_k\right]\Sigma_{\boldsymbol{\beta}}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{x}_i^{\prime}\boldsymbol{\beta}\right)\beta_k\right] \qquad [54]$$

where $\Sigma_{\boldsymbol{\beta}}$ is the variance-covariance matrix of the parameter vector $\boldsymbol{\beta}$.

The variance of the AME can be consistently estimated by

$$\hat{\sigma}^2_{AME} = \frac{\partial}{\partial \boldsymbol{\beta}'}\left[\frac{1}{n}\sum_{i=1}^{n} g\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}\right)\hat{\beta}_k\right]\hat{\Sigma}_{\boldsymbol{\beta}}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{n}\sum_{i=1}^{n} g\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}\right)\hat{\beta}_k\right] \qquad [55]$$

This estimate of the variance can afterwards be used to conduct hypotheses tests involving the average marginal effects estimate.

The methods presented so far are valid for marginal effects associated to continuous regressors. To evaluate the marginal effect of a dummy covariate, the individual marginal effect is no longer computed as a partial derivative of $G(.)$ in respect to x_k, being instead, computed by comparing the value of $G(.)$ when the covariate x_k has the value of one to the same function when the covariate has the value of zero, that is

$$\begin{aligned} ME_i &= G\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)_{x_k=1} - G\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)_{x_k=0} = \\ &= G\left(\beta_0 + \beta_1 x_{i1} + ... + \beta_k + ... + \beta_p x_{ip}\right) - \\ &\quad G\left(\beta_0 + \beta_1 x_{i1} + ... + \beta_1 x_{ik-1} + \beta_1 x_{ik+1} ... + \beta_p x_{ip}\right) \end{aligned} \qquad [56]$$

Similarly to the case of a continuous covariate, also with dummy regressors, the marginal effect depends on all independent variables in addition to the parameter vector $\boldsymbol{\beta}$, which means that it also varies across individuals. In this case, the average marginal effect is given by

$$AME = \frac{1}{n}\sum_{i=1}^{n}\left[G\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)_{x_k=1} - G\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)_{x_k=0}\right] \qquad [57]$$

Therefore, the framework presented above is also valid when one aims at assessing the effect of a dummy variable on a generic function $G\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)$.

One alternative to complement the analysis of the impact of a covariate on a generic aspect of the conditional distribution of y_i, is to investigated the full distribution of the marginal effects using, for instance, graphical methods, such as histograms or box-plot charts.

When the researcher proposes at analysing only the direction of the impact, disregarding the magnitude of the effect, the analysis of the sign of the parameters are, under some circumstances, sufficient to cast light on that impact. The requirement is that $G\left(\boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)$ to be monotonic in the argument $\mathbf{x}_i^{'}\boldsymbol{\beta}$. As it is was presented in equation [47], the marginal effect of a covariate x_k is given by $ME_i = \frac{\partial G\left(\mathbf{x}_i^{'}\boldsymbol{\beta}\right)}{\partial x_{ik}} = \frac{\partial G\left(\mathbf{x}_i^{'}\boldsymbol{\beta}\right)}{\partial \mathbf{x}_i^{'}\boldsymbol{\beta}}\beta_k$. Therefore, it is immediate that when $G\left(\boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)$is monotonic increasing then the marginal effect equals the sign of the parameter, on the contrary, when $G\left(\boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)$ is monotonic decreasing then the marginal effect has the opposite the sign of the parameter. On the cases where $G\left(\boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)$ does not have that properties, then the analysis of the sign of the parameters are of no value to appraise the impact of the covariate.

Chapter 3
Some remarks about the Portuguese health system

Some level of information and familiarity about the organizational and institutional arrangements of the Portuguese health care system will help in the discussion, debating and on the understanding of the empirical results presented in this thesis, namely in Chapters 5 and 6.

There are number of dimensions from which the institutional setting of a health care system can be described, however, because the substantive goals of the empirical applications developed are closely related to the access to medical care, in this chapter we emphasize the description of the health care system from the delivery perspective. However, we do not rule out other important dimensions like coverage and financing. For those who would like to learn more about the Portuguese health care system, Bentes *et al.* (2004) provide an excellent and quite comprehensive description of it. Furthermore, the reports issued by the Portuguese Observatory of Health Systems also contains, in Portuguese language, a detailed and complete characterization of the Portuguese health care system (Observatório Português dos Sistemas de Saúde 2003; Observatório Português dos Sistemas de Saúde 2004; Observatório Português dos Sistemas de Saúde 2005).

In the Portuguese health care system, a public-private mix accomplishes both the funding and medical care delivery functions. According to the most recent data, presented in the OCDE 2006 database, in 2004, the private financing accounted for 26.8% of total health care spending, with the remaining 73.2% of health expenditures being provided by the public purse (OECD 2006). Back in the year of 1999, the year that one of the datasets analysed has been collected, private financing accounted for 32.9% of total health care spending and public financing 67.1% of total health expenditures.

In terms of health care provision, the private sector, again, plays an important role in terms of service delivery, with a tendency to dominate in areas like the visits to specialized physicians, elective surgery and diagnostic tests while the public sector dominates in the supply of GP visits, non-elective inpatient care and maternity care (Barros 1999; Oliveira 2005).

The level of health insurance coverage in Portugal is universal, as everyone is entitled to coverage, being possible to identify two main types of health insurance schemes provisioned by the two major co-existing systems. First, the coverage provided by the statutory National Health Service (NHS), and second, the health insurance supplied by special public and private insurance schemes whose membership is based on professional or occupational category, henceforth referred to as health subsystems. Dixon and Mossialos (2000) and Bentes *et al* (2004) mention the existence of a third coverage system, the health insurance provided by commercial insurance companies, fulfilling a supplementary role to the NHS rather than providing an alternative to it. In this essay, we will emphasize the two major systems of health care coverage that co-exist in the Portuguese health care system, the National Health Service and the health subsystems.

The Portuguese National Health Service was born in 1979 and was created to guarantee the right of all citizens to health protection, universal free health care through the NHS and access to health care for all citizens regardless of their economic or social condition. The NHS is mainly financed through the national budget, and provides complete coverage for health care. At least in theory, the NHS neither incorporate any relevant cost sharing mechanism nor exclude categories of services from coverage. The Portuguese citizens covered only by the NHS will be referred to as 'NHS-only' individuals throughout this dissertation.

Under the compulsory national health insurance provided by the NHS, the first point of contact to gain access to health care in public facilities is through the General Practitioner (GP) working in health care centres (HC). There is a large network of

around 360 public health care centres, dispersed across the territory, employing about 30,000 professionals, of which 25% are GPs and 20% nurses, both paid on a salaried basis. The care provided by these health care centres includes general medical care for adults, children's care, women's health, family planning, prenatal and perinatal care, as well as first aid. Furthermore, they perform bureaucratic tasks such as certification of incapacity to work and capacity for some jobs. The health care centres also provide home visits and preventive services (Bentes, Dias *et al.* 2004 233). Each GP is responsible for a given number of patients, ranging from 1,000 to 2,000 patients per GP.

To visit a public specialist, generally in the out patient wards in the hospitals, individuals should have a prior visit to a GP to get the appropriate referral to the specialized physician, thus, individuals (supposedly) have no direct access to secondary care as GPs are expected to act as gatekeepers to the secondary care provided by the public system. However, to a large number of people, for instance, those who first seek health care in the health centre and do not see their health care requirements satisfied, or those who disregard the health centres as medical care providers, the emergency departments in hospitals are on the top of alternatives as the first point of contact to enter the system. Therefore, the gate-keeping system operates imperfectly in the Portuguese health care system (Oliveira 2005).

In terms of cost-sharing mechanisms, the 'NHS-only' beneficiaries are required to make a small co-payment to the public provider at the moment of consumption, usually on a fee-for-service basis. The co-payment required, while equal across beneficiaries, usually varies by type of service. However, not all individuals have to contribute with a fee when consuming medical services at the NHS facilities. To provide extra protection to some population groups, the law exempts from the mandatory payment the following cohorts of individuals: pregnant women, children aged 12 or below, pensioners with a pension below the minimum national wage, the

unemployed and their dependents, workers earning a wage not superior to the minimum national wage, individuals with chronic diseases, etc. (Portugal 1992).

These rules are applicable to the 'NHS-only' beneficiaries who seek medical care at wide network of NHS providers, constituted mostly by health care centres and hospitals. As expected, whenever the 'NHS-only' beneficiary decides to demand the medical services of a physician in the private sector, the individual can freely choose to see a physician from a wide range of medical providers, all supplying their services in the 'quasi'-free health care market. When this is the case, the individual must support the full cost of the service, typically on a fee-for-service basis, not being entitled to reimbursement by the NHS. Nevertheless, some share of this out-of-pocket expenditure can be recouped through the tax-system, as the private health care expenditures are tax-deductible.

Therefore, the statutory public health insurance provided by the Portuguese state through an NHS type arrangement provides their citizens with comprehensive benefits. By law, there are no services explicitly excluded and the financial cost-sharing mechanisms have no real expression, because they are very low and, in addition, a large proportion of the population is exempt from its payment. Moreover, there are no explicit constraints in the quantity of services demanded per time-period. However, the real story in terms of the completeness of the Portuguese NHS is not as clear as one may think, as a short analysis will immediately unveil.

The above description corresponds to the statutory paradigm, that is, corresponds to the way that things should have been happening in the Portuguese NHS and not to what actually happens. The reality of the public health care system is quite different, and the comprehensiveness objective turns out to be a hard to find achieved objective. First, although the NHS insurance scheme does not impose the explicit exclusion of any type of medical care from the set of services that beneficiaries would be entitled to utilize, there are actually some categories of services where the public provision is null or very limited, for instance oral and ophthalmology health care (Dixon and Mossialos

2000; Bentes, Dias *et al.* 2004). To illustrate the lack of public supply of these types of services, we did some calculations with data taken from the National Health survey, 1998/99 (the survey will be presented in full detail in chapters 4 and 5). The data revealed that, taking into account all people that visited a dentist in a three months period, roughly 90% of them had the medical appointment in a private provider. Second, while the NHS insurance scheme does not impose explicitly any type of cost sharing nor any limitations on the quantity of services demanded, the true is that important cost-sharing mechanisms do in fact exist. As we aforementioned, at the point of delivery the patient should pay a monetary value, although small. Therefore, the monetary co-payment, due to its smallness, can be ruled out as an effective cost sharing mechanism. However, as it is well-known, in the utilization of medical care services, besides the monetary costs other types of costs can be important, the so-called non-monetary costs. As examples of non-monetary costs we may cite travel costs (both in financial and time terms), waiting time in the medical office, waiting time in a waiting list, etc. In health systems characterized by the provision of free health care at the point of delivery, the monetary cost does not come out as a rationing device, however, other demand rationing mechanisms emerge, for example, time-costs in the form of appointment delays, waiting in the medical office waiting rooms etc. (Acton 1975; Propper 1995). Hence, these class of non-monetary costs can be thought as a form of cost to the individual. It turns out that in the Portuguese National Health Service these non-monetary costs can be quite considerable. For example, Cabral (2002) found that more than 54% of the people that demanded GP services in a public health centre had to wait more than two weeks to get an appointment. Moreover, more than 48% of the demand for secondary care in the hospital, after being referred by the GP, is only satisfied in more than one month. In addition, aiming to evaluate the extent of the time costs involved in seeing a public GP, Lourenço and Ferreira (2005) found that the average travel time to the health centre is approximately 20 minutes and also that the total time spent in the health centre is 2 hours, on average.

The same authors have reported that 17% of the patients spent more than 3 hours in the health centre and that approximately 45% of the patients self-reported being unsatisfied with the time spent waiting. Therefore, visiting public health care provision facilities imposes important time-costs.

This short discussion shows that although in theory the statutory public health insurance is comprehensive, in the sense that it provides a complete health insurance contract, the reality shows a different picture. The individuals actually face several constraints to access the public health services, like e.g difficulties of access due to time costs or geographical barriers, absence of GPs and specialists in rural areas and the lack of certain services. Therefore, those restrictions convert the completeness of the NHS health insurance contract into an incomplete insurance, with implicitly imposed non-monetary cost sharing mechanisms as well as other limitations on utilization of the public health services. The description that we completed corresponds to the functioning of the NHS in the present, which has been evolving since its creation, back in 1979.

However, before the establishment of the NHS, the situation of the population in terms of health care protection was rather different. Before the revolution that instituted the democracy in Portugal, in April 1974, the commitment of the Portuguese State to the provision of health protection to the citizens was very low. In consequence of this low level of health protection, the financing of health care expenditures were generally left to the individual and his or her family. The State only provisioned limited items of medical care such as preventive care, some maternal and child health care having also some interventions in the areas of mental health as well as in the control of infectious diseases (Bentes, Dias *et al.* 2004). Therefore, at that time, the contribution of the state in the provision and financing of health care for the general population was residual, as shown in Table 1, which presents the share of the population eligible to receive health care through public financing for the years 1960 to 1975.

	1960	1965	1970	1971-1975
Coverage (% population)	18%	32%	40%	58%

Table 1 — Population covered by public health insurance arrangements before 1975

Source: OECD, health data, 2006

Despite the rights of the general population to health care protection had been neglected by the state, it did not impede the same state to set off (decree-law nº 45 002, April 27, 1963) a health insurance scheme to protect all civil servants and respective household dependents in terms of health care financing. Furthermore, some other public institutions, financed by the public purse, for example the military, have also created health insurance schemes in order to finance the health care costs of their staff and dependents. Understandably, these initiatives promoted health inequalities in access and utilization of health care services. Some private companies, for example in the banking sector, the electrical energy sector, also set off their own health insurance schemes, hence putting their employees and respective families under the protection of a health insurance.

As a result, prior to the creation of the NHS, the situation of the Portuguese citizens in terms of health protection can be summarized as follows. An important share of the population was not covered by any health insurance arrangement, while the other share was covered by health insurance schemes for which membership was based on professional category, provided by private or public institutions to whom the insured worked. These alternative forms of health care protection are known as health subsystems, so this term will be used throughout this dissertation.

It would have been predictable that after the creation and subsequent evolvement of an integrated public health system (the NHS), the health subsystems beneficiaries would have been integrated in the new public health care system, at least the subsystems financed by the state, like, for example, the health subsystem for the public

servants. Nevertheless, contrarily to the expected, all these alternative health insurance schemes were kept, have developed considerably in a way that nowadays they are important players in the financing and provision of health care to a non negligible number of Portuguese citizens. Bentes et al. (2004) point out that these alternative health insurance schemes were kept because trade unions were not willing to quit the easy access they enjoyed to a wide range of quality medical providers. An extensive discussion about the reasons for the non-integration of the health subsystems into the NHS is beyond the scope of this dissertation, however Gouveia (1999) and Batista (1999) discusses the subject in some detail.

Summing up, the health subsystems are organizations linked to a public or private institution, whose aim is to fund and/or provide medical care to their employees and respective dependents with compulsory membership.

Some figures help to illustrate the size of the health subsystems, showing at the same time their importance and relevance within the Portuguese health care system. The first and most obvious indicator is the extent of their coverage. Unfortunately, precise figures regarding the number of individuals covered by the health subsystems are often missing, with the estimates varying in function of the source of information. Some OECD publications as well as other publications issued by the Portuguese and the European Observatory of Health Care Systems have estimated the health subsystems coverage to be roughly 25% of the Portuguese population (Bentes, Dias *et al.* 2004; Observatório Português dos Sistemas de Saúde 2004; OECD 2004). However, we believe that this last figure overestimates the coverage by health subsystems. Using alternative sources of information, we systematically obtain lower estimates.

On the one hand, using data from both versions of the National Health Survey, 1995/96 and 1998/99, which is a nationwide and representative health related survey, we estimate that only 16% of the Portuguese population is covered by a health subsystem. However, this estimate may be too low because some people might omit their membership. To illustrate the reasons that may drive this omission, suppose an

individual covered by a health subsystem that always chooses to get medical care in public providers[6]. Therefore, such an individual when asked about his coverage system may respond NHS instead of the correct answer. The reason is that from his point of view, he is always using the NHS services and not any other health subsystem, despite the third party payer being the health subsystem. And on the other, summing the total number of enrolees of each health subsystem we reach a figure close to the 2.3 million people (Batista 1999), about 23% of the Portuguese population. Once more this figure can be a biased estimate for the coverage by a health subsystem, and unlike the estimation obtained through the National Health Survey dataset, this one could be overestimated as double-counting may arise due to people belonging to more than one health subsystem. Therefore, combining the information from both sources, the National Health Survey and from Batista's (1999) calculations, it is possible to conclude that the extent of coverage of the health subsystems ranges from 16 to 22 percent, and not the 25% usually presented in studies about the Portuguese health care system.

An additional indicator that can be used to contextualize the extent of the activity of the subsystems as health insurance providers are the expenditures in financing the health care needs of their enrolees. The National Statistical Institute, in its Satellite Account of Health, provides the most recent figures, reporting the total health expenditures of the health subsystems for the year 2004. The total expenditure, at current prices, on health services of the public health subsystems was 941,907 million euros while the total expenditure of the private health subsystems was 303,011 million, reaching an overall expenditure of 1,244,918 million euros.

---

[6] As it will be described in detail later, the beneficiaries of subsystems can receive medical care in public providers.

These figures, showing the health subsystems extent of coverage and health expenditures, are sufficiently illustrative of the importance and potential impact of these health care financing organizations within the Portuguese health care system.

Within the group of all health subsystems, it is interesting to distinguish between two types of funds (being the distinction essentially useful in Chapter 5). From all the health subsystems, 16 according to Batista (1999), one of them differs from the remaining, mainly because of its extent in terms of the number of beneficiaries. The largest health subsystem, known as ADSE, is controlled by the Ministry of Finance, and provides health insurance coverage for all civil servants, NHS workers included, and relatives that depend on them. According to data supplied directly by the ADSE fund, upon request, they covered, by the end of the year 2000 more than one million and three thousand individuals (1.375.701). From this total number of beneficiaries, roughly 60% (823.610) obtain coverage in their own name, while the remaining 40% (552.091) enjoy the ADSE coverage through a family member. Therefore, the ADSE fund covers about 13.1% of the Portuguese population and approximately 60% of the total number of individuals covered by a health subsystem contract.

The main dissimilarity between the ADSE fund and the remaining funds is, among other aspects, its scale of operation, driven essentially by the high number of individuals covered by the ADSE (civil servants fund) when compared to the individuals covered by the other health subsystems. Another distinction that can be made regards to the characteristics of the beneficiaries and the way of assuring the provision of medical care. Some health subsystems, other than the ADSE, cover populations with special characteristics, for example the military personnel, whereas other subsystems funds have their own provision capacity aimed at delivering health care directly to the beneficiaries. Moreover, the health subsystems organizations encompass a wide array of arrangements similar to those in managed care, where managed care is applied in its broader sense that encompasses a diverse range of institutional arrangements, from Preferred Provider Organization (PPO) to Health

Maintenance Organization (HMO), among other types of organizational and hybrid strictures. For instance, the civil servants fund, the ADSE, can be framed into a structure that integrates Fee-For-Service (FFS) components as well as PPO arrangements. On the other hand, the health subsystems for the banking sector and for the military, besides containing FFS and PPO structures, can also be thought as HMOs because they have the functions of insurance and service delivery integrated.

All health subsystems schemes share a feature that will be very important to the empirical application whose results are presented, and analysed in Chapter 5. The membership to these alternative health insurance groups is mandatory and based on professional or occupational category, being, consequently totally independent of the risk of falling sick. Thus, though the likely differences in the distribution of health risks across health subsystems beneficiaries, they are not selected according to that risk of falling sick. We will return to this issue later in Chapter 5, nevertheless, we refer the reader to Barros et al. (2005) who present a more detailed discussion regarding the exogeneity of these health insurance schemes.

In what concerns the financing sources, a mixture of employer and employee contributions usually finances the health subsystems. To most schemes, the contribution is compulsory and each member contributes with one percent of their remuneration, regardless of the health status and the number of members of the household that are covered by the health insurance contract (Alberto 1999).

In what follows, we attempt characterize the insurance contrast supplied by the health subsystems. This is a challenging task, however, because the different health subsystems provide a diverse array of medical care insurance arrangements to finance and/or provide health care. Regarding the insurance schemes provided by the health subsystems, we put the following questions:

- Are these health insurance contracts complete?
- Do they have attached any cost-sharing mechanisms or any other demand limiting strategies?

- How do they provide medical care to the beneficiaries?

To begin with, the subsystems provide medical care either directly in own providers or by contract with private or public providers, and in some cases, by a combination of both forms. Therefore, the health subsystems enrolees are entitled to get medical care in the following types of providers:

- NHS providers,
- private providers,
- facilities run by the health subsystem organization.

When the subsystems beneficiaries demand care in NHS providers, they have equal rights and obligations relative to the NHS counterparts (Direcção Geral da Saúde 2005). Therefore, when the health subsystems members choose to be seen by a doctor working to the NHS, they are subject to the same cost sharing mechanisms, explicit or implicit, and service exclusions, as the NHS counterparts. Naturally, they are also exposed to the same access problems faced by the 'NHS-only' beneficiaries. This observation implies that from the individual's point of view, in the access to the NHS services, the NHS insurance contracts are similar in all respects to the insurance contract provided by all health subsystems. Note however that, the law requires that subsystems pay the NHS for the medical care services utilized by their enrolees at the rates instituted by the Ministry of Health.

Additionally, the health subsystems insurance design is such that it allows the holder of the contract, and household members, to demand medical care in all private providers. When this is the case, the type of the relationship between the three agents involved in the transaction (the patient, the doctor and the health subsystem) depends on whether or not a contract exists between the private provider and the health subsystem. Most of the health subsystems incentive their beneficiaries to choose a provider from a list of providers which they have contracted with, generally with a lower price relative to the market price. This incentive is in the financial form by setting the patient's coinsurance or co-payment rates lower for in-network-providers

relative to out of list providers. The beneficiary, at the point of delivery, has to support a small payment per service, which can be seen as a co-payment in the form of a deductible, and afterwards the health subsystem pays directly to the provider the remaining cost of the medical act.

On the contrary, when health care demand is done in private providers without contract with the subsystem, that is, the demand is made in private providers out of a catalogue, the patient is responsible for the payment of the full cost of the service delivered and, afterwards, the patient puts a claim to the subsystem and is partially reimbursed. The extent of the reimbursement varies across the different health subsystems. For instance, in doctor visits, the ADSE fund reimburses 80% of the cost of the visit, in a maximum of approximately 20 €. Other subsystems reimburse only 75% of the cost of the visits, however with diverse upper limits. Moreover, in this category of health care providing, the co-insurance rates to which the beneficiary is subject depend on the type of service provided as well as on the type of provider. A wide variation across health subsystems is found. As abovementioned, the array of arrangements is quite diverse being difficult to describe them exhaustively.

Finally, some health subsystems schemes, for instance the health subsystems for the banking sector and for the armed forces, have the capacity to provide medical care services directly in their own facilities and with their own doctors. Again, the use of this type of services might be, at least for some funds, subject to an array of demand limiting policies that aims at containing costs and utilization with questionable marginal utility.

Regarding the comprehensiveness of the health insurance plans provided by the two-coexisting health protection systems, the NHS and the health subsystems, we might ask, who is better off in terms of coverage? Are the NHS-only beneficiaries or, on the contrary, the health subsystems beneficiaries are those who enjoy from a more generous system to protect them from the financial risks of becoming ill? The answer to this question is straightforward, as we will show in the next few paragraphs.

To begin with, suppose that the health subsystems beneficiaries demanded all medical care exclusively in public providers, ignoring the private providers. Therefore, in this virtual scenario, from a point of view of access and opportunities for consumption, 'NHS-only' beneficiaries and subsystems beneficiaries, would be in similar situations, that is, they would be exposed to the same explicit prices (consumer co-payments) as well as to the same implicit prices (time costs, exclusion of some services, dental care for instance).

However, as shown above, the access to NHS providers may be subjected to important time-related costs, thus private providers emerge as viable alternatives to obtain health care without facing the high time costs imposed by the NHS. Nevertheless, when visiting a private provider the individual faces a financial cost, which might dissuade most people from demanding health care in private providers, unless the individual enjoys from a health insurance contract that gives him privileged access to private medical amenities. As we saw earlier, the health subsystems beneficiaries enjoy such privileged access. This privileged access gives them the advantage of the health insurance provided by the health subsystems relative to the NHS protection. In conclusion, the health subsystems beneficiaries are covered by a health insurance contract that is more generous than the insurance provided under the shield of NHS. In conclusion, these health subsystems schemes offer more comprehensive health protection plans, offering a greater range of providers accessible at lower costs than is available under the NHS plan.

Regarding the health insurance status, in this dissertation, we identify three types of access groups. The first group includes individuals covered only by the health insurance scheme provided by the National Health Service, henceforth referred to as 'NHS-only' individuals. The second group includes those people covered by the ADSE fund, which encompasses all public servants and respective household counterparts, from now on referred to as ADSE beneficiaries. The last health insurance class consists of individuals who benefit from a health insurance contract provided by a subsystem

other than the ADSE, which will be referred to as OSS (Other Subsystems) beneficiaries.

In general, in countries whose health systems are characterized by the existence of a National Health Service, the alternative health insurance arrangements can be classified in either complementary or supplementary. According to Mossialos and Thomson (2002) the alternative health insurance arrangements are complementary when they provide full or partial coverage for services that are excluded or not fully covered by the statutory health care system. On the other hand, they are classified as supplementary when it aims at covering services that are also funded by the public service, increasing consumer choice and access to different health services. Supplementary health insurance is sometimes referred to as 'duplicate coverage' (Vera-Hernandez 1999; Mossialos and Thomson 2002) or 'additional health insurance' (Barros, Machado *et al.* 2005).

Having in mind this taxonomy for the alternative health insurance schemes that co-exist with the NHS, how to classify the health subsystems? In terms of general health care services, the health subsystems have been acting both as complementary and supplementary types of health insurance (Oliveira 2001). The subsystems are complementary because they guarantee health care coverage for services almost not provided by the NHS, most notably, dental care and ophthalmology. They are also, and mostly, supplementary because cover they provide coverage for the utilization of private care, by reimbursing the patient's expenses, for services that are provisioned in the public sector. In addition, as it was mentioned before, some health subsystems provided medical care services on their own facilities.

In this thesis, we are particularly interested in one specific type of medical care, doctor visits. We defend that for this type of medical care service the supplementary role of the health subsystems is much more important than the complementary one. Notice that from the full range of doctor visits types that one can consider, the health

subsystems are complementary only for dental and ophthalmologic care, being supplementary for all other types of visits

Chapter 4
Econometric analysis of health care utilization: An alternative hurdle specification using LCMs♣

4.1 — Introduction

Over the last thirty years two major classes of econometric specifications have dominated the empirical literature on medical care utilization analysis[7]: one-part and two-part models.

On the one hand, one-part models are defined through a linear or non-linear reduced form equation, where health care utilization is explained as a function of a set of medical care determinants[8]. On the other, two-part models analyses medical care utilization by estimating two equations; the first identifies the factors that distinguish between users and non-users and the second unveils the determinants of the intensity of medical care utilization, conditional on a positive utilization[9].

The debate over the merits of each approach to explain health care utilization has been intense and interesting. The choice of one specification over another has usually been an empirical question. We contribute to this discussion by presenting a hurdle

---

♣ This chapter is an extended version of a paper submitted to the Health Economics Journal, being currently under revision.

7 The econometric methodology has had a tendency to dominate the filed of health care demand, yet, earlier analysis have resorted on statistical methods other than regression models, e.g. analysis of variance and analyses of covariance. A review of early empirical methods for medical care utilization study can be found on Duan *et al.* (1983).

8 In the count data framework, the Poisson, the NB and the LCM regression models are specifications that can be framed on the family of one-part models.

9 In count data contexts, the hurdle model frames into this category of the two-part models to study the demand for medical care services.

formulation that departs from the standard specifications in the way individuals with positive utilization are modelled. We see it as closer to the underlying decision process. We propose a specification close to Bago d'Uva's (2005; 2006) model, based on latent class models, explicitly assuming that there are unobservable characteristics specific to the population of health care users. This assumption results from the underlying decision process by economic agents, and guides the definition of the econometric model. We found that the specification outperforms, from a statistical point of view, a number of competing count data models that have been used in the literature to study health care utilization.

The application uses data taken from the Portuguese National Health Survey (Ministério da Saúde - Instituto Nacional de Saúde 1999) in order to estimate and test statistical hypothesis on a wide variety of regression models specified in the spirit of one-part model and two-part models. We measure health care utilization as the total number of doctor visits in a period of three months.

The contribution of this application is two-fold: on the one hand, we suggest an alternative hurdle formulation, and compare its statistical performance with the most common count data models used to model count data using likelihood ratio tests (LR) tests, information criteria statistics and Vuong tests. On the other, we assess the robustness of the effect of various covariates, e.g. income, health insurance status, place of residence (rural/urban) and education, on the conditional mean of various competing count data models under evaluation. The results of these comparisons are relevant to evaluate whether the conclusions taken from models specified under different assumptions about the unobserved heterogeneity are stable, or, on the contrary, change across models.

This chapter is organized as follows: in section 2 we present the relevant literature regarding the comparison of hurdle and LCM's, followed, in section 3, by explaining the alternative hurdle formulation. In section 4 we present the dataset used in this application, exploring some details about the dependent variable as well as about the

covariates. Next, in section 5, we present and discuss the econometric results. Finally, in section 6, closing the chapter, presents some general conclusions and some ideas about our future research agenda in the area of econometric modelling of health care utilization data.

4.2 — The relevant literature

Some authors consider two-part models more appropriate to explain medical care utilization because, as they argue, they can be viewed as the empirical counterpart of the principal agent set-up, which represents well the actual decision process concerning medical care utilization (Pohlmeier and Ulrich 1995; Jimenez-Martin, Labeaga *et al.* 2002). Moreover, relative to simpler single index models, e.g. Poisson or NB regression models, they can be more enlightening because they allow the separation and the quantification of the medical care determinants concerning the decision to visit a physician and the frequency of visits, which can be significant for health policy making (Pohlmeier and Ulrich 1995).

From an empirical point of view, until 1997, with a "theoretical" support and good statistical properties, the hurdle framework was considered as the appropriate model to explain medical care data. In fact, when compared to simple one-part models, like Poisson or NB regression models, the hurdle framework generally arose as the specification that better fits the data (Manning, Newhouse *et al.* 1987; Grootendorst 1995; Deb and Trivedi 1997; Gerdtham 1997; Deb and Trivedi 2002; Van Ourti 2004).

However, Deb and Trivedi (1997; 2002) after analysing some of the hypothesis underlying the hurdle model, made some criticisms against it[10], proposing instead a latent class model as better alternative to study medical care utilization. These authors defended that LCMs present statistical properties that make it more appealing to study

10 See Section 2.6.1 (page 42) for details.

utilization data. In effect, to support their view, they used a number of health care utilization measures and estimated hurdle and LCMs, finding that the LCM outperforms the hurdle model. Other authors studied the same issue, finding that latent class models fit the data better than the competing hurdles estimated (Deb and Holmes 2000; Gerdtham and Trivedi 2001; Sarma and Simpson 2006). Therefore, the empirical evidence seems to agree that the LCM is a better framework to analyse health care utilization data.

Despite the dominance of the LCM over the hurdle, there are studies, although only a few, pointing towards the statistical gains of the hurdle over the LCM specification. For instance, Jiménez-Martín et al. (2002), using data from the European Community Household Panel, estimated hurdle and LCMs to evaluate the determinants of individual medical care utilization. In their application, they used several indicators of utilization, namely, the number of visits to a GP and the number of visits to a specialized doctor, both indicators measured during a 12-month period. Regarding the results they found mixed evidence. The LCM specification were found to be more suitable than hurdle to analyse the number of visits to a GP, the opposite was reported for visits to specialist physicians. Another example of the statistical superiority of the hurdle specification over the LCM is given by Winkelmann (2004). The author analysed the total number of visits to a doctor during a three-month period and found that the hurdle outperforms the LCM. To finish, Bago d'Uva (2006) using RAND data, analysed the number of doctor visits in one year period and found that a model of the hurdle family, adapted to panel data, proved to be statistically superior relatively to LCM specification, also specified to a panel data context[11]. Therefore, under some

11 In fact, the model by Bago d'Uva uses insights from the two families of models, hurdle and LCM, to purpose a new specification.

conditions, the hurdle framework may arise as appropriate to explain the health care data.

It is important to note that all evidence that have reported the superiority of LCM over the hurdle, compared the LCM specification defined by Equation [9] with one very specific formulation of the hurdle, the one that relies on a binary model for the first stage, and for the second on a truncated-at-zero Poisson or NB model. Notice that Santos-Silva (2003) argues that in the popular hurdle model the unobserved heterogeneity is incorrectly modelled, causing the NB based hurdle model to be misspecified.

4.3 — An alternative hurdle formulation using LCMs

The motivation to evolve from the Poisson to the NB and, from the latter to the LCM regression model was statistical and not economic reasoning. The goal was to formulate a count data model with good fit to the data. The Poisson, NB and the LCM are in the spirit of one-part models, sharing the underlying assumption that the data generating processes[12] governing both the decision to seek health care and the frequency of health care utilization, conditional on some utilization, are the same. In these models, as is well known, the impact of individual and doctor inputs to the decision making process regarding the number of visits, are entangled, therefore, non-separable. Even in the LCM, assuming that P stochastic processes govern health care choices, individual and doctor contributions to the choice are mixed in each stochastic process. In this context, it is difficult to know which one, doctor or patient, contributes the most to the decisions regarding the number of visits.

---

[12] We use the plural "data generation mechanisms" because the LCM framework assumes that two or more stochastic processes are involved in the overall specification.

As we pointed out in Chapter 2, it has been recognized that, in the context of health care, the decision involves two stages, with different key decision makers in each step (Zweifel 1981; Pohlmeier and Ulrich 1995). In the first stage, the decision to seek medical care is assumed to be mostly patient-based; in the second, decisions are mainly doctor based, possibly including patient preferences. Accordingly, it has been argued that the specification of empirical models should recognize this two stage decisional structure.

We see the first-stage decision, to initiate the contact with the physician, to be essentially patient-driven. Based on his/her characteristics and private information, the patient decides to go to the doctor. Here, the unobserved heterogeneity results from draws from the same distribution for all members of the population. Moreover, patients decide based on the coarse signal about their health condition and on the expected value of going to the doctor. After the contact with the physician, the medical expert obtains a finer signal about the true condition of the patient. In this decision, the crucial unobserved heterogeneity will be related to the doctor. Based on the information collected, the physician decides the course of treatment, and patients then become either "healthy" patients in the sense they require few or none treatment resources, or "sick" patients, requiring more medical care (namely visits). This leads naturally to a latent class approach based on the second stage decision-process characteristics.

In Section 2.8 — Specification of latent class models for truncated samples —, we pointed out that in the specification of models in truncated-at-zero datasets the unobserved heterogeneity can be specified in the actual population or in the truncated population (see equations [39] and [40]). Therefore, one reasonable question to ask is which mixture specification should be used to model the second stage of the hurdle.

The response to the question depends critically on the population that the researcher is interested in. In the case of hurdle models, it is natural to estimate the

model for the positive counts by making assumptions concerning the distribution of the unobserved heterogeneity in the truncated population (Santos-Silva 2003).

Thus, if one plans to incorporate the unobserved heterogeneity through the utilization of the latent class framework, the LCM for the positives should be specified according to density defined by Equation [40], that is,

$$f_s\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \sum_{j=1}^{P} f_j^s\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)\hat{\pi}_j = \\ = \sum_{j=1}^{P}\left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)}{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)}\right)\hat{\pi}_j \qquad [58]$$

However, the popular hurdle formulation, the model that uses the negative binomial distribution as parent distribution as a means to account for the unobserved heterogeneity in the second part of the model, specifies this stage according to the continuous counterpart of [39], which is defined by the density that appears next,

$$f_s\left(y_i \mid \mathbf{x}_i\right) = \frac{\int_{v_i} f\left(y_i \mid \mathbf{x}_i, v_i\right) h\left(v_i \mid \mathbf{x}_i\right) dv_i}{1 - \int_{v_i} F\left(0 \mid \mathbf{x}_i, v_i\right) h\left(v_i \mid \mathbf{x}_i\right) dv_i} \qquad [59]$$

Thus, under specification [59] the researcher is making assumptions regarding the unobserved heterogeneity in the actual population. Consequently, according to Santos-Silva (2003) the investigator is making inference to the actual population and not to the truncated population as it should be in the case of hurdle models.

From this discussion, we can conclude that in hurdle contexts the inclusion of the unobserved heterogeneity in the model for the positives should follow clear hypothesis specification. For the health care users, the data generating mechanism should use density [40] as the preferred specification, implying that those models that assume specification [39] are misspecifed, being therefore inappropriate. Consequently, one can argue that the popular hurdle specifications based on the NB distribution, when the use of this density is justified as a means to account for unobserved heterogeneity in the second part of the model, is misspecified.

Mullahy (1986), Pohlmeier (1995), Gerdtham (1997), Jimenez-Martin (2002) and Grootendorst (2002) are some examples of applications of hurdle models specified in such, supposedly, inadequate way. Even the semi parametric hurdle model proposed by Gurmu (1997) uses the density [59] to model the positives.

Van Ourti (2004) and Winkelmann (2004) moved towards the appropriate specification proposing to model the second stage of the hurdle according to the continuous counterpart of the density given by Equation [40], that is, by

$$f_s\left(y_i \mid \mathbf{x}_i\right)=\int_{v_i} \frac{f\left(y_i \mid \mathbf{x}_i, v_i\right)}{1-F\left(0 \mid \mathbf{x}_i, v_i\right)} h\left(v_i \mid \mathbf{x}_i\right) dv_i \qquad [60]$$

Thus, the specifications proposed in both papers correctly point towards the truncated population, as it should be in the case of hurdle models.

On the one hand, Winkelmann proposed a probit model to identify the factors that distinguish between health care users and non-users, on the other, Van Ourti suggested a logit model. The model for the second stage, both authors used the truncated-at-zero Poisson as starting point. Next, to include the (assumed) continuous unobservables in the model, they assumed them as a draw form the normal distribution, and present in the truncated population, as it should be in the case of hurdle models.

For example, Winkelmann assumes that, conditional on the unobservables, let υ_i represent them, and on the regressors, $\mathbf{x}_i$, the probability model for the positives is defined as a truncated Poisson, that is,

$$f_s\left(y_i \mid \mathbf{x}_i, \upsilon_i\right)=\frac{\mathrm{e}^{-\lambda_i} \lambda_i^{y_i}}{y_i!\left(1-e^{-\lambda_i}\right)} \qquad [61]$$

where the conditional mean is specified as $\lambda_i=\mathrm{e}^{\mathbf{x}_i'\beta+\upsilon_i}$, thus including a random variable υ_i that is the individual unobserved heterogeneity. Furthermore, the unobservables, υ_i, are assumed to be normally distributed in the truncated population with mean one

and variance σ^2. Therefore, the marginal probability of y_i, conditional on $y_i > 0$ and $\mathbf{x}_i$, is given by

$$f_s\left(y_i \mid \boldsymbol{x}_i, \beta\right) = \int_{\upsilon_i} \left[\left(\frac{e^{-e^{\mathbf{x}_i'\beta+\upsilon_i}} \left(e^{\mathbf{x}_i'\beta+\upsilon_i}\right)^{y_i}}{y_i!\left(1-e^{-e^{\mathbf{x}_i'\beta+\upsilon_i}}\right)} \right) h_s\left(\upsilon_i \mid \mathbf{x}_i\right) \right] d\upsilon_i \qquad [62]$$

where $h_s\left(\upsilon_i \mid \mathbf{x}_i\right)$ is the normal density function. Assuming independence between the unobservables in both parts of the model, as in the classic hurdle, it is possible to separate the estimation into two parts. However, the integral that appears in the likelihood function for the second part (see expression [62]) does not have an analytical solution therefore Winkelmann evaluates the likelihood using Gauss-Hermite integration (Winkelmann 2004).

It would have been interesting to compare the statistical performance of these alternative hurdles with the popular hurdle and with the LCM. Contrarily to Van Ourti, Winkelmann reported some findings concluding that the new hurdle offers a substantial improvement over all other models[13]. We consider that these findings are evidence in favour of Santos-Silva's thesis that, in the case of hurdle models, the hypothesis regarding the distribution of the unobservables should be made in the truncated population and not in the actual population.

Note, however, that even Winkelmann's and Van Ourti's specifications are not immune to criticism. Both proposals are fully parametric, hence somewhat arbitrary. The point is that, due to the lack of an economic theory to enlighten them about the true form of the unobserved heterogeneity distribution, they may fail the true

---

[13] The competing models estimated by Winkelmann were the Poisson, the Negative Binomial, the Poisson log-normal, the Hurdle negative binomial, the Probit-Poisson-log-normal, the two components finite mixture negative Binomial and a multi-episode Poisson logarithmic regression model.

(unknown) distribution for the data. In that event, inconsistent parameters will result from the estimation process because pseudo-maximum likelihood methods do not apply (Wedel, Desarbo *et al.* 1993). Hence, in this chapter we defend that a specification less dependent on strong distributional assumptions are fully desirable. As presented in a previous chapter, an alternative to account for the presence of unobserved heterogeneity, making, at the same time, less demanding distributional assumptions about the unmeasured heterogeneity is the use of latent class models. Therefore, we advocate that latent class models are a better framework to specify the hurdle model's second stage. Note that a few pages above we have defended the view that the LCM approach is appealing from a decisional point of view, representing well the actual decision structures involved in the second stage of the hurdle. Therefore, the LCM to explain the utilization of the patients with a positive utilization has a two-fold motivation, it provides a good approximation on the patient behaviour, presenting also statistical advantages.

Given the rationale presented above it is reasonable to estimate the model for the positive counts according to Equation [40]. For a formulation of the full hurdle model, let y_i denote the dependent variable and $\mathbf{x}_i^{'} = \left[x_{i1}, x_{i2}, \cdots, x_{ik} \right]$ denote a $(1xk)$ vector of covariates. Furthermore, assume that both $f_0(.)$ and $f_j(.),\ j = 1,2,...,P$ are discrete probability functions adequate to model count data. In addition, suppose that $f_0(.)$ governs the hurdle part and a finite mixture of truncated-at-zero count distributions governs the process after crossing the hurdle. Then, in our alternative model, the probability of being a non-user is given by

$$P\left(y_i = 0 \mid \mathbf{x}_i, \beta_0 \right) = f_0\left(0 \mid \mathbf{x}_i, \beta_0 \right) \qquad [63]$$

while the probability of being a user is defined by

$$P\left(y_i > 0 \mid \mathbf{x}_i, \beta_0 \right) = 1 - f_0\left(0 \mid \mathbf{x}_i, \beta_0 \right) \qquad [64]$$

Notice that the last two equations form the first part of the model being similar to the first part of hurdle shown in Section 2.6. Contrarily to the specification presented

in that section, the second stage equation in this new context is defined by a LCM of P truncated distributions, and is given by,

$$f_s\left(y_i \mid \mathbf{x}_i, \beta\right) = \sum_{j=1}^{P}\left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \beta_j, \upsilon_i^j\right)}{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \beta_j, \upsilon_i^j\right)}\right)\hat{\pi}_j \qquad [65]$$

Therefore, the conditional probability function of the count, y_i, is given by,

$$f\left(y_i \mid \mathbf{x}_i; \boldsymbol{\beta}_0, \boldsymbol{\beta}_j\right) =$$
$$= \begin{cases} f_0\left(0 \mid \mathbf{x}_i; \boldsymbol{\beta}_0\right) & \text{if} \quad y_i = 0 \\ \left[1 - f_0\left(0 \mid \mathbf{x}_i; \boldsymbol{\beta}_0\right)\right] * \left[\sum_{j=1}^{P}\left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)}{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)}\right)\hat{\pi}_j\right] & \text{if} \quad y_i = 1, 2, \ldots, \end{cases} \qquad [66]$$

In the first step a binary model is estimated, while in the second step, constraining our sample to the individuals with a positive number of visits, we estimate a model that assumes that the density in the truncated sample is given by a LCM applied to truncated data. Henceforth, our alternative hurdle will be referred to as LCH-Cross (latent class hurdle for cross-section data).

Regarding the descriptors of this model, its conditional mean function for the overall population is given by

$$E\left(y_i \mid \mathbf{x}_i, \beta_0, \boldsymbol{\beta}_j\right) = \left[1 - f_0\left(0 \mid \mathbf{x}_i, \boldsymbol{\beta}_0\right)\right] * \sum_{j=1}^{P}\left[\hat{\pi}_j * E_j\left(y_i \mid \mathbf{x}_i, \beta_j, y_i > 0\right)\right] \qquad [67]$$

The summation that appears in the previous equation, is the conditional mean function of y_i in the users population, that is,

$$E\left(y_i \mid \mathbf{x}_i, \beta_j, y_i > 0\right) = \sum_{j=1}^{P}\left[\hat{\pi}_j * E_j\left(y_i \mid \mathbf{x}_i, \beta_j, y_i > 0\right)\right] \qquad [68]$$

It is easy to conclude that, in Equation [68], each $E_j\left(y_i \mid \mathbf{x}_i, \beta_j, y_i > 0\right),\ j = 1, 2, \ldots, P$ represents the conditional mean function for the latent class j in the users population, which is given by,

$$E_j\left(y_i \mid \mathbf{x}_i, \beta_j, y_i > 0\right) = \frac{\exp\left(\mathbf{x}_i' \beta_j\right)}{P_j\left(y_i > 0 \mid \mathbf{x}_i, \beta_j\right)} = \frac{\exp\left(\mathbf{x}_i' \beta_j\right)}{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \beta_j\right)} \quad [69]$$

In what follows, we give an example of a practical specification of the model. For a model based on the NB, whose density function was presented in

[6], the two equations presented above define the first part of the alternative hurdle model,

$$P\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_0\right) = f_0\left(0 \mid \mathbf{x}_i, \boldsymbol{\beta}_0\right) = \left(\frac{\eta_{i0}}{\lambda_{i0} + \eta_{i0}}\right)^{\eta_{i0}} \quad [70]$$

$$P\left(y_i > 0 \mid \mathbf{x}_i, \beta_0\right) = 1 - \left(\frac{\eta_{i0}}{\lambda_{i0} + \eta_{i0}}\right)^{\eta_{i0}} \quad [71]$$

On the other hand, the specification to explain the health care utilization, conditional on utilization being positive, that is, the model for the second stage, is given by,

$$f_2\left(y_i \mid \mathbf{x}_i, \beta_1, ..., \beta_j, y_i > 0\right) =$$

$$= \sum_{j=1}^{P} \left(\frac{\Gamma\left(y_i + \eta_{ij}\right) \eta_{ij}^{\eta_{ij}} \left(\lambda_{ij} + \eta_{ij}\right)^{-\left(\eta_{ij} + y_i\right)} \lambda_{ij}^{y_i}}{y_i! \Gamma\left(\eta_{ij}\right) \left(1 - \left(\frac{\eta_{ij}}{\lambda_{ij} + \eta_{ij}}\right)^{\eta_{ij}}\right)} \hat{\pi}_j \right) \quad y_i = 1, 2... \quad [72]$$

where $\lambda_{ij} = \exp\left(\mathbf{x}_i' \boldsymbol{\beta}_j\right)$ $j = 0, 1, ..., P$ and, $\eta_{ij} = \left(\frac{1}{\alpha_j}\right) \lambda_{ij}^k$ $j = 0, 1, ..., P$.

Therefore, combining equations [70] to [72], the conditional probability function of y_i can be written as

$$f_2\left(y_i \mid \mathbf{x}_i, \beta_0, ..., \beta_j\right) =$$

$$= \begin{cases} \left(\dfrac{\eta_{i0}}{\lambda_{i0}+\eta_{i0}}\right)^{\eta_{i0}} & y_i = 0 \\ \left[1-\left(\dfrac{\eta_{i0}}{\lambda_{i0}+\eta_{i0}}\right)^{\eta_{i0}}\right] * \sum_{j=1}^{P} \left(\dfrac{\Gamma\left(y_i+\eta_{ij}\right)\eta_{ij}^{\eta_{ij}}\left(\lambda_{ij}+\eta_{ij}\right)^{-\left(\eta_{ij}+y_i\right)}\lambda_{ij}^{y_i}}{y_i!\Gamma\left(\eta_{ij}\right)\left(1-\left(\dfrac{\eta_{ij}}{\lambda_{ij}+\eta_{ij}}\right)^{\eta_{ij}}\right)} \hat{\pi}_j \right) & y_i = 1,2... \end{cases} \qquad [73]$$

Like in the popular hurdle specification, since the first stage process only contains binary information about the dependent variable, the parameters of the first part of the model — the parameters $(\boldsymbol{\beta}_0, \alpha_0)$— are not separately identifiable, thus, to identify the model we set $\alpha_0 = 1$.

For a regression model whose conditional probability function of the dependent variable is defined by [73], the conditional mean for the overall population is given by

$$E\left(y_i \mid \mathbf{x}_i, \beta_0, \boldsymbol{\beta}_j\right) = \left[1-\left(\frac{\lambda_{i0}^{k}}{\lambda_{i0}+\lambda_{i0}^{k}}\right)^{\lambda_{i0}^{k}}\right] * \sum_{j=1}^{P}\left[\hat{\pi}_j * \frac{\exp\left(\mathbf{x}_i^{'}\beta_j\right)}{1-\left(\dfrac{\eta_{ij}}{\lambda_{ij}+\eta_{ij}}\right)^{\eta_{ij}}}\right] \qquad [74]$$

Having as reference the users population, then, the conditional mean function is given by

$$E\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p, y_i > 0\right) = \sum_{j=1}^{P}\left[\hat{\pi}_j * \frac{\exp\left(\mathbf{x}_i^{'}\beta_j\right)}{1-\left(\dfrac{\eta_{ij}}{\lambda_{ij}+\eta_{ij}}\right)^{\eta_{ij}}}\right] \qquad [75]$$

And the conditional mean function for the latent class j in the users population, which is given by,

$$E_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j, y_i > 0\right) = \frac{\exp\left(\mathbf{x}_i' \beta_j\right)}{1 - \left(\frac{\eta_{ij}}{\lambda_{ij} + \eta_{ij}}\right)^{\eta_{ij}}} \quad j = 1, 2, ..., P \qquad [76]$$

Bago d'Uva (2006) suggested that the models to analyse medical care utilization are not constrained to be only hurdle or only LCM, being possible to combine the features of both formulations. This author, in fact, puts into practice a general model, that contains the hurdle we propose, combining the hurdle and the LCM frameworks. The model proposed by Bago d'Uva (2006) uses the LCM to specify both stages involved in the hurdle framework, therefore relaxing the independence assumption between the two-parts of the model. However, the full model is not identified in cross-section data contexts, in special the first stage (Bago d'Uva 2005). One strategy for identification, the one that we have relied upon, consists in specifying the model for the first part as a simple binary model, and not as a latent class model. The second part of our model and Bago d'Uva's model coincides, with the difference that our model applies to cross-section data and Bago d'Uva's to panel data.

The model proposed by Bago d'Uva (2006) is richer than ours because it permits to model, using LCMs, the unobserved heterogeneity on both stages of the hurdle specification. The estimation of her model is, however, data demanding and requires panel data which was not our case. Therefore, our modelling strategy of disregarding the LCM specification on the first stage is not an option; we followed the constrains imposed by the cross-section nature of the available dataset. Moreover, our specification assumes independence between the unobservables on both stages of the model. Regarding the independence assumption, it should be noted, however, that this assumption is somehow week as we are dealing with conditional and not unconditional independence. The inclusion of common covariates in the two stages can control for some sources of dependence. Winkelmann (2004) pointed out, in

addition, difficulties in identify the correlation parameter in a model with correlation among the unobserved factors in the two parts.

As mentioned elsewhere on this dissertation, Winkelmann (2004) and Van Ourti (2004) also proposed hurdle specifications to model the number of doctor visits. In their formulation, the first stage is specified as usual, using a binary model and the model for the positives was developed assuming the (continuous) unobservables in the users population. Therefore, our alternative LCH-Cross model can also be viewed as the finite mixture counterpart of their models.

Despite the possibility of estimating this model in two steps, one can obtain the overall maximum likelihood estimator by maximizing the following log-likelihood function,

$$LogL = \sum_{i=1}^{N} \left[\begin{array}{l} (1-d_i) * \left[\log\left(f_0(0 \mid \boldsymbol{x}_i; \boldsymbol{\beta}_1) \right) \right] + \\ d_i * \log\left[\left[1 - f_0(0 \mid \boldsymbol{x}_i; \boldsymbol{\beta}_1) \right] * \sum_{j=1}^{P} \hat{\pi}_j \left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j \right)}{1 - P_j\left(y_i = 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j \right)} \right) \right] \end{array} \right] \quad [77]$$

where $d_i = \begin{cases} 1 & if \quad y_i > 0 \\ 0 & if \quad y_i = 0 \end{cases}$, $i = 1, 2, ..., n$.

In sum, relative to the common hurdle specifications, in this chapter we propose a modification in the approach to deal with the individuals with a positive number of visits, actually, the users. We suggest using the LCM framework to specify a density for the strictly positive utilization providing a closer approximation to the underlying decision process, and, at the same time, gaining statistical flexibility to the second stage of the hurdle.

4.3.1 — LCM for the health care users and posterior class analysis

Our alternative hurdle specification is based on the assumption that the population of health care users consists of P latent classes of individuals. Therefore, similar to what was done for the standard LCM model and presented in Section 2.5.1, also in this

specification, after estimating the mixing probabilities, allows us to classify the individuals probabilistically.

For a known distribution of the mixing probabilities of the LCM specification in the users population, $\hat{\pi}_j \ (j = 1, 2..., P)$, the posterior probability that health care user *i (i=1,…,n)* belongs to the latent class *j* (j = 1, 2,…,P) is given by

$$P\left(I_i \in C_j \mid Y_i = y_i, \mathbf{x}_i, y_i > 0\right) = \\ = \frac{\hat{\pi}_j * \dfrac{f_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right)}{P_j\left(y_i > 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right)}}{\sum_{k=1}^{P}\left[\hat{\pi}_k * \dfrac{f_k\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right)}{P_k\left(y_i > 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j\right)}\right]} \quad j = 1, 2, ..., P \qquad [78]$$

The assignment of the health care users to each latent class is made according to what was explained in Section 2.5.1, with a slight difference. The probabilities that appear in inequality [15] should be replaced by the probability computed as shown above.

4.4 — Data and Variables

This chapter begins by presenting some general features of the dataset used to estimate the models, namely, the number of individuals interviewed, the type of information it contains, followed by the description of the dependent variable analysed, presenting as well, the covariates considered as health care determinants. In addition, we describe the way that some of the covariates were created from the raw data present in the dataset.

The empirical results presented in this chapter were estimated using a cross-section data taken from the National Health Survey, henceforth referred to as NHS98_99. Trained interviewers applied the questionnaire, between October 1998 and September 1999, to 48.606 individuals of all ages and genders, belonging to 21.808 households

located at the five health regions, thus generating a representative sample of the Portuguese, non-institutionalized, population.

The NHS98_99 survey provides a wide range of information, at an individual level, about socioeconomic variables, life styles and health status indicators, ranging from chronicle diseases to morbidity indicators. In addition, the questionnaire includes also a large array of variables capturing medical care utilization. Moreover, the dataset also offers some information about the type of health insurance coverage of the individuals. Some data taken from external sources was imputed to the original dataset.

After dropping the 4.8% of individuals reporting voluntary health insurance subscription and deleting observations with missing values on any of the interest variables, the final sample comprises 42.501 observations. Voluntary health insurance introduces issues of selection we opt to side-step, given the relatively small number of observations involved. On the other hand, the elimination of observations with missing values may raise sample selection issues. They may occur whenever one estimates models using a sub-sample and there are correlated unobservables that influences inclusion in the sub-sample and the dependent variable (Vella 1998). If this deletion is non-random then standard procedures applied to the final sample will result on incorrect inference regarding the impact of the observables on doctor visits (Vella 1998; Wooldridge 2002). In our study, about 6% of the individuals did not report 'income', however, only a small share (10%) of those did it deliberately. The remaining 5,4% declared not knowing the household income. In this situation, the event of correlated unobserved factors influencing both the decision to respond and the number of doctor visits seems unlikely. Nonetheless, to test whether this reduction in the sample is random we performed the statistical test suggested by Wooldridge (2002) (*procedure 17.2, page 568*). Relatively to its result, the interest coefficient (the 2SLS parameter on $\hat{\lambda}$) is 0,516, presenting a standard error of 0,604, thus a t-statistic of 0,85. Therefore, the lack of statistical significance of the parameter suggests that the deletion induced by the missing values on the covariate income is random, giving an

indication that sample selection bias is not an issue in our dataset, at least from this source. It is prudent to regard this finding only as preliminary, essentially due to the low quality of the instruments used to conduct the test, which were the following,

- the purchasing power in the region in which the individual resides,
- the number of families living in the household,
- a dummy variable that indicates whether the individual visited a dentist in the last year,
- a dummy variable that indicates if the individual is single,
- a dummy variable that indicates if the individual is a widow.

The dependent variable in the application presented here — the variable 'visits' — measures the total number of visits to physicians in the three months prior to the interview. The interviewer asked each individual about how many times she/he had visited a physician in the last three months. Therefore, the number answered by the individual encompasses all types of doctor consultations; e.g. consultations to general practitioners, consultations to specialized physicians, consultations to dentists, consultations to hospital emergency rooms, etc., both to public and to private providers.

Table 2 (below) presents the empirical distribution of the dependent variable, 'visits',

'visits'	Frequency	Relative Frequency	Accumulated relative frequency
0	19.556	46,01%	46,01%
1	10.096	23,75%	69,77%
2	5.268	12,40%	82,16%
3	3.907	9,19%	91,36%
4	1.380	3,25%	94,60%
5	817	1,92%	96,52%
6	683	1,61%	98,13%
7	168	0,40%	98,53%
8	137	0,32%	98,85%
9	45	0,11%	98,96%
10	170	0,40%	99,36%
11	13	0,03%	99,39%
12	119	0,28%	99,67%
13	11	0,03%	99,69%
14	6	0,01%	99,71%
15	40	0,09%	99,80%
16	5	0,01%	99,81%
18	1	0,00%	99,81%
20	4	0,01%	99,82%
21	33	0,08%	99,90%
22	3	0,01%	99,91%
23	1	0,00%	99,91%
24	1	0,00%	99,91%
25	7	0,02%	99,93%
27	2	0,00%	99,93%
28	1	0,00%	99,94%
30	1	0,00%	99,94%
Mean		1,29	
Variance		4,24	
Variance/mean		3,29	
N		*42.501*	

Table 2 - Empirical distribution of the dependent variable 'visits'

From this table we conclude that approximately 46% percent of the individuals reported 0 visits to a physician in a period of three months, while approximately 70% reported zero or one visit. It is noteworthy to highlight the high percentage of zeros in the empirical distribution, about 46%, being this an indicator that alerts us to the need to use flexible count data models capable of appropriately accommodate this high percentage of zeros. Regarding other summary statistics describing the empirical distribution of the dependent variable, the maximum number of visits reported is thirty, with an average of 1,29 and a standard deviation of 2,06. Notice that the unconditional variance is almost four times the unconditional mean, meaning that, even after conditioning on the regressors, the conditional variance will probably continue to be substantially higher relative to the conditional mean, which is a clear sign of overdispersion.

To confirm the existence of overdispersion, we implemented an overdispersion test suggested by Cameron and Trivedi (1998). The result of the test (not presented here) indicated that, in fact, overdispersion is a characteristic of our data.

As covariates, we selected those that have been found to influence medical care utilization in similar studies. The covariates were clustered into five groups, encompassing socioeconomic and demographic covariates, health status indicators, a supply side determinant. We also included a group of covariates reflection the health insurance status of the individuals and, finally, three a group of three seasonality variables. Table 3 presents a description of these variables.

Variable	Variable Definition
Socioeconomic and demographic	
Age [/10]	Age in years, divided by 10
sqAge	Square of age[/10]
Female	= 1 if the individual is female
Married	= 1 if the individual is married
Education	Number of years of schooling. In the case of child, the

Variable	Variable Definition
	variables measures the number of years of schooling of most educated adult living in the household
Not_work	= 1 if the individual did not work in the two weeks previous to the application of the survey
Retired	= 1 if the individual is retired
Unemployed	= 1 if the individual is unemployed
(log) Income	Logarithm of equivalised monthly real income in hundreds of Euros
North	= 1 if the individual resides in the north region
Centre	= 1 if the individual resides in the centre region
LTV	= 1 if the individual resides in the Lisbon and Tagus Valley region
Alentejo	= 1 if the individual resides in the Alentejo region
Rural_Area	= 1 if the individual resides in a rural area
Health Status	
Diabetes	= 1 if the individual has diabetes
Insulin	= 1 if the individual is insulin dependent
EBP	= 1 if the individual has elevated blood pressure
Asthma	= 1 if the individual has asthma
Bronchitis	= 1 if the individual has bronchitis
Allergy	= 1 if the individual has an allergy
Back pain	= 1 if the individual has back pain
Ill_long_Run	= 1 if the individual has an illness for more than 3 months
Ill_Short_run	= 1 if the individual reported an illness in the two weeks prior to the application of the survey
Limited	= 1 if the individual has some sort of physical handicap that impedes him to execute certain basic physical daily activities
Stress	= 1 if the individual has been taking sleeping pills in the last two weeks
NeverSmoked	= 1 if the individual never smoked during her/his lifetime
not_physical_activity	= 1 if the individual's daily activities do not require physical activity
mild_exercise	= 1 if the individual engages in mild sports activities at least four hours a week.
Supply side	
Phy_1000_residents	Total number of licensed physicians per 1000 inhabitants
Health Insurance Status	
NHS-only	= 1 if the individual is covered only through the NHS
ADSE	= 1 if the individual is covered by the civil servants health

Variable	Variable Definition
	insurance scheme
	Seasonality
Winter	= 1 if the period of observation was in the Winter
Spring	= 1 if the period of observation was in the Spring
Summer	= 1 if the period of observation was in the Summer

Table 3 — Definition of covariates

We put off to Chapter 5 the presentation of the arguments that lead us to pick out these covariates as determinants of health care. In the current section, the objective is to describe how some of the explanatory variables were generated from the raw data present in the dataset. In this respect, we do not cover all the covariates, as the process that lead to the generation of some of them is obvious, thus, not deserving further explanations explanation.

The covariate 'education' measures the total number of years in school. In the case of individuals aged less than 14, education is no longer the number of years of schooling of the individual, being alternatively measured as the maximum number of years of schooling among the adults living in the household.

Another variable included in the model is the individual 'income'. Our dataset, as it is common, does not include a question to measure income as a continuous variable, neither at individual nor at household level. In alternative, income is measured by a categorical ordinal variable (with ten thresholds) that indicates the category of the disposable net household monthly income, earned in the month prior to the interview. The reported income includes regular wages, retirement pensions, and all sort of social security subsidies enjoyed by all members of the household. It is well known that when survey design measures income in such a way it is usual to control for the income effects including in the regression model a set of dummy variables, one for each income category considered (*e.g.* (Gerdtham 1997; Barros 1999)). Yet, in our empirical application, in order to keep the specifications parsimonious in terms of the

number of parameters to estimate, we controlled for the income through the inclusion of the monthly equivalent disposable income, which, using the appropriate methodology, can be computed from our data. According to Pereira (1995), when surveys include disposable income data as a categorical ordinal variable, with each category representing the income interval at which the household fits, two adjustments have to be made before the computation of equivalent income: first, the interpolation of grouped data, and second the normalization to differences in household characteristics. To perform the interpolation of grouped data, we have assumed that the midpoint of the interval at which the family belongs is the income of the household[14]. To make the second correction — a normalization to account for the family characteristics — we used the modified OECD scale. The scale assigns a weight of 1.0 to the first adult in the household, and for each additional adult (persons aged 14 years and over) a weight of 0,5, and for each child a weight of 0,3 (OECD 2004).

We also included the dummy variable that indicates whether the individual resides in a rural area, the 'rural_area' covariate. To create it we combined information contained in our dataset with information gathered from external sources, namely, from the Portuguese National Statistics Office (Instituto Nacional de Estatística 1999). The variable classifies each individual's place of residence as predominantly rural, predominantly urban and medially urban. After the initial classification in these three categories, we merged the predominantly urban and medially urban categories, creating in this way two dummy variables, referred to as 'rural_area' and 'urban_area'.

The remaining variables comprised in the socioeconomic category are self-explanatory.

Another category of variables included as covariates, are those intended to reflect the individual health status. The first seven dummy covariates reflect the presence or

---

[14] And assuming a value of 2 494 € for the last, open-ended, income bracket.

absence of chronic condition, and were capturing by asking each patient whether she/he had been suffering from each the condition in the last year. The variable 'Ill_Short_run' equals one if the individual reported to an illness in the two weeks prior to the application of the survey, and zero otherwise. In turn, the variable 'Ill_long_Run' is a dummy variable that equals one if the individual reports being ill for more than three months. The variable 'Limited', may be considered as a measure of disability; equals one if the individual has some type physical disability that incapacitates him from executing certain basic physical activities, e.g., get up from the bed, get up from a chair, run, go up and downstairs, pick-up an item from the floor, etc. On the other hand, the covariate referred to as 'Stress' is also a dummy variable that equals one if the individual had been taking sleeping pills in the two weeks prior to the application of the survey, zero otherwise. The last three variables included in the health status group, 'NeverSmoked', 'not physical activity' and 'mild_exercise' are entertained as proxies for health status. 'NeverSmoked' is a equals one 1 the individual never smoked during her/his lifetime, zero otherwise. The remaining two variables indicate the amount of physical activities made by the individual as part of their daily activities or as part of sports activities performed for fitness purposes. On the one hand, 'not physical activity' equals one if the daily activities made by the individual do not demand any type of strong physical effort, zero otherwise, on the other, 'mild_exercise' equals one if the individuals engages in activity mild sports[15], zero otherwise.

One popular form of measure health status in surveys is by asking directly the patients to assess his/her general health status, usually in a five point Likert scale, oftenly, ranging from 'excellent' health status to 'very bad' health status. Most of the

---

[15] Defined as light activity sports, e.g., walking, riding a bicycle or other mild activities, performed at least four hours a week

empirical studies of medical care utilization rely heavily on this variable to control for health status, which is included in the regression models as a set of dummies. Despite the NHS98_99 includes individual self-assessed health (henceforth referred to as SAH) we decided not to use it at all in thas part of the group of covariates. The major reason to leave it out was its inclusion would lead to the elimination of 36% of observations, decreasing to 27.044 observations. Surely, the 27.000 observations provide enough degrees of freedom to estimate the model parameters with good precision. However, we hypothesized that the loss of these individuals may cause sample selection bias. It can be argued that individuals in worse (unmeasured) health status are more reluctant to self-assess their health, or because they are unable to do it or because they do not want express their opinion about it. This suggests the existence of unobserved factors influencing both, the non-response and the number of visits, thus causing sample selection bias. To test whether the elimination of these individuals is random, we implemented the statistical test suggested by Wooldridge (2002), already mentioned above in the treatment of a similar issue.

The result of the test evidences that sample selection bias can, in effect, be an issue. The relevant coefficient of the test is -0,68, with a standard error of 0,15, thus a t-statistic of -4,55. To implement the test we used as instruments some behavioural variables considered as health determinants. We implemented the test using the following instrumental variables:

- the quantity of beer intake per day,
- the quantity of wine intake per day,
- the quantity of gin intake per day,
- the quantity of liqueurs intake per day,
- the number of meals intake per day,
- a dummy variable that indicates whether the individual smokes on a daily basis.

This result offers a first indication that the inclusion of SAH in our regression models as covariate may cause the working sample to depart from a random sample.

Undoubtedly, the result reported should be viewed prudently and not taken as definitive. We fully recognize that the instruments used to perform the statistical test may not be the most appropriates. We view the result as an important one, that should not be ignore, as it raises issues methodological that can spill over the findings of the models, possibly influencing the health policies designed based on such findings. Therefore, the issue will be on the forefront of our future research agenda.

Our strategy to avoid the sample selection bias was, as referred above, to exclude SAH from the models. In this way, we avoid the potential for the occurrence of sample selection bias. However, the approach of excluding SAH can have a cost also leading to potentially misleading results (Cameron and Trivedi 2005), nevertheless, we believe that in our application this is less likely to occur because we have a large array of variables to capture health status, measuring it sufficiently well.

The next variable, 'Phy_1000_residents', was created by adding external data to our sample. Assigning each individual's area of residence to a region and computing the total number of physicians per 1000 residents in that region.

Finally, regarding the health insurance status, Portugal provides health insurance with two main types of insurance schemes. The first, is the one provided by the statutory National Health Service (NHS), covering all the population; the second, is the health insurance provided by public and private insurance funds whose membership is based on professional or occupational category, referred to as Health Subsystems, which provide double coverage for circa 25% of the population. Among these, the fund covering all civil servants, usually referred to as ADSE, is different from the remaining mainly because the scale of operation. Because membership to these funds comes with the profession or occupation, it is reasonable to argue that the variables capturing health insurance are exogenous in our models (Barros, Machado *et al.* 2005).

Table 4 reports the summary statistics mean, standard deviation maximum and minimum of the independent variables.

Variable	Mean	S.D.	Max.	Min.
Socioeconomic and demographic				
Age [/10]	4,240	2,331	0	10,3
Female	0,527	0,499	0	1
Married	0,540	0,498	0	1
Education | age > 17	5,459	4,304	0	24
Not_work	0,589	0,492	0	1
Retired	0,202	0,401	0	1
Unemployed	0,030	0,171	0	1
Income/100	3,656	2,718	0,231	24,939
North	0,315	0,464	0	1
Centre	0,200	0,400	0	1
LTV	0,246	0,431	0	1
Alentejo	0,119	0,324	0	1
Rural_area	0,170	0,376	0	1
Health Status				
Diabetes	0,056	0,231	0	1
Insulin	0,006	0,079	0	1
EBP	0,178	0,383	0	1
Asthma	0,062	0,241	0	1
Bronchitis	0,030	0,170	0	1
Allergy	0,144	0,351	0	1
Back pain	0,407	0,491	0	1
Ill_long_Run	0,009	0,096	0	1
Ill_Short_run	0,344	0,475	0	1
Limited	0,045	0,207	0	1
Stress	0,113	0,317	0	1
NeverSmoked	0,629	0,483	0	1
not physical activity	0,609	0,488	0	1
mild_exercise	0,149	0,356	0	1
Supply side				
Phy_1000_Inhabitants	2,774	2,220	0,579	9,152
Health insurance status				
NHS-only	0,848	0,359	0	1
ADSE	0,095	0,293	0	1
Seasonality				
Winter	0,248	0,432	0	1
Spring	0,252	0,434	0	1
Summer	0,244	0,429	0	1

Variable	Mean	S.D.	Max.	Min.
N = 42.501				

Table 4 — Summary statistics

Thereafter, in Chapter 5, Section 5.3.2.1 (page 165) we provide a thorough analysis of these summary statistics, which useful in order to characterize the sample.

4.5 — Analysis and discussion of the results

This section goes through an analysis of the econometric results found in the application. We analyse the model estimates under two different perspectives, thus this section is organized into two sub-sections: in the first one, subsection, 4.5.1 — Evaluation of fitted models: Does the LCH-Cross outperform the competing alternatives? we address the issue of determining if the LCH-Cross hurdle specification outperforms, from a statistical point of view, the competing specifications specified and estimated in the present chapter. In the next subsection 4.5.2 — Do different models provide different evidence? we estimate the effect of selected covariates (e.g. 'income', 'education', 'rural_area' and 'NHS-only') on the conditional mean function of several competing models estimated. The motivation to conduct this analysis is to evaluate the robustness of some potentially relevant health policy indicators across models.

4.5.1 — Evaluation of fitted models: Does the LCH-Cross outperform the competing alternatives?

In the current chapter, we have defended a LCM approach to find the determinants of medical care utilization, conditional on utilization being positive. Accordingly, we proposed the specification of the model that we referred to as LCH-Cross. Therefore, in this subsection, we take up verifying whether the alternative hurdle proposal

outperforms some of the count data specifications that have tended to dominate the empirical studies, namely, the hurdle based on the NB probability and the LCMs, with NB as baseline densities.

To begin with, Table 5 shows a description of the 12 models estimated, in addition to the respective number of parameters estimated. Finally, the table also reports the maximum log-likelihood values of the different models.

Acronym	Model Description	Param	LogL
NB1	Negative Binomial (type 1) regression model	36	- 61.837
H_NB1	Popular Hurdle: First and second parts based on the NB1 distribution. In the first part the dispersion parameter was set to one.	71	- 61.470
LCM_NB1	Two class LCM. NB1 density as component distribution.	73	- 61.171
R-LCM_NB1	LCM_NB1 with slopes constrained to be equal across latent classes.	39	- 61.241
LCH_Cross(NB1)	Alternative Hurdle: first part based on the NB1. The second part is a 2 LCM with truncated NB1 as baseline distribution. In the first part the dispersion parameter was set to one.	108	- 60.938
R-LCH_Cross(NB1)	LCH_Cross(NB1) with slopes constrained to be equal across latent classes in the second stage.	74	- 60.990
NB2	Negative Binomial (type 2) regression model	36	- 62.201
H_NB2	Popular Hurdle: First and second parts based on the NB2 distribution. In the first part the dispersion parameter was set to one.	71	- 61.275
LCM_NB2	Two class LCM. NB2 as component distribution.	73	- 61.389
R-LCM_NB2	LCM_NB2 with slopes constrained to be equal across latent classes.	39	- 61.473

Acronym	Model Description	Param	LogL
LCH_Cross(NB2)	Alternative Hurdle: first part based on the NB2. The second part is a 2 LCM with truncated NB2 as baseline distribution. In the first part the dispersion parameter was set to one.	108	- 60.741
R-LCH_Cross(NB2)	LCH_Cross(NB2) with slopes constrained to be equal across latent classes in the second stage.	74	- 60.797

Table 5 -— Description of the competing models estimated

All models were estimated using the same number of observations, 42.501 cases, therefore the analysis of the maximum log-likelihood values provide the first insight regarding the model that better adjusts the data. In general, as expected, increased model complexity leads to improvements in the value of the log-likelihood. Note that both single index models, NB1 and NB2, present the lower log-likelihood values when compared to all competing specifications. In addition, among the specifications based on the NB1 density, the LCM model presents a higher log-likelihood than the log-likelihood of the popular hurdle specification. Additionally, the LCH-Cross based offers a substantial improvement over all competing alternative specifications.

A similar pattern emerges in all specifications based on the NB2 density, however, we note a slight difference. In this case, the LCM is "*log-likelihood dominated*" by the popular hurdle specification, the model labelled H_NB2 in Table 5.

The analysis of the results discussed above, although based only on the log-likelihood values offer the first evidence towards the statistical superiority of hurdle specification that relies upon the LCM framework to deal with the individuals with a positive utilization. Nevertheless, before definitive conclusion one has to implement and analyse the results of more formal model selection tests. In Chapter 2, in section 2.10 — Model Comparison and Selection Procedures — we explained the methods that we have used to compare the competing specifications considered in the chapter. We

remind that to select the model that better fits the data we used LR tests to compare the nested alternatives, for the non-nested alternatives we used information criteria and Vuong tests.

Table 6 reports the results of LR tests carried out to compare the nested alternatives estimated in this application, which are the following,

- NBi is nested in H_NBi *(i = 1, 2),*
- R-LCM_NBi is nested in LCM_NBi *(i = 1, 2)* ,
- R-LCH-Cross(NBi) is nested in LCH-Cross(NBi) *(i = 1, 2),*
- NBi is nested in LCM_NB_i *(i = 1, 2),*
- NBi is nested in R-LCM_NBi *(i = 1, 2),*
- H_NBi is nested in LCH-Cross(NBi) *(i = 1, 2),*
- H_NBi is nested in R- LCH-Cross(NBi) *(i = 1, 2),*

Test	Null	Alternative	LR	D.F	Decision
LR1	NB1	H_NB1	734,1	36	H_NB1
LR 2	NB1	LCM_NB1	1332,7	37	LCM_NB1
LR 3	NB1	R-LCM_NB1	1191,4	3	R-LCM_NB1
LR 4	R-LCM_NB1	LCM_NB1	141,3	34	LCM_NB1
LR 5	H_NB1	LCH-Cross(NB1)	1064,0	37	LCH-Cross(NB1)
LR 6	H_NB1	R-LCH-Cross(NB1)	958,9	3	R- LCH-Cross(NB1)
LR 7	R-LCH-Cross(NB1)	LCH-Cross(NB1)	105,1	34	LCH-Cross(NB1)
LR 8	NB2	H_NB2	1851,1	36	H_NB2
LR 9	NB2	LCM_NB2	1622,3	37	LCM_NB2
LR 10	NB2	R-LCM_NB2	1456,0	3	R-LCM_NB2
LR 11	R-LCMNB2	LCM_NB2	166,3	34	LCM_NB2
LR 12	H_NB2	LCH-Cross(NB2)	1068,6	37	LCH-Cross(NB2)
LR 13	H_BN2	R-LCH-Cross(NB2)	956,6	3	R- LCH-Cross(NB2)
LR 14	R- LCH-Cross(NB2)	LCH-Cross(NB2)	111,9	34	LCH-Cross(NB2)

Table 6 — Likelihood ratio tests results[16]

The first two columns in Table 6 present, respectively, the model under the null hypothesis, and the model under the alternative. The column labelled 'LR' show the value of the likelihood ratio statistic (see statistic [41] in section 2.10.1) while the degrees of freedom of the test are presented in the next column. In every case, the critical value for the test is taken from a $\chi^2(D.F.)$[17] distribution. Finally, the last column of Table 6 shows the result of the test.

We begin by noting that all LR hypotheses tests presented in Table 6 reject the null models, that is, the data seems to provide support to the more general models. In spite of the fact that the tests LR2, LR3, LR5, LR6, LR9, LR10, LR12, and LR13 involve parameter restrictions in the boundary of the parameter space, thus the use of the usual χ^2 distribution is likely to under-reject the null hypothesis, all results points towards the rejection of the null.

As previously referred, the dimension of the sample is about 42.501, causing probably the large LR values reported in the table. This may lead to the well-known difficulty of testing hypothesis in large samples. In these circumstances, there is a tendency for the tests to reject always the null. To guard against this tendency to over rejection in large samples, we performed the same tests for a smaller sample ($N = 10.625$) taken at random from the complete working sample. This experiment allowed us to reach similar conclusions, that is, the rejection of all null hypotheses.

---

[16] Some of these tests are made in the boundary of the parameter space, so the standard testing procedure may not be valid. Deb and Trivedi (1997) showed, however, that the test has excellent power under the alternative, even when the usual critical value is used.

[17] D.F. stands for degrees of freedom.

The value of the LR statistic has suffered, however, a substantial reduction, as expected.

Both versions of the single index NB regression models were rejected in favour of the popular hurdle model (tests LR1 and LR8), showing that the parameter estimates of the two parts of the hurdle are significantly different. This provides some support to the view that the decisions regarding medical care, in fact, may involve more than one stochastic process. The standard LCM, models LCM_NB1 and LCM_NB2, are found to be preferred specifications when compared to the uni-component NB models (tests LR2 and LR9). Moreover, the constrained versions of the standard two-component models, models R-LCM_NB1 and R-LCM_NB2, are also found to outperform the uni-component NB models (tests LR3 and LR10). These findings suggest that individual behaviour concerning the utilization of doctor consultations is not well represented by models of the single index BN family. Therefore, we may conclude that the researcher must specify more general count data distributions to explain adequately the individual behaviour.

Within the family of LCMs, the results do not support the constrained versions model of LCM (tests LR4 and LR11), showing that, when considered jointly, the vectors of slopes are different across the latent classes.

Table 6 also reports that the LR test of H_NB1 (H_NB2) against LCH-Cross(NB1) (LCH-Cross(NB2)), rejects the null hypothesis (tests LR5 and LR12), suggesting that our alternative hurdle model outperforms the popular NB based hurdle. This seems to indicate that, conversely to the common practice of specifying a simple truncated-at-zero negative binomial model to find the determinants of utilization, conditional on utilization being positive, the second part of the model is better described by a two-component LCM model specified as a finite mixture of two truncated NB distributions. An alternative interpretation for this finding is that the popular hurdle model accounts incorrectly for the unmeasured factors, supposedly present in the users population, being needed approaches more sophisticated to account for the unobserved effects. In

addition, the LR test of R- LCH-Cross(NB1) against the LCH-Cross(NB1) model rejects the null, with similar conclusions for the alternative hurdles based on the NB2 distribution.

Taking together the results of the LR tests showed in Table 6, the tests have lead to the rejection of the following specifications:

- NBi — models rejected when compared against the models H_NBi, LCM_NBi and R-LCM_NBi, *i = 1, 2* (for i = 1, tests LR1, LR2 and LR3. For i = 2, tests LR8, LR9 and LR10),
- H_NBi — models rejected when compared against the models LCH-Cross(NBi) and R- LCH-Cross(NBi), i = 1, 2 (for i = 1, tests LR5 and LR6. For i = 2, tests LR12 and LR13),
- R-LCM_NBi — models rejected when compared against the model LCM_NBi, i = 1, 2 (for i = 1, test LR4. For i = 2, test LR11),
- R-LCH-Cross(NBi) — models rejected when compared against the models LCM_NBi, i = 1, 2 (for i = 1, test LR7. For i = 2, test LR14)

We now examine the results of the Vuong tests and BIC and CAIC statistics to provide further insights about the comparison of the non-nested considered.

We begin by reporting, in Table 7, the results of several Voung tests implemented,

Test	Model 1 F_{β_1}	Model 2: G_{β_2}	$V\left(\hat{\beta}_1, \hat{\beta}_2\right)$	Result
V1	LCM_NB1	LCM_NB2	12,1	LCM_NB1
V2	LCM_NB1	LCH-Cross(NB1)	-10,0	LCH-Cross(NB1)
V3	LCM_NB1	LCH-Cross(NB2)	-15,6	LCH-Cross(NB2)
V4	LCM_NB2	LCH-Cross(NB1)	-13,3	LCH-Cross(NB1)
V5	LCM_NB2	LCH-Cross(NB2)	-19,1	LCH-Cross(NB2)
V6	LCH-Cross(NB1)	LCH-Cross(NB2)	-12,0	LCH-Cross(NB2)
V7	LCM_NB1	H_NB1	6,3	LCM_NB1
V8	LCM_NB2	H_NB2	-2,2	H_NB2 *
V9	LCM_NB1	H_NB2	2,1	LCM_NB1*

Test	Model 1 F_{β_1}	Model 2: G_{β_2}	$V\left(\hat{\beta}_1, \hat{\beta}_2\right)$	Result
V10	LCM_NB2	H_NB1	1,6	Inconclusive
V11	NB1	NB2	6,7	NB1
V12	H_NB1	H_NB2	-6,1	H_NB2

Table 7 — Results of various Vuong tests implemented to discriminate among non-nested alternatives[18]

To begin with, test V1 compares the two versions of the standard latent class model. It contrasts the LCM based on the NB1 density against the LCM based on the NB2 density. The test statistic for these test is 12,1, hence the null hypothesis of equivalent models is rejected in favour of LCM_NB1 being better than LCM_NB2.

Another group of tests presented in Table 7, tests V2 through V5, are those that aim at comparing the standard latent class model specification, model LCM_NBi, against our alternative hurdle specification, the model LCH-Cross(NBi). The test statistic for the LCM_NB1 against the LCH-Cross(NB1) model is -10,0, for the LCM_NB1 against the LCH-Cross(NB2) model is -15,6, for the LCM_NB2 against the LCH-Cross(NB1) model is -13.3, and finally, the test statistic for the model LCM_NB2 against the model LCH-Cross(NB2) model is -19,1. In conclusion, all these four tests rejects the null hypothesis of equivalency between LCM_NBi and LCH-Cross(NBi) (i=1,2) in favour of the alternative hurdle being better that the standard latent class model. In sum, the results of all tests provide clear evidence about the better statistical performance of our LCM based hurdle.

On the other hand, test V6 compares the two alternative versions of the LCH-Cross(NBi) model, the versions based on the NB1 density against the version based on the NB2 density. The test statistic for the LCH-Cross(NB1) model against the model

---

[18] All the tests are significant for 1%, except the two cases signalled with the *: are significant for $0,01 \leq \alpha \leq 0,05$.

LCH-Cross(NB2) is -12,0, thus the null hypothesis of equivalent models is rejected, with statistical significance, in favour of the LCH-Cross(NB2) being better than the LCH-Cross(NB1). This result shows that the specification with NB2 as parent distribution is the preferred model to fit our data.

Turning now to the analysis of the results of the model selection methods based on information criteria statistics, Table 8 presents values of the statistics BIC and CAIC for all models estimated.

Model	BIC	CAIC
NB1	124.057	124.093
H_NB1	123.696	123.767
LCM_NB1	123.119	123.192
LCH-Cross(NB1)	123.027	123.135
NB2	124.785	124.821
H_NB2	123.306	123.377
LCM_NB2	123.557	123.630
LCH-Cross(NB2)♣	122.632	122.740

Table 8 — Information Criteria results

♣ — Model preferred by BIC and CAIC

The figures presented in the Table 8 show that the single index models NB1 and NB2 perform poorly relative to all other econometric specifications, as is evidenced by the large BIC and CAIC values. Thus, both BIC e CAIC statistics overwhelmingly reject the NBi regression models. This observation is entirely in line with the conclusions previously presented. The comparison of LCM_NBi (i=1,2) against H_NBi (i=1,2) model provide mixed evidence regarding which family of models performs better. The standard latent class model performs better than the popular hurdle model, when the NB1 is the parent distribution, while the opposite result occurs when the NB2 is the base model to specify the LCM.

Given the objective of the application developed in this chapter, what is more relevant to emphasize from Table 8 is that, regardless the parent distribution (NB1 or NB2) used to specify the alternative LCM based hurdle, both the BIC and CAIC statistics present the lower values when compared with the BIC and CAIC of all other competing models. This result offers clear evidence that LCH-Cross(NB2) proposal outperforms all competing specifications estimated in this application. What is more, within each density class, the model with lower values for BIC and CAIC is the new hurdle based on the negative binomial 2 distribution, presenting in this case a BIC of 122.632 and a CAIC of 122.740.

Note that the conclusion found by using these model selection methods is a further input to add to Voung tests conclusion that supports the superiority of hurdle formulation whose second part is specified as an LCM.

In sum, taking together all the evidence presented above, the conclusion is that when compared to the standard count data models, the alternative proposed formulation for the hurdle model provides the most satisfactory fit among all the models considered in this chapter. Therefore, in line with Pohlmeier and Ulrich (1995) our results seems to evidence that the contact and the frequency decisions are governed by two different stochastic processes. Ignoring these differences leads to inconsistent parameter estimates and to economic misinterpretations. In addition, our results also show that for appropriately explain the frequency of healthcare utilization, conditional on frequency being positive, one has to rely upon flexible stochastic specifications able to properly incorporate the unobserved heterogeneity, supposedly present in the health care users population. As was shown, the latent class approach respond are a possible response to find these flexible specifications capable to account for the unobserved heterogeneity. Furthermore, our results are in accordance with Winkelmann's (2004) conclusions that hurdle specifications that deviate from the popular hurdle specification can outperform the count data alternatives based on standard latent class models. Our findings suggest that the evidence that has been

reported favouring standard latent class models over the popular hurdle specification (e.g. (Deb and Trivedi 1997; Deb and Holmes 2000; Deb and Trivedi 2002; Jimenez-Martin, Labeaga *et al.* 2002; Sarma and Simpson 2006) can only be interpreted as evidence against that particular specification of the hurdle but not against the general hurdle framework (Winkelmann 2004). Hence, our findings, in a way, may renew the somewhat old discussion about the better econometric strategy to explain the utilization of medical care by the individuals.

It t should be noted that all studies, at least to the extent of our knowledge, that have reported the statistical superiority of the LCM over the hurdle, measures utilization over the period of one year. This long period of measurement, surely, increases the probability of violating the single illness spell hypothesis on which these models are based (Santos-Silva and Windmeijer 2001). When the data regarding the number of doctor consultations are collected during a shorter time period (for example, 3 months), decreasing the probability of violating the single illness spell assumption, the hurdle structure may again emerge as the appropriate model to explain individual behaviour. In this data paradigm, if the second part of the hurdle correctly specifies the unobservables then the hurdle structure outperforms the LCM framework. This is precisely what our and Winkelmann's (2004) results have showed. In both empirical applications doctor visits are measured during a three months period, and the second part of the hurdle are correctly specified. The results of both applications presented similar conclusions regarding the preference for the hurdle structure.

Therefore, in our view, the performance of one specification over another, being probably dependent on the manner one specifies the model for the positives, may also depend on the characteristics of the data, namely, on the survey design as well as on the type of medical care utilization that is analysed, as shown by Jimenez-Martin *et al.* (2002).

One interesting question would be verify whether the LCH-Cross(NB2) model still continue to be the preferred specification if applied to medical care data gathered during the period of one year. We leave it for future work.

The next three tables report the parameter estimates of the models Poisson, NB1, H_NB2, LCM_NB1 and LCH-Cross(NB2).

Model	**Poisson**	**NB1**
Constant	-0,678 (0,064)	-0,683 (0,05)
Socioeconomic and demographic		
Age [/10]	-0,043 (0,021)	-0,073 (0,017)
sqAge	0,004 (0,002)	0,008 (0,002)
Female	0,119 (0,017)	0,153 (0,014)
Married	0,148 (0,02)	0,159 (0,016)
Education	0,006 (0,003)	0,009 (0,002)
Not_work	0,221 (0,021)	0,185 (0,017)
Retired	0,153 (0,024)	0,143 (0,02)
Unemployed	-0,014 (0,049)	-0,048 (0,04)
(log) Income	0,036 (0,013)	0,045 (0,011)
North	0,126 (0,028)	0,157 (0,022)
Centre	0,247 (0,029)	0,263 (0,023)
LTV	0,133 (0,029)	0,167 (0,023)
Alentejo	0,131 (0,032)	0,164 (0,025)
Rural_Area	-0,082 (0,020)	-0,086 (0,017)
Health Status		
Diabetes	0,228 (0,025)	0,254 (0,021)
Insulin	0,374 (0,077)	0,294 (0,06)
EBP	0,229 (0,018)	0,261 (0,015)
Asthma	0,131 (0,027)	0,125 (0,022)
Bronchitis	0,087 (0,032)	0,097 (0,028)
Allergy	0,178 (0,019)	0,172 (0,015)
Back pain	0,191 (0,018)	0,196 (0,015)
Ill_long_Run	0,743 (0,055)	0,717 (0,046)
Ill_Short_run	0,644 (0,016)	0,608 (0,013)
Limited	0,116 (0,034)	0,031 (0,028)
Stress	0,349 (0,02)	0,343 (0,016)
NeverSmoked	-0,125 (0,018)	-0,1 (0,014)

Model	Poisson	NB1
not physical activity	0,058 (0,025)	0,026 (0,02)
mild_exercise	-0,009 (0,03)	-0,038 (0,024)
Supply side		
Phy_1000_Inhabitants	0,012 (0,004)	0,01 (0,003)
Health insurance status		
NHS-only	-0,11 (0,032)	-0,107 (0,025)
ADSE	-0,105 (0,038)	-0,101 (0,029)
Seasonality		
Winter	0,006 (0,021)	0,027 (0,016)
Spring	0,093 (0,021)	0,1 (0,016)
Summer	0,02 (0,02)	0,038 (0,017)
α	---	1,130 (0,028)

Table 9 — Parameter estimates of the Poisson and NB1 regression models.

Model H_NB2	First part	Second part
Constant	-0,905 (0,088)	-0,488 (0,095)
Socioeconomic and demographic		
Age [/10]	-0,319 (0,031)	0,032 (0,029)
SqAge	0,035 (0,003)	-0,006 (0,003)
Female	0,322 (0,024)	0,024 (0,026)
Married	0,314 (0,029)	0,082 (0,028)
Education	0,022 (0,004)	-0,004 (0,004)
not_work	0,153 (0,031)	0,217 (0,031)
Retired	0,305 (0,044)	0,112 (0,033)
Unemployed	-0,072 (0,067)	0,079 (0,07)
(log) Income	0,119 (0,02)	-0,003 (0,019)
North	0,265 (0,037)	0,025 (0,042)
Centre	0,42 (0,04)	0,144 (0,043)
LVT	0,274 (0,04)	0,034 (0,043)
Alentejo	0,25 (0,044)	0,012 (0,046)
Rural_Area	-0,166 (0,031)	-0,05 (0,029)
Health Status		
Diabetes	0,842 (0,064)	0,145 (0,036)
Insulin	0,882 (0,238)	0,431 (0,099)
EBP	0,699 (0,035)	0,113 (0,024)
Asthma	0,303 (0,05)	0,115 (0,036)

Model H_NB2	First part	Second part
Bronchitis	0,24 (0,072)	0,081 (0,042)
Allergy	0,372 (0,033)	0,137 (0,027)
Back pain	0,316 (0,026)	0,15 (0,025)
Ill_long_Run	1,542 (0,17)	0,742 (0,068)
Ill_Short_run	1,002 (0,025)	0,506 (0,022)
Limited	-0,125 (0,067)	0,257 (0,044)
Stress	0,875 (0,044)	0,294 (0,026)
NeverSmoked	-0,106 (0,025)	-0,137 (0,028)
not physical activity	0,026 (0,036)	0,11 (0,039)
mild_exercise	-0,087 (0,042)	0,07 (0,047)
Supply side		
Phy_1000_Inhabitants	0,018 (0,005)	0,009 (0,005)
Health insurance status		
NHS-only	-0,22 (0,048)	-0,073 (0,048)
ADSE	-0,205 (0,057)	-0,076 (0,058)
Seasonality		
Winter	0,071 (0,03)	-0,038 (0,03)
Spring	0,185 (0,03)	0,06 (0,031)
Summer	0,066 (0,03)	0,007 (0,03)
α	1	1,122 (0,067)

Table 10 — Parameter estimates of the Hurdle based on the NB2

On the other hand, Table 11**Error! Reference source not found.** reports the parameter estimates and robust standard errors of the LCM_NB1 model, while Table 12**Error! Reference source not found.** reports the parameter estimates of the model refereed to as LCH-Cross(NB2).

Latent Class model: NB1 as parent distribution		
	Latent class I (13,8%)	Latent class II (86,2%)

Latent Class model: NB1 as parent distribution

	Latent class I (13,8%)	Latent class II (86,2%)
Constant	-0,282 (0,309)	-0,826 (0,07)
Socioeconomic and demographic		
Age [/10]	0,044 (0,129)	-0,095 (0,028)
sqAge	-0,013 (0,015)	0,012 (0,003)
Female	-0,065 (0,074)	0,202 (0,019)
Married	0,085 (0,103)	0,177 (0,022)
Education	0,009 (0,012)	0,009 (0,003)
not_work	0,397 (0,12)	0,149 (0,027)
Retired	0,261 (0,13)	0,118 (0,029)
Unemployed	0,212 (0,159)	-0,115 (0,053)
(log) Income	0,069 (0,059)	0,042 (0,014)
North	-0,067 (0,11)	0,211 (0,031)
Centre	0,185 (0,113)	0,294 (0,031)
LVT	-0,046 (0,125)	0,221 (0,032)
Alentejo	-0,03 (0,118)	0,205 (0,033)
Rural_Area	-0,062 (0,093)	-0,091 (0,023)
Health Status		
Diabetes	0,276 (0,151)	0,262 (0,027)
Insulin	0,929 (0,197)	0,178 (0,062)
EBP	0,167 (0,097)	0,285 (0,02)
Asthma	0,338 (0,121)	0,094 (0,028)
Bronchitis	-0,021 (0,237)	0,123 (0,044)
Allergy	0,278 (0,086)	0,156 (0,02)
Back pain	0,189 (0,095)	0,206 (0,021)
Ill_long_Run	0,967 (0,19)	0,693 (0,051)
Ill_Short_run	0,971 (0,09)	0,551 (0,023)
Limited	0,503 (0,103)	-0,074 (0,036)
Stress	0,548 (0,089)	0,315 (0,021)
NeverSmoked	-0,218 (0,081)	-0,084 (0,018)
not physical activity	0,139 (0,1)	-0,001 (0,026)
mild_exercise	0,005 (0,128)	-0,051 (0,031)
Supply side		
Phy_1000_Inhabitants	0,041 (0,017)	0,004 (0,004)
Health insurance status		
NHS-only	-0,145 (0,133)	-0,106 (0,031)

Latent Class model: NB1 as parent distribution

	Latent class I (13,8%)	Latent class II (86,2%)
ADSE	-0,199 (0,163)	-0,09 (0,037)
Seasonality		
Winter	-0,088 (0,083)	0,052 (0,021)
Spring	0,063 (0,085)	0,112 (0,021)
Summer	-0,16 (0,096)	0,081 (0,023)
α	4,262 (0,351)	0,507 (0,035)

Table 11 — Parameter estimates of the model LCM_NB1

Hurdle Latent class model: NB2 as parent distribution
(LCH-Cross(NB2) model)

	First part	Second part	
		Latent class I (17,8%)	Latent class II (82,2%)
Constant	-0,905 (0,088)	0,743 (0,257)	-0,596 (0,092)
Socioeconomic and demographic			
Age [/10]	-0,319 (0,031)	0,062 (0,061)	0,055 (0,029)
sqAge	0,035 (0,003)	-0,014 (0,007)	-0,005 (0,003)
Female	0,322 (0,024)	-0,132 (0,058)	0,101 (0,024)
Married	0,314 (0,029)	-0,026 (0,062)	0,115 (0,025)
Education	0,022 (0,004)	-0,001 (0,009)	-0,004 (0,004)
not_work	0,153 (0,031)	0,185 (0,069)	0,206 (0,03)
Retired	0,305 (0,044)	0,229 (0,086)	0,045 (0,03)
Unemployed	-0,072 (0,067)	0,173 (0,144)	0,001 (0,066)
(log) Income	0,119 (0,02)	0,01 (0,042)	-0,01 (0,018)
North	0,265 (0,037)	-0,104 (0,089)	0,104 (0,039)
Centre	0,420 (0,04)	0,082 (0,093)	0,175 (0,039)
LVT	0,274 (0,04)	-0,129 (0,09)	0,128 (0,04)
Alentejo	0,25 (0,044)	-0,137 (0,1)	0,111 (0,044)
Rural_Area	-0,166 (0,031)	-0,068 (0,068)	-0,034 (0,027)
Health Status			
Diabetes	0,842 (0,064)	0,052 (0,097)	0,153 (0,031)
Insulin	0,882 (0,238)	0,632 (0,182)	0,193 (0,078)

Hurdle Latent class model: NB2 as parent distribution
(LCH-Cross(NB2) model)

	First part	Second part	
		Latent class I (17,8%)	Latent class II (82,2%)
EBP	0,699 (0,035)	0,034 (0,062)	0,126 (0,023)
Asthma	0,052 (0,097)	0,052 (0,097)	0,052 (0,097)
Bronchitis	0,24 (0,072)	-0,067 (0,109)	0,128 (0,041)
Allergy	0,372 (0,033)	0,143 (0,061)	0,104 (0,024)
Back pain	0,316 (0,026)	0,048 (0,059)	0,176 (0,024)
Ill_long_Run	1,542 (0,17)	0,675 (0,157)	0,629 (0,053)
Ill_Short_run	1,002 (0,025)	0,452 (0,051)	0,466 (0,022)
Limited	-0,126 (0,067)	0,439 (0,096)	0,091 (0,041)
Stress	0,875 (0,044)	0,285 (0,063)	0,245 (0,024)
NeverSmoked	-0,106 (0,025)	-0,161 (0,063)	-0,103 (0,025)
not physical activity	0,026 (0,036)	0,156 (0,085)	0,039 (0,036)
mild_exercise	-0,087 (0,042)	0,083 (0,102)	0,018 (0,043)
Supply side			
Phy_1000_Inhabitants	0,018 (0,005)	0,014 (0,012)	0,005 (0,005)
Health insurance status			
NHS-only	-0,220 (0,048)	-0,096 (0,107)	-0,05 (0,042)
ADSE	-0,205 (0,057)	-0,083 (0,127)	-0,053 (0,052)
Seasonality			
Winter	0,071 (0,03)	-0,102 (0,067)	0,009 (0,028)
Spring	0,185 (0,03)	0,024 (0,069)	0,082 (0,027)
Summer	0,066 (0,03)	-0,139 (0,069)	0,089 (0,028)
α	1	1,242 (0,193)	0,158 (0,035)

Table 12 — Parameter estimates of the model LCH-Cross(NB2)

4.5.2 — Do different models provide different evidence?

An important health policy theme, in Portugal and on the other OECD countries, is the access to health care. There is a wide consensus in Portugal that access to health care should only depend on the need for medical care and not on socioeconomic or demographic variables, like, for instance, area of residence, income, health insurance

status etc. Thus, an empirical test to examine whether those in equal health care need have equal access to medical care is relevant for health authorities. However, as it was clearly asserted by Deb and Holmes (2000), the estimates of medical care utilization may be dependent on the empirical specification used to examine the data, thus, if health care utilization models do not properly reflect the behavioural structures then estimates will not reflect actual use, and suggested policies may have unexpected consequences. In order to cast some light on the extent to which policy relevant measures depend on the form of the empirical specification used to analyse health care utilization, we report the effect of income, education, place of residence and health insurance status on the mean function for a number of alternative count data models.

Table 13 reports the estimates of the impact of such covariates on the conditional mean function of the overall population for five competing specifications.

Model Covariate	Poisson	NB1	H_NB2	LCM_NB1	LCH-Cross(NB2)
Income (elasticity)	0,036 (0,013)	0,045 (0,010)	0,053 (0,013)	0,048 (0,012)	0,053 (0,013)
Education	0,008 (0,003)	0,011 (0,003)	0,007 (0,003)	0,011 (0,003)	0,007 (0,003)
Rural_Area	-0,103 (0,024)	-0,107 (0,020)	-0,107 (0,024)	-0,106 (0,025)	-0,107 (0,024)
NHS-only	-0,147 (0,045)	-0,143 (0,035)	-0,152 (0,045)	-0,156 (0,04)	-0,152 (0,043)

Table 13 — Effect of selected covariates on the conditional mean for a number of models[19]

[19] To compute the figures presented in Table 13 we used the following procedure: To begin with, the descriptor of the conditional distribution that we are interested in is on the conditional mean function of the overall population. Therefore, using the notation presented in section 2.11, expression [7] $\left[\exp\left(\mathbf{x}_i'\beta\right)\right]$ gives the functional form of the $G(.)$ function for

the Poisson and NB1 specifications. On the other hand, for the models H_NB2, LCM_NB1 and LCH-Cross(NB2), the corresponding $G(.)$ depends on two linear indexes and differs across models, being defined, respectively, by $E_{H\_NB2}(y_i \mid \mathbf{x}_i) = \frac{1-(1+\lambda_{i0})^{-1}}{1-(1+\alpha_1\lambda_{i1})^{-\frac{1}{\alpha_1}}} * \lambda_{i1}$,

$E_{LCM\_NB1}(y_i \mid \mathbf{x}_i) = \sum_{j=1}^{P} \pi_j \lambda_{ij}$ and

$E_{\text{LCH-Cross(NB2)}}(y_i \mid \mathbf{x}_i) = \left[\frac{\lambda_{i0}}{\lambda_{i0}+1}\right] * \sum_{j=1}^{2} \left[\hat{\pi}_j * \frac{\lambda_{ij}}{\left(1-(1+\alpha_j\lambda_{ij})^{-\frac{1}{\alpha_j}}\right)}\right]$. Using the conditional mean functions presented above, we estimated the marginal effect for individual i according to expression [46] for the covariate 'education' and according to [56] for the dummy covariates 'rural_area' and 'NHS-only'. Afterwards, we obtained an estimated of the average of the individual marginal effects distribution, obtaining, in this way, an estimate of the Average Marginal Effects (AME). The numbers between parentheses in Table 13 are the standard errors of the AME estimates, computed using the Delta method. Regarding the income elasticity, the one corresponding to the Poisson and NB1 specifications comes out directly from the estimation results. In both models conditional mean of the overall population is specified in such a way that constrains the income elasticity to be invariant across individuals. Note that the income covariate is included in logs in the linear index of the mean function. For all other models analysed in the previous table, models H_NB2, LCM_NB1 and LCH-Cross(NB2), given the functional form of the conditional mean for the overall population, the income elasticity estimate varies across individuals. Therefore, for each model, we estimated the individual income elasticity of utilization and averaged across individuals, obtaining, in this way, an Average Elasticity (AE) for the population. The Standard errors are between parentheses. All estimates are statistically significant ($p < 0{,}01$).

The figures presented in Table 13 evidence that for each covariate considered all models estimate impacts of the same sign across models. Despite some variations in the magnitude of the point estimates presented in the table, all 99% confidence intervals (which are not presented in Table 13) generally overlap in a large extent, meaning that, after accounting for the variability, the estimates resulting from the different models are not statistically different. Therefore, these finding evidence that, at least for the covariates considered in our analysis, to identify the magnitude of the impact of such covariates in the conditional mean of the overall population it is indifferent to specify the data generating mechanism as a PRM, LCM_NB1 or as a model of the hurdle family, model H_NB2 or LCH-Cross(NB2).

The effects differ, however, across individuals, therefore to illustrate the variation of the estimated individual effects across individuals the four figures presented afterwards depict the full distribution of the individual marginal effects. It is worth to emphasize that the standard errors presented in Table 13, appearing above, are the standard errors of the average marginal effects estimates, calculated resorting on the Delta method, and not the standard deviation of the distribution of the individual marginal effects.

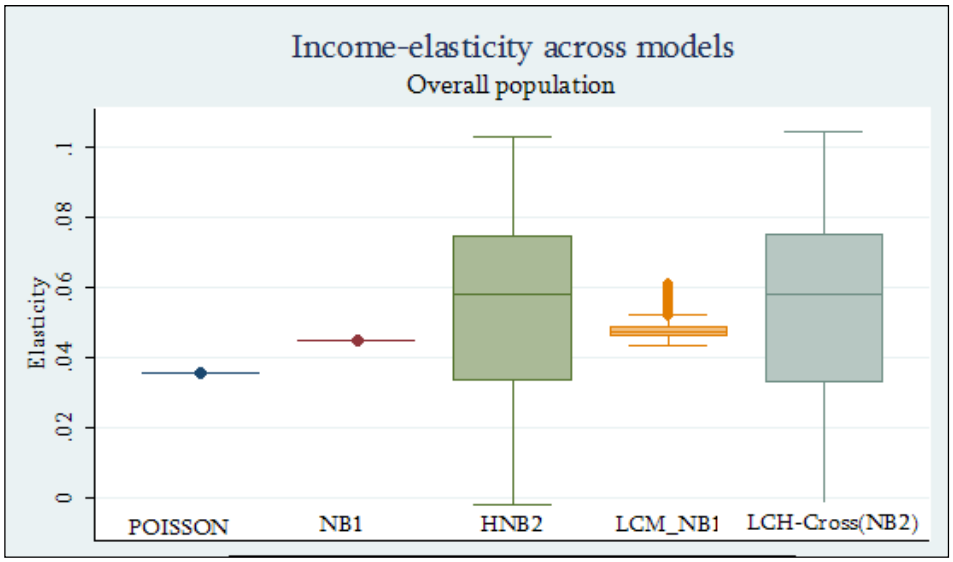

Figure 2 — Box plots of the distribution of the average elasticity effects of 'income' across various specifications[20].

Despite the apparent similarities in the estimates of the population income-elasticities of health care utilization, presented in Table 13, Figure 2 shows that the specifications of the hurdle family, models H_NB2 and LCH-Cross(NB2), estimate individual income-elasticities in a rather wide interval. The estimates range from a minimum of approximately 0 to a maximum of a little more than an income elasticity of 0.1. However, most of the estimates, about 75%, are located between 0,03 and 0,07 points of elasticity. Contrarily, the model LCM_NB1 estimates income-elasticities of utilization in a much narrow interval, ranging from 0,04 to 0,06, with approximately 75% of the estimates nearly collapsing into a single number. In conclusion, the estimate of the income-elasticity of utilization for the overall population seems to be

[20] For the Poisson and NB1 specifications, the functional form of the conditional mean constrains the income-elasticity to be invariant across individuals, thus, the box-plot for these two models collapses into a single point in the graph.

almost invariant across models, nonetheless, the individual response to income varies across individuals for the same model, and across models.

The next three figures depict the distribution of the individual marginal effects for the covariates 'education', 'Rural_area' and 'NHS-only'.

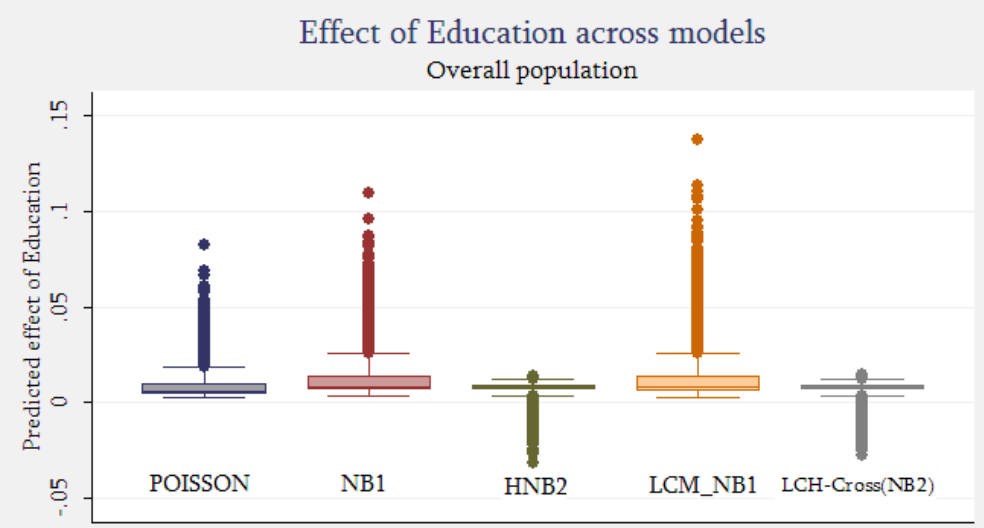

Figure 3 — Box plots of the distribution of the marginal effects of 'education' across various specifications

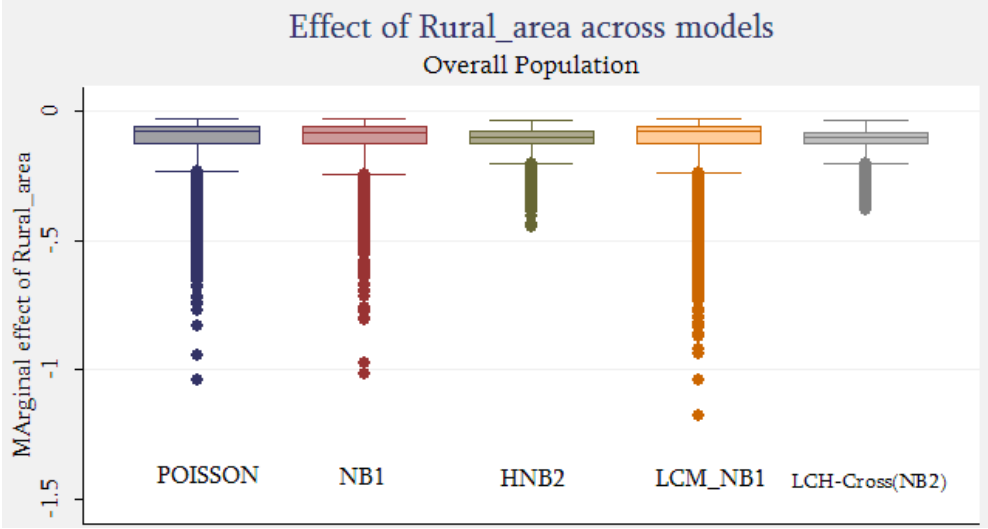

Figure 4 — Box-plots of the distribution of the marginal effects of 'rural_area' across various specifications

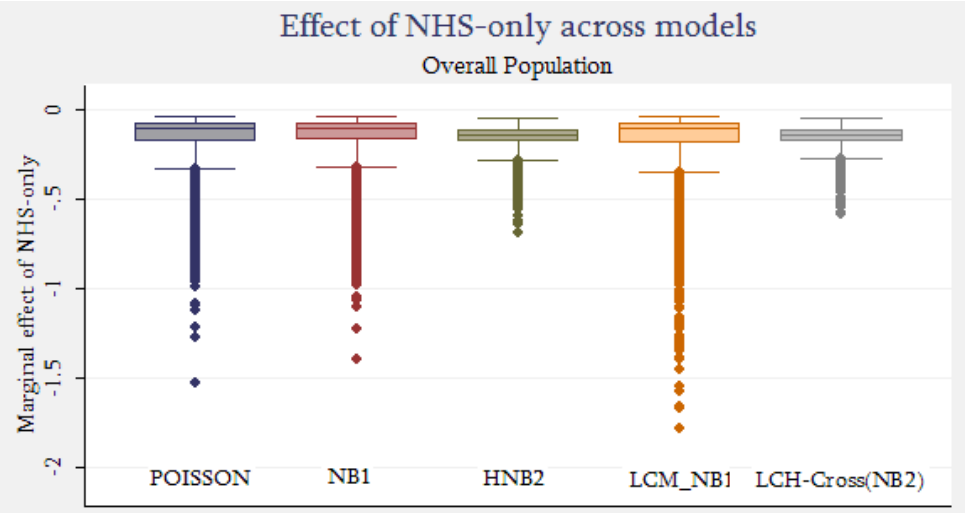

Figure 5 — Box plots of the distribution of the marginal effects of 'NHS-only' across various specifications

After the analysis of the figures presented above, we may conclude that, for each covariate, all competing specifications estimate marginal effects of the same order of magnitude and with a comparable variability.

In conclusion, all empirical evidence presented thus far indicates that whenever the goal of the research is to assess how the conditional mean of the overall population respond to a change in the covariates, then the extra effort to estimate models that are more sophisticated seems to be rather worthless. Apparently, these more sophisticated models do not disclose any new relevant results, when the estimates are computed for the overall population. In this case, a simple specification like Poisson or Negative Binomial will be sufficient to generate information regarding these aspects. More sophistication in the models permit, however, to enhance the data analysis, allowing for example to study the results from a point of view that are inaccessible when one specifies and estimates simple one-index models, like the Poisson and the negative binomial regression models.

We have also found, nonetheless, that the impact of a given covariate varies across individuals, presenting, for some covariates, wide variations. Therefore, it may occur that the different models generate different effects when the researcher's goal is to evaluate the impact of a covariate for a cohort of individuals with special characteristics. This is an important research issue that deserves to be investigated in future developments of this work.

In what follows, we enhance the data analysis by studying some of the characteristics of the distribution of the health care users population, for the family of hurdle models (H_NB2 and LCH-Cross(NB2)). In this context, we aim at comparing the effect of the covariates 'income', 'education', 'rural_area' and 'NHS-only' on the conditional mean function of the users population.

Table 14 shows the estimates of the effect of 'income', 'education', 'Rural_area' and 'NHS-only*'* regressors on the conditional mean function for the health care users population.

Model Covariate	H_NB2	LCH-Cross(NB2)
Income	-0,001	-0,002
(elasticity)	(0,010)	(0,010)
Education	-0,005	-0,004
	(0,005)	(0,005)
Rural_Area	-0,07	-0,07
	(0,038)	(0,038)
NHS-only	-0,103	-0,101
	(0,070)	(0,070)

Table 14 — Estimate of the impact, using AME, on the conditional mean function for the health care users population

After analysing the figures presented in the previous table, we conclude that average marginal effects estimated by the two models are very similar both in sign and in magnitude. Moreover, none of the estimates presents statistical significance at the standard levels. Accordingly, in the health care users population, factors like income, education, living place and insurance status do not influence the individual choices regarding the utilization of healthcare. That is, after crossing the hurdle, the utilization of health care does not react to changes in these covariates. This is the case for the average marginal effect for the population of health care users, however, the effect of the covariates varies across individuals. Consequently, the marginal effect for individuals with specific characteristics may be statistically significant and be substantially different from the average marginal effects presented in Table 14.

Figure 6 (below) compares the distribution of individual income-elasticity of utilization of the models H_NB2 and LCH-Cross(NB2) for the health care users population, while Figure 7 compares the distribution of individual marginal effects of the covariate 'NHS-only', for the same two models.

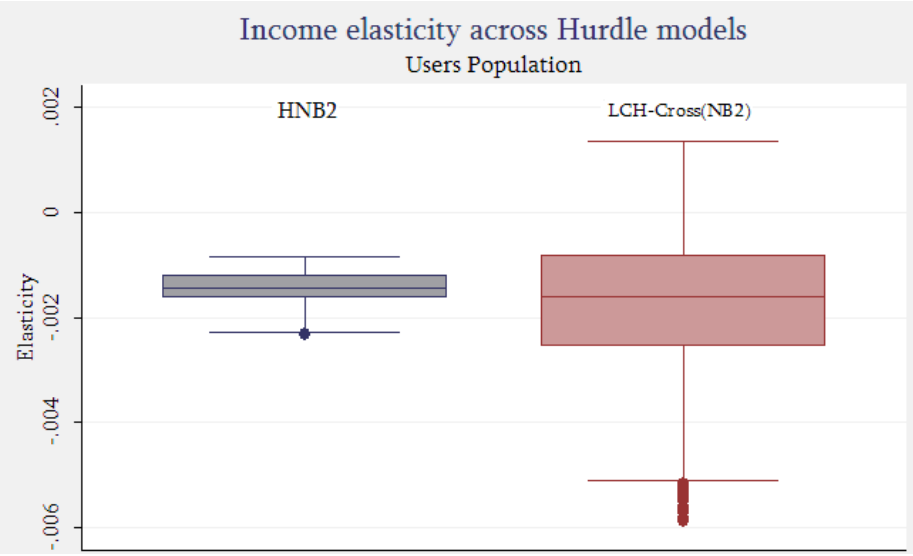

Figure 6 —Box plots of the distribution of the income elasticities for the health care users population of models HNB2 and LCH-Cross(NB2)

The figure above shows that the income-elasticity of utilization presents some degree of variation across individuals, however this variability is higher for the model LCH-Cross(NB2). However, for both models, none the individual based income-elasticity presents statistical significance, meaning that for the users population, income is not a factor that influences individual choices regarding health care utilization.

Concerning the marginal effects of the 'NHS-only' covariate, the conclusions are completely in line with the conclusions about the effect of income, presented above.

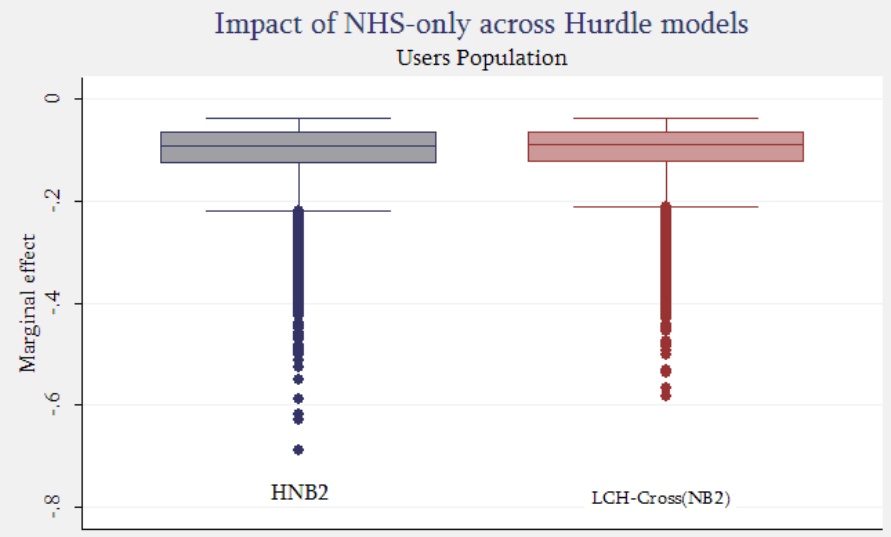

Figure 7 — Box plots of the distribution of the marginal effect of 'NHS-only' regressor for the models HNB2 and LCH-Cross(NB2).

Taking together all the findings regarding the purpose of this section, addressed in this , we may summarize that the calculations based on all models only evidence small difference on the AME estimates, possibly statistical similar. Therefore, our findings indicate that the different models estimated do not unveil different evidence about the effect of the covariates on the conditional mean function.

In our view, what out findings suggest is that the advantage of using structural models, like models of the hurdle or LCM family, is that they enhance the analysis,

allowing extracting more knowledge from the data, reaching populations that usually are seen as aggregated by the simpler models.

Moreover, most of the times, among the analyst's aims are the study of the effect of the covariates on model descriptors, other than the mean. For instance, the goal can be to study the probability of having at least one visit to the doctor, or the probability that the utilization exceeds a given value. We cannot exclude the possibility that, in these new contexts of analysis, it could be the case that different models generate different conclusions. Yet, we postpone the investigation of this possibility for future work.

4.6 — General Conclusions

The main goal of the investigation conducted in the chapter that now end was to contrast different econometric approaches to explain medical care utilization, when utilization is measured as a count variable. We compared the econometric specifications that have been most used to model count data (e.g. negative binomial regression model, 'popular' hurdle regression model and latent class regression model) with an alternative specification for the hurdle model. The alternative hurdle formulation suggested departs from the popular hurdle in the way it models the individuals with a positive number of visits, where we proposed a specification based on latent class models explicitly assuming that the unobservables are specific to the population of health care users. Adopting models from the LCM family in the second stage has the advantage of being a semi-parametric specification that frees the model from possibly strong, and rather arbitrary, distributional assumptions about individual heterogeneity. Moreover, the hurdle model proposed in this chapter relaxes the implicit assumption, on which the popular hurdle model is based, that a single stochastic process governs the choices of the users of medical care. This is tantamount to assume the existence of a single health care consumption group. On the contrary,

our proposal, assuming two consumption groups of users, has the potential to capture different responses to changes in a covariate in the right tail of the distribution.

We have evaluated the empirical performance of the alternative LCM based hurdle model, referred throughout this chapter to as LCH-Cross, against the standard count data specifications using data taken from the Portuguese National Health Survey. We adopted a battery of evaluation strategies to compare the competing models and all converged into the same conclusion: the hurdle formulation based on the LCM framework to model the positives does indeed fit the data better than all other competing count data specifications analysed.

The ultimate purpose of this type of empirical studies, and the current one is not the exception, is producing knowledge relevant for health policy making. However, some relevant policy measures may depend on the empirical specification adopted thus, thus we estimated the effect of income, education, place of residence and health insurance status on the conditional mean function for a number of competing models estimated. Regarding the subject, the calculations based on all models generally indicate small differences, possibly without statistical significant, on the estimates of the average marginal effects. Based on the findings we have defended that when the analyst's objective is to study the mean function that characterizes the overall population, the advantage of using structural models, like standard hurdle, standard LCM or the LCH-Cross, is that they enhance the analysis, thus the findings, allowing the researcher extracting more information from the data.

This study while providing answers to selected questions, raises new issues and challenges that may lead to new inquiries. The final remarks of this chapter briefly identify some of those issues, thereby pointing to what may be futures areas of investigation.

First, the comparison between the competing specifications has been made on the grounds of statistical performance. Of course, economic implications and underlying

assumptions concerning the data generating process are also relevant for the choice of the appropriate model.

Second, another issue is the replication of results on similar datasets, required to establish the robustness of the alternative approach proposed. Would the LCH-Cross continue to outperform the competitors when applied to similar datasets?

Third, equally interesting is the evaluation of the impact of the survey design on model performance. In particular, we may ask whether the semi-parametric hurdle model outperforms the standard count data models when medical care utilization is measured during a period of one year, leading to an increase on the probability of violating the single illness spell assumption that, as is known, underpinned the hurdle model. In this context, it would be important to test the assumption of the single illness spell (Santos-Silva and Windmeijer 2001).

Forth, we may ask how does our model compares to specifications that account for the unobservables in the second stage of the hurdle using continuous mixtures, specified in the spirit of Winkelmann (2004) and Van Ourti.

Fifth, a question investigated in this chapter, although superficially, was the effect of non-response on SAH (which causes dropping of many cases) on obtaining a final random sample. We found, unfortunately based on poor instrumental variables, that the inclusion of SAH in the regression models would render sample selection issues, therefore, we excluded it from our regressors, in the hope that this procedure did not bias our findings. Thus, this issue provides opportunities for further methodological work that revolves around the covariate SAH, one of the most often used variable to reflect the individual health status of the individual. For instance, it would be relevant to test for sample selection bias using other, with more quality, instrumental variables as well as using other datasets.

If the sample selection bias is confirmed, it is necessary further work to specify count data models with sample selection. An additional issue around the SAH variable is that it may be endogenous when used as a covariate in health care utilization models

(Windmeijer and Silva 1997; Van Ourti 2004; Barros, Machado *et al.* 2005), being this another relevant area for future research. To estimate the impact of the endogeneity of SAH on relevant estimates for policy making.

Chapter 5
The Utilization of Medical Care in Portugal
Effect of supplementary health insurance and place of residence

5.1 — Introduction

In the Portuguese health care system, the achievement of an equitable delivery of medical care services ranks high among its objectives (Pereira 1990; Dixon and Mossialos 2000). The principle of equity underlined here is the horizontal one, stating that people in equal need for medical care should be treated equally, irrespective of characteristics such as income, health insurance status, place of residence, among others[21] (Wagstaff and van Doorslaer 2000).

Despite the health system to be based on a National Health Service, providing universal coverage for medical care, nearly free at the point of delivery, there is scope for horizontal inequity, for instance, related to income, health insurance status, place of residence, among other factors[22]. This is the case mainly due to the actual

---

[21] It should be noticed that it is beyond the scope of this thesis a critical appraisal about the theoretical foundations of the principle of equity

[22] Notice that a number of authors have addressed the issue of evaluating income-related iniquity in the Portuguese system (van Doorslaer and Jones, 2004, van Doorslaer and Koolman, 2004, van Doorslaer et al., 2004, Jimenez-Martin et al., 2002). Therefore, in the current chapter, we relegate the issue of income-related equity to a second plan, concentrating, instead, on the implications of health insurance status and place of residence.

Moreover, it is not among our purposes a discussion about the issues related with the empirical verification of whether equity in medical care delivery has been achieved. A number of authors discussed those aspects (Wagstaff and van Doorslaer, 2000, Culyer et al., 1992, Wagstaff et al., 1991, Mooney et al., 1991; Gerdtham, 1997; Van Ourti, 2004). In this chapter

limitations, principally in the access to doctor visits, of the coverage provided by the statutory public health service.

On the one hand, those people covered solely by the NHS (referred to as 'NHS-only' individuals), relative to those benefiting, in addition, from the coverage of a health subsystem, have lower access to doctor consultations. This, obvious, lower access may undermine the equity principle discussed above. On the other, the place of residence — principally on the urban *vs* rural categorization — may also threat the principle of horizontal equity owing to the disparities in the distribution of medical care resources between the different types of regions. The concentration of resources in urban regions in detriment of rural regions may encourage different patterns of medical care utilization by the individuals residing in the different regions, regardless of medical care need.

The purposes of the current chapter can be summarized as follows;

- to evaluate the effect health insurance status and of the place of residence (urban vs rural), on the utilization of medical care services in Portugal. The impact of such factors by latent class will be emphasized in our analysis,
- to unveil latent classes of health care users, and analyse them both in terms of expected health care utilization and in terms of the characteristics of the individuals that each latent class encompasses,
- to estimate the impact of other factors, e.g., income, health status and so on, on the utilization of medical care services. Again, we give special interest to the effects by latent class.

---

we rely on the regression methodology, which simply amounts to investigate if, conditional on medical care need, the socioeconomic covariates influence medical care utilization (Wagstaff et al., 1991, Wagstaff and van Doorslaer, 2000, Gerdtham, 1997, Bago d'Uva, 2005; van Doorslaer et al., 2004).

In order to fulfil those objectives we rely upon the analysis of the parameters of the preferred count data model of the previous chapter, the model LCH-Cross(NB2)[23]. As explained in the last chapter, the model LCH-Cross is a specification of the hurdle family, consisting on a binary model to find the determinants of some utilization and, given some utilization, the model analyses the positive counts adopting a two point of support LCM that assumes the unobserved heterogeneity on the population of health care users. The data was taken from the National Health Survey (1998/99)[24], and we measured the utilization of medical care services as the total number of doctor visits in a three-month period, encompassing, thus, all types of ambulatory visits, to all sort of providers.

This study can contribute to the literature on several counts:

First, the econometric specification adopted to respond to our research questions, although not completely innovative in the literature, as explained in the previous chapter, departs from the methodologies generally adopted to address similar issues. Actually, to the extent of our knowledge, this study is the first to use this type of specification to address empirically this theme. Some work has been conducted to split the overall population into latent classes, however, this is the first study that analyses the effect of several covariates by the latent classes that the health care users population might encompass.

Second, we defended in Chapter 3 that the health insurance plans provided by the health subsystems could be seen as a form of supplementary health insurance (henceforth referred to as SHI). In consequence, our first research question amounts at

---

[23] Henceforward, we will refer to this model simply as LCM-Cross, avoiding the utilization of NB2.

[24] A detailed explanation of this survey was presented at Section 4.4, page 98.

investigating the impact of supplementary health insurance[25] on the use of doctor consultations. Many authors have been investigating the subject for long time (Manning, Newhouse *et al.* 1987; Cameron, Trivedi *et al.* 1988; Chiappori, Durand *et al.* 1998; Holly, Gardiol *et al.* 1998; Vera-Hernandez 1999; Harmon and Nolan 2001; Schellhorn 2001; Buchmueller, Couffinhal *et al.* 2004; Jones, Koolman *et al.* 2004)[26], in, an attempt to produce evidence on the magnitude of the individual's reaction to different health insurance plans characterized by different levels of coverage. Almost invariably, the preferred methodology was the regression analysis. In all those studies, however, there is a non-neglected methodological challenge that must be resolved. Usually, the data for these studies come from non-experimental settings, meaning that, most likely, the decision to buy a (supplementary) health insurance plan is not random, being possibly dependent on (unobservable) individual characteristics, such as the individual's level of risk aversion, preferences, etc, giving rise to adverse selection problems. For example, those individuals that, based on private information, anticipate greater expected medical care use, are also those who are more likely to buy extra health insurance (Newhouse, Phelps *et al.* 1980; Cameron, Trivedi *et al.* 1988; Chiappori, Durand *et al.* 1998; Vera-Hernandez 1999; Buchmueller, Couffinhal *et al.* 2004). When these selection problems are disregarded in the specification of the regression models, the insurance parameter measures two untangled effects. On the one hand, the actual influence of SHI (moral hazard), on the other, the adverse selection effect. Thus, if one ignores the selection effect, then the impact of health insurance will be biased upwards, leading to an overestimation of the impact of

---

[25] Alternatively, the terms 'duplicate coverage' (Vera-Hernandez, 2000) or 'additional health insurance' (Barros *et al.*, 2005) are also used.

[26] To be accurate, some of the referred studies have addressed the impact of co-insurance on the utilization of medical care.

additional health insurance on the utilization of medical care (Chiappori, Durand *et al.* 1998; Buchmueller, Couffinhal *et al.* 2004; Barros, Machado *et al.* 2005). Vera-Hernandez (1999), Holly et al. (1998) and (Deb and Trivedi 2006) suggested different methodologies to handle the endogeneity issues in this context[27]. Our empirical application is, nonetheless, free from this type of methodological challenge because, as argued in Chapter 3, in the Portuguese health system, the supplementary health insurance is provided through the workplace, meaning that they are unrelated to the expected future health care utilization. Underpinning this claim is the assumption, acceptable we believe, that the choice of the occupation was not made due to the health insurance it provides. Thus, we can take the covariates reflecting SHI as exogenous, which permits us to estimate with precision, and rely on ML estimation methods, the actual impact of SHI without the running the risk of capturing confounding adverse selection effects.

The third area where our study can contribute is on the estimation of the impact of residing in a rural area on the medical care utilization. The issue is, nonetheless, difficult to deal with owing to the quality of the data reflecting the living place. Usually, data on regional information represents some broad regional division of the country, large areas of the territory, which includes both urban and rural areas, therefore there is no clear distinction among individuals living in rural and urban areas. A clear disadvantage of using an ample geographical classification is that in the event of finding differences on medical care utilization, it is difficult to justify such findings. As an example, Oliveira (2004) reported the existence of geographic variations in hospital utilization between five (wide-area) regions in Portugal. The author struggled to justify the finding acknowledging that it was not obvious why

[27] Regarding the treatment of endogeneous covariates in count data models, it is worth to refer to the work of Windmeijer and Santos-Silva (1997).

northern and centre regions have a lower propensity to use hospital care. Therefore, in order to study the effect of place of residence on the utilization of medical care services one has to rely on data that classifies the living place of each respondent either as urban or rural. In this regards, we used a classification of the Portuguese National Institute of Statistics of each "freguesia"[28] on 'predominantly rural', 'medium rural' and 'predominantly urban'. After merging these data with our dataset, we end-up with information classifying the living place of each individual either as rural or urban.

The fourth perspective from which this chapter may be considered relevant is that it will reveal the contribution of a number of other factors to the utilization of medical care. It turns out that unveiling such determinants can be significant for health policy makers because with this information they will understand, for example, why the medical care utilization is potentially different, even when people have the same health status, and thus the same need for medical care.

Fifth, it can also reveal relevant information to the administrators of the health subsystems. Among the health subsystems, we consider two types of health insurance plans: on the one hand, we consider the health insurance plans provided to the civil servants (referred to as ADSE), on the other, the health insurance schemes provided by all other health subsystems (referred to as OSS group). Therefore, the results reveal the level of medical care utilization of one type of health insurance relative to the other, and the differences can be related, possibly, to moral hazard, and ultimately, the

---

[28] For some purposes, the Portuguese territory is divided in three nested levels: 'Distritos', 'Concelhos' and 'Freguesias', with 'Concelhos' nested in 'Distritos' and 'Freguesias' nested in 'Concelhos'. Moreover, in 1999 the Portuguese mainland territory was divided in approximately 4030 'Freguesias', with an average population of 2.445 inhabitants, a maximum of 81.800 and a minimum of 39 inhabitants.

findings can signal the need for adjustments on the characteristics of the contracts provided by the subsystems.

Finally, the concluding contribution of this chapter is closely connected to the specification chosen to analyse our data. We have used, as explained in the previous chapter, a model from the LCM family to understand the medical care utilization of those individuals with a positive utilization. Surely, this modelling strategy qualifies us to identify latent classes of health care users, in addition to their characterization, both in terms of expected health care utilization and in terms of (observed) characteristics of the individuals comprised in each latent class. This can be significant for policy making because it may permit the identification of target populations for which particular policy measures may be designed.

This chapter contains 5 sections. In Section 2 we present a short review of the literature, focusing on the findings relative to the effect of supplementary health insurance and place of residence. Next, in section 3 we remind how we measured health care utilization, present the list of covariates used in the regression model, along with the rationale to use them. Still in the third section, we present a preliminary analysis of the data. Following the data section, section 4 presents and discusses the findings of the empirical analysis. Finally, some conclusions are drawn in Section 5.

5.2 — A personal roadmap to the most relevant literature

This section aims at presenting the main findings, reported in the literature, relative to our two primary research questions: the effect of SHI and the effect of place of residence on the utilization of medical care.

In regards to the first question, a large number of studies have investigated the theme, yet examining the effect of voluntary health insurance, therefore, in contexts where the confounding effects related to adverse selection may arise. Unfortunately, the second subject, although relevant from a health policy point of view, has received

much less attention on the part of the researchers, thus, the existing literature addressing the theme is somewhat scarce. In substitution, however, we seek for studies reporting evidence about the effect of population density on the medical care utilization. The justification for such procedure is the assumption that the less populated areas correspond to rural settings. In fact, we note that in our application, the variable indicating residence in a rural area is defined using as main criterion the population density on a given area.

The criterion to select the studies was to search for researches conducted in health system contexts similar to ours; therefore, we leave out of consideration most of the evidence provided on the subject that comes from the United States, thus, excluding from our review the important evidence offered by the Rand Health Insurance Experiment (RHIE) (Manning, Newhouse *et al.* 1987).

We begin by noticing that the research from several European countries have found a positive effect of SHI on the health care utilization, mainly on outpatient care (Pohlmeier and Ulrich 1995; Deb and Trivedi 1997; Holly, Gardiol *et al.* 1998; Vera-Hernandez 1999; Harmon and Nolan 2001; Schellhorn 2001; Buchmueller, Couffinhal *et al.* 2004; Jones, Koolman *et al.* 2004; Sarma and Simpson 2006)[29].

We onset by citing the research by Pohlmeier and Ulrich (1995), that reported evidence on the effect of voluntary private health insurance on the utilization of general practitioner visits and specialist visits, analysed individually, for a sample of employed individuals in German. The authors specified and estimated a NB hurdle regression model in a count data context. In what concerns the findings, they reported

---

[29] In fact, the study by Sarma and Simpson (2006) reports the impact of supplemental health insurance for Canada, and the one by Deb and Trivedi (1997) for the United States. Our option for citing the results of those studies, offering evidence for countries outside Europe, is because the methodologies adopted resort on models of the LCM family.

that, for both measures of medical care utilization, the coefficient of the dummy variable reflecting private health insurance is only significant at the first stage of the hurdle model, with a negative coefficient for the case of GP visits and a positive one for the case of specialists visits. Regarding the difficulties due to possible selection problems, the authors pointed out that they do not seem relevant for their model.

Another study motivated by the topic is the one conducted by Holly *et al.* (1998) who provided results on the impact of (voluntary) SHI on the utilization of health services in Switzerland. In order to account for the endogenous nature of the SHI covariates, the authors specified and estimated a simultaneous two equations model to explain the propensity to have at least one inpatient stay, given that the insured has used some medical treatment. The findings reported suggest that the effect of SHI is to increase the likelihood of a patient to have at least one inpatient stay.

Using other type of medical care utilization indicator and other methodology to deal with the endogeneous nature of the SHI covariates, Vera-Hernandez (1999), analysed the effect of duplicate coverage on the total number visits to a specialist doctor, for Spain. To deal with the endogeneity of the health insurance variables, the author estimated the equation of the number of visits using the GMM technique. Regarding the findings, the author reported that benefiting from duplicate health insurance coverage implies an increase, on average, of about 27% in the number of visits to specialist physicians.

Relative to French data, Buchmueller *et al.* (2004) estimated a model to analyse the determinants of the probability of having at least one doctor visit as a function of supplemental health insurance plus other control variables. The authors also estimated models to explain the doctor utilization, conditional on some utilization. Their evidence points toward a positive and significant effect in what concerns the effect of supplemental health insurance on the probability of having at least one doctor visit. In opposition, conditional on having at least one visit, the effect of insurance on the

number of visits was found to be irrelevant, despite the warning about possible caveats[30] issued by the authors.

Finally, Sarma and Simpson (2006) offered evidence on the effects of supplementary health insurance for the Canadian health care system. The authors analysed, among other measures, the total number of doctor visits (in one-year period) as a function of several covariates, SHI covariates included. A wide array of alternative count data models were specified and estimated, yet, only the estimates of the preferred model — two-support point LCM, NB2 as parent distribution — were presented. The evidence offered points towards a positive and statistically significant effect of supplemental insurance on medical care utilization among the latent class of low users, and a non-significant effect for the latent class of high users. Two alternative explanations are suggested; either it results from the impact of supplemental health insurance on the utilization for preventive care or it is simply the result of moral hazard.

Bringing together all the empirical evidence just presented, all seems to make the case that those individuals who enjoy from SHI tend to increase their medical care utilization. As explained above, all the investigations presented thus far refers, mostly, to European countries, however, more relevant to our work is the evidence for Portugal. Accordingly, in what follows we review some of the most relevant evidence for Portugal.

A number of authors developed work with the purpose of studying the impact of supplemental health insurance on the utilization of medical care services by the

[30] The period over which utilization is measured is too short (one month), therefore, the number of individuals with multiple visits is very small.

Portuguese population (Barros 1999; Lourenço 2003; Jones, Koolman *et al.* 2004; Pereira 2004; Barros, Machado *et al.* 2005)[31].

Jones *at al.* (2004) evaluated the effect of (voluntary) SHI on the probability of visit a specialist physician at least once during the last year. To deal with the potential endogeneity of the dummy covariate used to reflect SHI the authors relied upon a number of strategies. Regarding the results, that it is what we are interested in, the study found a positive effect of SHI on the probability of using visits to specialized doctors. Moreover, the authors argued that the results suggest that the increased probability of utilization is not due to selection effects.

The remaining authors cited above evaluated the effect of SHI, however, that provided by the health subsystems, therefore, with exogenous covariates reflecting SHI.

To begin with, Pereira (2004) used data taken from the National Health Survey, the version conducted in 1995/96, to evaluate the effect of SHI provided by the health subsystems[32] on the decision about the type of provider to consult. Their set of alternatives were visit a specialist, visit a GP or do not go to the doctor. Regarding the estimation strategy, the adoption of an ordered logit model was justified on the grounds that a propensity latent variable governed the decisions, with the choice indexed, in some way, to the propensity variable. The main and most relevant (to us) conclusion of that study is that SHI is linked to a higher propensity to use medical care.

---

[31] The study by Jones et al. (2004) evaluated the effect of (voluntary) SHI for a number of countries, not only for Portugal: Set aside Portugal, the research has included Ireland, Italy, Spain and UK.

[32] In their study, the author, contrarily to what we have done, aggregated all health subsystems on a sole group. Thus, their covariate that measures SHI encompasses the fund of the civil servants (ADSE) in addition to all funds provided by the other health subsystems.

Lourenço (2003), on the other hand, specified a bivariate probit model to estimate the effect of health subsystem membership[33] (SHI) on the choice of the type of provider, categorized as *GP vs Specialist / Public vs Private*. The data to estimate the model was taken from the National Health Survey (1998/99), and the target population were those patients who had at least one visit to the doctor in the last three months. Concerning the results, the most relevant reported is that those NHS-only individuals presented both, a lower probability of visiting a specialist physician and a lower probability of visiting a private doctor. Moreover, comparing the behaviour of those covered by the ADSE health insurance scheme with those covered by a plan provided by other health subsystem (the OSS group), it was found that those individuals covered by the ADSE present a lower probability of visiting a specialist physician.

To sum up, the evidence presented thus far for Portugal, shows that the health insurance status is an important determinant on the choice to see the doctor, and that, once that decision has been made, differences on the health insurance status triggers different preferences about the type of doctor to visit.

All studies cited thus far regarding the Portuguese reality have evaluated the effect of SHI on the decision to visit the doctor, or, given that such decision has been made, evaluated the impact of SHI on the choice of a type of provider. Nevertheless, more interesting to the ongoing work in this thesis is the evaluation of the effect of SHI on the total utilization of medical care services.

The first author to address the research question at issue here, using health care utilization measured as the total number of physician visits was Barros (1999). The author adopted a Zero Inflated Poisson (ZIP) model to explain the number of visits in a three month period as a function SHI (provided by a health subsystem), in addition to

---

[33] In this study, the plans provided by the civil servants (ADSE) and by the remaining health subsystems were considered separated.

many other medical care determinants. It was found that the overall effect of SHI on the total number of visits was not statistically significant, thus suggesting that coverage by a health subsystem does not render, on average, an increase on the number of visits. The same author, along with colleagues (Barros, Machado *et al.* 2005), conducted a study with similar purposes, however with a different methodology — a matching estimator approach — redefining, however, the supplementary health insurance variable. On that investigation, SHI was defined as being a beneficiary of the civil servants fund (ADSE)[34]. Concerning the findings reported, they were in line to those reported in the previous study, which is: being a beneficiary of a SHI contract does not alter significantly the individual behaviour regarding the total number of doctor consultations consumed.

Thus far, we shortly reviewed the literature offering evidence about the impact of SHI on medical care utilization, however, among our purposes in this chapter it is too the evaluation of whether those individuals living in rural areas behave differently relative to the urban cohorts, regarding the utilization of doctor visits. There is, nonetheless, a scarcity of studies with such main purposes, thus, alternatively, we report the findings about the impact of population density on the utilization of health care.

In general, all studies found that those individuals living in rural areas, on average, tend to use less frequently medical care services (Pohlmeier and Ulrich 1995; Gerdtham 1997; Zhang, Tao *et al.* 2000; Gerdtham and Trivedi 2001; Yawn, Mainous *et al.* 2001; Hauenstein, Petterson *et al.* 2006; Sarma and Simpson 2006).

---

[34] The individuals with membership to the other health subsystem (those of the OSS group) were deleted. In addition, all individuals that have reported to own a voluntary health insurance contract were also deleted.

Beginning with Pohlmeier and Ulrich (1995), the authors offered evidence on the effect of population density on the utilization of GP and specialist visits, for a sample of employed individuals in German, using a hurdle model. Regarding the evidence offered by the study: living in a less populated area is associated with a higher probability of visiting a GP and with a lower probability of visiting a specialist physician. On the other hand, regarding the impact of the covariate on the second stage of the hurdle model, for the case of GP demand, does not showed any impact, on the contrary, for the utilization of specialist visits, those living in less populated areas tend to use less consultations.

Another study that we can refer is that conducted by Gerdtham (1997). In line with Pohlmeier and Ulrich's choice regarding the model specification, Gerdtham estimated a NB hurdle model to explain the number of doctor visits in a twelve month period for a sample of adults. The study found that residing in a big city do not render any effect on the probability of seeking medical care, however, for those who sought medical care, those residing in a big city present a tendency to increase the frequency of utilization.

On the other hand, Gerdtham and Trivedi (2001) adopted a two support point LCM to study the frequency of doctor visits. The effect of living in a big city was found to be associated with a lower utilization on the latent class of infrequent users and with no effect on the frequent users latent class.

The last study that we refer is that conducted by Sarma and Simpson (2006), who estimated a two support point LCM to explain the frequency of doctor visits. Among the covariates, three of them are included to reflect the place of residence, 'rural_area', 'urban area' and 'metropolitan area'. The findings reveal that residing in rural areas is associated with a lower utilization of doctor visits, being this true in the two latent classes implied by the model.

In sum, all evidence presented suggests that the individual health care utilization is responsive to supplementary health insurance and to the place of residence.

We continue this chapter by presenting some details about dependent variable and the regressors included in the models. Moreover, we also conduct a preliminary analysis of the data.

5.3 — Dependent variables, covariates and preliminary analysis of the data

As referred in the introduction of the current chapter, its purpose is to analyse the results of the LCH-Cross model in order to respond to some research questions relevant from a health economics perspective. As in any empirical application, one essential component in the model estimated in Chapter 4 was the dataset, nonetheless, there, we omitted some information about that issue that can be important at this point. Thus, this section aims at filling the gaps about the data and variables not covered in the previous chapter.

The current section begins by briefly recalling the variable selected as indicator of medical care utilization, followed the revision of the complete list of covariates included in the regression model. Next, we use arguments, either borrowed from the Grossman's model or based on intuition (which are admittedly ad-hoc), to justify the choice of the covariates. After presenting such arguments, we continue by conducting a preliminary and exploratory analysis of the data.

5.3.1 — Health care utilization measure and covariates

In Section 4.4 — Data and Variables, page 98 — we presented the variable chosen to reflect utilization of medical care: it was the total number of consultations made to licensed doctors in a period of three months. The interviewer asked each person how many times he/she visited a physician (any type of physician) in the last three months.

Hence, the answer includes all types of doctor consultations; primary health care doctors in the public and in the private sector, consultations to specialized physicians in public hospitals and in private clinics, consultations to dentists, consultations to hospital emergency rooms, etc. The design of the survey in this area makes it impossible to disentangle the doctor visits by type of visit, therefore, we are constrained to analyse the total number of visits.

The empirical distribution of the number of visits (henceforth referred to as 'visits') variable was presented in Table 2, page 101. There, we concluded that approximately 46% percent of the individuals reported zero doctor consultations, in a period of three months, and that the average number of consultations was 1,29. Moreover, the average number of consultations of those that have at least one visit is found to be 2,39 visits. Figure 8 presents a bar chart of the empirical distribution of the dependent variable 'visits'.

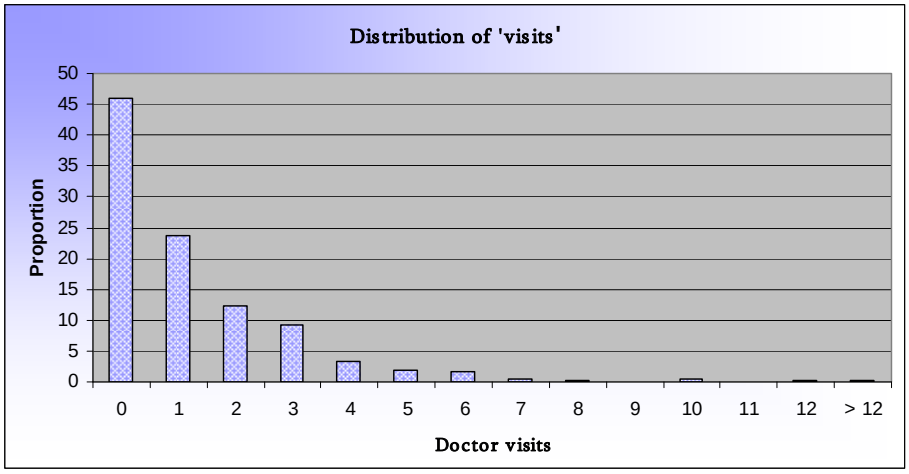

Figure 8 — Empirical distribution of the dependent variable 'visits'

Turning our focus to the covariates included in the model as determinants of medical care utilization, Table 3, page 104 lists all the covariates included. As mentioned there, the 34 explanatory variables selected were clustered into five groups, encompassing socioeconomic and demographic covariates, covariates that reflect the

individual's health status and a supply side determinant. In addition, we have also considered a group of covariates to reflect supplementary health insurance and, finally, a group of covariates to control for seasonality effects.

To facilitate the reading of the remaining of this chapter we reproduce Table 3 below, which, as it was shown in Chapter 4, includes the list of the covariates considered as determinants of medical care.

Variable	Variable Definition
Socioeconomic demographic	
Age [/10]	Age in years, divided by 10
sqAge	Square of age[/10]
Female	= 1 if the individual is female
Married	= 1 if the individual is married
Education	Number of years of schooling. In the case of child, the variables measures the number of years of schooling of most educated adult living in the household
Not_work	= 1 if the individual did not work in the two weeks previous to the application of the survey
Retired	= 1 if the individual is retired
Unemployed	= 1 if the individual is unemployed
(log) Income	Logarithm of equivalised monthly real income in hundreds of Euros
North	= 1 if the individual resides in the north region
Centre	= 1 if the individual resides in the centre region
LTV	= 1 if the individual resides in the Lisbon and Tagus Valley region
Alentejo	= 1 if the individual resides in the Alentejo region
Rural_Area	= 1 if the individual resides in a rural area
Health Status	
Diabetes	= 1 if the individual has diabetes
Insulin	= 1 if the individual is insulin dependent
EBP	= 1 if the individual has elevated blood pressure
Asthma	= 1 if the individual has asthma
Bronchitis	= 1 if the individual has bronchitis
Allergy	= 1 if the individual has an allergy
Back pain	= 1 if the individual has back pain

Variable	Variable Definition
Ill_long_Run	= 1 if the individual has an illness for more than 3 months
Ill_Short_run	= 1 if the individual reported an illness in the two weeks prior to the application of the survey
Limited	= 1 if the individual has some sort of physical handicap that impedes him to execute certain basic physical daily activities
Stress	= 1 if the individual has been taking sleeping pills in the last two weeks
NeverSmoked	= 1 if the individual never smoked during her/his lifetime
not_physical_activity	= 1 if the individual's daily activities do not require physical activity
mild_exercise	= 1 if the individual engages in mild sports activities at least four hours a week.
Supply side	
Phy_1000_residents	Total number of licensed physicians per 1000 inhabitants
Health Insurance Status	
NHS-only	= 1 if the individual is covered only through the NHS
ADSE	= 1 if the individual is covered by the civil servants health insurance scheme
Seasonality	
Winter	= 1 if the period of observation was in the Winter
Spring	= 1 if the period of observation was in the Spring
Summer	= 1 if the period of observation was in the Summer

Definition of covariates (table 3 repeated)

We continue this section by briefly providing the justification and intuition about why these variables are thought to influence the individuals regarding the number of doctor visits. As claimed above, we base our reasoning primarily on models developed in the spirit of the integrated framework, developed by Grossman, to analyse the demand for health an medical care services (Grossman 1972; Wagstaff 1986). In addition, the selected covariates have been shown, by the related empirical literature, to be important determinants of medical care utilization. Details about the generation of the covariates from the raw data were the subjected of Section 4.4.

The first factor that appears as a socio-economic determinant of health care utilization is age (included in the model in tens of years). We also include the square of

age as a regressor. This covariate is thought to capture the depreciation of health capital that, according to Grossman, increases with age. The Grossman's demand for health and health care model predicts that if health depreciates at higher rates as the individual gets older, then the optimal stock of health decreases as the individual gets older, which in turn affect the utilization of medical care, predicting that older people consume more medical care (Grossman 1972; Jones, Rice *et al.* 2006). In addition, age may also influence the individual preferences as well as the efficiency with which individuals produce health (Winkelmann 2004). The inclusion of the age in regression models are sometimes taken as an extra means to control for the imperfect measurement of health status, signifying that age may be considered as a health status proxy (Bago d'Uva 2005).

Gender ('female') and marital status ('married') are two additional factors included that may influence the rate at which the health stock depreciates as well as the efficiency of the health production function (Wagstaff 1986). On one hand, in the case of gender, the differences may be due to biological, life style as well as attitudes towards (health) risk differences (Mocan, Tekin *et al.* 2004). And on the other, in the case of marital status, the differences may be also due to different life styles, attitudes towards health risks and possibly due to the better level of information that married people may have. For example, married individuals may benefit from the spouse's advice regarding the best way to combine resources to produce health. The differences induced in the health depreciation rate and in the health production function by these factors may affect the utilization of doctor consultations.

The next factor included as covariate in the socioeconomic and demographic group is the level of education of the individual ('education' covariate) that is defined as the total number of years in school completed with success. In the case of minors (individuals that have less than 14 years old), however, the covariate is redefined and measured as the maximum number of years of school among the adults living in the child's household. We have decided to adopt this definition as we consider that, in the

case of minors, what influences the decision about the number of visits is the education of the decision maker in the household, which, in the case of children, is clearly the most educated adult that resides in the household, usually the parents. The effect of education on the utilization of doctor visits can be viewed as a by-product of its effect on the health production function. The argument to use such covariate is that more educated individuals are more efficient health producers, therefore, they tend to use less medical care services (Jones, Rice *et al.* 2006).

Next, we include as medical care determinants three dummy variables ('Not_work', 'Retired' and 'Unemployed') to capture the occupation status of the person. Clearly, the individuals that do not work present lower opportunity costs (both time and monetary related) when visiting the doctor, relative to the individuals that have a job. In accordance, we anticipate that these covariates tend to increase the utilization of doctor visits. Moreover, unemployed individuals may also present a higher rate of health depreciation because being unemployed can induce a high a high level of stress. These differences in the rate of depreciation may be another source to explain the different medical care utilization.

A further variable included as a socio-economic determinant of health care utilization is the individual income. As referred in the data and variables section of Chapter 4, we have included the monthly equivalent disposable income. According to Grossman (1972), the utilization of medical care presents a tendency to rise with wage, which, in our application, is represented by equivalent income. Basically stated, the argument used by the author is that the demand for health capital increases with wage, thus, the individuals with higher wages need to invest more on his/her health, including more medical care related inputs. One must recognize, however, that, on the other hand, the utilization of medical care entails time-costs to the individuals, which should be considered in the form of opportunity costs. It is easy to argue that the higher the individual's wage the higher the opportunity costs of the time, which creates an incentive to decrease the utilization of medical care visits. Still related to

income, Gerdtham (1997) points out that the current income is not the only factor that influences the willingness to pay for health care. The stock of wealth probably presents a stronger effect, especially for the elderly. Unfortunately, our dataset do not include any wealth indicator, thus, we are confined to use the equivalent income.

The final set of covariates included in the socioeconomic and demographic determinants is the place of residence. It includes four dummy variables[35] to reflect the wide-area region of residence, in addition to a covariate to represent, as well, the place of residence, but in an urban *vs* rural categorization (the covariate 'rural_area'). These covariates were incorporated to control for possible behavioural differences in individuals and/or doctors living in the different regions, in addition to control for the differences in the supply of medical care services[36]. We recognize that, at the wide region level, it is challenging to match a given region with a specific behavioural pattern on the part of the individuals or doctors. Moreover, the different regions also encompass very different situations in terms of medical care supply. Thus, in the event of finding differences on the utilization of doctor visits among the patients residing in different regions, surely, it will be hard to justify those findings. Notice however, that this is not the case for the 'Rural_area' covariate as one can argue that the rural life style is different from the urban and also that the supply of medical care services differs between rural and urban regions. On the one hand, we may argue that those individuals living in urban regions are subjected to a systematic diverse array of environmental factors — from pollution to stressful life style— that influence the rate of health depreciation (Wagstaff 1986). On the other, concerning the supply of

---

[35] The dummy regressors 'north', 'centre', 'lvt', 'alentejo' and 'Algarve' ('Algarve' is the excluded category).

[36] Note that among our covariates is one to reflect the doctor availability 'Phy_1000_residents', however, this covariate is only a rough proxy for the supply of medical care services.

medical care services, it is well known the uneven distribution of health resources, both public and private, across the country (Santana 1999; Oliveira and Bevan 2003; Bentes, Dias *et al.* 2004). In Portugal, the most deprived areas in terms of medical care resources are the poorer and isolated geographical areas, which mainly coincide with the rural areas. Therefore, individuals living in rural areas, characterized by less supply of medical care services, a visit to the doctor entails more time-costs, mainly travel time, thus, we expect, they will tend to lower the utilization of consultations.

The next group of variables included as covariates is the health status of the individuals. The first sort of variables included as health status indicators are those reflecting the presence of some of the most common chronic conditions, from diabetes to back-pain. The next two variables included in the health status group are indicators that intend to capture the short-run and long run health status. The variable 'ill_short_run' equals one if the individual reported to had been felling ill in the two weeks prior to the application of the survey, and zero otherwise. In turn, the variable 'ill_long_run' is a dummy variable that equals one if the individual reports being ill for more than three months, zero otherwise. Under the group of health status indicators, the variable 'limited', may be considered as a rough measure of disability and the variable 'stress' is a dummy variable that is used as a proxy to measure the individual's level of exposure to stress. The rationale to include it as medical care determinant are as follows: the individuals that intake sleeping pills are probably exposed to high levels of daily stress, which in turn, may influence the rate of health depreciation, changing in this way the consumption of medical care. The last three variables included under the umbrella of health status covariates, 'NeverSmoked', 'not_physical_activity' and 'mild_exercise' are proxies of health status.

The next covariate included, 'Phy_1000_residents', captures the availability of medical care services in the individual's area of residence, thus a supply side medical care determinant. Wagstaff (1986) argues that if excess demand for medical services varies from area to area, then the time-costs involved in visiting the doctor also varies

across the areas, therefore, the supply of doctors will tend to affect the individual's utilization of medical care services. Individuals living in areas with less supply of medical care services have more time-costs to visit the doctor, thus, lowering the utilization of medical care. Under the context of a hurdle model, Pohlmeier and Ulrich (1995) used the second stage parameter of this covariate to test for the hypothesis of supplier-induced demand. This variable, we remind, represents the total number of licensed physicians per 1000 residents in the area of residency.

Another group of variables that we consider as determinants of medical care utilization, which are very important for our application, are those reflecting the type of health insurance coverage of the individuals. The health insurance covariates have been found by many authors to be an important explicative factor on the utilization of medical care, influencing the behaviour of patients and doctors (Manning, Newhouse *et al.* 1987; Cameron, Trivedi *et al.* 1988; Vera-Hernandez 1999). As several authors pointed out, the level of health care utilization and the choice of the type of health insurance contract may be jointly determined, causing the health insurance variables to be endogenous. Thus, in standard econometric models that do not account for the endogenous nature of the covariate, the estimates that measure the effect of the health insurance are biased (Newhouse, Phelps *et al.* 1980; Cameron, Trivedi *et al.* 1988; Vera-Hernandez 1999). This endogeneity bias only emerges, however, when the choice whether to sign, or not, a given insurance contract is made by the individual. As it was explained in Chapter 3, in the Portuguese health care system the individuals are not free to choose the health insurance contract they hold, in consequence, health insurance status and expected future medical care consumption are two uncorrelated

variables[37]. All health insurance contracts considered in our application are either based on the occupation of the individual (on the case of the health insurance contracts provided by the health subsystems) or on the NHS. In addition, there is no reason to expect that health subsystems beneficiaries have chosen their jobs because the supplementary health insurance coverage that the type of occupation supplies. This means that all covariates included to reflect health insurance status are exogenous in our analysis. The health insurance covariates consist of two dummy variables; the covariates 'NHS-only' (equals one if the individual is covered only by the statutory NHS, zero otherwise) and the covariate 'ADSE' (equals one if the individual is entitled to be financed by the civil servants health insurance scheme, 0 otherwise). The group of other subsystems (OSS) is the excluded category.

Finally, we included a group of variables to control for the period of the year when the measurement of the health care utilization took place; the covariates 'Winter', 'Spring' and 'Summer'. These covariates are included to control for seasonal differences that may occur on the health status of the individuals, which are not capture properly by our health status covariates.

In previous experiments, we run regression models using some extra independent variables. For instance, we included a number of interaction variables to try to disentangle the influence of a given variable across different groups. As an example, we tested models with health insurance variables interacting with (log)income, however, none of the average marginal effects of these interaction variables turned out to be statistically significant, thus, we dropped such interaction variables from the regression models.

---

[37] We emphasize that we have deleted from our dataset all individuals that have reported to benefit from a voluntary health insurance, usually bought in a private health insurance company.

5.3.2 — Preliminary and exploratory analysis of the data

In this sub-section, we conduct a preliminary analysis of the data, built upon the calculation of some basic summary statistics, aiming at addressing the two main research questions that motivate the current chapter. We begin, nonetheless, by reporting some basic summary statistics of the covariates to characterize the full sample, regardless of the health insurance status and place of residence[38]. After such sample characterization, we present some summary statistics by health insurance status and by place of residence.

5.3.2.1 — Characterization of the full sample

Summary statistics of the independent variables considered in the analysis were reported in Table 4 (page 110), which, for the sake of easy the reading, is reproduced below. The table provides some basic statistical information useful to characterize the full sample.

Variable	Mean	S.D.	Max.	Min.
Socioeconomic and demographic				
Age [/10]	4,240	2,331	0	10,3
Female	0,527	0,499	0	1
Married	0,540	0,498	0	1
Education | age > 17	5,459	4,304	0	24
Not_work	0,589	0,492	0	1
Retired	0,202	0,401	0	1
Unemployed	0,030	0,171	0	1
Income/100	3,656	2,718	0,231	24,939

[38] Hereafter, whenever we refer to the place of residence of the individual we are designating the split rural *vs* urban.

Variable	Mean	S.D.	Max.	Min.
North	0,315	0,464	0	1
Centre	0,200	0,400	0	1
LTV	0,246	0,431	0	1
Alentejo	0,119	0,324	0	1
Rural_area	0,170	0,376	0	1
Health Status				
Diabetes	0,056	0,231	0	1
Insulin	0,006	0,079	0	1
EBP	0,178	0,383	0	1
Asthma	0,062	0,241	0	1
Bronchitis	0,030	0,170	0	1
Allergy	0,144	0,351	0	1
Back pain	0,407	0,491	0	1
Ill_long_Run	0,009	0,096	0	1
Ill_Short_run	0,344	0,475	0	1
Limited	0,045	0,207	0	1
Stress	0,113	0,317	0	1
NeverSmoked	0,629	0,483	0	1
not physical activity	0,609	0,488	0	1
mild_exercise	0,149	0,356	0	1
Supply side				
Phy_1000_Inhabitants	2,774	2,220	0,579	9,152
Health insurance status				
NHS-only	0,848	0,359	0	1
ADSE	0,095	0,293	0	1
Seasonality				
Winter	0,248	0,432	0	1
Spring	0,252	0,434	0	1
Summer	0,244	0,429	0	1
N = 42.501				

Summary statistics (Table 4 repeated)

Beginning with the analysis of the socioeconomic and demographic covariates, the previous table indicates that the average age is 42 years old, presenting a standard deviation of about 23,3 years. In terms of gender, the data reveal that approximately

53% of the individuals in the sample are woman. Regarding the marital status, the majority of the individuals are married, about 54%. Turning our attention to the covariate 'education', considering only the adult population, the data reveal that the average number of years with success in school is about 5.5 years. This low figure clearly endorses the low levels of education of the Portuguese population. Regarding the covariates reflecting the individual's occupation status, the statistics reveal that about 59,0% of the individuals did not work in the two weeks before the application of the survey[39], about 20,0% of the individuals were be retired, while about 3,0% of the individuals, at the time of the questionnaire, were unemployed.

Yet in the socioeconomic and demographic group of covariates, the average equivalent monthly income was found to be roughly 365 €. Concerning the distribution of the population across the territory, the data shows that approximately 31% of the individuals reside in the North region, about 20% in the Centre and 25% of the individuals has established its place of residence on the Lisbon and Tagus Valley region. Finally, in the south regions of Portugal, Alentejo and Algarve, reside 24% of the population equally distributed. More relevant to this chapter is, nonetheless, the distribution of the population by urban *vs* rural regions. The data reveal that the Portuguese population tends to agglomerate in urban areas, about 83% of the population resides on urban areas, therefore, only 17% of the individuals have the place of residence on rural areas.

Turning our focus to the covariates used as health status indicators, the descriptive statistics shows that approximately 5,6% of the individuals have diabetes. In addition,

---

[39] It is worth to notice out that this figure was computed for the whole population, minors included. Restricting this statistic for the population aged between 18 and 65 years old, the proportion of individuals reporting not working in the two weeks before the application of the survey, naturally, decreases to about 35%.

given that an individual has diabetes, the probability of using insulin is about 0,11. Concerning the other chronic conditions, the most striking value comes from the percentage of individuals with elevated blood pressure, 'EBP' covariate, with almost 18% a suffering from this condition. In addition, we also highlight the percentage of individuals complaining from back-pain, about 41% of the individuals[40]. The statistics taken from the table also illustrate that about 1% of the individuals reported to be ill for three or more months, whereas about 34% reported to have been feeling ill in the two weeks before the interview. Yet in the health status covariates, approximately 4.5% of the population presents some form of physical limitation that restrains the execution of some basic daily activities, whilst 11% are exposed to stress. Concerning the covariates reflecting physical activity, taken as proxies to health status, a bit less than two thirds of the population (61% to be accurate) has a daily life that do not require any kind of physical activity, while only 15% reported to engage in light physical activity on a regular basis. To close this sort of variables related to the individual's health status, the data reveal that approximately 63% of the individuals declared that never smoked.

The next figure illustrates the availability of physicians at territorial level, which can be considered a rough proxy to capture the supply of medical care services. The figure clearly demonstrates the uneven distribution of doctors across the country. The average number of doctors per 1000 residents is 2,8, however, there is a region that presents a ratio doctors/1000 residents of about 0,5 doctors, while other has a ratio of about 9 doctors per 1000 residents. This disparity in doctor availability across the territory surely can encourage different levels of medical care consumption across the country.

---

[40] This figure increases to about 50% if we consider only the individuals aged 18 or more.

Finally, regarding the health insurance coverage, roughly 85% of the individuals are covered only through the NHS, while the remaining 15% enjoy from an insurance contract provided by a health subsystem. From those, about 9,5% are covered through the civil servants fund, the ADSE fund, while the rest 5,5% are covered by one health insurance contract provided by one health subsystem other than ADSE.

5.3.2.2 — Summary statistics by health insurance status and place of residence

The first approach, although rudimentary, to tackle the research questions of the current chapter[41] consists in conducting a preliminary analysis of the data. In this sub-section, we begin by exploring the distribution of the number of doctor visits, by type of health insurance coverage and by place of residence.

The next two figures present bar charts of the empirical distribution of doctor visits. The first one is by health insurance status and the second by place of residence.

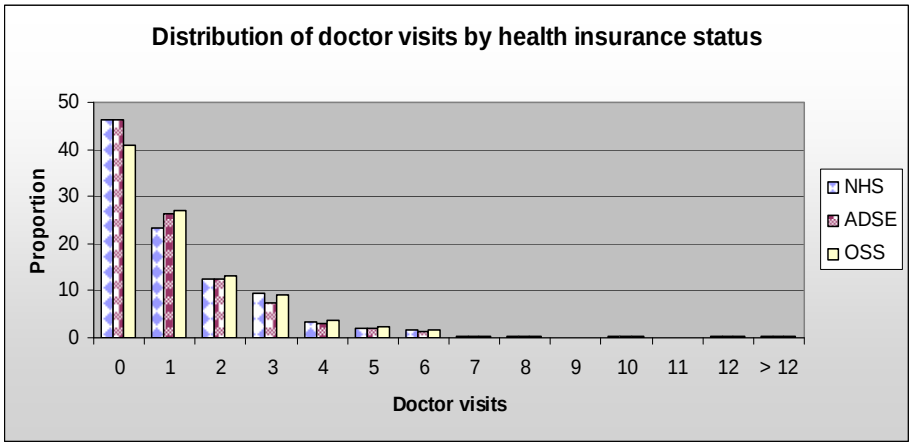

Figure 9 — Distribution of 'visits' by health insurance status

[41] The research questions mentioned are the effect of SHI and place of residence.

From the graphic shown in Figure 9, it is possible to conclude that the NHS-only and ADSE beneficiaries' present similar proportion of zero visits, and significantly higher than the OSS enrolees do.

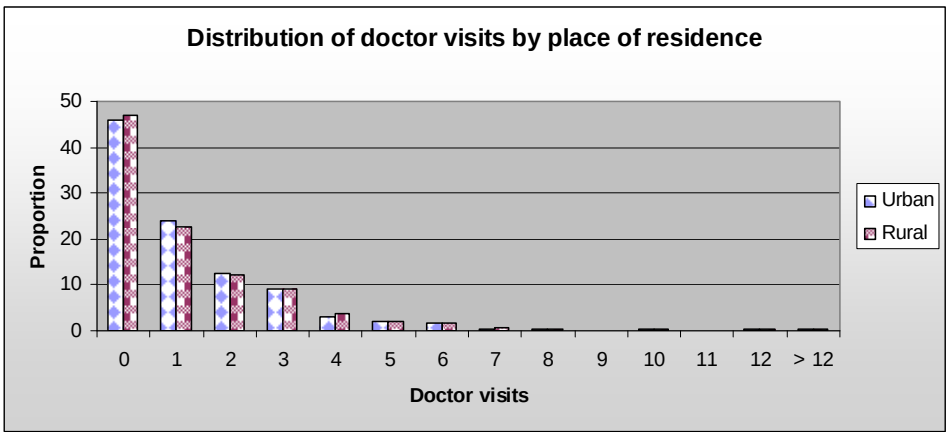

Figure 10 — Distribution 'visits' by place of residence

The graphical representation of the empirical distribution of doctor visits by place of residence does not reveal a clear different pattern of medical care utilization between urban and rural peers.

The former analysis was based only on the simple visual inspection of the bar charts of the empirical distributions. In spite of these results, to introduce robustness on the analysis, we should adopt inferential statistical methods to compare the empirical distributions.

Therefore, for now we begin by asking, and testing, whether the observed empirical distribution of doctor visits is statistically similar across the health insurance groups and across place of residence. To do so, we performed four two-sample Kolmogorov-Smirnov tests for equality of distributions; by health insurance coverage (three statistical tests) and by place of residence (one test). The p-values of the statistical tests are presented in Table 15, shown below.

H0: Equality of distributions	p-value
$f_{NHS}(y_i) = f_{ADSE}(y_i)$	0,001
$f_{NHS}(y_i) = f_{OSS}(y_i)$	< 0,0001
$f_{OSS}(y_i) = f_{ADSE}(y_i)$	< 0,0001
$f_{Rural}(y_i) = f_{Urban}(y_i)$	0,405

Table 15 — Kolmogorov-Smirnov tests for equality of the distribution of 'visits' by health insurance status and by place of residence

All p-values of the Kolmogorov-Smirnov tests, presented in the first three rows of the previous table, indicate that the distributions of the number of consultations are statistically different across the different health insurance groups. Contrasting this, the p-value relative to the Kolmogorov-Smirnov test of equality of the distribution of doctor visits by place of residence (p-value = 0,405), reveal that one cannot discard the hypothesis of equality of distributions.

These results suggest that, regardless of the individuals' characteristics, the type of health insurance coverage, indeed causes different patterns of health care utilization. Moreover, and contrarily to the previous conclusion, regardless of the living place, the individuals present a comparable utilization of medical care services. That is, taken overall, living in rural areas does not mean a different utilization of doctor visits.

We continue this preliminary data analyses by presenting some statistics of the medical care utilization by health insurance status and by place of residence. We begin by showing, in Table 16, the average number of visits by health insurance status.

	Mean	S.D.	Min	Max	N
NHS	1,29	2,06	0	30	36.037
ADSE	1,19	1,92	0	30	4.027
OSS	1,40	2,18	0	30	2.437
H_0: $\mu_{NHS} = \mu_{ADSE}$	|t-value| = 3,10			p-value = 0,002	
H_0: $\mu_{NHS} = \mu_{OSS}$	|t-value|= 2,42			p-value = 0,01	
H_0: $\mu_{ADSE} = \mu_{OSS}$	|t-value|= 4,05			p-value = 0,0001	

Table 16 — Summary statistics of 'visits' by health insurance status

The unconditional mean of the ADSE group is 1,19 visits, a value lower than the mean number of visits of the NHS-only group, which visits the doctor 1,29 times in each three month period. In addition, the health insurance group that visits the doctor more frequently is the OSS group, with an average utilization of 1,40 visits in every three months. Note that all unconditional averages are statistically different, as it is illustrated by the low p-values presented in the Table 16.

Another statistic that is worth to analyse is the average number of doctor consultations, but considering only the individuals with a positive number of visits. The results appear in Table 17.

	Mean	S.D.	Min	Max	N
NHS	2,41	2,29	1	30	19.343
ADSE	2,21	2,14	1	30	2.162
OSS	2,37	2,40	1	30	1.440
H0: $\mu_{NHS} = \mu_{ADSE}$	|t-value| = 3,82			p-value = 0,0001	
H0: $\mu_{NHS} = \mu_{OSS}$	|t-value| = 0,69			p-value = 0,4915	
H0: $\mu_{ADSE} = \mu_{OSS}$	|t-value| = 2,00			p-value = 0,045	

Table 17 — Summary statistics of positive 'visits' by health insurance status

Conditional on some medical care utilization, the NHS-only group presents a significantly higher medical care utilization relative to the ADSE cohorts. In addition, the figures suggest that the mean utilization of the NHS-only group is statistically

similar to the utilization of the OSS group, and, for a level of significance of 1%, the average medical care utilization of the ADSE group is similar to the mean utilization of the OSS group of beneficiaries. We may speculate that the long tail of the distribution for this group causes the higher average utilization of the NHS-only beneficiaries.

Table 18, on the other hand, presents the proportion of individuals that have visited the physician at least once in the last three months, or, expressed differently, provides an indication of the probability of visiting the doctor.

	Mean	S.D.	N
NHS-only	0,54	0,50	36.037
ADSE	0,54	0,50	4.027
OSS	0,59	0,50	2.437
H0: $\mu_{NHS} = \mu_{ADSE}$	|t-value| = 0,01	p-value = 0,99	
H0: $\mu_{NHS} = \mu_{OSS}$	|t-value| = 5,19	p-value < 0,0001	
H0: $\mu_{ADSE} = \mu_{OSS}$	|t-value| = 4,23	p-value < 0,0001	

Table 18 — Proportion of individuals with at least one visit by health insurance status

The statistics presented in the previous table suggest that the NHS-only beneficiaries cannot be statistically distinguished from the ADSE group, meaning that, ignoring the possible differences in the composition of the two groups, the probability of being a health care user is similar across these types of health insurance coverage. Statistically significant differences are found, however, between the NHS-only group and the OSS group, and between the two health insurance groups that split the health subsystems, ADSE and OSS. For the beneficiaries of the NHS-only and the ADSE groups we estimate a 0,54 probability of being a medical care user. This figure raises to 0,59 for the OSS health insurance scheme.

The evidence presented so far shows clear differences on the medical care utilization made by the beneficiaries of the different health insurance schemes considered.

In regards to the comparison of the utilization of medical care consultations by place of residence, Table 19, Table 20 and Table 21 show some summary statistics, in addition to the result of a t-test to test the null of equality of the means.

	Mean	S.D.	Min	Max	N
Rural	1,29	2,00	0	30	7.242
Urban	1,29	2,01	0	30	35.259
H0: Equal means	|t-value| = 0,033		p-value = 0,97		

Table 19 — Summary statistics of 'visits' by place of residence

	Mean	S.D.	Min	Max	N
Rural	2,43	2,18	1	30	3.841
Urban	2,38	2,30	1	30	19.104
H0: Equal means	|t-value| = 1,23		p-value = 0,21		

Table 20 — Summary statistics of positive 'visits' by place of residence

	Mean	S.D.	N
Rural	0,53	0,50	7.242
Urban	0,54	0,50	35.259
H0: Equal proportions	|t-value| = 1,78	p-value = 0,075	

Table 21 — Proportion of individuals with at least one visit by place of residence

All two-sample independent *t* tests conducted point towards the conclusion that individuals present similar medical care utilization regardless of place of residence. A rudimentary reading of this result would lead one to conclude that, only incidentally, the place of residence plays a role on the individual's decisions regarding the utilization of medical care. Note, however, that this is not as clear as it appears to be, essentially because there are several other factors, correlated with the place of

residence, that may affect medical care utilization and those factors must be accounted for in the analysis.

The same situation occurs for the health insurance variables. For example, it is known that individuals in the different insurance schemes may benefit from different health status, income, education. These differences can probably explain part of the different average utilization of health services found.

The next step is to show the differences in the individual's characteristics that each health insurance group encompasses. We begin by reporting, in Table 22, some summary statistics for the independent variables by health insurance status, 'NHS-only', 'ADSE' and 'OSS'.

	36.037 NHS		4.027 ADSE		2.437 OSS				
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	(1)	(2)	(3)
Socioeconomic and demographic									
Age [/10]	4,30	2,35	3,81	2,15	4,10	2,22	*	*	*
Female	0,53	0,50	0,55	0,50	0,47	0,50	*	*	*
Married	0,54	0,50	0,51	0,50	0,58	0,49	*	*	*
Education | age ≥ 18	5,27	3,84	9,60	5,27	8,20	4,42	*	*	*
not_work	0,59	0,49	0,54	0,50	0,65	0,48	*	*	*
Retired	0,21	0,41	0,14	0,35	0,19	0,39	*	ns	ns
Unemployed	0,03	0,18	0,003	0,05	0,01	0,09	*	*	*
Income[/100]	3,18	2,16	6,31	3,83	6,36	3,61	*	*	ns
North	0,33	0,47	0,21	0,41	0,21	0,41	*	*	ns
Centre	0,20	0,40	0,20	0,40	0,17	0,38	ns	*	*
LVT	0,23	0,42	0,30	0,46	0,38	0,49	*	*	*
Alentejo	0,12	0,32	0,16	0,37	0,11	0,32	*	ns	*
Rural_Area	0,19	0,39	0,08	0,27	0,08	0,28	*	*	ns
Health Status									
Diabetes	0,06	0,24	0,04	0,19	0,05	0,22	*	ns	ns
Insulin	0,006	0,08	0,005	0,07	0,01	0,09	ns	ns	ns
EBP	0,18	0,39	0,13	0,34	0,16	0,36	*	*	*
Asthma	0,06	0,25	0,05	0,21	0,05	0,22	*	*	ns
Bronchitis	0,03	0,17	0,02	0,15	0,02	0,15	*	ns	ns
Allergy	0,14	0,35	0,16	0,37	0,16	0,37	*	ns	ns

	36.037 NHS		4.027 ADSE		2.437 OSS				
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	(1)	(2)	(3)
Back pain	0,42	0,49	0,31	0,46	0,33	0,47	*	*	ns
sick_long_Run	0,01	0,1	0,01	0,08	0,005	0,07	ns	*	ns
sick_short_run	0,36	0,48	0,27	0,44	0,28	0,45	*	*	ns
Limited	0,05	0,22	0,02	0,13	0,02	0,14	*	*	ns
Stress	0,11	0,32	0,10	0,30	0,11	0,31	*	ns	ns
NeverSmoked	0,64	0,48	0,59	0,49	0,58	0,49	*	*	ns
Not_physical_activity	0,62	0,48	0,52	0,50	0,53	0,50	*	*	ns
mild_exercise	0,14	0,35	0,19	0,39	0,20	0,40	*	*	ns
Supply side									
Phy_1000_Inhabitants	2,68	2,16	3,22	2,46	3,40	2,37	*	*	*

Table 22 — Summary statistics of covariates by health insurance status

Note: The rightmost columns labelled (1), (2) and (3) contains;
(1) Result of the test $H_0: \mu_{NHS} = \mu_{ADSE}$ *vs* $H_1: \mu_{NHS} \neq \mu_{ADSE}$
(2) Result of the test $H_0: \mu_{NHS} = \mu_{OSS}$ *vs* $H_1: \mu_{NHS} \neq \mu_{OSS}$
(3) Result of the test $H_0: \mu_{OSS} = \mu_{ADSE}$ *vs* $H_1: \mu_{OSS} \neq \mu_{ADSE}$
ns – statistically non-significant
***** — p-value ≤ 0,01

The last column of the previous table (column 3) indicates that the ADSE and OSS health insurance groups present some differences in the socio-economic and demographic characteristics however, there are striking similarities in terms of health status. Notice that most of the *t* tests conducted to compare the means of the health status variables, for each of this group, are not statistically significant for a level of significance of 1%.

Regarding the composition of the two groups (ADSE and OSS) in the socio-economic and demographic covariates, the two groups present some noticeable differences. The OSS beneficiaries, relative to the ADSE group, are on average, older, comprising a higher proportion of males, presenting also a higher percentage of married people. Taking into consideration the adult population of the groups in comparison, the OSS members present, on average, lower levels of education relative

to the ADSE beneficiaries. The data still shows that both health insurance groups have alike monthly disposable equivalent incomes, around € 630. Finally, we did not find any statistical differences concerning the urban/rural place of residence. In sum, the OSS and ADSE health insurance groups comprise people with similar health status levels and slight socio-economic differences.

However, the differences between the NHS-only group and any of the two health insurance groups provided by the health subsystems are more marked than the differences found between the ADSE and OSS groups, reported earlier. The differences are on both types of covariates considered; socioeconomic and demographic and health status covariates. Compared with ADSE and OSS, the NHS-only group encompasses people that are, on average, older, much lower educated (a difference of 3 years compared with the OSS and 4 if compared with the ADSE) and significantly poorer. We highlight that the average monthly income of the health subsystems insurance groups are more than twice the monthly income of the NHS-only individuals. In addition, the proportion of retirees and unemployed individuals are higher in the NHS-only group than in any of the other health insurance groups. Regarding the place of residence, the regional distribution of the NHS-only group also differs from the regional distribution of the other groups in comparison. In fact, most individuals of the ADSE or the OSS groups tend to be concentrated in the region of the capital and less in the other areas of the country. It is also worth to note that the NHS-only beneficiaries present more tendencies to be located in rural areas, when compared to the other two health insurance groups, which are located mostly in urban areas. This is not surprising given that membership to this health insurance schemes comes with the occupation status (profession) and, usually, the workplace of such professions are mostly located in the urban areas of Portugal. Take for instance, the working place of the civil servants, of the militaries and the employers of the banking sector.

Regarding the differences of the groups in terms of health status, the NHS-only group, presents much lower health status levels when compared to any of the two

health subsystems insurance groups. This claim can be readily endorsed by the analysis of the statistics of covariates reflecting chronic conditions. The prevalence of chronicle conditions is higher in the NHS-only group than in any of the two other health insurance groups. In addition, among the NHS-only individuals, the proportion of those who felt ill in the two weeks prior to the application of the questionnaire is higher when compared to the other two groups. However, the variable that reflects the absence of smoking is favourable to the NHS-only group, as 64% of these individuals reported as all life non-smokers against the about 59% among ADSE and OSS groups.

Regarding the covariates reflecting physical activity, the NHS-only individuals tend to practice less exercise. Finally, the density of doctors is much lower in the areas where the NHS-only individuals reside, which is not surprising at all, because, as we have seen previously, the NHS-only beneficiaries' present a higher tendency to reside in rural areas, where, surely, the concentration of doctors is much lower.

To recapitulate, there are differences in the characteristics of the individuals of each health insurance group, however, the most dissimilar group is the NHS-only.

We now repeat the same exercise, but presenting, in Table 23, some summary statistics of the covariates by place of residence. The table also shows the p-value of a *t* test conducted to compare the means across place of residence.

	N = 7.242 Rural		N = 35.259 Urban		H_0 : Equality of means
Variable	**Mean**	**S.D.**	**Mean**	**S.D.**	**p-value**
Socioeconomic					
Age [/10]	4,80	2,40	4,12	2,30	< 0,0001
Female	0,52	0,50	0,53	0,50	0,05
Married	0,58	0,49	0,53	0,50	< 0,0001
Education | age ≥ 18	3,62	3,45	6,08	4,50	< 0,0001
not_work	0,64	0,48	0,58	0,49	< 0,0001
Retired	0,27	0,44	0,19	0,39	< 0,0001

Unemployed	0,017	0,13	0,03	0,18	< 0,0001
Income[/100]	2,52	1,73	3,90	2,82	< 0,0001
North	0,21	0,41	0,34	0,47	< 0,0001
Centre	0,35	0,48	0,17	0,37	< 0,0001
LVT	0,10	0,30	0,28	0,45	< 0,0001
Alentejo	0,25	0,43	0,09	0,29	< 0,0001
Health Status					
Diabetes	0,06	0,24	0,05	0,23	< 0,0001
EBP	0,22	0,42	0,17	0,37	< 0,0001
Asthma	0,08	0,27	0,06	0,24	< 0,0001
Bronchitis	0,035	0,18	0,028	0,17	0,002
Allergy	0,138	0,35	0,144	0,35	0,17
Back pain	0,49	0,50	0,39	0,49	< 0,0001
sick_long_Run	0,01	0,10	0,008	0,09	0,17
sick_short_run	0,41	0,49	0,33	0,47	< 0,0001
Limited	0,07	0,25	0,04	0,20	< 0,0001
Stress	0,10	0,31	0,11	0,32	0,01
NeverSmoked	0,69	0,46	0,62	0,49	< 0,0001
Not_physical_activity	0,66	0,48	0,60	0,49	< 0,0001
mild_exercise	0,13	0,34	0,15	0,36	< 0,0001
Supply side					
Phy_1000_Inhabitants	1,54	0,88	3,03	2,32	<0,0001
Insurance Status					
NHS-only	0,93	0,26	0,83	0,37	< 0,0001
ADSE	0,04	0,20	0,10	0,30	< 0,0001

Table 23 — Summary statistics of the covariates by place of residence

To begin with, the analysis of the differences between urban and rural residents, the summary statistics illustrate that, relative to urban residents, rural residents are older, present a higher tendency to be married and have lower levels of education. Actually, the level of education of rural residents is almost half the education of the urban ones. In addition, the proportion of retired individuals is significantly higher in rural regions. Another striking difference between the two groups is that those residing in rural areas are relatively poorer than those that reside in urban areas. The data still

disclose that there are no statistically significant differences on the distribution of men and women across rural and urban areas.

In regards to the health status of the two groups in comparison, the data indicates that the people living in rural areas present, on average, lower health status endowments relative to the urban residents. This can be readily corroborated by the analysis of the covariates reflecting chronic conditions (please see Table 23, above). Apart from the variables 'Allergy' and 'ill_long_run', where there are no statistically significant differences, all other variables are rated worst for the rural people than for those living in the urban areas. Moreover, the physical activity variables present much more favourable values for the urban group than for those living in rural areas. Finally, and somewhat, unsurprisingly, given the uneven distribution of medical care facilities across the territory, the density of doctors is much higher in urban than in rural areas.

Taken together all the evidence provided in this preliminary data analysis section, it suggests, on the one hand, the existence of differences in the medical care utilization across the health insurance groups, on the other, similar patterns of medical care utilization across rural and urban areas. These conclusions can only be taken as preliminary, because they are based on the analysis of only one variable, thus discarding the impact that individual's characteristics may exert on the medical care utilization.

In this regards, we provided evidence that the characteristics of the individuals differ across health insurance groups and place of residence. Therefore, to meaningfully address this chapter's research questions and evaluate correctly the impact of the health insurance status and of the place of residence, we have to interrogate our data more thoroughly, using the appropriate methodologies.

Therefore, to bring about the answer to our main research questions, posed in the introduction of this chapter, in the next section we present an analysis of the parameters that have come out from Chapter's 4 preferred model.

5.4 — Analysis of the regression results and discussion

The overall purpose of this section is to present an analysis of the medical care utilization of the Portuguese population, having in mind a couple of objectives. To achieve these goals, we will either analyse directly the parameters that came out from the preceding chapter's preferred hurdle model (the model LCH-Cross) or, in other cases, use such parameters to estimate the appropriate statistics that will help us to carry out that analysis.

Given that the second stage of our model is built upon the LCM framework, this section begins by studying the latent classes of health care users implied by the specification, in an attempt to unveil their main characteristics. Following this, the purpose of the next two sub-sections is to address the estimation of the effect of supplementary health insurance and place of residence in medical care utilization. Next, we investigate some other results of the model, which are, from our point of view, relevant to understand the contribution of other factors to explain the differences on the medical care utilization. Finally, in the last sub-section, we use the posterior class analysis methodology to cast light on the characteristics of the individuals that belong to each latent classes implied by the model.

As referred above, the model LCH-Cross, specified and estimated in Chapter 4, provides the parameter estimates for interpretation. Such estimates are reported in Table 12 (page 126), nonetheless, we have opted by reporting the table at this point.

Hurdle Latent class model: NB2 as parent distribution
(LCH-Cross(NB2) model)

	First part	Second part	
		Latent class I (17,8%)	Latent class II (82,2%)
Constant	-0,905 (0,088)	0,743 (0,257)	-0,596 (0,092)
Socioeconomic and demographic			
Age [/10]	-0,319 (0,031)	0,062 (0,061)	0,055 (0,029)
sqAge	0,035 (0,003)	-0,014 (0,007)	-0,005 (0,003)
Female	0,322 (0,024)	-0,132 (0,058)	0,101 (0,024)
Married	0,314 (0,029)	-0,026 (0,062)	0,115 (0,025)
Education	0,022 (0,004)	-0,001 (0,009)	-0,004 (0,004)
not_work	0,153 (0,031)	0,185 (0,069)	0,206 (0,03)
Retired	0,305 (0,044)	0,229 (0,086)	0,045 (0,03)
Unemployed	-0,072 (0,067)	0,173 (0,144)	0,001 (0,066)
(log) Income	0,119 (0,02)	0,01 (0,042)	-0,01 (0,018)
North	0,265 (0,037)	-0,104 (0,089)	0,104 (0,039)
Centre	0,420 (0,04)	0,082 (0,093)	0,175 (0,039)
LVT	0,274 (0,04)	-0,129 (0,09)	0,128 (0,04)
Alentejo	0,25 (0,044)	-0,137 (0,1)	0,111 (0,044)
Rural_Area	-0,166 (0,031)	-0,068 (0,068)	-0,034 (0,027)
Health Status			
Diabetes	0,842 (0,064)	0,052 (0,097)	0,153 (0,031)
Insulin	0,882 (0,238)	0,632 (0,182)	0,193 (0,078)
EBP	0,699 (0,035)	0,034 (0,062)	0,126 (0,023)
Asthma	0,052 (0,097)	0,052 (0,097)	0,052 (0,097)
Bronchitis	0,24 (0,072)	-0,067 (0,109)	0,128 (0,041)
Allergy	0,372 (0,033)	0,143 (0,061)	0,104 (0,024)
Back pain	0,316 (0,026)	0,048 (0,059)	0,176 (0,024)
Ill_long_Run	1,542 (0,17)	0,675 (0,157)	0,629 (0,053)
Ill_Short_run	1,002 (0,025)	0,452 (0,051)	0,466 (0,022)
Limited	-0,126 (0,067)	0,439 (0,096)	0,091 (0,041)
Stress	0,875 (0,044)	0,285 (0,063)	0,245 (0,024)
NeverSmoked	-0,106 (0,025)	-0,161 (0,063)	-0,103 (0,025)
not physical activity	0,026 (0,036)	0,156 (0,085)	0,039 (0,036)
mild_exercise	-0,087 (0,042)	0,083 (0,102)	0,018 (0,043)
Supply side			
Phy_1000_Inhabitants	0,018 (0,005)	0,014 (0,012)	0,005 (0,005)

Hurdle Latent class model: NB2 as parent distribution
(LCH-Cross(NB2) model)

	First part	Second part	
		Latent class I (17,8%)	Latent class II (82,2%)
Health insurance status			
NHS-only	-0,220 (0,048)	-0,096 (0,107)	-0,05 (0,042)
ADSE	-0,205 (0,057)	-0,083 (0,127)	-0,053 (0,052)
Seasonality			
Winter	0,071 (0,03)	-0,102 (0,067)	0,009 (0,028)
Spring	0,185 (0,03)	0,024 (0,069)	0,082 (0,027)
Summer	0,066 (0,03)	-0,139 (0,069)	0,089 (0,028)
α	1	1,242 (0,193)	0,158 (0,035)

Parameter estimates of the model LCH-Cross(NB2) (Table 12 repeated)

5.4.1 — Analysis of the latent classes of users

As established in the previous chapter, the model LCH-Cross outperformed a number of competing models. Given that the second stage of it is specified as a two-point of support LCM, then our data endorses the hypothesis that the population of users of doctor visits is composed by two different latent classes of patients.

As suggested by the mixing probabilities estimated by the model $(\pi = 0{,}178)$, a first latent class comprises about 17,8% of the users population while the remaining 82,2% of the patients belong to the other latent class[42]. Hence, the results evidence that the

[42] An alternative view is the split of the overall population (as opposed to the users population) implied by the model. In this context the model results suggest that the (overall) population can be divided into three classes of individuals: 46% of non-users, about

population of medical care users can be dichotomized into two latent classes of different sizes. In addition, as we show hereafter, the two latent classes present disparate characteristics in terms of the predicted health care utilization.

Figure 11 shows two box-plots depicting the distribution of the fitted means[43], by latent class, of the model LCH-Cross. It is noticeably evident from the charts that the predicted means of the patients in latent class I (left box-plot) show a clear tendency to be systematically larger than those that are found for the patients of latent class II (right box-plot). This offers the first sign that those individuals in latent class I, on average, tend to be more frequent users than those in latent class II.

To complement this analysis, Table 24, (please see below) reports some summary statistics of the distribution of fitted means by latent class.

$9.6\% = (54\% \times 17.8\%)$ are comprised in a second group and the remaining $44.4\% = (54\% \times 82.2\%)$ forms the other class.

[43] For each latent class considered, we estimated the distribution of the fitted means. The conditional mean function, by latent class j, for our preferred specification is given by

$$E_{(LCH-Cross)\,j}\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j, y_i > 0\right) = \frac{\exp\left(\mathbf{x}_i'\beta_j\right)}{1-\left(1+\alpha_j \exp\left(\mathbf{x}_i'\beta_j\right)\right)^{-\frac{1}{\alpha_j}}} \quad j = 1,2.$$

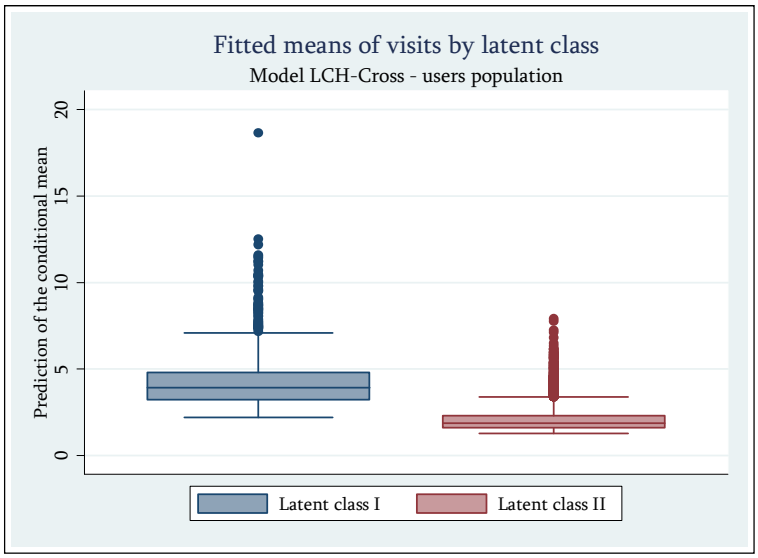

Figure 11 — Box-plots of the distribution of the fitted means by latent class for the model LCH-Cross

The values presented in Table 24 indicates that latent class I present an expected utilization of doctor visits of about 4,27 visits in each three month period, a maximum of about 19 and a minimum of a bit more than two visits. Moreover, roughly 80% of the predicted conditional means are within the interval [2,83 – 6,00] visits.

	Latent class I 17,8%	Latent class II 82,2%
Mean	4,27	2,03
Maximum	18,62	7,90
Minimum	2,20	1,29
Percentile	--	--
10%	2,83	1,49
25%	3,21	1,60
50%	3,93	1,87
75%	4,79	2,31
90%	6,00	2,80

Table 24 — Summary statistics of the distribution of fitted means by latent class on the users population

On the other hand, latent class II, encompassing 82,2% of the patients, presents an average utilization of roughly 2,03 doctor visits, a maximum of 7,9 visits and a minimum of 1,0 doctor consultation, per three month period. Moreover, about 80% of the predicted means belong to the interval [1,5 – 2,8].

Summing up, and bringing together all the evidence produced thus far by the analysis of the latent classes of health care users, clearly there is a class of high health care users, while the other can be considered as a class of low users.

Hence, the two underlying latent populations of users are different in respects that reach beyond the proportion of individuals it contains, namely, in the intensity of utilization of doctor visits. This provides a signal of a likely different behaviour on the part of the patients, or the doctors when making decisions on their behalf, in the different latent classes.

In what follows, the next two sub-sections study the impact of SHI and place of residence on the utilization of doctor visits. To uncover as much evidence as possible concerning the subject, we proceed by investigating the impact of the covariates on various descriptors implied by our preferred hurdle model. Namely, we estimate, and report, the effect of the covariates (´NHS-only', 'ADSE' and 'rural_area') on[44]:

First stage of the hurdle model:

- the probability of being a health care user[45];

Second stage of the hurdle model:

---

[44] This analysis is conducted following the methodology presented in Section 2.11 (please, see page 60). We begin by analyzing the distribution of the individual marginal effects relying on graphical means and then reporting the estimates of relevant average marginal effects (AME).

[45] $G\left(\mathbf{x}_i^t\beta_0\right) = P\left(y_i > 0 \mid \mathbf{x}\right) = \left(\frac{\exp\left(\mathbf{x}_i^t\beta_0\right)}{1+\exp\left(\mathbf{x}_i^t\beta_0\right)}\right)$

- the conditional mean function of the health care users population[46];
- the conditional mean of the users population by latent class[47];

Overall distribution:

- the conditional mean of the overall population[48];

5.4.2 — Analysis of the effect of SHI

This sub-section analyses the effect of supplementary health insurance on the number of visits to the doctor.

5.4.2.1 — Effect (of SHI) on the probability of visit the doctor

Figure 12 reports the effect of health insurance status on the probability of a person to have at least one visit to the doctor, holding other variables constant. The left box-plot depicts the distribution of the partial effects of NHS-only relative to the OSS group, while the right box-plot convey some information about the distribution of the effects of being beneficiary of the ADSE relative to the OSS.

---

[46] $$G\left(\mathbf{x}_i^t\beta_1, \mathbf{x}_i^t\beta_2\right) = E_{LCH-Cross}\left(y_i \mid \mathbf{x}_i, y_i > 0\right) = \sum_{j=1}^{2}\left[\frac{\exp\left(\mathbf{x}_i'\beta_j\right)}{1-\left(1+\alpha_j \exp\left(\mathbf{x}_i'\beta_j\right)\right)^{-\frac{1}{\alpha_j}}}\right]$$

[47] $$G\left(\mathbf{x}_i^t\beta_j\right) = E_{(LCH-Cross)\,j}\left(y_i \mid \mathbf{x}_i, y_i > 0, \beta_j\right) = \frac{\exp\left(\mathbf{x}_i'\beta_j\right)}{1-\left(1+\alpha_j \exp\left(\mathbf{x}_i'\beta_j\right)\right)^{-\frac{1}{\alpha_j}}}$$

[48] $$G\left(\mathbf{x}_i^t\beta_0, \mathbf{x}_i^t\beta_1, \mathbf{x}_i^t\beta_2\right) = E_{\text{LCH-Cross}}\left(y_i \mid \mathbf{x}_i\right) = \left(\frac{\exp\left(\mathbf{x}_i^t\beta_0\right)}{1+\exp\left(\mathbf{x}_i^t\beta_0\right)}\right) \times \sum_{j=1}^{2}\left[\frac{\exp\left(\mathbf{x}_i'\beta_j\right)}{1-\left(1+\alpha_j \exp\left(\mathbf{x}_i'\beta_j\right)\right)^{-\frac{1}{\alpha_j}}}\right]$$

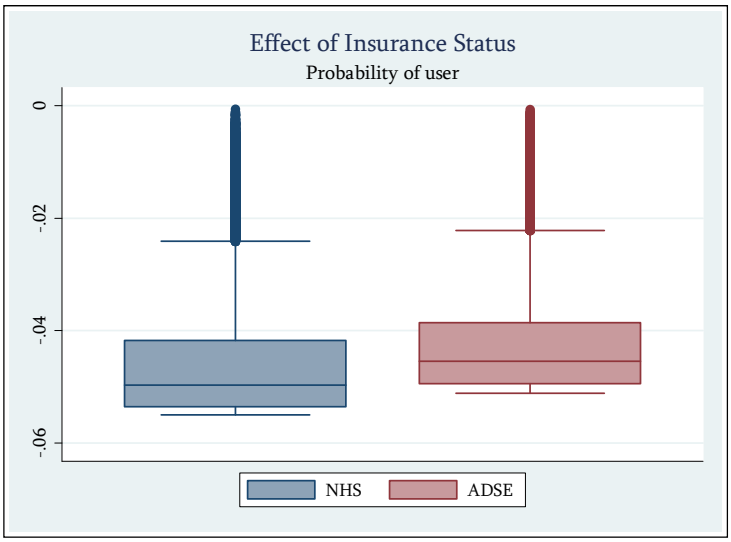

Figure 12 — Box-plots of the distribution of the effect of health insurance status (NHS and ADSE) on the probability of a doctor visit

It is noticeable from the previous Figure that those individuals covered only by the National Health Service (the NHS-only beneficiaries), assuming that the covariates used for control remain constant, have a lower probability of visiting a doctor, when compared to those individuals that enjoy from a health insurance contract provided by one of the insurance schemes of the OSS group. On the other hand, when one compares the impact of the 'ADSE' relative to the OSS group (the right box-plot) a similar result is obtained, that is, the effect of benefiting from the ADSE coverage is to lower the probability of a person having at least one doctor consultation[49]. It should be referred that the estimated standard errors (not shown here) of all individual effects reported earlier are such that the individual effects are statistically significant, for a level of significance of 99%. This signifies that, being member of the OSS group is advantageous in terms of the probability of making a doctor visit.

[49] It is important to emphasize that both box-plot charts are located below zero.

Another result that is worth to investigate refers to the estimation of the effect of being covered only by the NHS relative to the coverage provided to the civil servants by the respective fund (the 'ADSE' covariate). This can be carried out by comparing the two box-plot charts that appear in Figure 12. From it, it is perceptible that the differences between the two box-plot charts are slim, meaning that, probably, the two groups present a similar probability of making a visit the doctor. To substantiate this claim, and provide robustness, however, we estimated the magnitude of the individual effects of NHS-only relative to ADSE, along with the respective standard errors. The estimates found clearly support the absence of differences between the two groups in the probability of making a doctor visit, as they were, in fact, small in magnitude, besides presenting a large standard error relative to the magnitude of the effect. Therefore, this finding suggests that NHS-only beneficiaries and ADSE beneficiaries, after controlling for the differences in the characteristics of the individuals, do not present different behaviour regarding the utilization of the first physician visit.

Besides the analysis, built upon a graphical analysis, already presented, which investigated the distribution of the individual marginal effects, we also estimated the mean of such distributions, getting the average marginal effects (AME). Their magnitude was estimated in,

- -0,045 (s.e. 0.010), NHS-only vs OSS
- -0,042 (s.e. 0,012), ADSE vs OSS
- -0,003 (s.e. 0,008), NHS-only vs ADSE (statistically equal to zero)

In short, the results evidence that the beneficiaries of ADSE and NHS-only present a similar probability of visiting a physician, and, on the contrary, those enjoying from a health insurance contract provisioned by the OSS group have a higher probability of making a doctor visit.

We now turn to the investigation of the coefficients of the health insurance covariates of the second stage of the model.

5.4.2.2 — Effect (of SHI) on the average utilization for the users population

Figure 13 shows two box-plot charts illustrating the distribution of the effect of health insurance status (´NHS-only' and 'ADSE') on the conditional mean of the overall health care users population. The left box-plot illustrates the partial effects of NHS-only relative to the OSS group, while the one on the right shows the partial effects of ADSE relative to the OSS.

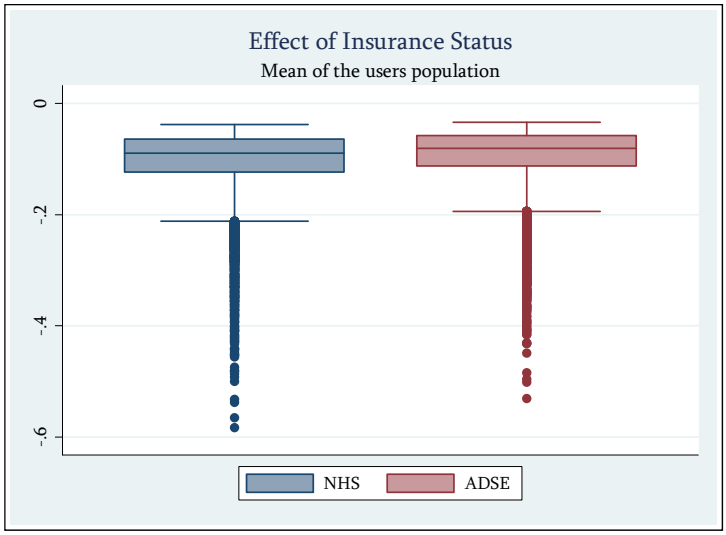

Figure 13 — Box-plot of the distribution of the partial effects of health insurance status (NHS-only and ADSE) on the conditional mean of the overall users population

The box-plots presented in Figure 13 illustrate that among those individuals who had at least one visit to the doctor, both covariates 'NHS-only' and 'ADSE' present a negative effect on the individual's demand for visits, as soon as the first visit has been made, as illustrated by the charts, entirely located below zero. Since the previous conclusion was based only on a simple visual examination of the box-plot charts, one cannot be definitive about the negativity of such impact. It may be the case that the estimates present a high degree of imprecision. In fact, after the estimation of the standard errors of the partial effects (figures not reported on this thesis) we concluded that, for a level of significance of 99%, they are not statistically different from zero. In

addition to this analysis, we estimated the average marginal effects (AME), getting the following results,

- -0,101 (s.e. 0,067): NHS-only vs OSS
- -0,092 (s.e. 0,071): ADSE vs OSS
- -0,009 (s.e. 0,054): NHS-only vs ADSE

Therefore, this indicates that among those individuals that crossed the hurdle, the health insurance status does not influence the number of visits to the doctor.

5.4.2.3 — Effect (of SHI) on the mean by latent class

We conducted a similar analysis for the latent class of high users and for the latent class of low users. The findings of such analysis, by latent class, pointed towards identical conclusions, that is, all individual partial effects associated to the health insurance variables, for each latent class, were found to be statistically insignificant, meaning that, they do not play a role on the individuals' decisions by latent classes. We estimated also the average marginal effects (AME) by latent class, getting the following results;

High users class

- -0,316 (s.e. 0,368): NHS-only vs OSS
- -0,256 (s.e. 0,378): ADSE vs OSS
- -0,060 (s.e. 0,291): NHS-only vs ADSE

Low users class

- -0,06 (s.e. 0,05): NHS-only vs OSS
- -0,061 (s.e. 0,05): ADSE vs OSS
- 0,0008 (s.e. 0,045): NHS-only vs ADSE

Therefore, this indicates that among those individuals who crossed the hurdle, the health insurance status does not influence the number of visits.

This evidences clearly that regardless of the latent class, on average, the individuals tend to use doctor visits with the same frequency regardless of the health insurance status of the patient.

5.4.2.4 — Effect (of SHI) on the mean for the overall population

Thus far, in this section, we investigated the effect of health insurance status on the different parts of the hurdle model. We begun by evaluating their effect on the first stage and after that, we checked its impact on the second stage of the hurdle model. These separated effects cannot, however, be extended to the overall population if, as it is the case, the covariates reflecting health insurance status enters in the linear index of both parts of the model. Therefore, to estimate the overall effect of the health insurance covariates, one has to combine the coefficients of both parts of the model.

Figure 14 shows the distribution of the individual partial effects of health insurance status on the conditional mean of the overall population, holding other variables constant. The left box-plot shows the distribution of the effects of 'NHS-only' relative to the 'OSS', while that on the right represents the distribution the effects of 'ADSE' relative to the 'OSS' (which is the excluded category).

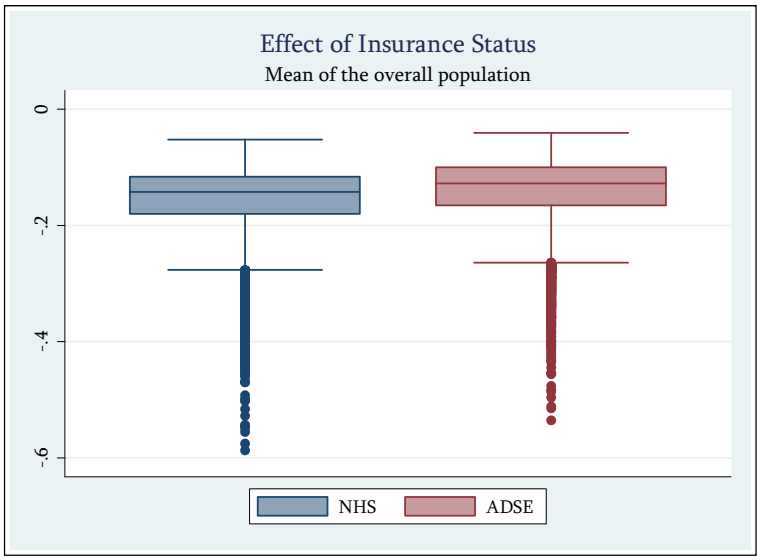

Figure 14 — Box-plots of the distribution of the effect of health insurance status (NHS-only and ADSE) on the conditional mean of the overall population

The box-plot charts depicted above reveal that, for every individual, the overall effect of the covariates 'NHS-only' and 'ADSE' is negative. The effects range from a minimum of -0,587 (-0,535) to a maximum of -0.053 (-0,041) for the NHS-only (ADSE) covariate. Moreover, 75% of the partial effects are within the interval [-0,116 — -0,80] ([-0,100 — -0,165]) for the 'NHS-only' ('ADSE') beneficiaries. Additionally, all estimates of the partial effects are statistically different from zero.

In regards to the evaluation of the impact of 'NHS-only' relative to the 'ADSE', the results (not shown in this dissertation) suggest that the beneficiaries of those insurance schemes present, on average, an overall identical utilization of visits to the doctor.

To complement the analysis of the distribution of the partial effects, we computed the average marginal effect (AME) of 'NHS-only' relative to the 'OSS', the AME of 'ADSE' relative to 'OSS' and, finally, the AME of 'NHS-only' relative to the 'ADSE' health insurance group, obtaining the following results,

- -0,152 (s.e. = 0,043): NHS-only *vs* OSS
- -0,136 (s.e. = 0,04): ADSE *vs* OSS
- -0,016 (s.e. = 0,033): NHS-only *vs* ADSE.

From the value of the ratio parameter/standard it follows immediately that the first and second AME are statistically different from zero, while the third is not. Therefore, the results support the view that, for the overall population, the beneficiaries of either NHS-only or ADSE insurance schemes tend to use, on average, less doctor visits than the individuals that benefit from a health insurance plan provisioned by the OSS group. On the other hand, on average, the frequency of doctor visits made by the ADSE beneficiaries is similar to the number of doctor visits made by the NHS-only individuals.

5.4.2.5 — Summary of the main findings and discussion (effect of SHI)

Here, we make a brief summary of the main findings regarding the effect of alternative health insurance arrangements, followed by a discussion of the results. To begin with, Table 25 summarizes the estimates of the AME along with the respective standard error.

Covariate / Model descriptor	NHS-only vs OSS	ADSE vs OSS	NHS-only vs ADSE
Impact on the First stage			
$P(y_i > 0 \mid \mathbf{x}_i)$	-0,045 (0,010)	-0,042 (0,012)	-0,003 (0,008)
Impact on the Second stage			
$E(y_i \mid \mathbf{x}_i, y_i > 0)$	-0,101 (0,067)	-0,092 (0,071)	-0,009 (0,054)
$E(y_i \mid \mathbf{x}_i, y_i > 0, High\,users)$	-0,316 (0,368)	-0,256 (0,378)	-0,060 (0,291)
$E(y_i \mid \mathbf{x}_i, y_i > 0, Low\,users)$	-0,06 (0,05)	-0,061 (0,057)	0,0008 (0,045)

Overall impact			
$E(y_i \mid \mathbf{x}_i)$	-0,152 (0.043)	-0,136 (0,040)	-0,016 (0,033)

Table 25 — AME of health insurance status; LCH-Cross model[50].

The findings can be abridged into three main points:

- The NHS-only and ADSE beneficiaries present a similar probability of visiting the doctor. Those from the group OSS are the better off.

$P(y_i > 0 \mid \mathbf{x}_i)$	NHS-only $\approx$ ADSE $\prec$ OSS

- After the initial decision of making a visit to the doctor, the three alternative health insurance groups present a similar frequency of doctor visits. These results extend to the latent classes of health care users.

$E(y_i \mid \mathbf{x}_i, y_i > 0)$	NHS-only $\approx$ ADSE $\approx$ OSS
$E(y_i \mid \mathbf{x}_i, y_i > 0, High\,users)$	NHS-only $\approx$ ADSE $\approx$ OSS
$E(y_i \mid \mathbf{x}_i, y_i > 0, Low\,users)$	NHS-only $\approx$ ADSE $\approx$ OSS

- Regarding the overall effect, the NHS-only and ADSE beneficiaries seek medical care, on average, with a similar frequency, while the OSS enrolees seek the doctor more often than the beneficiaries of the two other health insurance groups does.

[50] The grey cells refer to estimates statistically different from zero, for a significance level of 99%. All other values are not statistically significant, all with p > 0,05.

$E(y_i \mid \mathbf{x}_i)$	NHS-only $\approx$ ADSE $\prec$ OSS

As shown in Section 5.2, all international evidence regarding the impact of SHI on the number of doctor visits seems to agree upon the point that the individuals that enjoy from such double coverage, on average, tend to use more visits than those that do not benefit from such insurance. Hence, our results agree with such evidence, although only partially. As we defended previously, the ADSE and OSS funds provide supplemental health insurance to their beneficiaries, therefore, we would expected to find a higher intensity of utilization on the part of the beneficiaries of such funds, relative to the NHS-only peers. However, taking into account the overall effect of SHI on the number of doctor consultations, our results suggest a similar utilization on the part of the NHS-only and ADSE beneficiaries, and that the OSS enrolees use, on average, more doctor consultations. Nevertheless, when one disentangles the impact of SHI on the two stages of the decision process, it turns out that the observed higher utilization on the part of the OSS beneficiaries is due to its impact on the probability of making the first visit, and not on the number of visits after the first contact has been made. In fact, as soon as a doctor visit has taken place, our results show that the behaviour of the patients of each health insurance group considered (NHS-only, ADSE and OSS) is no longer different. The absence of differences is on the overall health care users population and across the latent classes that split such users population. This finding can be explained if one accepts that at this stage of the decision process it is the doctor's advice that matters the most. Thus, the patients fully comply with the doctor's advice, and because the physician does not respond to the incentives contained in the health insurance status of the individual, the allocation of visits are made according to the need of medical care and not according to the health insurance status of the individual. This result is not compatible with *ex-post* moral hazard.

On the other hand, the findings of the first stage of the hurdle model are somewhat puzzling, especially the similarities between the NHS-only and the ADSE beneficiaries

for which we, admittedly, do not have a fully explanation. Tentatively we may justify the finding on the following grounds: assume that, for the medical care services involved in our analysis, most of the ADSE beneficiaries choose a private doctor and that most of the NHS-only individuals, on the contrary, choose a public doctor[51]. Thus, both types of patients face costs to visit the doctor although of different nature. The NHS-only beneficiaries face, mostly, time-costs and, on the contrary, the ADSE enrolees face, mostly, monetary costs[52]. As a consequence, the similarities found between the NHS-only and the ADSE individuals on the probability of making the first visit to the doctor can be the result of the response to such costs.

The justification for the finding that the OSS beneficiaries present a higher probability of making a visit to the doctor, relative to the beneficiaries of the other health insurance groups, resorts on arguments related mainly to the policy of the institution that provides the health insurance, and not on the differences of the health insurance contracts. We recall that the OSS group encompasses the health subsystems of the militaries (five health subsystems, comprising about 27,5% of all individuals within the OSS group), of the banking sector (one health subsystem, encompassing about 20%) and of the electrical and telecommunications sector (two separated health subsystems). These institutions may incentive their employers to seek preventive care, on the form of doctor visits, and sometimes, it is even mandatory to make an annual check-up. These incentives to a preventive behaviour may spill over from the employer to the other family members covered by the fund. Therefore, we are

---

51 In fact, Lourenço (2003) has provided evidence pointing towards such kind of preferences of the holders of SHI.

52 In fact, as we explained in Chapter 3, the health insurance provided by the ADSE entitles the beneficiary to seek medical care in private providers, being reimbursed partially. However, in such circumstances, the co-payment demanded to the patient is usually high.

advocating that the higher probability of visiting a physician on the part of the OSS members are due to the more pro-active use of preventive medicine, encouraged by the policies of the institutions that provide the insurance policy.

Based on the findings reported and on their justification, we may conclude that the results do not support the hypothesis that health subsystems are a factor that causes inequity in the access to health care".

5.4.3 — The effect of place of residence on the utilization of medical care visits

The estimation of the impact of place of residence is among the research questions addressed in the chapter. The approach taken here to respond to it is similar to that we took in the previous sub-section.

5.4.3.1 — Effect (of rural) on the probability of visit the doctor

We begin by showing, in Figure 15, a box-plot that conveys information about the distribution of the partial effects of the 'rural_area' covariate on the probability of having at least one visit to the physician.

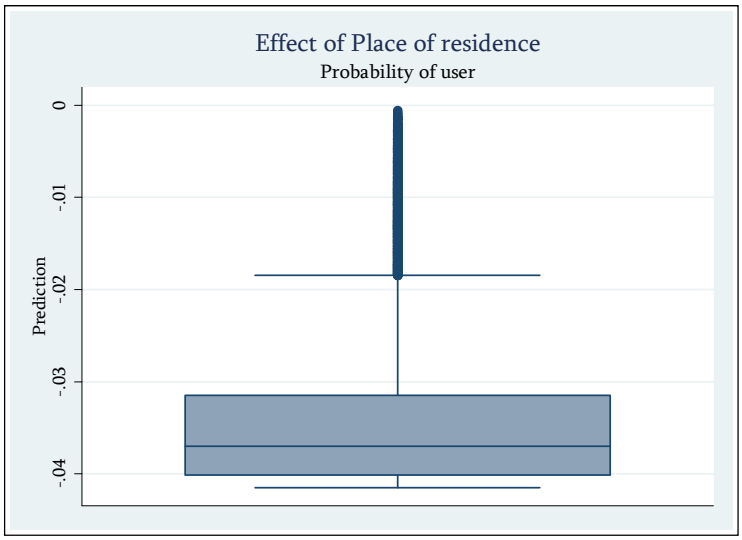

Figure 15 — Box-plot of the distribution of the effect of place of residence on the probability of at least one visit

As expected, given the negative sign of the 'rural_area' parameter, all individual effect are negative. In addition, all of them are statistically different from zero. Thus, this is evidence supporting that those individuals living in the rural regions of the country present, on average, a lower probability of seeing the doctor relative to those individuals residing in urban regions.

The AME of the covariate was estimated in -0,03 (s.e. 0,006), thus, being statistically different from zero. We take up this analysis, moving away from the first stage of the hurdle, and shifting the focus to the analysis of the parameters of the second stage of the LCH-Cross model.

5.4.3.2 — Effect (of rural) on the mean of the users population and by latent class

Figure 16 presents three box-plots depicting the distribution of the effect of place of residence on the conditional mean of the users population and by latent class. The left box-plot refers to the partial effects of the users population, the middle one refers to

the low users and, finally, the box-plot on the right represents the distribution of the effects of the population labelled as high users.

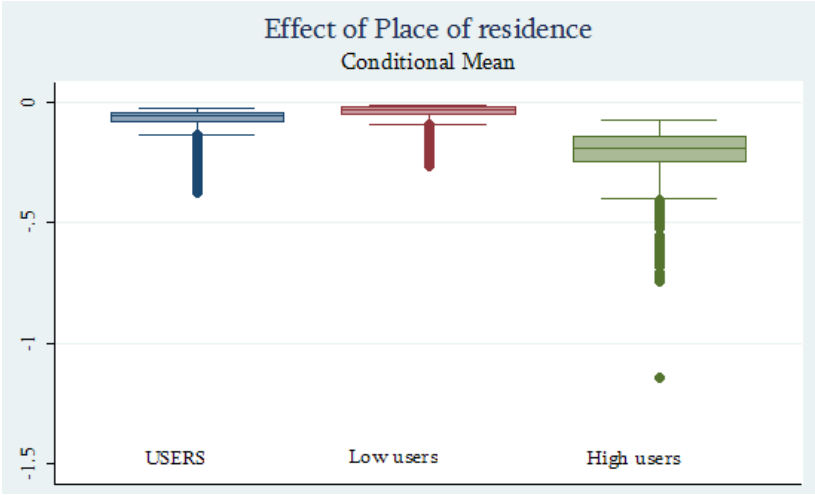

Figure 16 — Box-plots of the distribution of the effect of place of residence on the conditional mean of the users population, and by latent class.

The charts of the previous figure seem to suggest that individuals living in rural regions present, on average, a lower intensity of doctor visits, given some utilization, than their urban counterparts. The previous Figure conceals, however, important results that suggest that the place of residence does not contribute to explain the frequency of doctor visits. The standard errors of the estimates of the individual partial effects (not shown in the thesis) reveal that such partial effects are statistically zero, thus, suggesting that, contrary to the idea passed on by the box-plot charts, the 'Rural_area' covariate is not a determinant of the number of doctor visits.

Table 26 reports the AME of the covariate 'rural_area'.

Covariate Model descriptor	AME
Impact on the second stage	
$E(y_i \mid \mathbf{x}_i, y_i > 0)$	-0,07 (0,038)
$E(y_i \mid \mathbf{x}_i, y_i > 0, High\ users)$	-0,213 (0,209)
$E(y_i \mid \mathbf{x}_i, y_i > 0, Low\ users)$	-0,04 (0,03)

Table 26 — Estimates of the impact of the place of residence: LCH-Cross model.

All AME estimates shown, are negative, nonetheless, zero from a statistical point of view, therefore, after the initial decision to visit the doctor, on average, those who reside in rural regions present similar patterns of doctor utilization.

5.4.3.3 — Effect (of rural) on the mean of the overall population

We now shed light on the overall effect of the 'rural_area' on the number of doctor visits of the overall population.

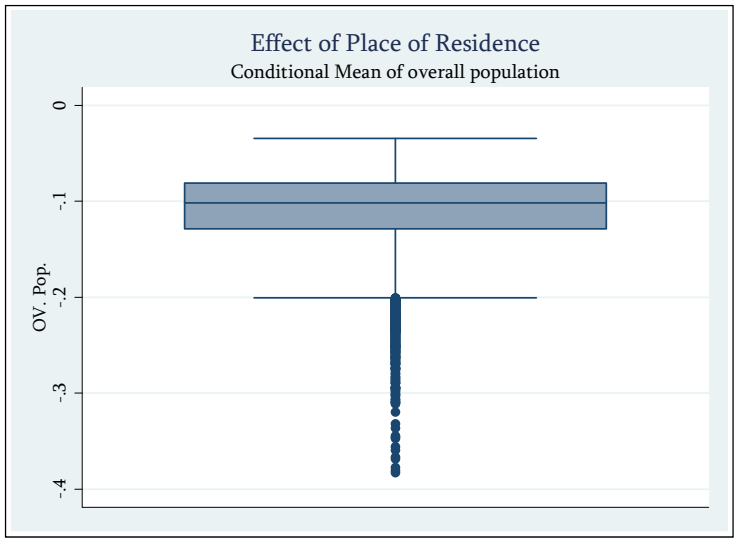

Figure 17 — Box-plot of the distribution of the effect of place of residence on the conditional mean of the overall population

From the previous figure it is clear that the individual partial effects are negative, with 75% of them located within the interval [-0,129 — -0,081]. In addition, the standard errors of the estimates indicate that they are statistically different from zero, for a level of significance of 99%. These finding suggest that, overall, the living place influences the number of times an individual seeks medical care. In regards to the AME estimate, its magnitude was estimated in -0,107 (s.e. = 0,024), therefore, with a z = 4,46. Taking into account the combined effect of the covariate 'Rural_area' on both stages of the LCH-Cross, the individuals residing in rural regions visit the doctor less frequently relative to the individuals that reside in urban regions. The average difference is small, less than 0,5 (4x0,107) visits a year.

5.4.3.4 Summary of the main findings and discussion (effect of rural)

To end up the presentation of the results regarding the impact of the 'Rural_area' covariate on the utilization of medical care services, and before enter in its discussion, Table 27 summarizes all AME estimates.

Model descriptor \ Covariate	AME
Impact on the first stage	
$P(user \mid \mathbf{x}_i)$	-0,03 (0,006)
Impact on the second stage	
$E(y_i \mid \mathbf{x}_i, y_i > 0)$	-0,07 (0,038)
$E(y_i \mid \mathbf{x}_i, y_i > 0, High\ users)$	-0,213 (0,209)
$E(y_i \mid \mathbf{x}_i, y_i > 0, Low\ users)$	-0,04 (0,03)

Overall impact	
$E(y_i \mid \mathbf{x}_i)$	-0,107 (0,024)

Table 27 — Estimates of the impact of place of residence on various descriptors taken from the LCH-Cross model[53].

Our global findings are in line with the international evidence on the subject where, overall, it has been found that those living in rural areas[54] tend to use less health care services. Our results point towards the same conclusion.

When we analysed the overall effect of residing in rural areas, we found that the rural cohorts tend to use less medical care services. However, the separation of the overall effect on the effect of making the first visit and, after making the first visit, on the effect of the number of visits, permits us to conclude that only the first effect presents statistical significance. Therefore, what contributes to the finding that overall intensity of utilization is lower for the individuals residing in rural areas is the individual's decision regarding the first visit and not the different referral behaviour of the doctors, in their advice to continuing visits, nor the individual's behaviour after the first visit was made. The result can be a combination of two interrelated and untangled effects. On the one hand, rural individuals may experience a lower rate of health depreciation due to the less stressful life style, thus they need less health care to preserve a given stock of health. On the other, people living in rural regions, the most

[53] The grey cells refer to effects statistically different from zero, for a significance level of 99%. All other values are not statistically significant, presenting a p-value > 0,05, except the effect of place of residence on the mean for the users population, which has a p-value of approximately 0,03.

[54] Some of the literature does not refer to rural area, however, we are using the population density as proxy to rural area. The lower populated areas are taken as rural regions. See Section 5.2 for a short review of the literature.

deprived regions in terms of medical care supply, face higher time-costs (not controlled for in the regression models due to the lack of data) to visit a doctor. The combination of these two factors, surely, contributes to explain the lower probability of visiting the doctor on the part of the rural residents. We believe, however, based in the results obtained on the second stage, that the supply factor plays a more prominent role than the health depreciation stock factor. The point is that, if the health stock of the rural individuals depreciates at lower rates, then, in the second stage, they would need fewer visits, and our findings did not point towards such conclusion. One can rightfully argue that the health care supply factor is also present, and may influence, the second stage decisions about the number of visits. However, the key difference at this stage is that the doctor's influence, and the information provided by him, is integrated in the decision process, thus cancelling the effect of different time-costs in the different regions. On other direction, our findings do not support the hypothesis of supplied-induced demand of the doctors practising in urban areas. As it was argued before, in Portugal the supply of doctors tend to concentrate on urban areas, with private supply naturally concentrating also on that urban regions. Therefore, the degree of competition is higher in urban areas than in rural areas. In addition, the payment scheme most often used by the private doctors is fee-for-service. Therefore, the conditions for the supplied induced demand behaviour are present in urban areas (Pohlmeier and Ulrich 1995). Regarding the results, given the lack of statistical significance of the 'rural_area' covariate, in the stage where the doctors are able to exert their influence, thus being able to induce demand, our results are not compatible with the supplied- induced demand on the part of the doctors.

5.4.4 — Analysis of the effect of the remaining covariates on the utilization of health care visits

This chapter has been emphasizing the evaluation of the effects of supplementary health insurance and residing in a rural area on the utilization of doctor visits. Nonetheless, it is worth to investigate the effect of the control covariates on the utilization of consultations. This is the purpose of the current sub-section.

The strategy is analysing, separately, the impact of the covariates in each part of the hurdle model. First, we address the issue of unveiling which factors contribute to explain the decision to seek the first doctor visit. Second, after the contact decision, we investigate the effect of the covariates on the conditional mean utilization by latent class[55].

Contrarily to the methodology followed thus far to interpret the model coefficients, built upon the estimation of the distribution of the marginal effects of the covariates, at the current sub-section, we study the effect of the covariates on the probability of crossing the hurdle through the analysis of the sign of the parameters[56]. On the other hand, to analyse the effect of the covariates by latent class, in the second stage of the model, to be technically accurate, we should have resorted on the estimation of the average marginal effects. This is the case because the impact of the covariates in the second stage of the hurdle, by latent class, is conducted using

---

[55] The analysis of the total effect of the covariates on the utilization of medical care is beyond the scope of our objectives in this chapter.

[56] The probability of crossing the hurdle, and observe a positive medical care utilization, is studied using $G\left(\mathbf{x}_i^t\boldsymbol{\beta}_0\right)=\frac{\exp\left(\mathbf{x}_i^t\boldsymbol{\beta}_0\right)}{1+\exp\left(\mathbf{x}_i^t\boldsymbol{\beta}_0\right)}$, where β_0 is the parameter estimates that come out from the first stage of the hurdle model. This function is clearly monotonic increasing in $\mathbf{x}_i^t\boldsymbol{\beta}_0$, therefore, the sign of the parameter coincides with the sign of the effect.

$G\left(\mathbf{x}_i'\boldsymbol{\beta}\right) = \dfrac{\exp\left(\mathbf{x}_i'\beta_j\right)}{\left(1-\left(1+\alpha_j\exp\left(\mathbf{x}_i'\beta_j\right)\right)^{-\frac{1}{\alpha_j}}\right)}$ $j=1,2$. It is possible to demonstrate that such $G(.)$ function is non-monotonic in $\mathbf{x}_i'\boldsymbol{\beta}$. Thus, the sign of the parameters may not coincide with the direction of the effect that we are interested in. Consequently, we computed the average marginal effects of the covariates (not reported anywhere in this thesis), and it turned out that, for our dataset, the sign of the coefficient coincided with the sign of the AME estimated. Thus, we continue the current analysis by evaluating the effect of the covariates on the average utilization, by latent class, resorting only on the signs of the coefficients.

We start by studying the parameter estimates of the first stage of the hurdle model, reported in the first column of Table 12**Error! Reference source not found.**, presented in page 126. An overview of the signs of the coefficients shows that all covariates present the expected impact on the probability of a positive number of visits. Moreover, almost all factors considered present statistical significance for a level of significance of 99%, implying that they have an effect on the individual decisions regarding the utilization of the first doctor visit.

The combination of the coefficients of the covariates 'age' and square of 'age' shows that age influences the probability of use through a parabola. The signs of the coefficients suggest that, given the control variables, the probability of an individual to visit the physician decreases approximately until 45 years old, and then begins to soar. This may be the consequence of an increase, with age, in the rate of health depreciation.

The positive and statistically significant coefficients of the covariates 'female', 'married', 'education' and 'retired' suggest that women, married people, those with more years of education and those that are retired are substantially more likely to seek doctor consultations. From all those covariates, the one that presents the lower impact

is education (a coefficient of 0,022), therefore suggesting that this factor plays only a minor role in influencing the probability of seeking medical care in the form of doctor visit.

The sign of the parameter connected to the covariate reflecting income, the covariate '(log) income', is positive and statistically significant, suggesting that, conditional on place of residence, insurance status, health status, and so on, those individuals with higher income tend to be more likely to seek medical care . Therefore, the results suggest that access to the health system depends positively on income. This is a clear indication of the presence of income-related inequity on access to medical care services.

The coefficients of the covariates reflecting the (wide-area) region of residence, are all estimated with precision, presenting also a positive value, with a similar magnitude. The exception is the coefficient of the covariate 'centre' that presents the highest value, about 0,42. In sum, the result indicates that the residents in the Algarve region are those that present the lowest probability of making a doctor visit, compared with those living in the remaining regions. Regarding the rationale for this result, admittedly, it is not obvious to us why the Algarve region have a lower propensity to use consultations. The result may be related to the supply of medical care in that region, or to the behaviour of the patients, or a combination of both situations. Possibly, the Algarve residents face higher times-costs to seek the doctor, thus, before seeking medical care they may let their health depreciate to such a point that to restore it, only impatient care can be used. In a rough way, our reasoning is supported by the results reported by Oliveira (2004) that found that, regarding the utilization of hospital care, the northern and centre regions have a lower propensity to use hospital care, thus, the Algarve region presents higher propensity to use such type of medical care.

Turning the focus to the covariates reflecting individual health status, the results indicate that lower health status endowments are associated to a higher probability of

making a visit to the doctor, as it is commonly found in similar studies. The coefficients associated to the health status covariates reflecting bad health, are positive and statistically significant. In addition, from all the chronic conditions included, the diabetes is the condition that causes the highest impact on the probability of making a doctor visit, followed by the hypertension condition. The coefficient of the covariate 'Never_Smoked' (a proxy for good health) is negative, implying that those that never smoked during their lifetime are substantially less likely to seek medical care. The coefficients of the covariates reflecting physical activity ('not physical activity' and 'mild_exercise') seem to indicate that they do not influence the probability of seeking doctor visits. Those individuals subjected to stress, capture by the 'stress' covariate, are far more likely to seek medical consultations, relative to those that are not exposed to stress. The result may stem from the impact of stress on the rate of health depreciation, which is higher for the stressed individuals. Finally, the parameter of the covariate reflecting physician density, the covariate 'Phy_1000_residents' has a positive and statistically significant value, although small in magnitude. Nonetheless, the result indicates that the individuals residing in regions where the supply of doctors is higher are more likely to seek medical care.

In regards to the analysis of the effect of the covariates on the number of physician visits, as soon as the hurdle has been crossed, the coefficients that allow such analysis are reported in the two right most columns of **Error! Reference source not found.**Table 12. On the one hand, the column labelled 'latent class I' report the parameter estimates for the latent class referred to as high users, on the other, the column labelled 'latent class II' reports the estimates of the latent class of low users.

The first thing that it is worth to mention is that some of the covariates play different roles according to the latent class, varying both in terms of statistical significance and on magnitude. Notice that we conducted an LR test to evaluate whether the vector of parameters that characterize medical care use by latent class are equal, concluding by the rejection of the null hypothesis (please, see test LR14 on

Table 6, page 114). Moreover, we conducted some other LR tests to assess the equality of parameters by latent class, for groups of covariates, e.g., socio-economic and demographic, health status and health insurance status, concluding for the rejection of the hypothesis of equality for all cases, except for the health insurance status. This last result indicates, as it was already discussed above, that after the contact decision, utilization is similar regardless the health insurance status of the individuals.

We now make an analysis of the results by latent class, starting out by latent class I, the class referred to as the class of high users. After the contact decision has been made, the results evidence that in the high users class, besides the statistically significance of most of the covariates reflecting health status[57], only the covariates 'Female' (-), 'Retired' (+) and 'not_work' (+) came out as statistically significant.

In this latent class of high users of health care, where possibly the role of the doctor's advice is of crucial importance to establish the process of care schedule, women tend to use significantly less care than men do. The result is opposite to what has been found in most empirical studies. One tentative explanation for this finding is that, probably, the males of this latent population may let their health status deteriorate more than females do before seeking care, meaning that their rate of health depreciation is higher relative to the women. Therefore, when male high users decide to see the doctor, they require a higher number of visits to face the illness, relative to the need of women.

On the contrary, the occupation status, being retired and not working, increases the utilization of services among the high users. Given our reasoning that the doctor plays the important role on the decisions at this stage, and more importantly, on this latent

---

[57] Nonetheless, some health status covariates did not come out as statistically significant on the latent class of high users class: namely, the covariates 'not physical activity', 'mild_exercise', 'Diabetes', 'Asthma', 'Bronchitis', 'EBP' and 'Back pain'.

class, the finding is not easily explained. Perhaps it can be rationalized by saying that this type of individuals presents lower opportunity costs of time, so they put more pressure on the doctor to schedule more visits, and the physician, may, somehow, respond to it, increasing the number of visits of the individual.

On the other hand, turning to the covariates reflecting the individual health status, the statistically significant covariates suggest, as it is often the case, that those individuals worse-off in terms of health status see the doctor more often. On the contrary, within this latent class, the absence of smoking (a life style variable considered as a proxy for good health) during the individual's lifetime is associated with fewer visits, relative to those that reported a smoking behaviour.

Overall, and summing up, these findings about the effect of the covariates on the latent class of high users are important because they suggest that, once within the system, regardless of income, place of residence and health insurance status, those individuals in similar need for medical care, receive, on average, a similar quantity of doctor visits, indicating equity related to income, place of residence and health insurance status.

The picture for the low users class is, however, somewhat different. In this latent class, the coefficients of the covariates used as health status indicators still show statistical significance, presenting the expected sign. However, on the socioeconomic and demographic group of covariates, some covariates became statistically significant, and others have changed the sign. The effect of female on this latent class is now positive, the covariates reflecting the (wide-area) region of residence became statistically significant, the same occurring with the covariate reflecting marital status. Summarizing, the variables of the socioeconomic and demographic group that are found to be statistically significant are: 'female' (+), 'married' (+), 'north'(+), 'centre'(+), 'LVT'(+), 'Alentejo' (+) and 'not_work' (+)

Therefore, in the latent class of low medical care users, contrarily to the case of the high users class, women tend to seek significantly more care than men do, as it is

generally found in the empirical analysis. Possibly, the rationalization of this finding encompasses the following justification: if we consider that the males and females included in this latent class present similar rates of health depreciation, thus, the health depreciation effect is ruled out, the effect that dominates is the precautionary care effect, that, as is known, women tend to have more developed than men do.

The effect of being married is positive, signifying that, married people tend to use more doctor visits than the non-married individuals do. The evidence about the effect of place of residence, in terms of the (wide-area) regional dummies, suggests that the Algarve citizens are those with lower utilization of doctor visits. It is not clear to us why the low users that reside at Algarve are those with the lowest health care utilization, after the contact decision has been made. Still on the socio-economic and demographic group of variables, those low users who did not work in the two weeks previous to the application of the survey (the covariate 'not_work') tend to seek significantly more care than those that have worked. Surely, the explanation for the result lies on the opportunity costs of the time of such individuals.

In regards to the covariates indicating health status, all of them are statistically significant, presenting, in addition coefficients with the expected sign.

Taking together all the evidence on the impact of the covariates in the second stage of our preferred hurdle model, one first finding that is worth to emphasize is that the covariates 'age', 'education', 'unemployed', 'income', 'rural_area', those covariates reflecting physical activity, physician density and SHI do not seem to be determinants of the number of times that the individual visit the doctor, as soon as the contact decision has been made.

Consequently, these findings suggest a conclusion relevant from a health policy perspective, which is related to the equity issue in the delivery of medical care. As presented above, a common interpretation of equity in medical care delivery is that utilization of medical care ought to be made on the basis of medical care need rather than on the basis of other factors not related to the health status of the individual, like,

income, place of residence health insurance status (Wagstaff and van Doorslaer 2000; Wagstaff and van Doorslaer 2000). According to this interpretation of the meaning of an equitable allocation of medical care (consultations), the results of our model suggest that within the Portuguese health care system, after the contact decision has been made, medical care services are delivered equitably in terms of income, health insurance status and place of residence (rural vs urban). Nonetheless, the statistically significant parameters of the covariates reflecting (wide-area) region of residence in the low users latent class leave some scope for inequalities related to the place of residence. This does not mean, however, that we can put forward the conclusion that these three factors (income, place of residence and SHI) are not important determinants in the number of times an individual utilizes medical care. In fact, it was reported in Table 13 (on page 127) that the overall effect of all three factors is statistically significant: with increases in income leading to an increase in the utilization of visits; rural areas residents and those without supplementary health insurance reduce the intensity of utilization. It is important, however, to pay attention where to look to reach such conclusions, because the place of the focus can suggest different conclusions, possibly leading to different the health policy recommendations. In our case, what is causing the observed overall effect of income, health insurance status and place of residence (rural/urban) is their effect on the entrance of the health care system, and not their effect on the intensity of utilization, for those inside the system. Therefore, the inequity that deserves concern and attention is that related to the probability of making the first visit and not the inequity in the overall utilization.

Notwithstanding these results about the equity related to income, health insurance status and place of residence (urban/rural), some regional inequities, in the second part of the hurdle, persist among the low users.

A second result that merits being underlined is that related to the effect of physician density on the utilization. Notice that the lack of statistical significance of

the covariate reflecting the density of physicians is evidence against the hypothesis of supplied induced demand.

5.4.5 — Analysis of class membership (model LCH-Cross)

The next logical step in the current analysis is the study of the composition of the two latent classes in terms of observed individual's characteristics, besides the analysis of their composition in terms of the utilization, already conducted. Unveiling the type of patients comprised in each latent class can be important from a health policy point of view. This will permit, for instance, to signal a health care user either as a low user or as a high user, based on readily observed characteristics, allowing the health authorities to design policies specific to the sub-populations and not to the overall population, as it is common to occur.

As it was presented in Section 2.5.1, the first step towards the description of the latent classes is the assignment of each individual to one and only one class. After the assignment step, we estimated a Probit model where the dependent variable indicates class membership to the low users and the independent variables are the covariates presented in Table 3, (see page 104). The results of such Probit model are presented in Table 28.

Covariate	Coef.	S. E.	Z
Socioeconomic and demographic			
Age [/10]	0,012	0,037	0,323
sqAge	0,005	0,004	1,367
Female	0,133	0,032	4,164

Covariate	Coef.	S. E.	Z
Married	0,040	0,037	1,081
Education	-0,004	0,005	-0,945
Not_work	-0,030	0,042	-0,722
Retired	-0,076	0,052	-1,454
Unemployed	-0,174	0,083	-2,100
(log) Income	-0,029	0,026	-1,123
North	0,112	0,049	2,275
Centre	0,034	0,052	0,661
LTV	0,131	0,052	2,518
Alentejo	0,111	0,060	1,845
Rural_Area	0,022	0,042	0,533
Health Status			
Diabetes	0,030	0,059	0,506
Insulin	-0,393	0,126	-3,124
EBP	0,009	0,038	0,246
Asthma	-0,061	0,052	-1,175
Bronchitis	0,040	0,076	0,522
Allergy	-0,022	0,037	-0,596
Back pain	0,020	0,034	0,570
Ill_long_Run	-0,050	0,114	-0,440
Ill_Short_run	-0,044	0,030	-1,491
Limited	-0,294	0,063	-4,656
Stress	-0,083	0,041	-2,038
NeverSmoked	0,033	0,034	0,961
not physical activity	-0,059	0,048	-1,229
mild_exercise	-0,043	0,058	-0,745
Supply side			
Phy_1000_residents	0,002	0,007	0,228
Health insurance status			
NHS-only	-0,012	0,060	-0,192
ADSE	0,062	0,072	0,862
Seasonality			
Winter	0,039	0,039	0,988

Covariate	Coef.	S. E.	Z
Spring	0,013	0,039	0,333
Summer	0,136	0,041	3,292
N = 22.945			

Table 28 — Parameter estimates, standard error and z-values of a Probit model to find the determinants of class membership to the low users class implied by the LCH-cross model

The results reported in the previous table evidence that among the group of covariates socioeconomic and demographic, the factors that contribute to explain the probability of belonging to the low users class are female(+), being unemployed(-), north(+) and LVT(+). Other socioeconomic and demographic factors did not come out as determinants of being classified as low user. Thus, we may conclude that the female patients and those residing either in the North or in the LVT regions, relative to the Algarve resident's, given that they have a positive utilization, show a higher propensity to be in the low users class. On the other hand, those health care users that have reported to be unemployed present a higher propensity to belong to the high users class.

Regarding the effect of health status on the assignment to the low users class, only three covariates are statistically significant: 'insulin' (-), 'limited' (-) and 'stress' (-), all of them being factors that decrease the probability of the individual to be classified as an low users.

Finally, those whose period of observation coincided with the summer have a higher probability of being assigned to the low users class.

Overall, the results suggest that, in line to what has been found by the related literature (Deb and Trivedi 2002; Atella, Brindisi *et al.* 2004; Bago d'Uva 2005), the health status of the individual is an important indicator of class membership, with those worse-off in terms of health status, naturally, presenting a higher probability of

belonging to the latent class of high users. Our results evidence, however, that, besides the health status determinants, one cannot ignore the role of other socioeconomic and demographic covariates on the probability of class membership, namely gender, employment status and the region where the individual resides.

5.5 — Conclusion

The purposes of the current chapter were stated at the introduction as follows: first, to evaluate the effect health insurance status and of the place of residence (urban vs rural), on the utilization of medical care services in Portugal. The impact of such factors by latent class were emphasized in our analysis; second, to unveil latent classes of health care users, and analyse them both in terms of expected health care utilization and in terms of the characteristics of the individuals that each latent class encompasses and third, to estimate the impact of other factors, e.g., income, health status and so on, on the utilization of medical care services.

We analysed the parameters estimated in the preferred count data model of the previous chapter, the model LCH-Cross. The model consists on a binary specification to find the determinants of some utilization and, given that the patient had at least one visit to the doctor, the positives are modelled through a two point of support LCM, based on the truncated NB2, that assumes the unobserved heterogeneity distributed on the population of health care users. The data to estimate the model was taken from the National Health Survey, the version conducted in 1998/99, and we measured the utilization of medical care services as the total number of doctor visits.

The most relevant findings of this chapter can be brought together in the following six points:

First, there is evidence that the population of the health care users can be split into two latent classes of users: a latent class referred to as high users, which comprises 17,8% of the individuals, presenting an average utilization of 4,27 visits per three

month period. On the other hand, the other latent class, that referred to as the low users, encompasses the remaining 82,2% of the patients, presenting an estimated average utilization of about 2 doctor consultations in each three month period.

Second, the analysis of the determinants of class membership shows that the propensity to be classified as a low user is associated mainly with good health status. Other covariates, however, are found as determinants of class membership, from which we highlight the gender, finding that the female factor soars the probability of classification as a low health care user.

Third, in what concerns the estimation of the effect of SHI on the medical care utilization, concentrating first on the overall medical care utilization, we found that the NHS-only beneficiaries present a similar frequency of doctor visits relative to those covered by the civil servants fund, the ADSE fund. However, those of the OSS group, present a higher frequency of doctor visits. Most relevant to us, given our modelling strategy, is, nonetheless, the decomposition of the impact of SHI on the probability of seeking a doctor visit and the effect on the frequency of consultations after at least one visit has been made.

This leads us to the fourth main conclusion drawn from our estimates. The NHS-only and ADSE beneficiaries present, on average, a similar probability of seeking the doctor. On the contrary, the beneficiaries of the OSS group, relative to the NHS-only and ADSE groups, are significantly more likely to visit the physician. Moreover, once within the health care system, the effect of being covered by a supplementary health insurance policy does not have a statistically significant effect on the overall frequency of utilization neither by latent class. On the one hand, we have claimed that the higher probability on the part of the OSS group is not triggered by the incentives contained in the health insurance contracts, but by the more pro-active use of preventive medicine, encouraged by the policies of the institutions that provide the health insurance policies of the OSS. On the other, the rationale for the second finding resorts on the believe that, at this stage of decision making process, it is the doctor that plays the major role

on the choice, and that he/she does not respond to the health insurance status of the individual. This is evidence, we argued, of absence of moral hazard behaviour on the part of the patients.

Fifth, in what concerns the effect of place of residence, in the urban *vs* rural characterization, the results evidence that the effect on the overall health care utilization is negative and statistically significant. Notice, however, that the decomposition of the overall effect on the impact of making the first visit and, after making the first visit, on the impact on the frequency of the visits, allowed us to claim that only the first effect presents statistical significance, being negative. Thus, as soon as the first visit is made, the place of residence does not play any role on the frequency decisions, overall, nor by latent class. We claimed that this finding is not compatible with the supplied-induced demand on the part of the doctors, namely, those practising in urban areas.

Sixth, implications to equity: our findings indicate that, after the contact decision, and controlling for other factors, medical care services are delivered equitably in terms of income, place of residence and health insurance status, thus, at this stage of utilization, our results are not compatible with inequity encouraged by these factors. Notice that that this finding extends to both latent populations that split the population of health care users. However, in terms of the impact of such factors (income, place of residence and health insurance status) on the entrance in the system, and obtain the first visit, income, living in urban areas and being covered by the OSS group are relevant factors, all increasing the likelihood of making the first visit to the doctor.

Regarding the weaknesses of the work conducted in this chapter, besides the potential sources of bias associated to the specification of the model adopted to analyse the data, we consider that the data limitations on the independent variables as more severe.

First, there is one type of variable on which we do not have information, the price of medical care, both monetary and time related. In the Portuguese system health context, these are two types of relevant costs, namely, the monetary ones when utilization is made in private doctors and time-costs in the case of utilization of public consultations. Second, another class of rather limited information is about health status. The complete group of health status covariates included in the analysis, do not measure accurately the health status of the individual. They provide only rudimentary measures of some health dimensions not reflecting the multidimensional nature of health status. In addition, due to sample selection concerns, related to the utilization of the SAH covariate, we excluded it from the analysis, being another possible source of bias.

Finally, we discuss some possible avenues for future research. Although improvements in the model specification are always possible, we believe that better data can be marginally more productive in enhancing our understanding of the factors underlying medical care utilization than model refinements. Namely, future surveys should include information on other important covariates like for instance, waiting time, travel time, price of private health care, besides, obviously, detailed information about the individual health status. Therefore, efforts to collect such type of data should be made. In addition, it would be important to have such type of data, like the data included in the National Health survey, in panel data contexts. This, surely, would improve our capacity to include complexity on the specifications, e.g, estimate models in the spirit of Bago d'Uva (2006), allowing us to increase the understanding about the factors that drive medical care utilization in Portugal.

Moreover, still on the area of the data, but on the theme of the measurement of utilization. It would be important that surveys included data about the number of visits for each type of visit — e.g. visits to GPs, visits to Specialists, outpatient care visits, etc — to different types of health professionals — e.g. for public and private doctors —. With such data, the researcher would have the possibility of specify

bivariate regression models addressing interesting health economics issues, like the estimation of the substitution and complementarity effects among different types of providers.

Chapter 6
Estimation of the effects of time-costs on the utilization of public primary care centres in Portugal[♣]

6.1 — Introduction

The estimation of the impact of time-costs on the utilization of medical care services is an important theme from a health policy perspective, being particularly relevant to those countries whose health systems are based on a National Health Service. In such countries, the monetary costs involved in consuming medical care services are generally null or very low, at least at the point of delivery. In these circumstances, times-costs naturally emerge as the price mechanism, thus functioning as a rationing device. Therefore, it is particularly relevant to learn how individuals react to that alternative form of cost. This subject has been debated for many years and by many authors, both at theoretical and empirical levels, as is witnessed by the large number of studies that dealt with it (Phelps and Newhouse 1974; Acton 1975; Colle and Grossman 1978; Coffey 1983; Lindsay and Feigenbaum 1984; Cauley 1987; Dor, Gertler *et al.* 1987; Janssen 1992; Propper 1995; Santana 1996; Goddard and Tavakoli 1998; Martin and Smith 1999; Blundell and Windmeijer 2000; Cullis, Jones *et al.* 2000; Gravelle, Dusheiko *et al.* 2002; Oliveira 2004; Barros and Olivella 2005; Bijan 2006).

The empirical literature that studies the pathways from which time-costs influence the individuals' choices regarding medical care services utilization, usually identifies

---

[♣] Preliminary versions of this chapter were presented at the first Workshop on the Utilization of Econometrics in Health Economics (held in Coimbra, Portugal, Faculty of Economics, April 2004) and at the 13th European Workshop on Econometrics and Health Economics (held in Venafro, Italy, September 2004).

two types of independent time-costs involved in the utilization of medical care. One first type of time-cost is the time spent in a waiting list, being usually implemented by measuring the gap of time elapsed between the appointment and the consultation, while the second type is the total time involved in completing a consultation, which encompasses the travel time to the doctor's office plus the total time spent in the physician's office. In the empirical application conducted in the present chapter, the waiting in a list time-cost category will be referred to as appointment delay (Colle and Grossman 1978; Cauley 1987) and the total time involved in a doctor visit will be referred to as physical waiting, a terminology borrowed from Cullis, Jones and Propper (2000).

Both categories of time-cost identified — appointment delay and physical waiting — obviously represent a cost to the health care user, though of a nature intrinsically different. How do these two types of time-costs function as a price mechanism? To begin with, waiting in a list, although not requiring the physical presence of the individual, is a cost to the individual because the underlying illness that has originated the demand for a doctor's visit may be getting worse while staying in the waiting list, being, thus, better to receive the medical care sooner rather than later. Regarding the physical waiting, we begin by noting that it requires the physical presence of the individual, therefore, that time has an opportunity cost to the patient in terms of either, forgone earnings or forgone leisure time.

The first author to emphasize that the time used to consume a good or service imposes a cost to the individual was Becker (1965) who developed a theoretical model to analyse the individual choices, explicitly incorporating in the model the time-costs involved in consuming the good or service in equal terms to monetary costs. Later, Grossman (1972), Phelps and Newhouse (1974) and Acton (1975), among others, proposed models to understand the individual's decisions regarding the consumption of medical care services. In such models, the physical waiting time-costs were explicitly considered as a part of the price the patient has to pay to use health care

services. Therefore, all models incorporated explicitly the physical time necessary to consume as an important decision variable. Concerning the time spent in a waiting list, Lindsay and Feigenbaum (1984), suggested an economic model that provides the general insights why this type of time-cost represents a cost to the individual.

In summation, the primary objective of this chapter is to provide evidence on this matter describing the results of an empirical analysis conducted to evaluate how the users of the public health care centres respond to the time-costs involved in visiting the general practitioner. To attain this goal we estimate the elasticity of health care centre utilization relative to the total time spent in visiting the health centre and provide evidence about the effect of an appointment delay on the utilization of the health care centre. Beyond the evaluation of the impact of time-costs, the estimated regression model also allows us to examine how a series of other individuals' characteristics affect their health care choices, e.g., education, age, health status, gender, occupation status, and so forth.

In the Portuguese Health System, based on a NHS, the extent of both types of time-costs may reach significant levels. Focusing on the time-costs involved in the utilization of primary health care, Cabral (2002) found that more than 54% of the people had to wait more than two weeks to get an appointment with a general practitioner. Moreover, based on data taken from the dataset analysed in this chapter, we found that the average delay to visit a general practitioner at the public health care centre was approximately 27 days and about 33% of the health care centre users experienced a delay in the consultation. Regarding the physical waiting dimension of time-costs, and using the data from the same dataset, we estimated the average total time spent to complete a consultation at the health centre in approximately two hours and a half. What is more, roughly 17% of the patients spent more than three hours in the health care centre and approximately 45% of them reported being unsatisfied with the time spent waiting in the waiting room. In addition, despite of the extent of the time-costs, as demonstrated above, and despite of being especially relevant to learn

how individuals react to them, only two authors have conducted analyses addressing this topic. However, they addressed the hospital sector and, more importantly, they used distance to the hospital as a proxy to the time-costs (Santana 1996; Oliveira 2004). Therefore, there is a complete lack of knowledge about the effect of time-costs on the utilization of public health care centres. On the other hand, despite the large amount of literature that has estimated the effects of time-costs on medical care utilization, the findings of those studies are of limited applicability to the Portuguese reality, and, more important, to the Portuguese primary health care utilization. Most of the literature on the effects of an appointment delay refers to the hospital setting (Lindsay and Feigenbaum 1984; Propper 1995; Goddard and Tavakoli 1998; Martin and Smith 1999; Blundell and Windmeijer 2000; Gravelle, Dusheiko *et al.* 2002), thus, being of little relevance to public primary medical care utilization in Portugal. In addition, the health reasons that drive the demand for primary health care are very different from those related to hospital care, thus precluding the extrapolation of the results obtained in these studies to the primary care utilization. Concerning the studies dealing with the effects of physical waiting on the utilization of medical services, it is also difficult to extrapolate their findings to the demand of GP services. Besides being somewhat outdated, almost all these studies analyse medical care categories different from those examined in this chapter, and more important, in health systems with different organizations (Phelps and Newhouse 1974; Acton 1975; Colle and Grossman 1978; Coffey 1983; Cauley 1987; Dor, Gertler *et al.* 1987; Gertler, Locay *et al.* 1987; Janssen 1992; Bolduc, Lacroix *et al.* 1996; Santana 1996; Oliveira 2004; Bijan 2006).

The present study can be important on several grounds;

First, because by providing empirical evidence on the subject, we contribute to the literature on the effect of time-costs on the utilization of medical care.

Second, it will shed extra light on a subject about which there is little empirical evidence for Portugal. The unveiling of such empirical evidence may be important from a health policy perspective, both for the health authorities and for the health care

centre managers. In particular, the results of the study may help to understand whether time-costs, and which class, can be used as an instrument to influence individual choices regarding the utilization of health care centres. Additionally, learning about the impact of other variables on the utilization of primary care may also help health centre managers and policy makers to understand why patients with comparable health conditions vary in the use of public GPs. This may be important to increase our understanding of the behaviour of users and may provide insights about organizational changes that could improve the effectiveness of this type of public providers, especially in countries with a gatekeeper system implemented.

In this study, the utilization of visits in the health care centre is measured using the number of consultations in the twelve months prior to the application of the survey, being thus a count data variable. The data for the empirical analysis were taken from the 2003/2004 Europep Survey, a health centre's users survey that is representative of the users of all public health care centres, which is administered in a non-regular basis (Ferreira, Raposo *et al.* 2005). To analyse the data, we specified and estimated various competing latent class models, appropriately tailored to handle the truncated count data that has resulted from the application of the survey.

The remainder of the current chapter contains five sections. In Section 2, we present a brief review of the literature about the effect of time-costs on the patient's choices and in section 3 we describe the methods used to collect the data, indicate and describe the dependent variable, in addition to the covariates included in our models. Still in this section, we report and briefly analyse some summary statistics. The next section, the forth, discusses the empirical methods used to analyse the data. In section 5, the empirical results are presented and discussed, and finally, section 6 finalizes the paper showing the main conclusions reached by the empirical analysis. We close section 5 by suggesting some policy implications suggested by our findings and present some unresolved issues related to our application, which unveil opportunities for future research.

6.2 — Effect of time costs on medical care utilization: short review of the literature

The objective of the current section is to briefly present and discuss some of the literature about the relevance of time-costs on individuals' medical care decisions, focusing mainly on the empirical part. As we will show in the next few pages, it is difficult to draw definitive conclusions about the impact of time-costs from past work since the empirical analyses conducted vary in many important respects. (1) the data used by the authors generally concern different countries and different years, thus describing different paradigms; (2) the variable used as indicator of medical care utilization often differs across studies; (3) the independent variables also differ across studies; (4) the econometric specifications chosen by the authors, surely influenced also by the type of dependent variable available to them, may also present some degree of variability. In what follows, we centre our attention on the results obtained by the authors, leaving the details about the modelling strategy to Table 29.

To begin with, one of the earliest authors that investigated the effect of time-costs on individual decisions regarding the utilization of medical care services was Acton (1975). The author, based on a choice model where physical waiting were incorporated explicitly as a decision variable, defended that time-costs are a determinant of health care choices, exerting a negative effect on the utilization of medical care services. The author specified an empirical model to study the effect of the distance (the proxy variable to incorporate time-costs) to the provider, along with other covariates, on the total volume of ambulatory care, measured by the number of visits. Acton's results, despite ignoring the count nature of the health care indicator, estimated a negative (-0,14), and statistically significant, elasticity of utilization of visits in physician clinics with respect to travel time. Thus, the empirical results found by Acton support that

the time-cost, measured by the distance to the provider, can function as a price mechanism.

Another study that we can refer is the one conducted by Colle and Grossman (1978). The authors specified a model to estimate the effect of both types of time-costs that we consider in this chapter, that is, physical waiting and appointment delay, in the annual number of paediatric office visits to physicians. The target population were preschool children ages one through five who utilized medical care provided by private practice paediatricians. Contrarily to Acton (1975), Colle and Grossman (1978) did not find evidence of a negative time-cost coefficient for both categories of time-costs, suggesting that parents are not responsive neither to the total (physical) time spent to visit the paediatrician nor to the existence of the existence of a delay.

Another study that has addressed this issue was conducted by Coffey (1983), who analysed the utilization of ambulatory female medical service, including gynaecological care, maternal health and family planning services. The study modelled three types of decisions: (1) the choice of the type of the provider, where the options are to seek public or private care; (2) the decision to have a visit; (3) the number of visits for those who had at least one visit to the doctor. In respect to the results, Coffey (1983) reported evidence only partially agreeing with Colle and Grossman's results. Coffey found that a 10% decline in the time-cost of public care relative to the private alternative increases the probability of choosing a public doctor by about 5%. Moreover, the author also found that an increase in the time-cost reduces the probability of using female ambulatory care. The reported elasticity is -0.09, with statistical significance. The same study also reports that for frequency decisions the time-price is not a determinant of the number of visits demanded in a year.

An additional analysis that it is worth to mention is the one carried out by Cauley (1987), which found a negative effect of time-costs on utilization. The author specified a Poisson model to explain the total number of outpatient visits made to a doctor in one-year period. Among the regressors were the two types of time-costs: the

appointment delay and total time spent to visit the doctor. Cauley (1987) found that increases in the appointment delay and on the physical time requirement were associated with decreases in the utilization of outpatient care, presenting statistical significance, making the case that both types of time-costs matter in health care decisions by the individuals.

All studies mentioned so far have specified econometric models to understand how time-costs influence the total utilization of a specific type of medical care service. Another class of studies that are usually adopted to address this topic are those that view health care utilization as a discrete choice among an exclusive set of alternatives. In this class of studies, the dependent variable is a binary or the multinomial choice among mutually exclusive medical care alternatives and the travel time to the doctor's office is usually used as a proxy of time-cost. In addition, all studies of this class specify some type of multinomial choice regression model. Examples of analyses made under this methodological framework are provided by Bijan (2006), Bolduc, Lacroix and Muller (1996), Mwabu, Ainsworth and Nyamete (1993), Janssen (1992), Dor, Gertler and Van Der Gaag (1987) and Gertler and Sanderson (1987).

To begin with, Janssen (1992) specifies a model to unveil which factors explain the probability of visiting a GP, distinguishing between patient initiated visits from doctor initiated ones. Each type of visits are modelled independently of each other, using two independent logit models, estimated using data from the Dutch population aged between 16 and 64. The time-cost associated with a visit comprises two components, the total time required to visit the GP (including travel time, waiting time and treatment time) and the value of time, which is valued by the individual's market wage, in some cases, while in others by the employment status. The study found time-price elasticities in the range of -0,09 to -0,14. For the patient initiated visit, all parameters have negative signs, with most presenting statistical significance for a level of 10%. For physician initiated visits, the time-cost variables still presenting the expected negative sign, however without statistical significance.

Another study that can be framed in the class of discrete choice models is the research conducted by Bijan (2006), who studied the health care choices among the rural population in India. The author considered that patients would have to choose from three types of providers: Public hospitals, private hospitals or from private doctors, without any reference to self-treatment. Bijan used a large array of medical care choice determinants, including the distance to the provider, which functions as a proxy for travel time, thus, a proxy to time costs in accessing the different health care providers considered. To assess the effect of the covariates on the individual choices, the author specified and estimated a mixed multinomial logit model, finding that the distance to the health provider plays a significant role on patient's decision about which type of health care facility to choose. In addition, the results showed that the importance of distance decreases with the health status of the individual.

These two studies, along with Bolduc, Lacroix and Muller (1996), Mwabu, Ainsworth and Nyamete (1993), Janssen (1992), Dor, Gertler and Van Der Gaag (1987) and Gertler and Sanderson (1987) have found that individuals when choosing whether to visit a doctor usually include time-costs as a determinant in the choice. Moreover, patients are also sensitive to the distance/travel time to the provider when choosing among alternative providers, thus showing that time costs are in fact a determinant in the health care choices. Concerning the direction of the impact, the studies present different evidence, with some supporting a positive effect of travel time (Bolduc, Lacroix *et al.* 1996), while others finding the opposite result (Gertler, Locay *et al.* 1987), that is, a significant negative travel time effect.

Regarding the evidence for Portugal, to the extent of our knowledge, only Santana (1996) and Oliveira (2004) have made some research on the subject, with both studies concluding that time-costs, with the distance to the health provider used as a proxy to time-costs, are a factor to which patients are sensitive when making choices regarding the utilization of medical care.

Santana (1996) analysed the effect of distance from the patient's residence to the hospital on the utilization of outpatient visits and on the utilization of emergency services. In the study, the utilization of health care is measured as the number of first outpatient visits and the number of visits to hospital emergency services per 10,000 inhabitants, at the county level. Using linear regression techniques, the author estimated the effect of the log of distance in each dependent variable, finding that, for both types of hospital services, the utilization decreases as distance increases. Using the parameters of Santana's study, Barros (2005) estimates the elasticity of outpatient visits utilization with respect to distance finding a high value. For those residing less than 50 kilometres away from the hospital the elasticity estimated was -4, while for the remainder, an elasticity of -2 was predicted. We notice that the findings reported by this study must be seen only as preliminary as the regression model estimated ignores that the dependent variables takes only positive values, and what is more, the study does not control for other factors that have been found to influence utilization, namely, health status, age, gender, income, etc. Despite these methodological weaknesses, the study signals that the distance to the provider is a factor that lowers the utilization of visits to hospitals, both outpatient and emergency visits (Barros 2005).

Contrarily to Santana's study, that presents some serious methodological drawbacks, the research by Oliveira (2004), who proposed a model of the demand for hospital care at the small area level, uses a sound methodology allowing her to assess the effect of time-costs on health care choices. The conceptual model used in the analysis is a Flow Demand Model (FDM), having a two-part (TPM) model as the empirical counterpart. The model explains the hospital utilization at small area level (the inpatient discharges per population) as a function of a series of determinants, where the covariate distance to the hospital is included as a proxy to capture time-costs. The model shows that the distance parameter is negative and statistically significant in the first part, thus increases in distance imply decreases in the probability of utilization. Regarding the

effect of distance on hospital utilization, given a positive utilization, the model estimates a negative and statistically significant effect, nevertheless, presenting a small value (Oliveira 2004).

To close this section, Table 29 gathers a summary of the empirical evidence regarding the effect of time-costs on individual health care decisions.

Reference	Health care utilization measure (Dependent variable)	Time cost measure	Econometric specification	Predictions	Additional information
(Acton 1975)	The total volume of ambulatory and impatient services demanded in the last year, encompassing three variables: Number of hospitalizations, the number of visits to a physician in a clinic and the number visits to a physician in his private office.	Distance to the provider. To the author, the variable distance measures both, the physical distance one has to travel and the money and time costs of travel.	The health care utilization by type of provider is estimated from a simultaneous equation system using two-stages least squares. The model does not have into account that the health care utilization variables are integer and non-negative.	The elasticity of utilization of visits in physician clinics with respect to travel time is -0.14. Summarizing, the model predicts that time costs, measured by the distance to the provider, function as a price mechanism.	The data used come from a 1965 survey of users of the outpatient departments of New York City municipal hospitals. Country: USA
(Colle and Grossman 1978)	Two health utilization measures: A dummy variable that indicates utilization of physician services during 1971 The annual number of paediatric office visits to physicians, in private practice, by preschool children ages 1 through	Two measures of physical waiting time: Travel time per visit to the usual source of private care and usual waiting time in the office per visit to the usual source of care. The covariate in the regression is the interaction between Wage and Total time (travel+waiting) Appointment delay: the	The binary variable (utilization) are specified to follow a logit model, while the number of visits is specified as a linear model and estimated by OLS.	The results do not show evidence of a negative time cost parameter in the binary variable that indicates use or non-use of paediatric care. The same result was obtained for the utilization of visits. In addition, The appointment delay did not present statistical significance, suggesting that	The dataset: The 1971 health survey conducted by the Centre for Health Administration Studies and the National Centre of the University of Chicago

Reference	Health care utilization measure (Dependent variable)	Time cost measure	Econometric specification	Predictions	Additional information
	5.	usual number of days that child has to wait to get an appointment (except for emergencies.		it is not an important rationing mechanism in the paediatric medical care services.	
(Coffey 1983)	The study analyses the utilization of ambulatory female medical service, including gynaecological, maternal health and family planning services. The paper considers three kind of decisions: Choice of type of provider, where the options are to seek public or private care. Decision to have a visit The number of visits for those with at least one visit to the doctor.	The time-cost measure is the value of time multiplied by the total minutes required to obtain care. Moreover, the value of time is measured by the market wage rate for women in the labour market, and by the reservation wage for those not in the labour market.	The binary variable is specified to follow a logit model, while the truncated number of doctor visits is modelled as a Tobit model and estimated accordingly. The integer nature of this last dependent variable is ignored.	Type of provider choice: The time price matters in the decision of the type of provider. A 10% decline in the time-price of public relative to the private alternative increases the probability of choosing a public doctor by about 5%. Entry decision: A increase in the time-price reduces the probability of entry for female ambulatory care. The reported elasticity is -0.09 with statistical significance.	Sample taken from a population of women aged 13-44 who resided in Dallas county, Texas. The data are relative to the year of 1977. The author reveals that the study was designed especially to develop improved time price measures and investigate their effect on the demand for medical care services. Country: USA

Reference	Health care utilization measure (Dependent variable)	Time cost measure	Econometric specification	Predictions	Additional information
				Frequency decision, given utilization: For those with a positive number of visits, the time-price is not a determinant of the number of visits demanded in a year.	
(Cauley 1987)	Analysis of outpatient care. The dependent variable is the number of visits made during a one year period of time.	Two types of time costs: The appointment delay, defined as the number of days between when an appointment is made and when the visit occurs. The time requirement defined as the sum of the office waiting time and travel time.	The empirical model assumes that the dependent variable follows a Poisson distribution. The mean parameter is parameterized as a linear function, and not as the exponential of the linear function, as has been usual. The linear mean function includes a series of time	Increases in the appointment delay and physical time were associated with decreases in the utilization of outpatient care.	Uses data relative to members of a prepaid practice medical care program (Kaiser Foundation Health plan). Country: USA

Reference	Health care utilization measure (Dependent variable)	Time cost measure	Econometric specification	Predictions	Additional information
			requirement interaction variables.		
(Dor, Gertler *et al.* 1987)	Models primary care as a multinomial variable , with 3 alternatives to treatment: Self-care Clinic Hospital	Travel time for the nearest available facility of each type.	Specifies a Nested Multinomial logit model. In the specification, travel time does not enter directly, entering instead via consumption and consumption squared. Similarly, the specification was by the maximization a conditional an expected conditional utility function.	Travel time plays a significant role in the choice of the medical provider. The health care choices among the poorer individuals are more travel time elastic than among richer individuals. In the conclusion the authors pointed out that travel time play a relevant role in rationing health care utilization	Sample taken from a population residing in rural Côte d'Ivoire, with data relative to the year 1985. Country: Côte dIvore
(Gertler, Locay *et al.* 1987)	Models primary care as a multinomial variable , with 3 alternatives to treatment: Self-care Public hospital Public clinic	Travel time was collected in discrete categories. In the specification the distance factor is included as an estimate of the probability of travelling more than one hour to the healthcare	Specifies a nested multinomial logit model. The framework to derive the empirical specification was a static model in which utility depends on health and	The parameter on the probability of travelling more than one hour is negative, meaning that increases in travel time (that is, increases in time-costs) reduce demand	The sample was taken from individuals living in urban areas of Peru, being relative to the year of 1984.

Reference	Health care utilization measure (Dependent variable)	Time cost measure	Econometric specification	Predictions	Additional information
	Private doctor	provider.	consumption of goods other than health care.		
(Janssen 1992)	Models the decision to whether to seek care in a GP. Two types of utilization are identified: The patient initiated visits and the doctor initiated	The time price associated with a visit. This comprises two components: the time required to visit the GP (including travel time, waiting time and treatment time) and the value of time, which is valued by the individual's market wage, in some cases, while in others by employment status	Uses simple logit specifications	The study found time-price elasticities in the range of -0.09 to -0.14. For the patient initiated visit, all parameters have negative signs, with most presenting statistical significance for a level of 10%. For physician initiated visits, the time variables continue to present the expected negative sign, however without statistical significance	Uses data from the Ducth population aged between 16 and 64. Country: Netherlands
(Bolduc, Lacroix *et al.* 1996)	Models ambulatory care as a multinomial variable , with 4 alternatives to treatment: Self-care Communal Health centre	Total time to receive care, which includes Travel time back and forth and waiting time	Specifies and estimates three alternative discrete choice models: Multinomial logit model, multinomial probit model and independent	Total time has a positive impact on the choice of a given alternative, but is only statistically significant in the multinomial probit specification.	Sample of a population residing in the rural district of Ouidah in Bénin. The data is relative to the year 1992.

Reference	Health care utilization measure (Dependent variable)	Time cost measure	Econometric specification	Predictions	Additional information
	Private clinic Hospital		multinomial probit model. The models were suggested by a conditional utility maximizing model		Country: Bénin
(Santana 1996)	Models the utilization of outpatient visits and utilization of hospital emergency services. The number of first outpatient visits and the number of visits to hospital emergency services per 10,000 inhabitants, at the county level	Distance from the patient's residence to the hospital.	Simple linear regression model estimated by ordinary least squares. The log of distances is the only regressor in the model.	For both dependent variables, the distance parameter is negative, meaning that as distances increases the utilization decreases.	The study uses data taken from a survey applied to a sample of users o of the Coimbra's teaching hospital, grouped by county. Country: Portugal
	Hospital utilization viewed as inpatient discharges per population point.	Distance to the hospital was used as a proxy time costs.	The conceptual model used in the analysis is a Flow Demand Model (FDM), having the two-part model (TPM) as	The impact of distance on the probability of a positive flow is negative, with statistical significance. Regarding the effect of	The dataset consists of 18700 statistical units with 275 local populations and 68 hospitals. The data was

Reference	Health care utilization measure (Dependent variable)	Time cost measure	Econometric specification	Predictions	Additional information
(Oliveira 2004)			empirical counterpart	distance on hospital utilization, given a positive utilization, the model estimates a negative impact, however, a small value.	extracted from multiple sources, with the utilization data obtained from the hospital discharges DGR database. Country: Portugal
(Bijan 2006)	Models ambulatory care as a multinomial variable, with 3 alternative providers: Public hospitals private hospitals private doctors (without any reference to self-treatment)	Distance to the provider. Included in the specification as a set of dummies	A Mixed Multinomial logit model suggested by a conditional utility maximizing model	The distance to the health provider plays a significant role on patient's decision about which type of health care facility to choose. In addition, the results unveiled that the importance of distance decreases with the health status of the individual.	Sample of the population of rural India. Obtained from the 52nd round of the National Sample Survey Organization (NSSO) Different estimations are made for children and adults. Because the self-treatment option is not considered, the study does not present any results about the impact

Reference	Health care utilization measure (Dependent variable)	Time cost measure	Econometric specification	Predictions	Additional information
					of distance on the health care utilization.

Table 29 – Summary of the empirical evidence regarding the effect of time costs on individual health care decisions.

Overall, and summing up, the findings of all studies reviewed in the previous table clearly show that the effect time-costs on health care decisions presents a wide degree of variation. The heterogeneity in the results can probably be justified because the different researches analyse different types of data, from different types of medical care services pertaining to different countries. In addition, to the utilization of different methodological tools adopted to estimate the impact of time costs.

6.3 — Dataset, dependent variable and covariates

The purpose of this section is threefold: first, we describe the questionnaire, the sampling mechanism and the type of information it contains; second, we present the dependent variable and the respective empirical distribution; third, we present the variables picked out as covariates. Moreover, we present the rationale for using such covariates as determinants of the medical care provided by the health centres, and, whenever possible, anticipate its effect on utilization.

6.3.1 — The questionnaire, dataset and sampling scheme

Our data comes from the Portuguese version of the Europep Survey, conducted, in 2003, by the Centre of Studies and Research in Health of the University of Coimbra (CEIS-UC) (Ferreira, Raposo *et al.* 2005). The survey was originally designed to obtain satisfaction scores from the health centre's users and was based on the Europep questionnaire, being administered in a non-regular basis. An international task force on patient evaluation of general practice care created the questionnaire (Ferreira 1999; Grol, Wensing *et al.* 2000), which initially involved researchers from seven countries (Denmark, Germany, The Netherland, Norway, Portugal, Sweden and United Kingdom). Afterwards, other nine countries joined the group (Austria, Belgium, Finland, France, Iceland, Israel, Slovenia, Spain and Switzerland). This measurement

instrument is currently assumed as a standard by WONCA-Europe, the European branch of the world association of general practitioners.

The measurement instrument encompasses 23 outcome questions grouped into five major dimensions, which are: (1) Patient-doctor interaction; (2) Medical care; (3) Information and support; (4) Continuity and cooperation; (5) Organization of services. Other questions complete the underlying conceptual model of this questionnaire, which may be grouped into three different dimensions: (6) Consultation, accessibility and appointment; (8) Professionals; (9) Health centre environment and services provided. In addition, the questionnaire also includes socio-demographic questions, about patients' health status and a question regarding the number of GP visits in the year prior to the application of the questionnaire.

The Portuguese version of the Europep questionnaire was validated in a nationwide sample and was firstly implemented in the Lisbon and Tagus Valley region as part of the tools used by the local Contracting Agency to monitor the activity of the health care centres (Ferreira 1999; Ferreira, Luz *et al.* 2001).

The process used to select the individuals was as follows: first, we obtained a database with information about all visits made to a public Health Centre during the last three months of 2003, obtaining in this way a list of 1,575,061 patients. The list contained information about the patient's address, gender and age. Because the patients with more than one consultation in the last three months of 2003 would appear multiple times on that list, we deleted the duplicated records related to visits made by the same individuals, ensuring in this way the non-occurrence of duplicated records. Following this deletion procedure, we extracted from the list a random sample of 67,000 patients. The random sample was selected to be proportional to the distribution of age and gender within each health centre. The questionnaires were directly mailed to the user's residency, being filled in and sent back by patients in a prepaid envelope. This scheme ensured the anonymity of the respondent.

Regarding the rate of response, we have received exactly 11.166 questionnaires, thus having a response rate of approximately 17%, however only 10.380 questionnaires were considered valid. This low response rate may lead one to question whether the sample continues to be representative of the users of public health care centres. To gauge the representativeness of the sample, we implemented chi-square tests to compare the age and gender of actual users, within each health district area, with the statistics of the same independent variables in our dataset. In fact, the χ^2 tests with both variables within each health district showed non-significance, thus, we might conclude that the resulting sample is representative by age and gender of the actual users.

The final working dataset does not comprise, however, information relative to the 10.380 individuals who sent valid questionnaires. To begin with, since our primary interest is in analysing the utilization of primary health care centres by adults we excluded all individuals aged 17 or less. In addition, all cases with missing values on at least one variable of interest to the empirical application were also excluded from the analysis. Hence, all parameter estimates reported in the current chapter are based on data for 7.502 individuals aged 18 through 98 years old. The data were collected in the year 2003.

As mentioned in Chapter 4 in reference to a similar issue, related to the existence of missing values, also here the removal of about 2.600 cases due to missing values may lead one to question whether the final working sample is a random sample of the interest population. To handle this difficulty, we proceeded differently from what we did in Chapter 4, thus, not conducting any formal hypotheses tests of sample selection. This different procedure is due to the lack of appropriate instrumental variables in the Europep dataset. The alternative procedure to handle the issue, although not so robust, was to compare the observed data with data taken from external data sources, namely, data taken from the National Health Survey 1998/99. From that survey, we selected

the users of public health centres, and using the variables age, gender, education, occupation status and some chronicle conditions we did not find any significant differences between the statistics produced by the different datasets. Therefore, we carry on the present empirical analysis assuming that the deletion of the cases does not introduce any bias in our results. Nonetheless, this is an open question.

6.3.2 — The indicator of health care centre utilization

Concerning the variable to reflect health centre utilization, the survey provides information about the individual's medical care utilization supplied by general practitioners, at the health care centre, asking the following question:

> "*How many times had you visited your family doctor in the last twelve months*"

Hence, in this study we have measured medical care utilization by the number of visits to the public general practitioner in one-year period, the variable henceforth referred to as 'hc_visits'.

Contrarily to the most common cases of health care utilization studies, where the researcher has a random sample of the overall population, in this application we do not have one. Under our sampling scheme, we did not observe the entire distribution of the dependent variable. We selected a sample from the population of individuals that visited the doctor in the last three months of 2003. Consequently, only the information about the individuals presenting a positive number of visits to the health care centre has the chance to be included in the sample. Therefore, the dependent variable is a truncated-at-zero variable.

The empirical distribution of 'hc_visits' is provided in Table 30 and the correspondent graphical representation is, also, depicted below, in Figure 18.

'hc_visits'	Frequency	Relative Frequency	Accumulated relative frequency
1	1.166	15,54%	15,54%
2	1,081	14,41%	29,95%
3	991	13,21%	43,16%
4	1.010	13,46%	56,62%
5	562	7,49%	64,12%
6	723	9,64%	73,75%
7	191	2,55%	76,3%
8	330	4,4%	80,7%
9	83	1,11%	81,8%
10	443	5,91%	87,71%
11	31	0,41%	88,12%
12	408	5,44%	93,56%
13	18	0,24%	93,8%
14	33	0,44%	94,24%
15	98	1,31%	95,55%
16	18	0,24%	95,79%
17	4	0,05%	95,84%
18	24	0,32%	96,16%
19	4	0,05%	96,21%
20	115	1,53%	97,75%
21	3	0,04%	97,79%
22	7	0,09%	97,87%
24	16	0,21%	98,09%
25	12	0,16%	98,25%
26	7	0,09%	98,35%
27	3	0,04%	98,39%
28	4	0,05%	98,43%
30	41	0,55%	98,99%
32	3	0,04%	99,03%
35	6	0,08%	99,11%
36	4	0,05%	99,15%
40	12	0,16%	99,32%
>=41	51	0,68%	100%
N = 7,502	Mean = 5,88	Max = 80.0	
	s.d = 6,64	Min = 1.0	

Table 30 — Empirical distribution of the dependent variable 'hc_visits'

The figures presented in the previous table reveal that about 16,0% of the individuals consult the health care centre once a year and that about 56,6% of health centre users seek medical care less than five times a year. Moreover, the maximum number of visits observed was 80,0 (only two individuals reported this impressive number of consultations), the average number of visits was estimated in 5,88, with a standard deviation of 6,64. Finally, the median of the empirical distribution was 4,0 visits in a year.

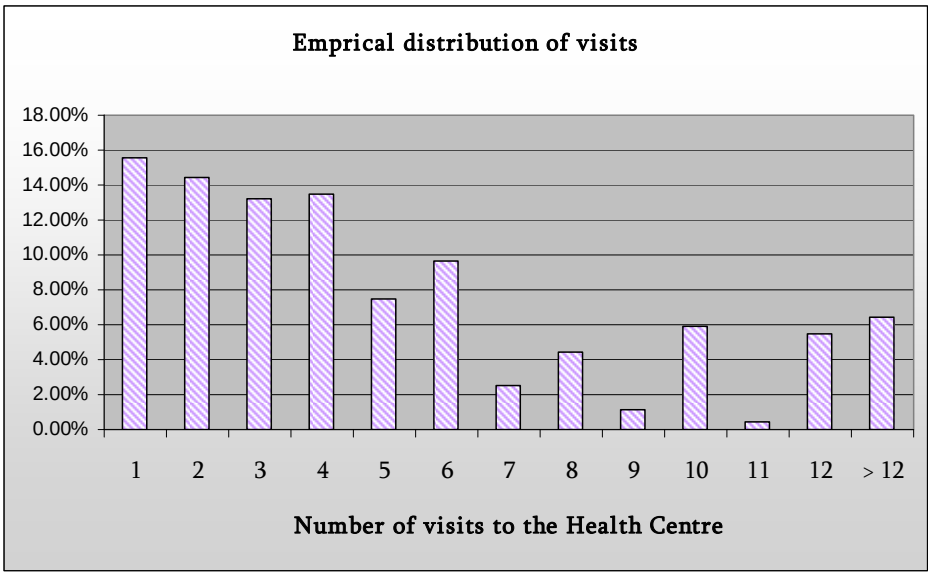

Figure 18 — Empirical distribution 'hc_visits'

6.3.3 — The determinants of health centre utilization

With respect to the explanatory variables selected as determinants of public GP utilization, besides the covariates included as time-cost indicators, we included those covariates generally selected as determinants in empirical models specified to understand the individual health care utilization. Both, their choice and their expected effect on the health care centre utilization are based on theoretical choice models developed by other authors, namely, Acton (1975), Phelps and Newhouse (1974),

Grossman (1972) and Wagstaff (1986), among others. In addition, we also rely on some arguments, admittedly *ad-hoc*, to suggest the utilization of some of the covariates.

In the remainder of this section, we present, and define, the independent variables considered in the empirical analysis, commenting briefly about its expected effects of the health care centre utilization. Concerning the evaluation of this impact, it is important to notice out that the analysis of the expected impact of the covariates must consider the fact that we are developing an empirical model to explain the utilization of a specific type of medical care (primary medical care), provided in a specific type of provider, health care centre general practitioners. This is to say that, notwithstanding the high degree of commitment that the Portuguese health centre users may have with the care provided in health care centre, they always have the possibility of substituting the health care centre consultations by consultations provisioned by other doctors. To exemplify this, the users of health care centres have always the chance of substitute consultations in the health centre by consultations in the private sector or consultations provisioned in hospital emergency departments. Therefore, the expected effect of a covariate in the utilization of health care centres is the composition of two effects that, with the available data, are impossible to disentangle. First, we must consider the impact of the covariate on the total utilization of medical care, independently of the provider, and second its effect on the substitution of public health care centres for private physicians or emergency care provided at hospitals. It is important to have this in mind at the time of interpretation of the coefficients of the econometric model.

In respect to the covariates used as determinants of health care centre utilization, we grouped them into four categories: a first category that includes the time-cost covariates and a second group comprises the socioeconomic and demographic variables. The third group of covariates is those that reflect the individual's health

status and finally, the closing class includes explanatory variables related to the characteristics of the health care centre and/or general practitioner.

In what follows, we present a short description of the variables included in each category considered, beginning with the ones that are central in this empirical application, the time-cost related factors.

Time-Costs covariates

The Europep survey provides detailed information about the last visit to the health care centre, including information about the two categories of time-costs involved in visiting it, the physical waiting and the waiting in a list.

Beginning with the physical waiting time-cost category, the questionnaire asked, on the one hand, how many minutes it takes to the individual to travel from home to the health care centre. On the other, how many minutes, in total, the patient spent in the health care centre, which, naturally, entails the waiting time in the waiting room plus the consultation time. Using the response to these two questions, we generated the covariate 'Total_time' that is the sum of travel time and the total time spent in the health care centre.

What is the expected effect of this covariate? To begin with, the total time spent to visit the general practitioner can be thought as a cost to the individual because, as the visit involves the physical presence, this time could have been used to earn income, for leisure or some other alternative utilization. As is stated by Martin and Smith (1999) the physical waiting requires the presence of the patient and, because time is a scarce resource, which is valued by individuals, this kind of waiting imposes an opportunity cost in the form of wasted time representing.

Acton (1975), on the other hand, shows that the necessary assumptions that make money function as a price in determining the demand for health care are sufficient to make time function as a price, as well.

Consequently, we anticipate that the total time spent in seeing the physician is a factor that, on average, decreases utilization of the health centre: Thus, the we anticipate a negative sign of the covariate 'total_time'. An additional argument that supports the negative coefficient is provided by Acton (1975) and Phelps and Newhouse (1974). Both authors claim that the impact of this kind of time-costs is stronger when the monetary cost is low relative to the time-cost, which is precisely the case in the public health care centres. In the Portuguese NHS, the extent of the co-payments at the delivery moment are low (many people are completely exempt from any payment) when compared to the extent of the time-costs.

Regarding the other time-cost dimension, we define appointment delay as the number of days between when an appointment is made and when the visit actually occurs (Colle and Grossman 1978; Cauley 1987). Concerning the practical issues related to the measurement of the appointment delay, the Europep survey has a hierarchical two-level question: the first question asks the patient whether he had, or not, an appointment delay, an in the case of a positive answer in the first level, the second question inquires the patient about the extent of such delay. Unfortunately, a non-neglected number of individuals responded yes in the first level, thus reporting the occurrence of a delay in the consultation, however, they did not mention the extent of such delay. Therefore, the utilization of the total number of days as indicator of an appointment delay causes a large number of missing values. Thus, we have opted for using the dummy variable 'App_delay', which equals one if the individual experiences a delay in the consultation, zero otherwise. Still regarding the generation of this covariate, it is extremely important to emphasize that the occurrence of a delay means that either the doctor or the health care centre imposed the waiting in a list, meaning that, all situations in which the visit was pre-scheduled by the doctor in a continuity of care process were not considered as appointment delays.

Which is the rationale to consider that waiting in a waiting list is a cost to the individual? To begin with, in general, the value of a good is lower if received later rather than sooner (Lindsay and Feigenbaum 1984). In the case of medical care, however, other factors can come into play, thus Propper (1995) partially agreed with the above claim. She claims that in the case of medical care, this class of time-cost, in fact, entails a cost to the patient because, as the patient waits in a list, the medical condition may deteriorate, subjecting the patient to pain and anxiety (losses in welfare), in addition to possible losses in income. When the health condition is such that the individual is unable to work, this might lead to losses in income, even in Portugal that has a Social Security System providing sick leave policies. The sick leave policies provided by the Portuguese authorities, however, do not guarantee the payment of the full salary to the ill worker. In general, the total payment is indexed to the number of days of illness. For instance, if the illness impedes the individual from working less than 30 days, the social security only reimburse 55% of the individuals' salary, and, more importantly, this payment is made only after the fourth day of illness (National Social Security 2005). This means that in the first four days of sick leave the worker loses income. What is more, visits to the doctor also may mean the loss of income as the time needed to see the physician is not reimbursed by the Social Security System (National Social Security 2005).

Moreover, in the event of a long waiting in a list, the relative cost of seeing a private physician is becoming lower, and accordingly people will be more willing to substitute public for private care or emergency care at hospitals.

Given this reasoning, we expect that the occurrence of an appointment delay imply fewer visits to the health care centre, therefore, a negative coefficient, either because people give up of seeing the doctor at all, opting for self-treatment and the illness heals without medical intervention or because they seek alternative providers.

Table 31 presents the covariates used to measure the time-costs and included as covariates in the regression models.

Time-costs	Variable definition
'Total_Time'	Total time (minutes) to visit the GP, which includes travel time, waiting in the waiting room and treatment time. (Log included in the regression)
'App_delay'	= 1 if in the individual had an appointment delay, 0 otherwise

Table 31 — Definition of the time-cost covariates

To soundly estimate the effect of the physical time-costs one should include in the regression not only the time requirement to visit the doctor but also opportunity cost of time, which should be multiplied, by the time requirement (Colle and Grossman 1978; Coffey 1983). The most common variable used to measure the opportunity cost of time is the individual's wage rate, however, unfortunately our dataset does not include any variable, or a proxy, to measure the wage rate. Therefore, we are forced to include only the time requirement, implicitly assuming that the opportunity cost is constant among the individuals. A violation of this assumption means that the model underestimates the time-price elasticity (Coffey 1983). Nevertheless, given the characteristics of our population we believe that omitting the opportunity cost of time does not cause bias in the results. Notice that retired (about 60% of the population) and low educated (about 77%) individuals mostly constitute our population, therefore, we believe that the variations in individual wage rates are small. Therefore, we proceed assuming that all health centre users present similar shadow prices for their time.

Our specification does not rule, however, out the possibility that the time-cost effect varies with some observed individual factors, like for instance, activity status, age and education. To evaluate whether the impact of the 'total_time' varies with these

individual characteristics, we follow Cauley (1987) and include a set of covariates that are the interaction of the total time required to visit the GP and the following independent variables;

- 'Non-active' (equals 1 if the individual does not work because he is retired or he is unemployed): we are assuming that those individuals that are retired or unemployed, thus are non-active, have a lower opportunity cost than those in the labour force. Thus, we anticipate that those in the labour force are more reactive to time costs than those that do not work. Undoubtedly, it would be more insightful if we have had the chance to categorize the individuals by employment status, however, our survey does not have that kind of information,
- 'Age': we are assuming that the reaction to the time-price and age are related through an inverted parabola. That is, as the individual gets older the impact of time on the GP utilization increases (the value of time increases), however, after a given age, the value of time begins to decrease. This is the expected behaviour because of the opportunity cost of time. The youngest and the eldest are the ones with lowest opportunity cost, thus, the ones presenting the lowest response to physical waiting,
- 'Education' (equals one if the individual has at most 9 years of formal education): the motivation to include this interaction term is because we are assuming that the higher the education the higher the value of time, which amounts to say that the higher the education the higher the response to the physical time-costs.

Thus, we created interaction variables, the multiplication of the (log) total waiting time in the health care centre and each one of the three variables abovementioned, so that separate elasticities of utilization with respect to the (physical) time-cost could be obtained. Moreover, to account for the possibility that this elasticity varies non-

linearly with age, we included an extra interaction between the time-cost and the square of age.

Table 32 presents information about the interactions to control for the value of (physical) waiting time

Interaction variable	Variable definition
(log)Total_time* non-active	Interaction between log of total time and non-active.
(log) Total_time * age	Interaction between log of total time and age.
(log) Total_time * sqare of age	Interaction between log of total time and age squared.
(log) Total_time * education	Interaction between log of total time and education.

Table 32 — Interactions between 'time-cost' and some other covariates

Concerning the time-costs related to queuing, we also consider the possibility that the time-cost effect varies with some observed individual factors. In this case, we consider factors like health status, age, activity status and gender. Thus, similarly to what we did for the (physical) time-cost, also in this situation we include a set of interaction variables. The variables that are interacted with the covariate 'app_delay' are;

- 'Bad_Sah' (equals one if the individual self-rated her/his health as bad): the motivation to include this interaction resorts on the assumption that for individuals in bad health status, the cost of waiting in a list is higher because the rate of depreciation of his/her health capital is also higher relative to other individuals. Therefore, for these individuals the threshold from which they quit from public care and substitute for alternatives is lower relative to the threshold of the other individuals (Barros and Olivella 2005),

- ‘Age’; Grossman (1972) claims that as the individual gets older the rate of health depreciation increases, thus, under this assumption, the cost of waiting — in terms of loss of well-being — in a list also increases with age. On the other hand, however, as the individual ages the loss of income due to waiting in a list (the individual is sick and possibly cannot work) decrease making elder people less responsive to delays. Thus, we have two forces pushing in opposite directions. The impact of age on the response to the ‘app_delay’ covariate is an empirical issue, we believe.
- ‘Non-active’; the reason why active and non-active individuals may put different estimates in the cost of waiting in a list is due to the possibility that individuals that work may lose income if the health condition impedes them from working,
- ‘Male’; as stated by Acton (1975) it is more probable that males may let their health deteriorate further than females do before seeking healthcare. Therefore, when males firstly seek care, their health condition is probably worse than the condition of females, thus, the cost of waiting (pain, anxiety, etc) for males may be higher than the cost of waiting for females, therefore we anticipate that the males respond more markedly to a delay in the consultation than the females.

Interaction variables	Variable definition
App_delay*non-active	Interaction between appointment delay and non-active
App_delay*age	Interaction between appointment delay and age
App_delay*male	Interaction between appointment delay and gender
App_delay*Bad_SAH	Interaction between appointment delay and health status

Table 33 — Interactions between ‘app_delay’ and some other covariates

We now turn our attention to the other group of covariates included in the analysis, the socioeconomic and demographic variables.

Socioeconomic and demographic variables

The utilization of this type of covariates, measuring individual characteristics, is standard for this type of literature. The motivation to use most of them as medical care determinants was provided in Chapter 5, therefore, here, we skip the full explanation why, and in which manner, they influence the utilization of GPs. Table 34 presents a description of the socioeconomic and demographic variables used in the regression models specified in this chapter.

Variable	Variable definition
Age	Age in years, divided by 10
Sqage	Square of age[/10]
Male	= 1 if the individual is male, 0 otherwise
Education (lower level of))	= 1 if the individual has at most 9 years of formal education
Non-active	= 1 if the individual reported that does not work
Married	= 1 if the individual is married
Widow	= 1 if the individual is a widow
North	= 1 if the individual resides in the North
Centre	= 1 if the individual resides in the Centre
LTV	= 1 if the individual resides in Lisbon and Tagus Valley area
Urb_Pop	Percentage of urban population present in the county where the individual resides [log in the linear index]

Table 34 — Socioeconomic and demographic covariates

We included the variables 'age', the square of age 'sqage'. As it was argued elsewhere, according to Grossman's (1972) model the depreciation rate of health capital increases with age, thus it is predicted that the utilization of health services increases as the individual ages. In the context of our analysis, however, this does not immediately signify that age is a factor that increases the utilization of public GPs because age can be thought as a factor that influences the substitution of visits at the

health centre for visits elsewhere, namely, private physicians and emergency rooms. In this regards, Rodriguez and Stoyanova (2004) found that older age is positively correlated to the probability of visiting a private GP. Although this evidence is for the Spanish population, we may assume that, at least, part of the health care centre users may present a similar behaviour. Therefore, under this hypothesis, because age influence substitution between types of providers, we anticipate that age presents a negative influence on health care centre utilization. The possible two contradictory impacts of age on utilization do not allow us to anticipate the effect of this covariate.

The next two covariates that appear in the table are 'male' and 'education'. The variable 'male' controls for the differences in behaviour between males and females, e.g attitudes towards risks, differences in life styles, utilization of preventive medical care, etc. Regarding the variable 'education', it is important to begin by noting that this is a dummy variable that equals one for the individuals with low levels of education, zero otherwise. The argument to use such covariate is that more educated individuals are more efficient health producers, therefore, they tend to use less medical care services (Jones, Rice *et al.* 2006). Thus, under this rationale, and given our definition of education, its coefficient is expected to be positive, meaning that the lower educated tend to use more often the care provided by the health centre. We cannot disregard, however, the impact of education on the substitution of visits at the health centre for visits elsewhere. In this regards, because the less educated individuals' present higher costs to acquire information about the private alternatives, it is possible to argue that they substitute less often the health care centre visits by the private alternatives. Moreover, the fact that the Portuguese health care system is based on a gate-keeping system to access specialized care combined with the finding (in own unpublished work) that less educated people present a lower probability to visit specialized doctors in the private sector, permit us to reason that the less educated people, contrarily to the more educated, rely more on the public GP to gain access to the specialized care.

Taken together this two effects, we expect that lower educated people to have more visits at the health centre.

A further socioeconomic determinant of medical care included in our model is a variable to account for the occupation status of the individual. The variable's designation is 'non-active' and equals one if the individual has declared that does not have a work in the labour market, either because he is retired or unemployed. We anticipate that this factor, to be non-active, increases the utilization of health care centres, basically because this class of individuals are less willing to substitute public care for private. We anticipate this effect occurs primarily because the opportunity cost of the non-active individuals is lower than the opportunity cost of the active ones in visiting the doctor at the times the health care centres provide the visits, usually at working hours.

The next two covariates included in the model are the dummy variables 'married' and 'widow' that control for the marital status of the individual. This type of covariates attempt to reflect possible differences in the individual's behaviour. The omitted category in this group of variables includes all individuals that are divorced or single.

The next three covariates incorporated in the model are intended to account for the place of residence. We consider the place of residence by using three dummy variables, 'North', 'Centre' and 'LVT'. These covariates equal one if the individual resides, respectively, in the North, Centre or Lisbon and Tagus Valley region. In this case, all individuals that reside in the south of Portugal (in the regions of Alentejo and Algarve) belong to the excluded category.

The last variable included in the socio-economic and demographic category is 'Urb_Pop' that represents the proportion of urban population present in the county where the individual resides. This variable was created with data provided by the Portuguese National Statistics Office and it was merged in our dataset (Instituto Nacional de Estatística 1999; Instituto Nacional de Estatística 2001).

The variable enters in the model to control for the urban *vs* rural status of the county where the individual resides. The variable can reflect differences in the supply of all types of medical care across urban and rural areas. In our view, in this case, the contribution of the substitution effect dominates over the effect of this variable on total demand. This result may be caused by the large geographic inequalities in the distribution of private medical care facilities, which tend to be concentrated in urban and coastal areas (Oliveira and Bevan 2003). Therefore, we anticipate that individuals residing in counties with higher proportions of urban population seek less care from health care centres.

Health status covariates

Another class of covariates that are rather standard in medical care utilization studies are the health status variables. Typically, health status is a variable that shifts the demand for medical services, with shifts occurring downwards for good health status and upwards for bad health status. We anticipate, naturally, that this pattern also replicate in the utilization of public GPs. We included nine covariates to control for the individual health status. Seven of them indicate the presence of a chronic health condition. Table 35 presents a description of the health status covariates included in the regression model.

Variable	Variable definition
Ill	= 1 if the individual has been felling ill in the days before the last visit to the GP
Bad_Sah	= 1 if the individual rated his self-assessed health as bad
Diabetes	= 1 if the individual has diabetes
Asthma	= 1 if the individual has asthma
Heart	= 1 if the individual has a heart disease
Digest	= 1 if the individual has a digestive disease
Hypertension	= 1 if the individual has high blood pressure
Osteo	= 1 if the individual has osteoporosis
Other_cd	= 1 if the individual has other chronicle diseases

Table 35 — Health Status covariates

We included the dummy variable 'ill' that equals one if the individual has been felling ill in the days before the last visit to the GP. In addition to the covariates reflecting chronic conditions, we also included the covariate 'Bad_Sah' that equals one if the individual self–rated his health as bad, 0 otherwise.

In the Europep questionnaire, the self-assessed health is measured by asking the patient about her/his general health: "*In general, you would say that your health is…"* . It uses a five-category Likert scale, with the following alternatives: Excellent health, very good, good, fair and bad. Contrarily to the case of the National Health Survey, where a high proportion of patients left the SAH question unanswered, in the Europep survey this was not the case, only a very low percentage of patients skip the SAH question, therefore, we do not run the risk of selection bias, at least from this source.

The existence of a chronic condition is measured through a set of dummy variables that indicates whether the individual suffers from one of the chronic conditions referred in the previous table. The regressor 'Other_cd' was included to signal the individuals that suffer from less frequent chronic conditions, such as, for example, cancer, psychiatric condition and headaches.

Characteristics of the Health Centre vs Family Doctor

The last group of covariates included in the model are those related to the Health Centre and/or to the Family doctor. They are presented in Table 36.

Variable	Variable definition
Walk	= 1 if the individual's mode of transportation is walking
Car	= 1 if the individual's mode of transportation is the car
Not_Enought_Time	= 1 if the individual felt that the GP doesn't give him enough time
Unsat_time_wr	= 1 if the individual is dissatisfied with the waiting time in the waiting room
Diff_book	= 1 if the individual thinks it is difficult to book a GP visit
D_spe_gp_phone	= 1 if the individual thinks it is difficult to speak with the family doctor by phone

Table 36 — Health Centre/Family Doctor covariates included in the time-costs regression model

The first two covariates listed in the previous table, 'Walk' and 'Car', are included to control for the transportation mean used to travel to the health centre. We note that when the mode of transportation is walk, it means that the person resides near the health care centre, not incurring, thus, in high monetary costs to access the centre, relative to all other categories. The omitted class is travel to the health care centre by taxi, by public transportation, by motorbike or by ambulance. Thus, we anticipate that the variable Walk exerts a positive influence in the utilization of public health care centres.

The next four covariates evaluate the patient's level of satisfaction about a specific aspect of the health care centre or the family doctor. In the Europep Questionnaire, the degree of satisfaction was measured through the utilization of a five point Likert scale with five categories. The categories range from *bad* to *excellent*. In this empirical application, we classified an individual as dissatisfied if she/he responded in the

leftmost two categories of the scale. Because all of these variables reflect dissatisfaction with some dimension of care, we expected that all of these to have a negative effect on the utilization of public GPs. It is easy to argue that dissatisfied people will be more willing to substitute the care provided in the health care centres by the medical care provisioned by the private physicians.

To facilitate the reading of this chapter we show, in Table 37, the complete set of covariates included in the regression models estimated in this chapter.

Variable name	Variable definition
Socioeconomic and Demographic	
Age	Age in years, divided by 10
Sqage	Square of age[/10]
Male	= 1 if the individual is male,
Education (lower level of)	= 1 if the individual has at most 9 years of formal education
Non-active	= 1 if the individual reported that does not work
Married	= 1 if the individual is married
Widow	= 1 if the individual is a widow
North	= 1 if the individual resides in the North
Centre	= 1 if the individual resides in the Centre
LTV	= 1 if the individual resides in Lisbon and Tagus Valley area
Urb_Pop	Percentage of urban population present in the county where the individual resides [log in the linear index]
Health Status	
Ill	= 1 if the individual has been felling sick in the days before the last visit to the GP
Bad_Sah	= 1 if the individual rated his self-assessed health as bad
Diabetes	= 1 if the individual has diabetes
Asthma	= 1 if the individual is asthmatic
Heart	= 1 if the individual has heart disease
Digest	= 1 if the individual has a digestive disease
Hypertension	= 1 if the individual has high blood pressure
Osteo	= 1 if the individual has osteoporosis

Variable name	Variable definition
Other_cd	= 1 if the individual has other chronicle diseases
Time-costs	
Total_Time	Total time (minutes) to visit the GP. [*travel time+waiting in the waiting room+treatment time*]. (Log included in the regression)
App_delay	= 1 if in the individual had an appointment delay, 0 otherwise
Characteristics of the Health Centre vs Family Doctor	
Walk	= 1 if the individual's mode of transportation is walking
Car	= 1 if the individual's mode of transportation is the car
Not_Enought_Time	= 1 if the individual felt that the GP doesn't give him enough time
Unsat_time_wr	= 1 if the individual is dissatisfied with the waiting time in the waiting room
Diff_book	= 1 if the individual thinks it is difficult to book a visit in the health centre
d_spe_gp_phone	= 1 if the individual thinks it is difficult to speak with the family doctor by phone
Interaction Variables	
(log)Total_time*non-active	Interaction between log of total time and non-active.
(log) Total_time*age	Interaction between log of total time and age.
(log) Total_time*square of age	Interaction between log of total time and age squared.
(log) Total_time*education	Interaction between log of total time and education.
App_delay*non-active	Interaction between appointment delay and non-active
App_delay*age	Interaction between appointment delay and age
App_delay*male	Interaction between appointment delay and gender
App_delay*Bad_SAH	Interaction between appointment delay and health status

Table 37 — Definition of complete set of explanatory variables selected to enter into the time-costs model

As is well known, multicollinearity problems may arise in utilization related models using a high number of covariates, which is the situation we have. To handle the issue, we inspected visually the correlation matrix of the covariates founding that the correlation between the covariates *age* and *non-active* was about 0,6. Moreover, most of all correlations were less than 0,3, with the majority having values less than 0,1. In preliminary experiments using other regressors, we found that the variable 'density of physicians' had a correlation of 0,6 and 0,7 respectively with 'logarithm of purchasing power' and 'logarithm of urban population'. We proceeded as Oliveira (2004) and eliminated the covariate that represented the density of physicians and the logarithm of purchasing power. After eliminating these variables, the problem of high correlations was overcome.

6.3.4 — Summary statistics and sample characterization

Before going any further on our analysis, we present a thorough characterization of the sample, with size N = 7.502. Table 38 gives the most common summary statistics, (e.g. average, standard deviation, minimum and maximum) for all explanatory variables involved in the analysis.

Variable	Mean	Std. dev.	Min.	Max.
Socioeconomic and demographic				
Age	5,515	1,882	1,8	9,8
Sqage	33,959	20,130	3,2	96,0
Male	0,344	0,475	0	1
Education (lower level of)	0,770	0,421	0	1
Non-active	0,605	0,489	0	1
Married	0,688	0,464	0	1
Widow	0,129	0,335	0	1
North	0,287	0,452	0	1
Centre	0,302	0,459	0	1
LTV	0,276	0,447	0	1

Variable	Mean	Std. dev.	Min.	Max.
Urb_Pop	0,778	0,259	0,142	1
Health status				
Ill	0,375	0,484	0	1
Bad_Sah	0,192	0,394	0	1
Diabetes	0,141	0,348	0	1
Asthma	0,132	0,339	0	1
Heart	0,208	0,406	0	1
Digest	0,330	0,470	0	1
Hypertension	0,181	0,385	0	1
Osteo	0,454	0,498	0	1
Other_cd	0,703	0,457	0	1
Time costs				
Total_Time	152,61	94,89	8	550
App_delay	0,324	0,468	0	1
Characteristics of the HC vs GP				
Walk	0,352	0,478	0	1
Car	0,457	0,498	0	1
Not_Enought_Time	0,215	0,411	0	1
unsat_time_wr	0,497	0,500	0	1
Diff_book	0,484	0,500	0	1
d_spe_gp_phone	0,414	0,493	0	1
[N = 7,502]				

Table 38 — Summary statistics of the explanatory variables

The average age of the respondents is about 55 years old. Figure 19 clearly illustrates that in the range 20 — 50 the empirical distribution of the covariate age does not depart too much from a uniform law, however, above 50 the distribution gains weight, revealing that, disregarding the contribution of other medical care determinants, the elderly tend to use the health centre more often than the younger cohorts.

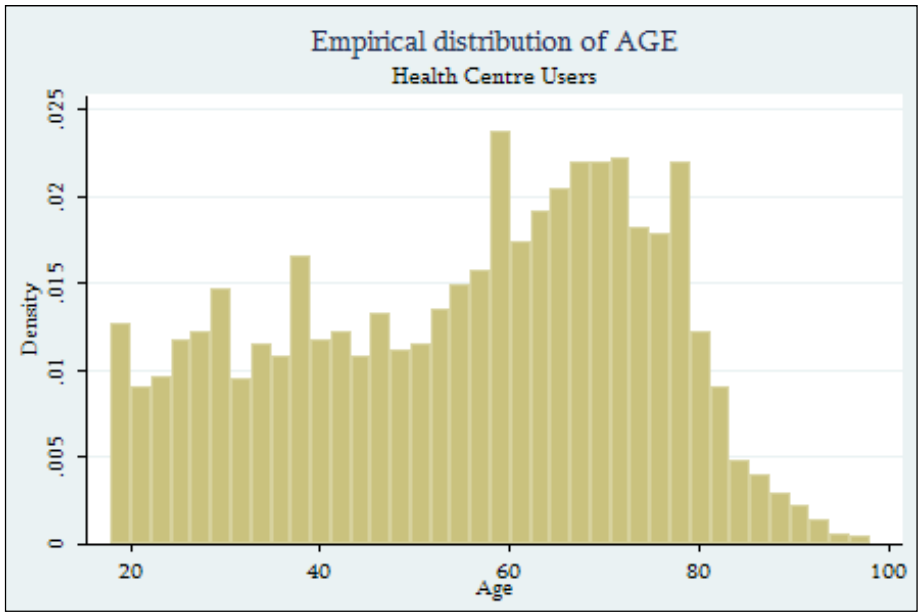

Figure 19 — Empirical distribution of 'age' of the health centre users

Males made up approximately 34% of the sample, which suggests that females are much more willing to seek medical care in public health care centres. Regarding education and working status, the data reveal that 77% of the individuals in the sample have a maximum of nine years of formal education and about 60% do not work, either because are retired or unemployed. Concerning the marital status of the individuals, the most common state is married, about 69%, with 13% of the individuals reporting widow as the marital status. The remaining 18% of the sampled individuals are divorced or single.

As regards to the region of residence, the numbers indicates that 28% of the individuals reside in the North of Portugal, about 30% in the Centre and 27% in Lisbon and Tagus valley region. Obviously, the remaining 15% of the individuals reside in the south of Portugal, either in Alentejo or in Algarve. The explanatory variable 'Urb_Pop' shows that health centre users tend to reside in counties more urbanized rather than in the rural parts. To sum up, in terms of socioeconomic and demographic characteristics, the average user of the health centre is a married female,

who resides in an urban county in the Centre region of Portugal, aged 60 —80 and presenting a low degree of education.

Turning our attention to the description of the health care centre users population in terms of health status, this group of variables show that, in general, the health centre users have much lower levels of health status compared with the health status of the general population. For example, Table 4 (which presents health status statistics for the general population, taken from the NHS98_99) shows that about 5% of the general population has diabetes. Notice that this figure climbs to 14% in the case of health centre users ($p < 0{,}001$). The same pattern is found for the Asthma, with 6,2% of the Portuguese population against 13,2% for the health centre users. The Elevated Blood pressure variable is the exception, presenting very similar figures in both populations, around 18%.

The figures in Table 38 also reveal that 37% of the individuals sought care in the health centre because she/he was felling ill, and that roughly 20% of the individuals self-rate her/his health as bad, which corresponds to the worst category available in the Likert scale used to assess individual health status.

The covariates reflecting time-costs variables show that, on average, each individual spent two and a half hours to visit the GP in the health centre. This number encompasses, as referred above, the travel time, plus waiting time in the waiting room and the time necessary to conclude the treatment with the doctor. Most of these 2,5 hours corresponds, however, to time spent in the health centre, both waiting and spent with the GP as the average travel time (one way) is about 20 minutes (Lourenço and Ferreira 2005). In addition, Table 38 shows that only about one-third of the patients experienced an appointment delay, showing that for the majority of the health centre users, their health care needs are timely satisfied on the health centre.

Regarding the last group of covariates, by far, the most usual transportation mean to the health centre is the car; about 46% of the individuals in the sample used this type

of transportation to the health centre, followed by walking, which was the choice of 35% of the patients. The variables that measure the degree of dissatisfaction with some aspect of the health centre indicates that, on average, 21% ('Not_Enought_Time') of the individuals think that the GP finishes the consultation too early and about 50% ('unsat_time_wr') of them are not satisfied with the physical waiting time in the waiting room. Moreover, 49% ('diff_book') of the patients consider that it is difficult to schedule a visit to their family doctor, while 41% ('d_spe_gp_phone') evaluates as difficult to talk with the family doctor by phone.

In the following section, we present the methodology adopted to address our research questions.

6.4 — Specification of a model to estimate the effect of time-costs on the utilization of health centres

In this section, we present and discuss some issues related to the specification of a suitable model to analyse and respond to the research questions that are essayed in the beginning of the current chapter.

Surely, the selected specification must account for the specificities of the sampling scheme chosen to collect the data, as well as the integer nature of the dependent variable, which is the number of visits made to the health centre in one-year period.

In the majority of the studies dealing with the specification and estimation of models to explain health care utilization, the researcher has available a random sample of the underlying population of interest. In our application we do not have one. Therefore, it is important that the sampling issues raised by the design protocol used to collect the data are not ignored, otherwise the results and, thus the conclusions, that come out from the models, would run the risk of being inappropriate and misleading.

As is pointed out by Santos-Silva (2003) and Wooldridge (2002), when the probability of being selected to be included in the sample depends on the endogenous

variable, it is fundamental to integrate the non-random features of the data in the specification process. The case is that under endogenous sampling the data do not represent appropriately the population of interest.

As explained in the previous section, our sampling design excludes the individuals with zero visits, thus we did not observe the entire distribution of the dependent variable. We recall that we selected a sample from the population of individuals that visited the doctor in the last three months of 2003. Hence, only the information about the individuals presenting a positive number of visits are included in the sample. Also note that in this sample scheme context, inclusion in the sample requires that the individual has had at least one visit to the health centre in the last three months of 2003. Therefore, the information about the individuals that did not use the health centre is totally lost. For this reason, to get unbiased estimates of the impact of time-costs on GP utilization, we must specify and estimate econometric models specialized to deal with truncated data. In this section we skip the details about the corrections that standard count data models have to go through to adequately handle truncated(-at-zero) data, as this was the subject of Section 2.7 (please, see page 45 for details).

Another issue that we must address, and cannot disregard in the specification of the econometric models, is the presence of unobserved heterogeneity. Most likely, the set of covariates selected as determinants of GP utilization do not fully account for all individual heterogeneity, hence, the models specified to analyse our research question should explicitly incorporate the individual factors that remain unobserved after the inclusion of the explanatory variables.

In Section 2.8 (please, see page 47) we explained some of the challenges that the empirical researcher faces when aims at specifying an econometric model to explain truncated(-at-zero) data in the presence of unobserved heterogeneity. Assuming the presence of such unobserved heterogeneity of discrete nature, in that section we explained, and defended, that the specification of a LCM in a truncated-at-zero dataset

could be made using two competing probability models, given by equations [39] (page 50) and [40] (page 50).

Moreover, there, we defended that the two different specifications correspond to two different assumptions regarding the unobserved factors. On the one hand, specification [39] corresponds to the assumption that the unobserved factors are relative to the individuals belonging to the actual population, on the other, the specification given by [40] corresponds to the assumption that the unobserved factors incorporated in the model are relative to the individuals in the truncated population. Therefore, at this point we should ask, and find an answer, about which specification, [39] or [40], should be used to model our data, in order to adequately respond to the research question?

As already explained previously in this dissertation, the response to the question depends critically on the population that the researcher is interested in, as well as the population that the data allows one to reach. Throughout this dissertation we have been defending that the hurdle framework is appropriate to explain individual choices regarding health care utilization. The question that now arises[59] is whether the hurdle model constitutes the appropriate structure to understand which factors explain the utilization of public health care centres. In our view, the answer is clearly positive; the hurdle framework suits perfectly as the appropriate framework to represent the utilization of public health care centres in Portugal. The reasons why we take this view are spelled out in the next paragraphs.

In a previous chapter we defended that one assumption implicit in the hurdle framework is that the overall population under study is a mixture of two subpopulations: the subpopulation of non-users of health care and the subpopulation

[59] The question is relevant to help overcoming the difficulty of choosing the appropriate specification to analyse the Europep data.

of users, each being described by a different statistical process. In the context of the Portuguese primary health care system, the assumption of a division between users and non-users are in complete accordance with the reality. To attest to this note that there is a segment of the Portuguese population that do not use the public health care centres, at least in the consultations dimension. Those are individuals covered by a health subsystem (either the ADSE or OSS) or by a voluntary health insurance schemes. These individuals are usually characterized as the better-off both in financial and in health status condition. Generally, due to his/her more comfortable economic status, tend to prefer the consultations supplied by doctors practising in the private sector. Those features might imply that the population that uses public primary care consists of individuals without supplementary health insurance coverage, which, combined with the, possible, lower economic capacity, do not have many alternatives outside the public sector.

Therefore, in our view, and supported by the arguments already presented, the hurdle model is a suitable model in our context. The first stage of the model would unveil the factors influencing the decision to be a health care centre user, while in the second part, would unveil the factors influencing the frequency of utilization for those who have decided to use the health care centre.

To estimate the two parts of the hurdle, however, a random sample of the entire population of interest should be available, though this is not our case, as it was previously described. Our sampling scheme only allows us to observe the population of health centre users, therefore the estimation of the two parts of hurdle model are not possible. We are constrained, thus, to estimate only the second part of the hurdle, the part that explains the intensity of health care utilization of the health centre users.

This reasoning leads us to the conclusion that the proper specification to address whether time-costs influence health care centre utilization or not is the specification that assumes that the unobserved factors incorporated in the model belong to the

individuals present in the health centre users population, thus, making assumptions about the distribution of the unobserved heterogeneity in this smaller population.

To review, in this chapter we specify models defined by a finite mixture of P truncated distributions, whose conditional probability function is given by

$$f_s\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right) = \sum_{j=1}^{P}\left(\frac{f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right)}{P_j\left(y_i > 0 \mid \mathbf{x}_i, \boldsymbol{\beta}_j, \upsilon_i^j\right)}\right)\pi_j \qquad [79]$$

where $f_s\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}\right)$ is the distribution in the truncated sample, and $f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right),\ j = 1,2,...,P$ are the parent distributions of the LCM. Moreover, the $\mathbf{x}_i$ is a vector of explanatory variables, defined in Table 37, and $\boldsymbol{\beta}_j,\ j = 1,2,...,P$ is a vector of parameters to be estimated.

The conditional mean function of y_i in the overall health care centre users population is defined as,

$$E\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j, y_i > 0\right) = \sum_{j=1}^{P}\left[\frac{\exp\left(\mathbf{x}_i'\beta_j\right)}{P_j\left(y_i > 0 \mid \mathbf{x}_i, \beta_j\right)}\pi_j\right] \qquad [80]$$

where $\exp\left(\mathbf{x}_i'\beta_j\right)$ is the assumed specification for the conditional mean of the parent distribution $f_j\left(y_i \mid \mathbf{x}_i, \upsilon_i^j, \boldsymbol{\beta}_j\right),\ j = 1,2,...,P$ and $P_j\left(y_i > 0 \mid \mathbf{x}_i, \beta_j\right)$ represents the probability of y_i being greater than zero.

Moreover, this specification also allows one to obtain the conditional average of the truncated distribution by latent class, given by,

$$E_j\left(y_i \mid \mathbf{x}_i, \beta_j, y_i > 0\right) = \frac{\exp\left(\mathbf{x}_i'\beta_j\right)}{P_j\left(y_i > 0 \mid \mathbf{x}_i, \beta_j\right)} \qquad [81]$$

Regarding the form of parent distribution and the number of latent classes, we use respectively, both versions of the NB and only two latent classes. It is worth to mention that the decision about these aspects was not made based on educated guesses, but after making a number of experiments estimating several competing models. For instance, we estimated models with P = 2, using, however, different parent

distributions as representative of each latent class, a combination of the Poisson with the negative binomial. Moreover, we also estimated latent class models setting P = 3. Using information criteria instruments to choose among the competing models, namely BIC and CAIC statistics, we conclude that these more complicated models did not present any substantial improvement relative to the others. In fact, Lourenço and Ferreira (2005) provided the results for models with three latent classes and pointed out that, based on the full sample both BIC and CAIC indicators support the two-component negative binomial II latent class model. Thus, rejecting the models specified using three latent classes.

Because we decided to use the NB probability model as baseline distribution to specify the LCM, the competing models estimated in this chapter are given by,

$$f_s\left(y_i \mid \mathbf{x}_i, \beta_1, ..., \beta_j, y_i > 0\right) =$$

$$= \sum_{j=1}^{2} \left(\frac{\Gamma\left(y_i + \eta_{ij}\right) \eta_{ij}^{\eta_{ij}} \left(\lambda_{ij} + \eta_{ij}\right)^{-\left(\eta_{ij} + y_i\right)} \lambda_{ij}^{y_i}}{y_i! \Gamma\left(\eta_{ij}\right) \left(1 - \left(\frac{\eta_{ij}}{\lambda_{ij} + \eta_{ij}}\right)^{\eta_{ij}}\right)} \pi_j \right) \quad y_i = 1, 2... \qquad [82]$$

Thus, under our choices for the parent distributions, the conditional mean function of $\left(y_i \mid y_i > 0\right)$, is given by

$$E\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p, y_i > 0\right) = \sum_{j=1}^{2} \left[\pi_j * \frac{\exp\left(\mathbf{x}_i' \boldsymbol{\beta}_j\right)}{1 - \left(\frac{\eta_{ij}}{\lambda_{ij} + \eta_{ij}}\right)^{\eta_{ij}}} \right] \qquad [83]$$

Accordingly, the conditional mean function for the latent class *j* in the users population, is given by,

$$E_j\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_j, y_i > 0\right) = \frac{\exp\left(\mathbf{x}_i' \boldsymbol{\beta}_j\right)}{1 - \left(\frac{\eta_{ij}}{\lambda_{ij} + \eta_{ij}}\right)^{\eta_{ij}}} \quad j = 1,2 \qquad [84]$$

In the above expressions, $\lambda_{ij} = \exp\left(\mathbf{x}_i' \boldsymbol{\beta}_j\right)$ $j = 1,2$ and, $\eta_{ij} = \left(\frac{1}{\alpha_j}\right)\lambda_{ij}^k$ $j = 1,2$. By setting k = 0 or k = 1, as explained in Chapter 2, Section 2.4, we obtain two different, ands alterative, specifications of the NB model.

For now, we leave the specification issues and proceed to the analysis of the results, which is the purpose of the next section.

6.5 — Empirical results and discussion

The current section presents and discusses the main empirical results that came out from the estimation of various competing models estimated in this chapter. We start by presenting and discussing the results of the model selection methods, followed by the presentation of an economic interpretation of the estimates found, naturally emphasizing the analysis of the results of the time-cost covariates.

Four sub-sections constitute the current section: the first subsection (6.5.1 — Evaluation of fitted models and model comparison) shows the results of a set of tests to select the specification presenting the better performance in fitting the data; the next one (6.5.2 — Analysis of the latent classes of health centre users) provides a global analysis of the LCM estimates while the third subsection, (6.5.3 — Analysis of the effect of the time-cost) presents a thorough analysis to verify whether time-costs matter in the health care centre utilization; finally, the last subsection (6.5.4 — Analysis of the impact of diverse covariates) briefly summarizes the effect of other explanatory variables in the utilization of health care centres.

6.5.1 — Evaluation of fitted models and model comparison

The purpose of this sub-section is to select the model with better goodness-of-fit. The model comparison and selection techniques adopted are the ones presented in Chapter 2, which are the likelihood ratio tests (LR tests) to choose among nested competing models, Vuong tests and information criteria statistics to select among the non-nested specifications.

Table 39 shows the acronyms of the six competing models specified and estimated, in addition to a short description of each model and the respective number of parameters that each model involves. The table also provides the log-likelihood values of the different models.

Acronym	Model Description	Param	LogL
TNB1	Truncated negative binomial (type 1) regression model	38	-19.600
LCM_TNB1	Two component LCM with TNB1 as baseline model	77	-19.284
R-LCM_TNB1	Constrained version of LCM_TNB1, obtained by imposing that the slopes to be equal across latent classes	41	-19.322
TNB2	Truncated negative binomial (type 2) regression model	38	-19.576
LCM_TNB2	Two component LCM with TNB2 as baseline distribution	77	-19.264
R-LCM_TNB2	Constrained version of LCM_TNB2, obtained by imposing that the slopes to be equal across latent classes	41	-19.288

Table 39 — Description of the competing models estimated

We set out by noting that the specifications based on the latent class approach present a much higher log-likelihood values than the uni-component/single index

models. This result provides the first indication, though weak, that the LCM specifications may outperform the simple one-component truncated models, however, some formal statistical tests are necessary to reach a more robust conclusion.

Beginning with the formal statistical tests to compare the nested specifications, Table 40 reports the results of LR tests conducted to contrast the uni-component model versus the two-component models — the tests labelled LR1 and LR2 — and the comparison of the most general form of the LCM specification with its constrained version — the tests labelled LR3 and LR4.

Test	Null	Alternative	LR Stat	D.F.	Decision
LR1	TNB1	LCM_TNB1	632	39	LCM_TNB1
LR2	TNB2	LCM_TNB2	624	39	LCM_TNB2
LR3	R-LCM_TNB1	LCM_TNB1	76	3	LCM_TNB1
LR4	R-LCM_TNB2	LCM_TNB2	48	3	LCM_TNB2

Table 40 — Results of likelihood ratio tests

The results of the tests LR1 and LR2 clearly point to the rejection ($p < 0.01$) of the uni-component truncated specifications (TNB1 and TNB2) in favour of the models based on the latent class framework, despite that in those cases the distribution of the LR test is non-standard and using the usual χ^2 distribution is likely to under-reject the null hypothesis (Bohning 1995; Deb and Trivedi 1997).

Focusing now on the comparisons of the models based on the latent class framework, another statistical hypothesis tested, whose results are shown in the tests labelled LR3 and LR4, are that that evaluates whether the slopes of the conditional mean functions are equal across latent classes. Based on the results of the tests, the null hypothesis is rejected in both situations, for a $p < 0.01$. This finding suggests that,

when considered jointly, the explanatory variables considered in the model exert a different effect in the two latent classes of health care users considered.

Summing up, the LR tests rejected both the uni-component models (models TNB1 and TNB2) in addition to the constrained versions of the latent class models (specifications R-LCM_TNB1 and R-LCM_TNB2), evidencing that to fit adequately the data one must use the most general LCM specification, either based on the NB1 or on the NB2 distribution.

In what follows, we continue by showing the results of the model selection techniques based on information criteria. The goal is to choose the best model out of the four competing models estimated, specifications TNB1, LCM_TNB1, TNB2 and LCM_TNB2. Table 41 provides the values for the BIC and CAIC information criteria.

Model	BIC	CAIC
TNB1	39.539	39.577
LCM_TNB1	39.255	39.332
TNB2	39.491	39.529
LCM_TNB2 ♣	39.215	39.292

Table 41 — Information criteria statistics

The figures presented in the previous table are completely consistent with the findings of the LR tests shown above. As is readily seen in Table 41, the single index models TNB1 and TBN2 generate much larger BIC and CAIC criteria functions than the LCM based alternatives. Therefore, the reliance on the information criteria

♣ Model preferred by BIC and CAIC criteria.

techniques also leads to the indication that the single index models TNB1 and TNB2 are not the preferred specifications.

Regarding the selection between the models LCM_TNB1 and LCM_TNB2, the information criteria statistics shown in the table clearly point towards the superiority of the LCM specification based on the NB2 distribution, as for this model, both the BIC and the CAIC criterion functions present a lower value. In short, the model selection based on information criteria evidences that the single index models are outperformed by the more complex specifications. In addition, it also indicates that the latent class alternative based on the NB2 is the preferred model when compared to the LCM based on the NB1 probability model.

In spite of the evidence presented above, which has favoured the LCM based on the NB2 over the LCM based on the NB1 probability model, this conclusion is not entirely consistent with the result of a Vuong test contrasting the two competing specifications. In fact, the value of the Voung's test statistics for the hypothesis LCM_TNB1 against the LCM_TNB2 is -1,48. Therefore, for the standard level of significance of 1% the null hypothesis of equivalent models is not rejected, meaning that, given the data available, this testing methodology is unable to discriminate between the two models. We note, however, that the Vuong statistics value of -1,48 corresponds to a *p* value of approximately 0,07. Hence, if one is willing to accept a less conservative standard significance level of 10%, then the null hypothesis of equivalent models is rejected in favour of LCM_TNB2 being better than LCM_TNB1.

Summing up the results of this sub-section, we may conclude the following: taking together all the results analysed, and discussed above, the conclusion is that the two-component LCM with TNB2 as parent distribution, referred to as LCM_TNB2, is the one that is supported by the data. Therefore, these results seem to evidence that at least two different stochastic processes are necessary to describe adequately the utilization of the GP consultations in the Portuguese health care centres.

From another perspective, the model comparison and selection results also suggests that to appropriately explain the frequency of GP visits for the overall population of health centre users, one has to adopt flexible stochastic specifications able to properly incorporate the inherent unobserved heterogeneity, supposedly present in the health centre users population. In line with the arguments provided in Chapter 4, here too, we highlight that the motivation to choose the negative binomial distribution as baseline model in the LCM specification is not due to its capability of incorporating the presence of unobserved heterogeneity. Our framework of analysis resorts on the assumption that the unobserved heterogeneity, presumably present in the health care centre users population, is adequately treated in the LCM approach adopted. Our choice for models of the negative binomial family was merely because this class of distributions is one of the most widely used and flexible to deal with count data (Deb and Holmes 2000).

Since the two-component latent class model with TNB2 as baseline distribution (model LCM_TNB2) appears as the one that better fits the data, in the sub-sections that follows we only present and discuss the estimates that resulted from the estimation of this model. Table 42 reports the parameter estimates, the corresponding robust standard errors and the z-statistic for the preferred specification.

Model LCM_TNB2	Latent Class I			Latent Class II		
	π = 0,80 (se = 0,04)			**1 - π** = 0,20 (se = 0,04)		
N = 7502, - Logl = 19,264	**Coef.**	**S.E**	**Z**	**Coef.**	**S.E.**	**z**
Constant	1,878	0,870	2,160	5,300	2,854	1,860
Socioeconomic and demographic						
Age	-0,094	0,377	-0,250	-2,630	1,192	-2,210
Sqage	-0,004	0,034	-0,130	0,250	0,106	2,350
Male	-0,113	0,037	-3,050	-0,286	0,123	-2,330
Education (lower level)	-0,172	0,353	-0,490	1,432	1,107	1,290
Non-active	0,280	0,282	0,990	0,652	0,818	0,800
Married	0,036	0,045	0,810	0,052	0,139	0,370
Widow	0,050	0,058	0,850	-0,031	0,190	-0,160
North	0,008	0,049	0,170	0,046	0,152	0,300
Centre	0,067	0,047	1,420	0,247	0,152	1,630
LTV	-0,022	0,048	-0,460	0,062	0,163	0,380
Urb_Pop (log)	-0,130	0,033	-3,930	0,100	0,114	0,880
Health Status						
Sick	0,047	0,027	1,710	0,054	0,094	0,570
Bad_Sah	0,289	0,047	6,140	0,096	0,163	0,590
Diabetes	0,135	0,040	3,400	0,055	0,139	0,400
Asthma	0,192	0,036	5,390	0,236	0,142	1,660
Heart	0,216	0,033	6,640	0,209	0,117	1,790
Digest	0,070	0,035	2,010	-0,019	0,127	-0,150
Hypertension	0,133	0,029	4,560	-0,115	0,113	-1,020
Osteo	0,039	0,036	1,080	-0,166	0,115	-1,440
Other_cd	0,160	0,040	3,960	0,301	0,136	2,210
Time-costs						
Total_Time (log)	-0,140	0,183	-0,770	-0,849	0,572	-1,480
App_delay	-0,378	0,105	-3,600	0,188	0,360	0,520
Health Centre vs Family Doctor						

Model LCM_TNB2	Latent Class I			Latent Class II		
	π = 0,80 (se = 0,04)			**1 - π** = 0,20 (se = 0,04)		
N = 7502, - Logl = 19,264	**Coef.**	**S.E**	**Z**	**Coef.**	**S.E.**	**z**
Walk	-0,036	0,039	-0,940	-0,141	0,132	-1,070
Car	-0,104	0,038	-2,760	-0,188	0,126	-1,490
Not_Enought_Time	-0,154	0,043	-3,620	0,080	0,121	0,660
Unsat_time_wr	0,007	0,035	0,210	-0,006	0,124	-0,050
Diff_book	-0,123	0,034	-3,610	-0,031	0,129	-0,240
D_spe_gp_phone	-0,093	0,032	-2,950	0,092	0,118	0,790
Interaction Variables						
(log)Total_time* non-active	-0,014	0,058	-0,240	-0,089	0,166	-0,540
(log) Total_time * age	0,018	0,080	0,220	0,539	0,243	2,220
(log) Total_time * sqage	0,0002	0,007	0,030	-0,051	0,022	-2,340
(log) Total_time * education	0,074	0,076	0,980	-0,252	0,227	-1,110
App_delay*non-active	-0,075	0,080	-0,930	0,302	0,250	1,210
App_delay*age	0,054	0,023	2,360	-0,064	0,070	-0,920
App_delay*male	-0,071	0,063	-1,120	-0,264	0,212	-1,250
App_delay*Bad_SAH	-0,057	0,074	-0,760	0,023	0,259	0,090
α	0,337	0,042	7,960	3,214	1,124	2,860

Table 42 - Parameter estimates, robust standard error and the z-statistic for the two-component LCM based on the truncated negative binomial (type 2) model.

Regarding the number of latent groups, our results are similar to those obtained in the analysis that has adopted specifications of the LCM family to fit the data. In general a number of studies have found that two latent class provided enough flexibility to fit data with quality (Deb and Trivedi 1997; Deb and Holmes 2000;

Gerdtham and Trivedi 2001; Deb and Trivedi 2002; Jimenez-Martin, Labeaga *et al.* 2002; Bago d'Uva 2005; Sarma and Simpson 2006). Concerning the parent probability models to define the LCM, still in line with other studies, we have also found that the densities from the NB family emerge as the most appropriate distribution. In fact, we specified and estimated latent class models using the Poisson probability function as component distribution (the results are not shown in this dissertation) and concluded that those models were outperformed by the LCM specifications with the NB used as baseline distribution.

6.5.2 — Analysis of the latent classes of health centre users

Before enter specifically in a detailed analysis and discussion regarding the subject that has primarily motivated this empirical study, we set out by presenting and discussing the results relative to the estimates of the mixing probabilities of the LCM estimated. These mixing probabilities, combined with an analysis of the expected health care centre utilization by latent class, surely, allow us to verify the feasibility of our prior reasoning that the overall population of health centre users can be divided into two latent classes. Thereafter, in a separated section, we present the results of a posterior analysis of class membership that we have conducted in an attempt to characterize the type of individuals belonging to each latent class.

Turning now to the analysis of the results, the mixing probabilities, as estimated by the model LCM_TNB2, are $\pi = 0{,}8$ both presenting a standard error of 0,04, thus being statistically different from zero[60]. This result, combined with the findings

[60] Despite the problems with testing the null hypothesis $\pi = 0$ using standard tests. Notice that the test is conducted in the boundary of the parameter space (Deb and Trivedi, 1997).

reported in the previous sub-section[61], reinforces the conclusion that the overall health centre users population comprises two latent sub-populations. The finding of this section adds, however, that the latent classes comprise different proportions of individuals, one largest latent class encompassing about 80% of the health care centre users and a second latent class, much smaller than the first one, encompassing the remaining 20% of the global population of patients of the health care centres.

A question that one can raise at this point is whether the two latent classes, already established, are different in respects other than its size, for example: do the individuals in each latent class present different behaviours concerning the intensity of utilization of the care provided at health care centres? To address this question, we estimated the distribution of the predicted conditional means of each latent class.

Table 43 shows the estimated sample average of the distribution of the fitted means, for each latent class and for the overall health care centre users population, along with some extra statistics conveying information about such distribution.

	Latent Class I (80%)	Latent Class II (20%)	Population
Mean	5,0	9,8	5,8
Minimum	2,0	3,5	2,4
Maximum	16,7	29,7	17,2
Percentile			
10%	3,1	6,1	3,7
25%	3,8	7,4	4,4
50%	4,7	9,6	5,5
75%	5,9	11,5	6,8
90%	7,5	14,0	8,4

[61] The finding that we emphasize from the previous section is the evidence that the model LCM_TNB2, a model based on latent classes, outperformed all single index competing specifications.

Table 43 — Summary statistics of the distribution of fitted means by latent class and population

From the analysis of the estimates presented in the previous table, we conclude that the patients of latent class I, about 80% of the population, present an expected health care utilization of approximately 5 visits a year, while the remaining 20% visits the doctor, on average, almost twice that number, about 9,8 visits in each year. Hence, the two underlying latent populations identified are different in respects other than the number of individuals contained in it, namely, in the intensity of utilization.

Notice that from Figure 20, which shows three box-plots of the distribution of the expected means by latent class and population, it is noticeable that the distribution of the fitted means of latent class I (the leftmost box-plot in Figure 20) shows a clear tendency to predict systematically lower fitted means than those that are found in the latent class II (in Figure 20 the middle box-plot).

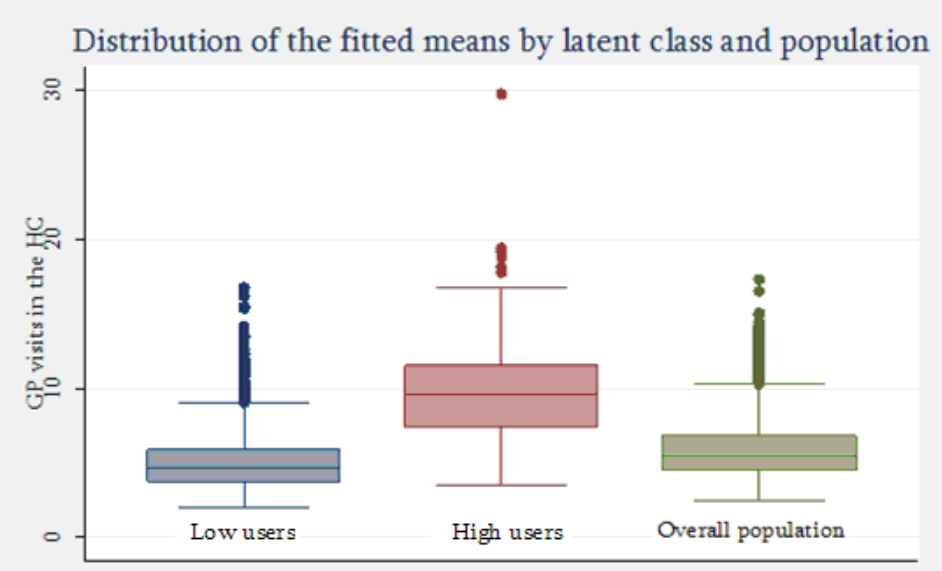

Figure 20 — Box-plots of the distribution of fitted means, by latent class and population (model LCM_TNB2)

Therefore, we may conclude that the overall population of users health care centres can be split into two latent clusters of individuals. On the one hand, one first cluster, which encompasses about 80% of the population, presents an average utilization of about 5 visits, on the other, a second cluster, comprising the remaining 20% of the individuals, visiting the doctor a little less than 10 times a year. Therefore, exclusively based on the expected utilization predicted by the model, our proposal is to label the users of latent class I as the low health care centre users and the individuals in the second latent class as the high health care centre users. In the analysis that follows, we adopt this terminology to describe the latent classes implied by the model.

6.5.3 — Analysis of the effect of the time-cost covariates

This subsection addresses in detail the main research question of this chapter, by analysing, at the outset, the effect of the existence of an appointment delay in the consultation ('App_delay' covariate), moving afterwards to the analysis of the impact of waiting time ('Total_time' covariate) on the utilization of visits in public health care centres.

6.5.3.1 Analysis of the effect of 'app_delay'

The analysis conducted here builds upon the estimation and on the analysis of the distribution of the marginal effects implied by this covariate[62].

[62] We recall that we calculated the individual's response to the existence of an appointment delay as the difference between the average utilization when the 'app_delay' covariate varies from one to zero.

To begin with, Figure 21 depicts two box-plot charts showing the distribution of the individual effects of the 'App_delay' in the utilization, for the low and for the high health care centre users.

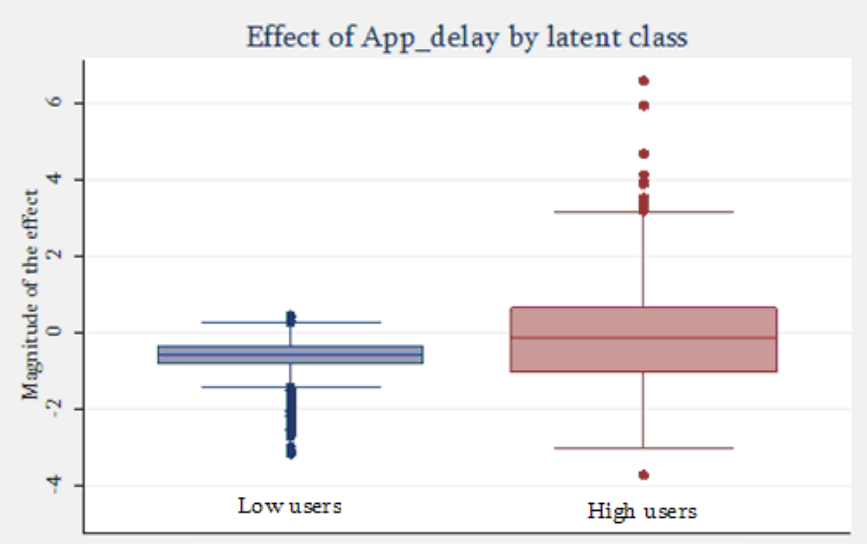

Figure 21 — Box-Plot of the distribution of the marginal effects of 'App_delay' by latent class

The simple visual inspection of the distribution for the low users latent class indicates that the magnitude of the individual's reactions to the delay in the appointment is mostly located below zero, meaning that most of the patient in the low users react negatively in face of a delay in the consultation appointment. Actually, after some further calculations, we have found that roughly 98% of the low users present such negative impact. It is worth to report that the marginal effects estimated as positive do not present statistical significance at the standard levels. On the contrary, approximately 75% of the estimates located within the percentiles 25th and 75th present a $p < 0.05$, thus being statistically different from zero. Given these results, we may conclude that the majority of the individuals in the low users latent population that experience a delay in the consultation tend lower the utilization of health care centres.

Concerning the response of the high users to the occurrence of a delay in the consultation, the right-most box-plot in Figure 21 illustrates the distribution of the marginal effects. The box-plot chart evidences that approximately 50% of the individuals of this latent sub-population, apparently, increase the number of consultations in face of an appointment delay, what can be considered somewhat surprising and totally contrary to the economic reasoning. The strangeness of the result, however, vanishes after computing and analysing the standard errors of the individual responses (not reported in this dissertation). We have observed that, in general, the magnitude of the standard errors are big comparatively to the magnitude of the effect, originating, obviously, p-values > 0,05. Accordingly, the individuals' responses to the appointment delay are, from a statistical point of view zero, evidencing that the high user's utilization of health care centres is independent of the delay imposed by the health centre.

After the analysis of the individual responses to the lack of timely consultations, we turn our interest to the analysis of the average marginal effect (AME). Table 44 shows the average marginal effect (AME) of the low users latent population, of the high users population and, finally, of the overall population of health centre users.

	Low users (80%)	High users (20%)	Population
AME	-0,64 (0,13)	-0,001 (0,95)	-0,56 (0,20)

Table 44 — Estimates of the overall effect of 'App_delay' in the utilization of health care centres by latent class and overall population[63]

[63] The estimates in the shaded cells are statistically significant (p < 0,01)

As expected, given the information released by the distribution of the marginal effects already analysed, for the low users population we found a negative average marginal effect of -0,64 with a standard error of 0,13, thus the estimate of the AME is statistically different from zero. Therefore, this finding indicates that overall, this latent sub-population, reacts negatively to the delays in the appointment, lowering utilization. A similar conclusion is obtained for the global health care centre population, where we found that the 'app_delay' variable cause the individuals to lower utilization in 0,56 consultations, being this estimate statistically significant for a $p < 0.01$.

On the contrary, for the high users population, despite the negative sign of the average marginal effect (-0,001), their magnitude decreased to the neighbourhood of zero, losing also the statistical significance. This result is, hence, a sign that the occurrence of an appointment delay in the consultation is not a factor that the high health care centre users consider when deciding about the intensity of utilization of the health centre.

As we have pointed out in Section 6.3 (at page 241), all specifications consider the possibility that some individual (observed) characteristics might influence the individual's response to the occurrence of an appointment delay, such characteristics being occupation status, gender, health status and age.

The figure presented below, depicts several box-plots conveying information about how activity status ('non-active' = 1 vs 'non-active' = 0), gender ('male' = 1 vs 'male' = 0), health status ('bad_sah' = 1 vs 'bad_sah' = 0) and age ('age' = 20 [10] 90) influences the individual's reaction to a delay in the health care centre consultation, for the low users population.

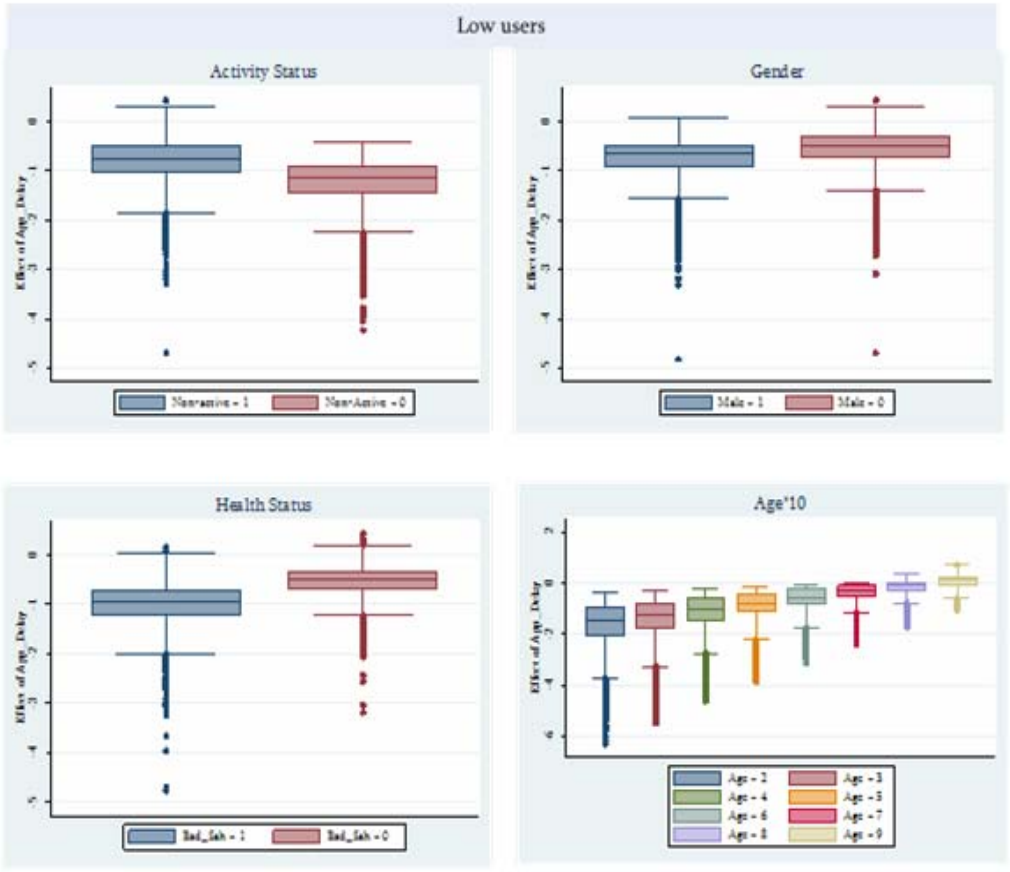

Figure 22 — Distribution of the impact of activity status, gender, health status and age on the effect of 'app_delay' for the low users population

The figure displayed in the upper left position of suggests that the individuals whose occupational status is active (*non-active* = 0, box-plot at right) present a clear tendency to respond more markedly to delays in the health care centre consultations relatively to the non-active (*non-active = 1*) cohorts. Despite the lack of statistical significance of most of the effects of *non-active* on the individuals' response to a delay, the result it still interesting because it provides evidence that the cost of waiting in a waiting list is higher for the employed individuals. Concerning the impact of gender ('male') on the individual's response to the 'app_delay', the graph does not reveal any noticeable differences between genders, possibly meaning that males and females evaluate the cost of waiting in a waiting list in similar terms. Turning our interest to

the impact of health status ('Bad_Sah') on the individual's response to the 'app_delay', the graphics displayed in the lower left position at, indicates that the worse-off individuals ('Bad_Sah' = 1, box-plot at left) tend to respond more noticeably to delays in the health care centre consultations relatively to the individuals in better health status. This finding shows that the cost of waiting in a waiting list is, probably, higher for the individuals in bad health status, which endorses our claim made previously.

To close the analysis of the immediately previous figure, we address the empirical question of how age influences the individual's response to an alteration in the 'app_delay' variable. The graph presented in the southeast corner depicts the way age influences the response to a delay in the consultation, as age varies from 20 year old to 90 year old, with a pace of 10 years. From the graph it is clear that as the individual gets older it becomes progressively less reactive to delays in the consultations. Moreover, it is worth to note that at old ages the reaction to delays are statistically zero, while in general, the young cohorts present negative reactions with statistical significance.

We carried out a similar analysis to the one presented above, but for the high health centre users population. We did not find any relevant differences.

Summing up, the analysis of the effect of 'app_delay' on the health care centre utilization allowed us to draw the following conclusions:

For the low health centre users;

- overall, those that cannot schedule a timely appointment at the health care centre lower, in average, their health centre utilization,
- as the individual gets older the magnitude of the response to a delay decreases: to values around zero for the individuals aged approximately above 70,
- individuals that reported their health status as bad and those whose activity status are on tend to respond more noticeably to the occurrence of delays in the health care centre consultations,

- gender does not appear to influence the individual's response to the occurrence of a delay, signifying that, the response of man and woman to an appointment delay is similar.

For the high health centre users;

- the explanatory variable 'app_delay' does not appear to influence the decisions about the number of visits,
- finally, the covariates age, gender, health status and occupation status do not present any statistically significant effect on the response to a delay in the appointment.

We now start the discussion of the empirical results reported thus far. Elsewhere in this dissertation, we have argued that waiting in a list for medical care, even a short waiting, may impose a cost to the patient in terms of both pain, anxiety, in general loss of well-being, and may also lead to loss of income because the illness may impede the individual from working while is waiting. We have also argued that as the waiting time in a list increases the shadow cost of waiting also increases, thus the relative cost of contacting an alternative physician is becoming lower. Therefore, it is likely to observe shifts in the willingness to substitute medical care in public health care centre for private care or emergency care, hence, we expected the appointment delay parameter to be negative.

Empirically, our results agree partially with the predictions presented. We have shown that a segment of the health care centre population, the low users, evaluate the time spent in a list as a cost, while the other share of the population, the high users, do not respond, at all, to this time-cost variable. If the results for the low users are in accordance to what the theory has predicted, an issue now is to understand the apparent contradiction between the theory and empirical findings for the high users population.

One possible way to reconcile the theory with our empirical findings is to recognize that the individuals that compose this latent sub-population are those that present a high intensity of utilization, being most likely the worse-off in terms of health status. Moreover, probably they have less economic power, not having the financial capacity to substitute public care by the, usually expensive, care provided in private clinics. In addition, we may tentatively suggest that these individuals are those health care centre users that are in a continuity of care process, fully managed by the family doctor. Given their, probably, high level of involvement and experience with the care received in the health care centre, they tend to trust in the judgments made by the family doctor regarding the adequate amount of care to consume, either in the health centre or at the hospital outpatient wards. Taking together the last two arguments it seems to be entirely justifiable their lack of response to the occurrence of delays in the consultations.

The next step of this section is to present and discuss the elasticity estimates of health care centre utilization relative to the total time involved in visiting the GP.

6.5.3.2 Analysis of the effect of the time-cost covariate *total_time*

We take up the current analysis by exploring whether the total time involved in visiting the GP influence, or not, the utilization of health care centres. Recall that the covariate 'total_time' is the total time involved in visiting the doctor, encompassing travel time, waiting time in the health care centre and consultation time.

Figure 23 illustrates two box-plot charts of the distribution of the individual elasticities of health care centre utilization relative to the total time involved in completing a consultation. The graph on the left shows the estimated distribution for the low users and the one on the right the distribution for the high users latent class.

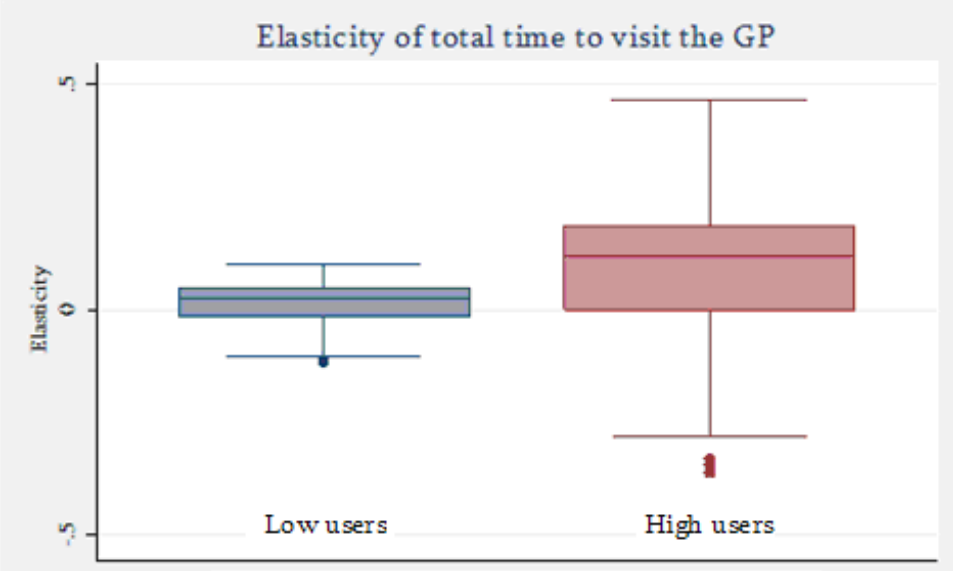

Figure 23 — Box-Plot of health centre utilization elasticities relatively to 'total_time', by latent class

The figure noticeably suggests that the elasticity of health care centre utilization relatively to total time is positive for most individuals, independently of the latent class to which they belong. Hence, this initial evidence clearly suggests that as the total time spent to visit the GP increases the same happens to the average utilization of the health centre, which is contrary to the rationale presented, stating that time-costs can act as a price, thus as a rationing device. A thorough analysis, however, shows that it is not the case, as the graphs conceal relevant information about the estimates, namely, the magnitude of the correspondent standard. It was found that all individual elasticities are zero from a statistical point of view (figures not reported on this dissertation). Therefore, the effects observed in the previous charts are not caused by any systematic causal relationship between total time spent to visit the GP and health centre utilization. The conclusions just presented are valid for the low users, for the high users as well as for the overall population of health care centre users.

In addition to the graphical analysis, already presented, we have also computed the average elasticity for each of the populations implied in our preferred specification. We obtained the following elasticity figures;

- 0,04 for the overall population of the health care centre users,
- 0,01 for the low users,
- 0,09 for the high users population.

All these estimates, besides being too small and possibly with no economic meaning, they are not statistically different from zero.

In short, taking together all the evidence regarding the effect of this explanatory variable, we conclude that that the total time an individual spends to visit the GP in the health care centre is not a factor that individuals have into account when making the decisions. In both latent classes, as well as in the overall population, the elasticity of utilization relative to total time is not statistically different from zero.

As explained in Section 6.3 we hypothesized that some observed characteristics of the individuals might cause health care centre users to value the physical waiting time differently, therefore, responding differently to that factor. In particular, there, we have assumed the covariates 'non-active', 'age' and 'education' are such individual features that may lead individuals to attach different value to the physical waiting time. The assumption was not confirmed by the data, as the elasticities of utilization relatively to 'total_time' did not prove to be dependent on the covariates activity status, age and education.

The failure to uncover a negative time-cost elasticity in the utilization of consultations can be considered, at least at first sight, puzzling, both because the theory undoubtedly predicts such negative effect and because many empirical applications have established that people respond to the waiting time by lowering utilization. However, we believe the finding was somewhat expected in the Portuguese context. In our view, the explanation for the absence of a negative time-cost elasticity

may reside on the characteristics of the population of health centre users. We note that approximately 25% of the Portuguese population benefits from occupational or private health insurance coverage and these are the most well-off and the ones with better access (including financial access) to care in the private sector. Those features imply that the population that uses public primary care sector are that without double coverage and that do not have much of a choice. Therefore, if there is no choice, utilisation might be partly inelastic to this class of time-costs. We note that our results are not unique, because other authors have found similar results, namely Coffey (1983) on ambulatory female medical care services along with others (Akin, Griffin *et al.* 1986; Mwabu, Ainsworth *et al.* 1993).

6.5.4 — Analysis of the impact of diverse covariates on the Health Centre utilization

In the previous section, we have concentrated on the two categories of time-cost covariates, making a detailed analysis about the subject. The model under analyses includes, however, a number of other covariates potentially contributing to understand the health care centre utilization. Therefore, the current section aims at analysing, and discussing, the impact of such explanatory variables on the health centre utilization.

We set out by the results of the covariates included in the socioeconomic and demographic category (Table 37, in page 262, presents a full description of the covariates). We begin by providing, in Table 45, point estimates of the average marginal effects (AME) of the explanatory variables included in the demographic and socioeconomic category.

	Low users (80%)	High users (20%)
Age	-0,136 (0,057)	0,141 (0,303)
Male	-0,574 (0,135)	-2,702 (0,841)
Education (less)	0,771 (0,180)	1,185 (1,175)
Non-active	0,749 (0,166)	2,90 (1,174)
Married	0,159 (0,196)	0,410 (1,093)
Widow	0,224 (0,268)	-0,244 (1,477)
North	0,037 (0,217)	0,368 (1,234)
Centre	0,300 (0,213)	2,041 (1,356)
LTV	-0,098 (0,213)	0,499 (1,337)
Urb_Pop (elasticity)	-0,112 (0,029)	0,080 (0,091)

Table 45 — AME of the socioeconomic and demographic variables[64]

The first aspect that is worth to refer is the difference in the statistical significance of the covariates across latent classes. For a 99% level of significance, only the gender factor influences the health care centre utilization among the high users, while for the

[64] The shaded cells correspond to estimates statistically significant ($p < 0,01$)

low users, besides gender, other factors like education, occupation status and place of residence appears to be factors to which the low users respond.

Regarding the 'age' covariate, despite the opposite signs of its AME across latent classes, a negative sign in the low users and a positive in the high users, the effects are not statistically different from zero, suggesting that age does not influence the utilization of GP visits in public health care centres. In addition to the calculation of the AME, we performed some extra analysis in order to pick up the possible non-linearities of age on utilization concluding, however, for the non-existence of such non-linearities[65]. The findings relative to the age factor (its lack of statistical significance) suggests that the substitution effect (lowers the health care centre utilization) counterbalances the impact of age on the utilization of medical care. Therefore, although only tentatively, we can conclude that the extra medical care (consultations) that the elder usually may need are obtained in doctors practising in the private sector or on emergency care units of hospitals.

Contrarily to 'age', the gender (the covariate 'male') of the individual is found to be an important explicative factor of utilization in both latent classes. The AME of 'male' is negative and statistically significant, both for the low users and for the high users. The 'male' gradient is steeper for the high users (-2,702) than for the low users (-0,574). Thus, the results suggest that the male individuals tend to use the care provided by health care centres less often than the females, being the differences more accentuated for the high users than for the low users. This result may be due to a more preventive behaviour of female individuals in what concerns health issues. Women with a more preventive attitude use the primary care more often than men does. The

---

[65] For both latent classes we estimated the distribution of the conditional mean for ages ranging from 20 to 90. After the analysis of the distributions, we did not observe any relevant parabolic effects in any latent class.

higher gradient observed on the high users may be explained by the different attitudes of males and females towards the continuity of care process, managed by the family doctor. We can hypothesize that male individuals who are high users are probably less willing to have their health care needs continuously supervised by the family doctor thus, may tend to pass over his/her advice, and use more often the medical care services provided elsewhere, namely, in private doctors or emergency care units in hospitals. In our view, another reasonable explanation for the big differences in utilization is because the pregnancy factor, which demands many consultations, therefore, provokes big differences in utilization among genders.

In regards to the effect of education, the educational dummy reveals that the lower educated[66] patients, relative to the higher educated, visit the health care centre doctor more often than the higher educated, as predicted before in Section 6.3. Table 45 shows that the AME of the covariate 'education' presents a positive value in both latent classes, however, the estimated AME of the high users is not statistically significant, meaning that in this class of users, low educated and high educated individuals presents similar levels of doctor visits. Surely, the positive coefficient for the low users can be justified on the grounds presented in Section 6.3, while the absence of the effect of education in the high users should not be observed as a deviation from our predictions, mainly after having into account the special characteristics of the individuals that compose the latent class under consideration.

Another factor accounted for in the model is the occupation status of the individual, represented by the covariate 'non-active'. As anticipated, the effect of the occupational dummy reveals that the non-active individuals visit the health care centre more often. We believe that this effect is primarily due to the opportunity cost of the non-active individuals is lower than the opportunity cost of the active ones in visiting the doctor

---

[66] We remember that the covariate 'education' captures lower levels of education.

at the times the health care centres provide the visits, usually at working hours. However, the identified positive effect is systematic only in the low users class, not presenting statistical significance in the high users. The individuals with a job may be more willing to substitute public for private care or emergency care units at hospitals because generally, the private care may be scheduled to the end of the day, therefore, imposing a lower opportunity cost in terms of loss of income to the working individuals. The non statistically significance high users class may be due to the fact that they are mainly individuals with a stringent need for continued care and so, more associated to pre-programmed plan of visits.

Regarding the effect of the covariates reflecting marital status (the dummy variables 'married' and 'widow') and place of residence (the dummy variables 'north', 'centre' and 'LVT'), the estimates of the respective AME, along with the respective standard errors, indicate that these factors are not determinants of consultations in the health care centre utilization, neither in the low users nor in the high users latent population. This means that the health care centre users tend to use a similar number of consultations, irrespective of the marital status and place of residence.

Finally, the covariate that captures the degree of urbanization of the county where the individual resides ('Urb_Pop') presents, as anticipated in Section 6.3, a negative and statistically significant elasticity (-0,112) for the low users population, and a statistically zero elasticity for the high users population. Therefore, within the low users population, a raise of 10% in the share of urban population in the individual's county of residence, causes the utilization to decrease by about 1,1%. Surely, the explanation for the negative elasticity found resides either in different behaviour rural *vs* urban individuals or in the large geographic inequalities in the distribution of private medical care facilities, which tend to be concentrated in urban and coastal areas (Oliveira and Bevan 2003), or a combination of both factors.

Turning to the analysis of the health status determinants on utilization, Table 46 reports the effect of several covariates reflecting the individual's health status. Nearly all dummy variables denoting health status have positive and statistically significant effect on the low users class. On the contrary, in the high users latent class, only the variable reflecting the existence of other chronicle diseases came out as statistically significant, with the remaining not presenting statistical significance.

	Low users (80%)	High users (20%)
Sick	0,208 (0,122)	0,427 (0,750)
Bad_Sah	1,310 (0,220)	0,836 (1,123)
Diabetes	0,628 (0,196)	0,445 (1,145)
Asthma	0,915 (0,181)	2,018 (1,319)
Heart	1,02 (0,165)	1,725 (1,024)
Digest	0,317 (0,160)	-0,146 (0,998)
Hypertension	0,600 (0,133)	-0,897 (0,904)
Osteo	0,171 (0,158)	-1,331 (0,972)
Other_cd	0,680 (0,166)	2,103 (0,933)

Table 46— AME of the health status covariates[67]

[67] The shaded cells correspond to estimates statistically significant ($p < 0.01$)

The results, thus, indicate that in the low users latent class, individuals reporting bad self-assessed health in addition to those that have reported suffering from chronic conditions, on average, increase utilization of health care centres. In sum, the variables that act as indicators of low health endowments have a positive effect in utilization, as many authors have found. Among the chronic conditions, that implying greater difference in utilization is the 'Heart' variable, with an estimated average marginal impact of about 1,0 visit a year.

Continuing this analysis of the average marginal effects of the explanatory variables, we move away from the health status covariates and pass on to the group of explanatory variables labelled as Characteristics of the Health Centre *vs* Family Doctor. Table 47 shows the average marginal effects of the covariates in that group.

	Low users (80%)	High users (20%)
Walk	-0,161 (0,171)	-1,106 (1,029)
Car	-0,460 (0,167)	-1,463 (0,985)
Not_Enought_Time	-0,651 (0,170)	0,642 (0,990)
Unsat_time_wr	0,033 (0,154)	-0,046 (0,988)
Diff_book	-0,541 (0,148)	-0,250 (1,029)
D_spe_gp_phone	-0,410 (0,138)	0,734 (0,945)

Table 47 — AME of the Characteristics of the Health Centre vs Family Doctor covariates[68]

[68] The shaded cells correspond to estimates statistically significant ($p < 0.01$)

A careful analysis of the ratio AME parameter/AME standard error lead us to the conclusion that the dummy covariates 'Walk', 'Car', 'Not_Enought_Time', 'Unsat_time_wr', 'Diff_book' and 'd_spe_gp_phone' do not play any role in allocating health care centre utilization amongst the high users latent class. On the contrary, among the low users, only the covariates 'Walk' and 'Unsat_time_wr' are statistically zero, with the others included in the group departing statistically from zero towards the negative direction.

The figures in the table reveal that the individuals that usually go by car to the health centre ('car'), those perceiving the consultation's duration as inadequate ('Not_Enought_Time'), those perceiving as difficult the task of booking a GP visit ('Unsat_time_wr'), and finally, those individuals that classify as difficult to speak with the family doctor by phone ('d_spe_gp_phone') tend to lower the utilization of the health care centres. Among these determinants found to be statistically different from zero, the perception that the length of the consultation 'Not_Enought_Time' is insufficient is the one that presents the largest negative effect, with an average marginal effect of -0,651. We believe that this variable reflects a quality of care dimension that is important to the patient, with poor quality usually associated to less utilization of care and to a higher willingness to substitute services. In the high users latent class this variable is not statistically significant perhaps because high users tend to be patients who are having their diseases monitored, thus, for them the consultation time is not as important as the continuity of care.

Finally, notice that the statistical insignificance of the covariate 'Unsat_time_wr'[69] is totally consistent with the previous finding that the total time involved in visiting the GP does not play a role in the individual's decisions about the number of doctor visits.

---

[69] It is a dummy variable, we recall, that equals 1 if the individual is unsatisfied with the waiting time in the waiting room, 0 otherwise

The results presented so far in the current chapter, built upon the estimation and on the analysis of the average marginal effects, permitted us to unveil at least one essential feature about the contribution of observed and unobserved factors in explaining health centre utilization across latent classes. Within the high health centre users, gender was found as the unique individual characteristic explaining health centre utilization, while in the low HC users latent class, many other factors have emerged as determinant to explain the differences in the expected individual utilization. This means that unobserved factors are much more important in the high users class than in the latent class of low users.

6.5.5 — Analysis of Class membership of the Health Centre users

At the current sub-section, we conduct an analysis to unveil the factors that explain class membership. To perform such analysis, we estimated a Probit model, where the dependent variable indicates class membership to the low users and the independent variables are those defined in Table 37, except the interactions. The regression results are reported in Table 48**Error! Reference source not found.**.

Covariate	Coef.	St. error	Z
Constant	1,672	0,240	6,970
Socioeconomic and demographic			
Age	0,189	0,092	2,04
Sqage	-0,014	0,008	-1,72
Male	0,087	0,059	1,48
Education (lower level)	-0,134	0,080	-1,67
Non-active	-0,056	0,069	-0,81
Married	0,050	0,075	0,67
Widow	0,091	0,107	0,85
North	0,031	0,089	0,35
Centre	0,013	0,086	0,15

LTV	0,042	0,092	0,45
Urb_Pop	-0,165	0,113	-1,46
Health Status			
Sick	-0,021	0,054	-0,4
Bad_Sah	-0,104	0,066	-1,56
Diabetes	-0,077	0,076	-1,01
Asthma	0,017	0,076	0,22
Heart	-0,201	0,067	-2,98
Digest	0,034	0,069	0,49
Hypertension	0,122	0,062	1,97
Osteo	0,056	0,067	0,85
Other_cd	-0,188	0,074	-2,53
Time-costs			
Total_Time (log)	-0,001	0,000	-3,56
App_delay	0,001	0,056	0,02
Health Centre vs Family Doctor			
Walk	-0,015	0,072	-0,2
Car	0,067	0,072	0,93
Not_Enought_Time	-0,139	0,064	-2,17
Unsat_time_wr	0,073	0,065	1,12
Diff_book	-0,090	0,065	-1,39
D_spe_gp_phone	-0,055	0,060	-0,92

Table 48 — Parameter estimates, standard errors and z-values of a probit model to find the determinants of class membership to the low users class

The parameter estimates reported in the previous table show that among the socioeconomic and demographic variables only the age of the individual contribute to explain the probability of belonging to the low users class. The effect of age is quite interesting as it is first increases that probability, and after some age, its effect of to decrease the probability of belonging to the low users class. The results revealed that other factors like, employment status, marital status and place of residence did not come out as determinants of being in the low users class. Summing up, we may

conclude that as patient gets older it becomes more likely to be assigned to the low users, however, after some age, that effect reverses.

On the other hand, the effect of health status points towards the direction that sicker individuals are more likely to be in the high users class, a result that conforms to the intuition that those individuals are, naturally, high intensity users of general health care, being necessarily also high users of GP consultations in the health centre.

From the other variables considered in the model, we only highlight the negative parameter of the covariate reflecting the individual's dissatisfaction with the time provided by the doctor during the consultation. It means that those individuals who are more dissatisfied with the consultation time are more likely to belong to the high users class.

Overall, the results suggest that the health status of the individual is an important indicator of class membership, also showing that one cannot ignore the role of age, and satisfaction with the consultation time as determinants of class membership.

6.6 — Summary of the conclusions, policy implications and future research

The main objective of the empirical application developed in the present chapter was to estimate the extent to which time-costs influence individuals' decisions regarding the utilization of primary health care consultations provided at public health care centres. We considered that health care centre users incurred into two different types of time-costs; on the one hand the total time involved in completing the consultation, on the other, the occurrence of a delay in the consultation. The dependent variable that we modelled was the number of primary health care consultations in public health care centres in a period of twelve months.

To achieve the goal of estimating the effect of time-costs, we specified and estimated various competing LCMs, appropriately tailored to handle truncated count data pertaining to the health care centre users population. After having carried out

some statistical tests on the competing specifications, we concluded that the specification that better served our purposes was the two-support point latent class model with truncated negative binomial II as parent distribution. Therefore, we interpreted the results of such specification, beginning by an attempt to characterize the users that compose the latent classes implied by the model, followed by the estimation of the effect of several covariates, by latent class, focusing on the time-costs covariates. The most relevant conclusions that can be drawn from this research are the following:

1. the users of the health care centres can be divided into two latent subpopulations: the population of the low users (80% of the patients, 5 visits per year) and the population of the high users (20% of the patients, 9,8 visits). Additionally, it was found that the patient behaviour differs according to the latent class to which the individual belongs.
2. for the low users:
    a. the patients who cannot schedule a timely appointment at the health centre lower, on average, their health centre utilization,
    b. as the patient gets older the magnitude of the response to an appointment delay decreases to values around zero for the individuals aged approximately above 70. This might reflect a null opportunity cost of waiting for a consultation.
    c. The patients who reported their health status as bad and those who work tend to respond more noticeably to the occurrence of delays in the health care centre consultations,
    d. The gender of the patient does not appear to influence the individual's response to the occurrence of a delay, signifying that, the response of man and woman to an appointment delay is similar,
    e. the elasticity of utilization relative to 'total_time' is statistically zero,

f. being a woman, less educated, without a professional occupation and residing in a rural area, increases, on average, the utilization of the health care centres,

g. patients in poor health status also tend to increase the number of visits to the health care centre,

h. Patients who evaluate as low some quality dimensions tend to reduce the number of visits to health care centre.

3. for the high users population:

a. the patients forced to wait on a list to visit the doctor present a similar number of visits relative to the patients who are offered an appointment on the same day,

b. the covariates 'age', 'male', 'bad_sah' and 'non-active' do not revealed any statistically significant effect on the response to a delay in the appointment,

c. the elasticity of utilization relative to 'total_time' is statistically zero,

d. From all covariates included as determinants, only the covariate reflecting gender ('male') and the covariate 'other_cd' showed a statistically significant effect on utilization. 'Male' tend to decrease utilization, while 'other_cd' tend to increase.

4. The analysis of class membership indicated that patient's health status is an important indicator of class membership. The patients worse-off tend to be classified as high users.

The main results summarized above hold a couple of lessons potentially relevant for health policy makers. Below, we summarize the four most relevant examples of policy implications that, in our opinion, can be derived from the results:

First lesson: the differences found on the latent populations identified suggest that, at least in the achievement of certain health policy goals related to the

utilization of medical care services, efficient policy making measures should be designed and directed to target populations and not to the general population. To do this, the policy maker must be aware that the individuals in each latent class have different characteristics, and might respond differently to different policies;

Second lesson: the elasticity of health care centre utilization relative to the total time involved in completing a consultation is zero, at least from a statistical viewpoint. Therefore, this time-cost variable cannot be viewed as a policy instrument capable of influencing individual behaviour regarding the utilization of medical care. Despite this result, obviously, we do not defend that this class of time-costs should be kept high or increase, only because this would not have impact on utilization. On the contrary, we advocate that the health care centres should attempt to reduce this time-costs as much as possible as it will not increase utilization, at least in the short run;

Third lesson: on the contrary, our results suggest that the time-cost category waiting in a list can be used as a sensible instrument for health policy capable of affect individuals' behaviour, at least the case of low health centre users, thus, influencing the utilization of the health care centres. If the objective of the policy maker were to increase the utilization of health care centres, the appropriate measure would be increase and/or reorganize the supply of medical care in the health care centres (e.g. more doctors, extended schedules). This action increases the number of visits provided without a delay, therefore, increasing the utilization of the health care centres, at least for about 80% of the users. One such action may encourage spill over effects on the utilization of emergency care provided in hospitals. Assume that most of the low users quitting from a health care centre consultation due to an excessive delay in the appointment of the visit demand the needed care in the hospital through the emergency unit. Therefore, under this assumption, increasing the access to the health care centres can have spill over effects on the utilization of hospital emergency care, reducing its demand. Undoubtedly, this last claim can be only seen as tentative

as we do not know anything about the behaviour of the individuals that quit from the health care centre consultations due to an occurrence of a delay. Notwithstanding the results indicate that the response of the low users to a delay varies with age, activity status and health status it would be hard to defend that these groups of individuals (the youths, the ones in bad health status and the ones whose activity status in on) should be discriminated positively in terms of timely consultations. Although this would have as effect increasing the number of consultations of these groups, it would also be unacceptable from an equality of treatment point of view.

Fourth lesson: the effect of the occupational dummy reveals that the non-active individuals (relative to the active ones) visit the health care centre more often. We believe that this effect is primarily due to the opportunity cost of the non-active individuals, which is lower than the opportunity cost of the active ones in visiting the doctor at the times the health care centres provide the visits, usually at working hours. Accordingly, this suggests that the health care centres should provide consultations to the active individuals off the working hours. Using a similar reasoning to that presented in lesson three, this measure may also have spill over effects to the utilization of hospital emergency care.

The application developed in this chapter has some potential caveats that ought to be presented and discussed. The identified weaknesses suggest some possibilities for future research.

The first possible drawback of the application revolves around our options regarding the specification of the model. The sampling scheme adopted in this project to collect the data leaves scope for alternative interpretations about the appropriateness of our specification to handle the special features of the data. The individuals selected to the sample were chosen randomly from a database containing information about all people who had at least one GP consultation in the health centre during the last three months of 2003. Therefore, under the selected sampling protocol only the individuals that

visited the health centre in the last three months of 2003 have a positive probability of being included in the sample, thus, all information about the non-users is truncated from the sample, generating a truncated (at-zero) sample. Given this data context, we moved away from the specification and estimation of standard count data models, adopting instead econometric specifications specialized to handle with the truncated count data. However, a careful analysis of the sampling scheme shows that truncation is not the only problem that we should have handled in the specification. In fact, a closer look reveals that the dataset resulted from sampling users randomly on-site. In such a context truncation is only part of the data features that ought to be accounted for in the model specification (Shaw 1988). The issue is that the probability of inclusion in the sample is based on the number of visits to the health care centre, and users who visit the health care centre more often present a higher likelihood to be included in the sample than people who visit the health centre occasionally. Summarizing, this signifies that the high users of the health centre are, likely, overrepresented in the sample (Shaw 1988; Englin and Shonkwiler 1995; Santos-Silva 1997). We recognize that to obtain unbiased estimates, perhaps it would have been more appropriate to resort on specifications specialized to deal with the endogenous stratification inherent in our sampling mechanism, besides the truncation problem, thus, estimating models adapted to handle the problems generated by sampling on-site (Shaw 1988; Santos-Silva 1997). In this regards we should mention, however, that in previous experiments with a similar dataset we estimated truncated and on-site sampling LCMs and did not find very different results. These experiments should be seen, however, as merely preliminary. So, as part of our future research agenda is the investigation of the impact that the on-site sampling has on the estimates of the effect of time-costs. Thus, to deal with the issue we propose the specification, estimation and interpretation of results coming from LCMs specialized to handle on-site sampling, and, naturally, its comparison with the results presented in the present chapter, to

evaluate at which extent on-site sampling is has biased the results present in this empirical application.

Second, issues related to the quality of the data. A challenge that every empirical health economist faces is finding data with enough quality to estimate the models. Usually, the big issue is that the researcher does not have any type of control over data collection process, therefore has to tailor the data to meet its objectives. In this chapter we have worked in such a context. The Europep dataset was collect initially to identify the quality of the health centres, thus, the dataset clearly presents some weaknesses when its data is used to respond to our research questions. First, the dataset does not include any variable reflecting the individuals' opportunity costs of time. Therefore, if, in fact, patients attach different opportunity costs to their time, the estimates of the impact of time-costs on utilization come biased downwards (Coffey 1983). Second, still related to data availability or quality, another possible drawback of the application relates to the covariates used as indicator of time-costs. Both covariates measure the time-costs involved in the last consultation, and, consequently, we take such time-costs as a good proxy to the time-costs involved in all consultations the individual have sought. Despite this could be a source of bias to our results, we should note, however, that in general the individuals have consultations in the same health centre and with the same doctor, therefore, we do not expect to find much variability in the time-costs across consultations. Third, yet in the lack of data context, the dataset does not measure variables like income, health insurance status, years of schooling, full information about the occupation status, and so on. The inclusion of such covariates would permit the improvement of our understanding about the manner patients choose the number of visits. In addition, one class of covariates that would reveal also particularly useful are those reflecting information about the family doctor, e.g., gender, age, etc., because it would allow estimating its contribution to the number of visits demanded.

Besides the opportunities for further investigation suggest by the above identified shortcomings of the study, some further investigating related to our work are possible.

First, it is possible to think on the improvement of the analysis of the characteristics of the latent classes (posterior class analysis) in order to be able to identify target populations for health policy making.

Second, in another direction, it would be quite interesting from a health economics perspective, to investigate how other types of medical care utilization interplay with the utilization of the public health centres. One very interesting health care indicator that clearly affects, and is affect by, the utilization of public primary care is the utilization of hospital emergency units. In this situation, the specification of a bivariate regression model would be appropriate. Such a model, besides explaining the utilization of the health centres, also included the demand for consultations at the emergency units in the hospital, making possible to estimate the substitution effects among the different types of providers. With this knowledge, policy makers could change the costs of access to the different providers, encouraging in this way patients to seek medical care in the more adequate provider.

Fifth, in another direction, another interesting extension for our work, is to estimate the effect of time-costs on the utilization of public health centres on the part of the overall Portuguese population, including the non-users (Propper 1990). This would permit us to extend our understanding about the effect of time-costs on the overall population, therefore, to predict more accurately the effect, on utilization, of an alteration of time costs. Without such knowledge, our predictions are always partial, as one can never know how the non-users of the public health centres react when the queue related time-costs to access the health centre descrease.

Chapter 7
Final Remarks

This final chapter presents a series of considerations that conclude this thesis.

The work developed throughout this dissertation has provided an investigation of issues relevant to identify patterns of health care utilization, when utilization is measured as a count variable. We consider the investigation conducted in this thesis relevant from two major points of view. First, the research is relevant from a health econometrics perspective, applied to the analysis of health care data, and second, it is also significant from a health economics perspective, as it has produced insightful knowledge regarding a number of issues of the Portuguese health system.

One first major avenue of investigation that we followed closely throughout this work was the exploration of the possibility of specifying models suited to split the actual health care users population into latent groups of patients. Identification of groups based on actual use of health care faces, however, methodological challenges, which have been addressed with econometric specifications adapted to handle the challenge. We proposed a latent class model specifically designed to account for the unobserved factors specific to the population of actual users of medical care services.

One main finding was the importance of identifying health care users with different levels of utilization, thus, different health care consumption groups. Another conclusion was that the different health care determinants considered presented, in general, different effects on the health care utilization of such groups. Still regarding this methodological issue, we have claimed that when two different stochastic processes govern the overall health care utilization, in the hurdle spirit, such specific LCMs provide a better approach to handle the utilization of the actual users. The results produced by our experiments, based on actual data, were quite encouraging. We found that the LCM resulting from looking differently at the individual unobserved heterogeneity, when combined with a binary regression model aiming at

explaining the probability of some medical care utilization, proved to outperform some of the most common competing econometric specifications adopted to study patterns of medical care utilization.

The second major avenue of investigation followed in this thesis regards the economics of health care utilization, specifically, in the area of producing insights about medical care utilization patterns in Portugal. Such insights contribute to inform policy debate on the efficient organization and delivery of medical care services. We analysed two types of medical care indicator, which have been examined by two econometric models. On one hand, the total number of visits to physicians, in a three-month period, of the Portuguese population, and on the other, the total number of visits made to the public health centre in one year period, for a sample of adult actual health centre users. From a health policy perspective, our results provided various insights potentially useful for the Portuguese health system, from which we highlight four of them, for their importance and uniqueness.

First, supplementary health insurance does not affect medical care utilization among the actual users. However, our findings evidenced that the beneficiaries of one of the health insurance groups considered in the application are more likely to visit the doctor. Nevertheless, we did not interpreted such result as moral hazard, on the contrary, we consider that higher likelihood as the result of the unique characteristics of populations that, possibly, are eager in the consumption of precautionary care. In short, our results do not evidence signs of moral hazard behaviour.

Second, actual health care users present a similar frequency of doctor visits regardless of residing in urban or rural areas. However, urban peers are more likely to have at least one visit. The result suggests, therefore, that health policy should encourage policy measures that discriminate favourably those rural residents that are not users of medical care.

Third, our findings showed that the actual users of primary medical care provided by public health centres are not reactive to the total time required to visit the doctor. The policy implication of such a result is that this variable cannot be adopted as a health policy instrument aiming at influencing the utilization of medical care.

Forth, the occurrence of a delay in the appointment for the low users population, can, possibly, be used as a tool to shift the utilization of hospital emergency units to public health centres. We argued in Chapter 6 that a decrease in the proportion of low users experiencing a delay in the appointment, increases the utilization of health centre of the low users population, spilling over the utilization of hospital care at emergency wards.

Throughout this thesis, we have considered only LCMs with two latent classes, thus, a priori assuming a segmentation of the target population into two latent health care groups. Notice that we did not conduct a careful analysis aiming at estimating the number of latent classes, sticking to models with only two component distributions. In some preliminary experiments, not reported in this dissertation, however, we endeavoured to estimate models with more than two latent groups facing, in such situations, difficulties in maximizing the respective log-likelihood function. The difficulties found in optimizing the log-likelihoods of the models built upon the LCM framework with more than two classes suggests that, probably, different estimation techniques should be adopted, such as, for instance, the EM algorithm, desirably combined with larger samples. Thus, an open avenue for further research in the line of our work is to investigate whether the populations analysed here can be clustered into more than two latent groups.

We close this dissertation by presenting two wide avenues to where this thesis can be extended, suggesting, thus, ideas for future research in this area of health econometrics applied to actual data.

The first suggestion concerns the specification of models to explain health care utilization. Although emphasizing throughout this thesis the adoption of latent class models defined by assuming the unobserved individual factors specific to the actual health care users population, we believe that innovative LCMs specified under the assumption that the unobserved heterogeneity is on the overall population can, in fact, outperform our proposal. As commonly seen, due to the lack of economic guidance to specification, the choice of the most appropriate specification is an empirical issue.

Therefore, based on intuition and statistical reasoning, we propose to model the medical care data using an LCM, which departs from the most common in the models adopted as parent distributions, that consists in mixing a NB density with a zero inflated model (ZI), thus gaining flexibility to accommodate the usually high proportion of zeros. In such a model, intuition lead us to anticipate that the component of the ZI comes as the model for the low users, because the zeros are in this class, and the NB density, on the other hand, is the model for the high users. Surely, some type of economic reasoning behind such specification would enrich the model. Another possibility for a mixture is for instance to combine the NB model with a with a hurdle specification, for instance.

The second suggestion to future research is related to the production of knowledge allowing overcoming the theoretical or, sometimes ad-hoc, conjectures reaching the facts that the data conveys.

To begin with, this thesis analysed only one type of medical care — visits to physicians — the total number of visits regardless of the characteristics of the provider, and the number of visits to public health centres. These are only rudimentary indicators of medical care services not reflecting the whole intensity of medical care utilization, or the quality dimension. Therefore, to be able to contribute with more insights on the behaviour of the individuals regarding medical care

choices, more studies explaining different medical care services are necessary. For instance, the utilization of preventive care, the utilization of hospital inpatient care and outpatient care, the utilization of diagnostic tests, the utilization of private doctors, constitute only some interesting examples.

Appendix A
Stata Programs

The current appendix aims at presenting the Stata programs written to estimate the parameters of the Chapters 4, 5 and 6. In addition, we present the programs written to implement the Voung tests, as well as some other programs used to make some calculations.

We begin by showing the log-likelihood programs of each specification considered in Chapter 4 and 6, followed by the presentation of one program developed to estimate each LCM repeatedly from different initial solutions.

After that, we present the program written to implement the Voung tests in addition to the code used to estimate the average marginal effects along with its standard errors.

Program to define the log-likelihood of the model NB1 — Negative Binomial (type 1) regression model

```
*******************
program define nb1
version 9.0
args lnf xb1 alfa1
local y = "$ML_y1"
tempvar m1 g1 c1
quietly generate double `m1' = exp(`xb1')
quietly generate double `g1'
=(exp(lngamma(`y'+(`m1'/`alfa1'))))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m1'/`al
fa1'))))
quietly generate double `c1' = `g1'*((1+`alfa1')^-
(`y'+(`m1'/`alfa1')))*(`alfa1'^`y')
quietly replace `lnf' = ln(`c1') if $ML_samp == 1
end
***************
```

Definition of the log-likelihood of the model H_NB1 — Popular Hurdle: First and second parts based on the NB1 distribution. In the first part the dispersion parameter was set to one.

```
*******************
program define h_nb1
* IN THIS SPECIFICATION WE SET ALFA = 1 IN THE FIRST STAGE.
  version 9.0
  args lnf xb1 xb2 alfa
  local y = "$ML_y1"
  tempvar m1 m2 g2 c2 ul y0 y1
  quietly generate double `y0' = `y' == 0
  quietly generate double `y1' = `y' > 0
  quietly generate double `m1' = exp(`xb1')
  quietly generate double `m2' = exp(`xb2')
  quietly generate double `g2'
=(exp(lngamma(`y'+`m2'/`alfa')))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m2'/`alfa'
))))
  quietly generate double `c2' = `g2'*((1+`alfa')^-
(`y'+`m2'/`alfa'))*((`alfa')^`y')
  quietly generate double `ul' = ((`alfa'+1)^-(`m2'/`alfa'))
  quietly replace `lnf' = `y0'*(-ln((2^`m1')-1))+ln((2^`m1')-1)-
`m1'*ln(2)+(`y1')*(ln(`c2')-ln(1-`ul')) if $ML_samp ==1
 end
*******************
```

Definition of the log-likelihood of the model LCM_NB1 — Two latent class LCM. NB1 density as component distribution.

```
************************************
program define lcm_nb1
 version 9.0
 args lnf xb1 alfa1 xb2 alfa2 p1
 local y = "$ML_y1"
 tempvar m1 m2 g1 g2 c1 c2
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2')
 quietly generate double `g1'
=(exp(lngamma(`y'+(`m1'/`alfa1'))))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m1'/`al
fa1'))))
 quietly generate double `g2'
=(exp(lngamma(`y'+(`m2'/`alfa2'))))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m2'/`al
fa2'))))
 quietly generate double `c1' = `g1'*((1+`alfa1')^-
(`y'+(`m1'/`alfa1')))*(`alfa1'^`y')
 quietly generate double `c2' = `g2'*((1+`alfa2')^-
(`y'+(`m2'/`alfa2')))*(`alfa2'^`y')
 quietly replace `lnf' = ln((`p1')*(`c1')+(1-`p1')*(`c2')) if $ML_samp == 1
end
-------------------------------------------------------------------------------
```

Definition of the log-likelihood of the model R-LCM_NB1 — LCM_NB1 with slopes constrained to be equal across latent classes.

```
*****************************************
program define r_lcm_nb1
 version 9.0
 args lnf xb1 a1 alfa1 a2 alfa2 p1
 local y = "$ML_y1"
 tempvar m1 m2 g1 g2 c1 c2
 quietly generate double `m1' = exp(`xb1'+`a1')
 quietly generate double `m2' = exp(`xb1'+`a2')
 quietly generate double `g1'
=(exp(lngamma(`y'+(`m1'/`alfa1'))))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m1'/`al
fa1'))))
 quietly generate double `g2'
=(exp(lngamma(`y'+(`m2'/`alfa2'))))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m2'/`al
fa2'))))
 quietly generate double `c1' = `g1'*((1+`alfa1')^-
(`y'+(`m1'/`alfa1')))*(`alfa1'^`y')
 quietly generate double `c2' = `g2'*((1+`alfa2')^-
(`y'+(`m2'/`alfa2')))*(`alfa2'^`y')
 quietly replace `lnf' = ln((`p1')*`c1'+(1-`p1')*`c2') if $ML_samp== 1
end
*********************************
```

Definition of the log-likelihood of the model LCH-Cross(NB1). First part based on the NB1. The second part is a 2 point of support LCM with NB1 as baseline distribution. In the first part the dispersion parameter was set to one.

```
***********************************
program define h_lcm_nb1
 version 9.0
```

```
 args lnf xb1 xb2 alfa2 xb3 alfa3 pi
 local y = "$ML_y1"
 tempvar m1 m2 m3 ul y0 y1 num2 num3 den2 den3 quo2 quo3
 quietly generate double `y0' = `y' == 0
 quietly generate double `y1' = `y' > 0
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2')
 quietly generate double `m3' = exp(`xb3')
 quietly generate double `num2'
=((exp(lngamma(`y'+`m2'/`alfa2')))*((1+`alfa2')^-
(`y'+`m2'/`alfa2'))*(`alfa2'^`y'))
 quietly generate double `num3'
=((exp(lngamma(`y'+`m3'/`alfa3')))*((1+`alfa3')^-
(`y'+`m3'/`alfa3'))*(`alfa3'^`y'))
 quietly generate double `den2'
=((exp(lngamma(`y'+1)))*(exp(lngamma(`m2'/`alfa2')))*(1-((1+`alfa2')^-
(`m2'/`alfa2'))))
 quietly generate double `den3'
=((exp(lngamma(`y'+1)))*(exp(lngamma(`m3'/`alfa3')))*(1-((1+`alfa3')^-
(`m3'/`alfa3'))))
 quietly generate double `quo2' = ((`num2')/(`den2'))
 quietly generate double `quo3' = ((`num3')/(`den3'))
 quietly replace `lnf' = `y0'*(-ln((2^`m1')-1))+ln((2^`m1')-1)-
`m1'*ln(2)+(`y1')*(ln(`pi'*(`quo2')+(1-`pi')*(`quo3'))) if $ML_samp ==1
end
**********************
```

Definition of the log-likelihood of the model R-LCH-Cross — LCH-Cross with slopes constrained to be equal across latent classes in the second stage.

```
***************************
program define r_hlcm_nb1
 version 8.2
 args lnf xb1 xb2 a2 alfa2 a3 alfa3 pi
 local y = "$ML_y1"
 tempvar m1 m2 m3 ul y0 y1 num2 num3 den2 den3 quo2 quo3
 quietly generate double `y0' = `y' == 0
 quietly generate double `y1' = `y' > 0
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2'+`a2')
 quietly generate double `m3' = exp(`xb2'+`a3')
 quietly generate double `num2' =(exp(lngamma(`y'+`m2'/`alfa2')))*((1+`alfa2')^-
(`y'+`m2'/`alfa2'))*(`alfa2'^`y')
 quietly generate double `num3' =(exp(lngamma(`y'+`m3'/`alfa3')))*((1+`alfa3')^-
(`y'+`m3'/`alfa3'))*(`alfa3'^`y')
 quietly generate double `den2'
=(exp(lngamma(`y'+1)))*(exp(lngamma(`m2'/`alfa2')))*(1-((1+`alfa2')^-
(`m2'/`alfa2')))
 quietly generate double `den3'
=(exp(lngamma(`y'+1)))*(exp(lngamma(`m3'/`alfa3')))*(1-((1+`alfa3')^-
(`m3'/`alfa3')))
 quietly generate double `quo2' = `num2'/`den2'
 quietly generate double `quo3' = `num3'/`den3'
 quietly replace `lnf' = `y0'*(-ln((2^`m1')-1))+ln((2^`m1')-1)-
`m1'*ln(2)+(`y1')*(ln(`pi'*`quo2'+(1-`pi')*`quo3')) if $ML_samp ==1
end
*******************************
```

Definition of the log-likelihood of the model NB2 — Negative Binomial (type 2) regression model

```
*************************
program define nb2
 version 8.2
 args lnf xb1 alfa1
 local y = "$ML_y1"
 tempvar m1 g1 c1
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `g1'
=(exp(lngamma(`y'+1/`alfa1')))/((exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa1'))))
 quietly generate double `c1' = `g1'*((1+`alfa1'*`m1')^-
(`y'+1/`alfa1'))*((`alfa1'*`m1')^`y')
 quietly replace `lnf' = ln(`c1') if $ML_samp == 1
end
************************
```

Definition of the log-likelihood of the model H_NB2 — Popular Hurdle: First and second parts based on the NB2 distribution. In the first part the dispersion parameter was set to one.

```
***************************
program define h_nb2
* NESTE MODELO, POR UMA QUESTÃO DE IDENTIFICAÇÃO, FAZEMOS ALFA1 = 1.
  version 8.2
 args lnf xb1 xb2 alfa
 local y = "$ML_y1"
 tempvar m1 m2 g2 c2 ul y0 y1
 quietly generate double `y0' = `y' == 0
 quietly generate double `y1' = `y' > 0
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2')
 quietly generate double `g2' =
(exp(lngamma(`y'+1/`alfa')))/((exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa'))))
 quietly generate double `c2' = `g2'*((1+`alfa'*`m2')^-
(`y'+1/`alfa'))*((`alfa'*`m2')^`y')
 quietly generate double `ul' = ((`alfa'*`m2'+1)^-(1/`alfa'))
 quietly replace `lnf' = -`y0'*`xb1'+`xb1'-ln(1+`m1')+(`y1')*((ln(`c2'))-(ln(1-
`ul')))
 end
***************************
```

Definition of the log-likelihood of the model LCM_NB2 — Two class LCM model with NB2 density as component distribution.

```
***************************
program define lcm_b2
 version 8.2
 args lnf xb1 alfa1 xb2  alfa2 pi
 local y = "$ML_y1"
 tempvar m1 m2 g1 g2 c1 c2
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2')
 quietly generate double `g1'
=(exp(lngamma(`y'+1/`alfa1')))/((exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa1'))))
 quietly generate double `g2'
=(exp(lngamma(`y'+1/`alfa2')))/((exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa2'))))
 quietly generate double `c1' = (`g1')*((1+`alfa1'*`m1')^-
(`y'+1/`alfa1'))*((`alfa1'*`m1')^`y')
 quietly generate double `c2' = (`g2')*((1+`alfa2'*`m2')^-
(`y'+1/`alfa2'))*((`alfa2'*`m2')^`y')
 quietly replace `lnf' = ln(`pi'*(`c1')+(1-`pi')*(`c2'))
end
***************************
```

Definition of the log-likelihood of the model R-LCM_NB2 — Model LCM_NB2 with slopes constrained to be equal across latent classes.

```
**************************
program define r_lcm_nb2
 version 8.2
 args lnf xb1 a1 alfa1 a2 alfa2 pi
 local y = "$ML_y1"
 tempvar m1 m2 g1 g2 c1 c2
 quietly generate double `m1' = exp(`xb1'+`a1')
 quietly generate double `m2' = exp(`xb1'+`a2')
 quietly generate double `g1'
=(exp(lngamma(`y'+1/`alfa1')))/((exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa1'))))
 quietly generate double `g2'
=(exp(lngamma(`y'+1/`alfa2')))/((exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa2'))))
 quietly generate double `c1' = `g1'*((1+`alfa1'*`m1')^-
(`y'+1/`alfa1'))*((`alfa1'*`m1')^^`y')
 quietly generate double `c2' = `g2'*((1+`alfa2'*`m2')^-
(`y'+1/`alfa2'))*((`alfa2'*`m2')^^`y')
 quietly replace `lnf' = ln(`pi'*`c1'+(1-`pi')*`c2')
end
**************************
```

Definition of the log-likelihood of the model LCH-Cross(NB2) — First part based on the NB2. The second part is a 2 class LCM with NB2 as baseline distribution. In the first part In the first part the dispersion parameter was set to one.

```
*****************************
program define hlcm_nb2
 version 8.2
 args lnf xb1 xb2 alfa2 xb3 alfa3 pi
 local y = "$ML_y1"
 tempvar m1 m2 m3 y0 y1 num2 num3 den2 den3 quo2 quo3
 quietly generate double `y0' = `y' == 0
 quietly generate double `y1' = `y' > 0
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2')
 quietly generate double `m3' = exp(`xb3')
 quietly generate double `num2'
=((exp(lngamma(`y'+1/`alfa2')))*((`alfa2'*`m2')^^`y')*((1+`alfa2'*`m2')^-
(`y'+1/`alfa2')))
 quietly generate double `num3'
=((exp(lngamma(`y'+1/`alfa3')))*((`alfa3'*`m3')^^`y')*((1+`alfa3'*`m3')^-
(`y'+1/`alfa3')))
 quietly generate double `den2'
=((exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa2')))*(1-((1+`alfa2'*`m2')^-
(1/`alfa2'))))
 quietly generate double `den3'
=((exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa3')))*(1-((1+`alfa3'*`m3')^-
(1/`alfa3'))))
 quietly generate double `quo2' = ((`num2')/(`den2'))
 quietly generate double `quo3' = ((`num3')/(`den3'))
 quietly replace `lnf' = -`y0'*(`xb1')+`xb1'-
ln(`m1'+1)+(`y1')*(ln(`pi'*(`quo2')+(1-`pi')*(`quo3'))) if $ML_samp== 1
 end
*****************************
```

Definition of the log-likelihood of the model R-LCH-Cross(NB2) — Model LCH-Cross with slopes constrained to be equal across latent classes in the second stage.

```
*****************************
program define r_hlcm_nb2
 version 8.2
 args lnf xb1 xb2 a2 alfa2 a3 alfa3 pi
 local y = "$ML_y1"
 tempvar m1 m2 m3 y0 y1 num2 num3 den2 den3 quo2 quo3
 quietly generate double `y0' = `y' == 0
 quietly generate double `y1' = `y' > 0
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2'+`a2')
 quietly generate double `m3' = exp(`xb2'+`a3')
 quietly generate double `num2'
=(exp(lngamma(`y'+1/`alfa2')))*((`alfa2'*`m2')^`y')*((1+`alfa2'*`m2')^-
(`y'+1/`alfa2'))
 quietly generate double `num3'
=(exp(lngamma(`y'+1/`alfa3')))*((`alfa3'*`m3')^`y')*((1+`alfa3'*`m3')^-
(`y'+1/`alfa3'))
 quietly generate double `den2'
=(exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa2')))*(1-((1+`alfa2'*`m2')^-
(1/`alfa2')))
 quietly generate double `den3'
=(exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa3')))*(1-((1+`alfa3'*`m3')^-
(1/`alfa3')))
 quietly generate double `quo2' = `num2'/`den2'
 quietly generate double `quo3' = `num3'/`den3'
 quietly replace `lnf' = -`y0'*(`xb1')+`xb1'-
ln(`m1'+1)+(`y1')*(ln(`pi'*`quo2'+(1-`pi')*`quo3')) if $ML_samp == 1
end
*****************************
```

Definition of the log-likelihood of the model TNB1 — Truncated negative binomial (type 1) regression model

```
************************************************
program define tnb1
 version 8.2
 args lnf xb alfa1
 local y = "$ML_y1"
 tempvar m1 num1 den1
 quietly generate double `m1' = exp(`xb')
 quietly generate double `num1' =
(exp(lngamma(`y'+`m1'/`alfa1')))*((1+`alfa1')^-(`y'+`m1'/`alfa1'))*(`alfa1'^`y')
 quietly generate double `den1' =
(exp(lngamma(`y'+1)))*(exp(lngamma(`m1'/`alfa1')))*(1-((1+`alfa1')^-
(`m1'/`alfa1')))
 quietly replace `lnf' = ln(`num1')-ln(`den1') if $ML_samp ==1
end
*******************************
```

Definition of the log-likelihood of the model LCM_TNB1 — Two component LCM with TNB1 as baseline model

```
*************************
program define mix_nb1_t
 version 9.2
 args lnf xb1 alfa1 xb2 alfa2 pi
 local y = "$ML_y1"
 tempvar m1 m2 num1 den1 quo1 num2 den2 quo2
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2')
```

```
 quietly generate double `num1' =
((exp(lngamma(`y'+`m1'/`alfa1')))*((1+`alfa1')^-
(`y'+`m1'/`alfa1'))*(`alfa1'^`y'))
 quietly generate double `num2' =
((exp(lngamma(`y'+`m2'/`alfa2')))*((1+`alfa2')^-
(`y'+`m2'/`alfa2'))*(`alfa2'^`y'))
 quietly generate double `den1' =
((exp(lngamma(`y'+1)))*(exp(lngamma(`m1'/`alfa1')))*(1-((1+`alfa1')^-
(`m1'/`alfa1'))))
 quietly generate double `den2' =
((exp(lngamma(`y'+1)))*(exp(lngamma(`m2'/`alfa2')))*(1-((1+`alfa2')^-
(`m2'/`alfa2'))))
 quietly generate double `quo1' = (`num1'/`den1')
 quietly generate double `quo2' = (`num2'/`den2')
 quietly replace `lnf' = ln(`pi'*`quo1'+(1-`pi')*`quo2') if $ML_samp ==1
end
************************
```

Definition of the log-likelihood of the model TNB2 — Truncated negative binomial (type 2) regression model

```
*****************************************
program define nb2_t
 version 9.2
 args lnf xb alfa1
 local y = "$ML_y1"
 tempvar m1 m2 num1 den1 quo1 num2 den2 quo2
 quietly generate double `m1' = exp(`xb')
 quietly generate double `num1' =
(exp(lngamma(`y'+1/`alfa1')))*((`alfa1'*`m1')^`y')*((1+`alfa1'*`m1')^-
(`y'+1/`alfa1'))
 quietly generate double `den1' =
(exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa1')))*(1-((1+`alfa1'*`m1')^-
(1/`alfa1')))
 quietly replace `lnf' = ln(`num1')-ln(`den1') if $ML_samp == 1
end
*****************************************
```

Definition of the log-likelihood of the model LCM_TNB2 — Two component LCM with TNB2 as baseline model

```
*********************************
program define mix_nb2_t
 version 9.2
 args lnf xb1 alfa1 xb2 alfa2 pi
 local y = "$ML_y1"
 tempvar m1 m2 num1 den1 quo1 num2 den2 quo2
 quietly generate double `m1' = exp(`xb1')
 quietly generate double `m2' = exp(`xb2')
 quietly generate double `num1' =
(exp(lngamma(`y'+1/`alfa1')))*((`alfa1'*`m1')^`y')*((1+`alfa1'*`m1')^-
(`y'+1/`alfa1'))
 quietly generate double `num2' =
(exp(lngamma(`y'+1/`alfa2')))*((`alfa2'*`m2')^`y')*((1+`alfa2'*`m2')^-
(`y'+1/`alfa2'))
 quietly generate double `den1' =
(exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa1')))*(1-((1+`alfa1'*`m1')^-
(1/`alfa1')))
 quietly generate double `den2' =
(exp(lngamma(`y'+1)))*(exp(lngamma(1/`alfa2')))*(1-((1+`alfa2'*`m2')^-
(1/`alfa2')))
 quietly generate double `quo1' = (`num1')/(`den1')
 quietly generate double `quo2' = (`num2')/(`den2')
 quietly replace `lnf' = ln(`pi'*`quo1'+(1-`pi')*`quo2') if $ML_samp == 1
end
```

Program to estimate the model from different initial solutions. The example is for the LCH-Cross(NB2) model.

```
*****************************
version 9.1
set more off, permanently
run
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Progs\defvars_conj4"   //Open an prepare the dataset
run "Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\defprogs"  //load the
programs into memory

local
armazena="Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergenci
a\Conjunto4\Resultados\tpm_fmnb2"

local n_iteracoes = 20
forvalues i = 1/`n_iteracoes' {
  preserve
  sample 60
  quietly probit $d1 $var1 $var2 $var3 $var4 if entra == 1
  matrix sol1 = e(b)
  restore
  preserve
  sample 85
  quietly nbreg numcons $var1 $var2 $var3 $var4 if entra == 1, robust
dispersion(constant)
  matrix sol2 = e(b)
  restore
  preserve
  sample 60
  quietly nbreg numcons $var1 $var2 $var3 $var4 if entra == 1, robust
dispersion(constant)
  matrix sol3 = e(b)
  restore
  noisily display `i'
  local pe1 = uniform()

   ml model lf hlcm_nb2 (equ1:numcons = $var1 $var2 $var3 $var4) (equ2: $var1
$var2 $var3 $var4) (alfa2:) (equ3:$var1 $var2 $var3 $var4)  (alfa3:) (p1:) if
entra == 1, missing robust technique(bfgs ) title ("hlcm_nb2")
   ml init sol1 sol2 sol3 `pe1', copy
   noisily ml search /p1 0.1 0.9 /alfa2 0.1 3 /alfa3 0.1 2 , repeat(100) trace
   capture log using `armazena', append text
   noisily display `i'
   noisily ml report
   capture noisily ml maximize, difficult iterate(20000)
   if _rc == 0  {
        noisily display "tpm_fmnb2 has converged"
        matrix s`i'tpm_fmnb2 = e(b)'
        svmat double s`i'tpm_fmnb2
        matrix v`i'tpm_fmnb2 = J(3,1,0)
        matrix v`i'tpm_fmnb2[1,1]=e(ll)
        matrix v`i'tpm_fmnb2[2,1]=e(N)
        matrix v`i'tpm_fmnb2[3,1]=e(k)
        svmat double v`i'tpm_fmnb2
        capture log close
   }
   else {
      noisily display "tpm_fmnb2 NAO CONVERGIU"
      capture log close
   }

}
*GUARDAR AS SOLUCOES NA BD
capture drop  reg-entra
save
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\tpm_fmnb2",    replace
   ******************************************
```

Program to read the MLE solutions stored in a dataset and, without any interactions, estimate the models.

```
 version 9.1
log using
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Post_estimation_Analysis\Con
junto4\todos_resultados", append
tempname sol
set more off
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\nb1.dta" , clear
mkmat nb1, matrix (`sol') nomissing
matrix nb1 = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\hurdle_nb1.dta", clear
mkmat hurdle_nb1, matrix (`sol') nomissing
matrix hurdle_nb1 = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\fmnb1.dta", clear
mkmat fmnb1, matrix (`sol') nomissing
matrix fmnb1 = `sol''

use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\fmnb1_rest.dta", clear
mkmat fmnb1_rest, matrix (`sol') nomissing
matrix fmnb1_rest = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\tpm_fmnb1.dta", clear
mkmat tpm_fmnb1, matrix (`sol') nomissing
matrix tpm_fmnb1 = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\tpm_fmnb1_rest.dta", clear
mkmat tpm_fmnb1_rest, matrix (`sol') nomissing
matrix tpm_fmnb1_rest = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\nb2.dta", clear
mkmat nb2, matrix (`sol') nomissing
matrix nb2 = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\hurdle_nb2.dta", clear
mkmat hurdle_nb2, matrix (`sol') nomissing
matrix hurdle_nb2 = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\fmnb2.dta", clear
mkmat fmnb2, matrix (`sol') nomissing
matrix fmnb2 = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\fmnb2_rest.dta", clear
mkmat fmnb2_rest, matrix (`sol') nomissing
matrix fmnb2_rest = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\tpm_fmnb2.dta", clear
mkmat tpm_fmnb2, matrix (`sol') nomissing
matrix tpm_fmnb2 = `sol''
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
```

```
o4\Resultados\tpm_fmnb2_rest.dta", clear
mkmat tpm_fmnb2_rest, matrix (`sol') nomissing
matrix tpm_fmnb2_rest = `sol''
run
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Progs\defvars_conj4"   //abre e prepara a base de dados
run "Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\defprogs"  //load the
programs in memory

** estimacao do nb1
ml model lf nb1 (equ:numcons = $var1 $var2 $var3 $var4) (alfa1:) if entra == 1,
technique(bfgs) missing robust
ml init nb1, copy
ml maximize, diff  iterate(0)
estimates store nb1
estimates stats nb1

*estimacao do hurdle_nb1
ml model lf hnb1 (equ1:numcons = $var1 $var2 $var3 $var4) (equ2: $var1 $var2
$var3 $var4) (alfa1:)  if entra == 1,   technique(bfgs) missing robust
cluster(keyfam)
ml init hurdle_nb1, copy
ml maximize, difficult iterate(0)
estimates store hurdle_nb1
estimates stats hurdle_nb1


**** ESTIMACAO DO fmnb1
ml model lf lcm_nb1 (equ1:numcons = $var1 $var2 $var3 $var4) (alfa1:)
(equ2:$var1 $var2 $var3 $var4)  (alfa2:) (p1:) if entra == 1, missing robust
technique(bfgs )
ml init fmnb1, copy
capture noisily ml maximize, difficult   iterate(0)
estimates store fmnb1
estimates stats fmnb1
test [equ1=equ2] : $var1 $var2 $var3 $var4
exit

constraint drop _all
  local r = 1
  foreach t in $var1 $var2 $var3 $var4 {
      constraint `r' _b[equ1:`t']=_b[equ2:`t']
      local r = `r'+1
  }
  local r = `r' -1
  constraint list _all
  ml model lf lcm_nb1 (equ1:numcons = $var1 $var2 $var3 $var4) (alfa1:) (equ2:
$var1 $var2 $var3 $var4)  (alfa2:) (p1:) if entra == 1, constraints(1-`r')
missing robust technique(bfgs ) title ("fmnb1_rest")
  ml init fmnb1_rest, copy
  ml maximize, difficult iterate(0)
  estimates store fmnb1_rest
  estimates stats fmnb1_rest

ml model lf hlcm_nb1 (equ1:numcons = $var1 $var2 $var3 $var4) (equ2: $var1 $var2
$var3 $var4) (alfa2:) (equ3:$var1 $var2 $var3 $var4)  (alfa3:) (p1:) if entra ==
1, missing robust technique(bfgs )
ml init tpm_fmnb1, copy
ml maximize, difficult  iterate(0)
estimates store tpm_fmnb1
estimates stats tpm_fmnb1
test [equ2=equ3] : $var1 $var2 $var3 $var4

constraint drop _all
  local r = 1
  foreach t in $var1 $var2 $var3 $var4 {
      constraint `r' _b[equ2:`t']=_b[equ3:`t']
      local r = `r'+1
  }
  local r = `r' -1
  constraint list _all
```

```
ml model lf hlcm_nb1 (equ1:numcons = $var1 $var2 $var3 $var4) (equ2: $var1 $var2
$var3 $var4) (alfa2:) (equ3:$var1 $var2 $var3 $var4)  (alfa3:) (p1:) if entra ==
1, constraints(1-`r') missing robust technique(bfgs )
ml init tpm_fmnb1_rest, copy
ml maximize, difficult iterate(0)
estimates store tpm_fmnb1_rest
estimates stats tpm_fmnb1_rest

** estimacao do nb2
ml model lf nb2 (equ:numcons = $var1 $var2 $var3 $var4) (alfa1:) if entra == 1,
technique(bfgs) missing robust
ml init nb2, copy
ml maximize, diff        iterate(0)
estimates store nb2
estimates stats nb2

ml model lf hnb2 (equ1:numcons = $var1 $var2 $var3 $var4) (equ2: $var1 $var2
$var3 $var4) (alfa1:)  if entra == 1,   technique(bfgs) missing robust
ml init hurdle_nb2, copy
ml maximize, difficult iterate(00)
estimates store hurdle_nb2
estimates stats hurdle_nb2

ml model lf lcm_nb2 (equ1:numcons = $var1 $var2 $var3 $var4) (alfa1:)
(equ2:$var1 $var2 $var3 $var4)  (alfa2:) (p1:) if entra == 1, missing robust
technique(bfgs )
ml init fmnb2, copy
capture noisily ml maximize, difficult iterate(0)
estimates store fmnb2
estimates stats fmnb2
test [equ1=equ2] : $var1 $var2 $var3 $var4


constraint drop _all
  local r = 1
  foreach t in $var1 $var2 $var3 $var4 {
      constraint `r' _b[equ1:`t']=_b[equ2:`t']
      local r = `r'+1
  }
  local r = `r' -1
  constraint list _all
  ml model lf lcm_nb2 (equ1:numcons = $var1 $var2 $var3 $var4) (alfa1:) (equ2:
$var1 $var2 $var3 $var4)  (alfa2:) (p1:) if entra == 1, constraints(1-`r')
missing robust technique(bfgs ) title ("fmnb2_rest")
  ml init fmnb2_rest, copy
  ml maximize, difficult  iterate(0)
  estimates store fmnb2_rest
  estimates stats fmnb2_rest

** ESTIMAÇÃO DO TMP_FMNB2
ml model lf hlcm_nb2 (equ1:numcons = $var1 $var2 $var3 $var4) (equ2: $var1 $var2
$var3 $var4) (alfa2:) (equ3:$var1 $var2 $var3 $var4)  (alfa3:) (p1:) if entra ==
1, missing robust technique(bfgs )
ml init tpm_fmnb2, copy
ml maximize, difficult iterate(0)
estimates store tpm_fmnb2      // SAVE THE ESTIMATES UNDER THE NAME TPM_fmnb2
estimates stats tpm_fmnb2
test [equ2=equ3] : $var1 $var2 $var3 $var4


 constraint drop _all
  local r = 1
  foreach t in $var1 $var2 $var3 $var4 {
      constraint `r' _b[equ2:`t']=_b[equ3:`t']
      local r = `r'+1
  }
  local r = `r' -1
  constraint list _all

ml model lf hlcm_nb2 (equ1:numcons = $var1 $var2 $var3 $var4) (equ2: $var1 $var2
$var3 $var4) (alfa2:) (equ3:$var1 $var2 $var3 $var4)  (alfa3:) (p1:) if entra ==
1, constraints(1-`r') missing robust technique(bfgs )
```

```
ml init tpm_fmnb2_rest, copy
ml maximize, difficult iterate(0)
estimates store tpm_fmnb2_rest
estimates stats tpm_fmnb2_rest
estimates stats nb1
estimates stats hurdle_nb1
estimates stats fmnb1
estimates stats fmnb1_rest
estimates stats tpm_fmnb1
estimates stats tpm_fmnb1_rest
estimates stats nb2
estimates stats hurdle_nb2
estimates stats fmnb2
estimates stats fmnb2_rest
estimates stats tpm_fmnb2
estimates stats tpm_fmnb2_rest
log close
```

Stata Program to conduct the Voung tests: Example of code to test model LCH-Cross(NB1) (hlcm_nb1) against model LCH-Cross(NB2) (hlcm_nb2):

```
tempname sol
set more off
*1)
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\tpm_fmnb1.dta" , clear
mkmat tpm_fmnb1, matrix (`sol') nomissing
matrix tpm_fmnb1 = `sol''
*2)
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\tpm_fmnb2.dta", clear
mkmat tpm_fmnb2, matrix (`sol') nomissing
matrix tpm_fmnb2 = `sol''

run
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Progs\defvars_conj4"   //
run "Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\defprogs"  //load the
programs in memory

** estimacao do tpm_fmnb1
ml model lf h_lcm_nb1 (equ1_m1:numcons = $var1 $var2 $var3 $var4) (equ2_m1:
$var1 $var2 $var3 $var4) (alfa2_m1:) (equ3_m1:$var1 $var2 $var3 $var4)
(alfa3_m1:) (p1_m1:) if entra == 1, missing robust technique(bfgs )
ml init tpm_fmnb1, copy
ml maximize, difficult     iterate(0)
local n = e(N)
display `n'
local logl_tpm_fmnb1 = e(ll)
display `logl_tpm_fmnb1'
local m1 = "exp(xb(equ1_m1))"
local m2 = "exp(xb(equ2_m1))"
local m3 = "exp(xb(equ3_m1))"
local a2 = "_b[alfa2_m1:_cons]"
local a3 = "_b[alfa3_m1:_cons]"
local p1 = "_b[p1_m1:_cons]"
local p2 = "(1-_b[p1_m1:_cons])"
local y = "numcons"
quietly generate y0 = `y' == 0
quietly generate y1 = `y' > 0
display `p1'
predictnl double ln_tpm_fmnb1 = y0*(-ln((2^`m1')-1))+ln((2^`m1')-1)-
`m1'*ln(2)+(y1)*(ln(`p1'*(((exp(lngamma(`y'+`m2'/`a2')))*((1+`a2')^-
(`y'+`m2'/`a2'))*(`a2'^`y'))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m2'/`a2')))*(1
-((1+`a2')^-(`m2'/`a2'))))))+(`p2')*(((exp(lngamma(`y'+`m3'/`a3')))*((1+`a3')^-
```

```
(`y'+`m3'/`a3'))*(`a3'^`y'))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m3'/`a3')))*(1
-((1+`a3')^-(`m3'/`a3')))))))  if entra == 1

ml model lf hlcm_nb2 (equ1_m2:numcons = $var1 $var2 $var3 $var4) (equ2_m2: $var1
$var2 $var3 $var4) (alfa2_m2:) (equ3_m2:$var1 $var2 $var3 $var4)  (alfa3_m2:)
(p1_m2:) if entra == 1, missing robust technique(bfgs )
ml init tpm_fmnb2, copy
ml maximize, difficult iterate(0) trace
local logl_tpm_fmnb2 = e(ll)
display `logl_tpm_fmnb2'

local m1 = "exp(xb(equ1_m2))"
local m2 = "exp(xb(equ2_m2))"
local m3 = "exp(xb(equ3_m2))"
local a2 = "_b[alfa2_m2:_cons]"
local a3 = "_b[alfa3_m2:_cons]"
local p1 = "_b[p1_m2:_cons]"
local p2 = "(1-_b[p1_m2:_cons])"

predictnl double ln_tpm_fmnb2 = -y0*ln(1+`m1')+y1*ln(`m1')-
y1*ln(1+`m1')+y1*ln(`p1'*(((exp(lngamma(`y'+1/`a2')))*((`a2'*`m2')^`y')*((1+`a2'
*`m2')^-(`y'+1/`a2')))/((exp(lngamma(`y'+1)))*(exp(lngamma(1/`a2')))*(1-
((1+`a2'*`m2')^-
(1/`a2')))))+`p2'*(((exp(lngamma(`y'+1/`a3')))*((`a3'*`m3')^`y')*((1+`a3'*`m3')^
-(`y'+1/`a3')))/((exp(lngamma(`y'+1)))*(exp(lngamma(1/`a3')))*(1-
((1+`a3'*`m3')^-(1/`a3'))))))  if entra == 1
local numerador = (`logl_tpm_fmnb1'-`logl_tpm_fmnb2')/sqrt(`n')
generate ee = (ln_tpm_fmnb1-ln_tpm_fmnb2)^2 if entra == 1
sum ee if entra == 1
local den_parte1 = r(mean)
display `den_parte1'
local den_parte2 = ((`logl_tpm_fmnb1'-`logl_tpm_fmnb2')^ 2)/(`n'^2)
local denominador = sqrt(`den_parte1'-`den_parte2')
local V= `numerador'/`denominador'
display `V'
```

Stata module to estimate the average marginal effect of the covariate education, in the low users population, implied by the LCM_NB1 specification. The program to estimate the average marginal effect of other covariates is similar to the presented below. The main difference is that in the stata ***predictnl*** command, where the appropriate expression should be used as argument.

```
version 9.1
set more off
tempname sol
use
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Resultados\fmnb1.dta", clear
mkmat fmnb1, matrix (`sol') nomissing
matrix fmnb1 = `sol''
run
"Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\Analisa_convergencia\Conjunt
o4\Progs\defvars_conj4"   //abre e prepara a base de dados
*** LEITURA DOS PROGRAMAS PARA MEMORIA
run "Z:\Douto\ProgStata\SubSistemas1_Contagem\Para_Tese\defprogs"  //load the
programs in memory
ml model lf lcm_nb1 (equ1:numcons = $var1 $var2 $var3 $var4) (alfa1:)
(equ2:$var1 $var2 $var3 $var4)  (alfa2:) (p1:) if entra == 1, missing robust
technique(bfgs )
ml init fmnb1, copy
```

```
ml maximize, difficult   iterate(0)
matrix V = e(V)
matrix U1 = J(73, 1, 0)
local p1 = "_b[p1:_cons]"
local p2 = "(1-_b[p1:_cons])"
local hu = "equ1"
local lu = "equ2"
local a1 = "_b[alfa1:_cons]"
local a2 = "_b[alfa2:_cons]"
local m1 = "exp(xb(equ1))"
local m2 = "exp(xb(equ2))"
local y = "numcons"
local fhu =
"(((exp(lngamma(`y'+(`m1'/`a1'))))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m1'/`a1'
)))))*((1+`a1')^-(`y'+(`m1'/`a1')))*(`a1')^`y')"
local flu =
"(((exp(lngamma(`y'+(`m2'/`a2'))))/((exp(lngamma(`y'+1)))*(exp(lngamma(`m2'/`a2'
)))))*((1+`a2')^-(`y'+(`m2'/`a2')))*(`a2')^`y')"
predictnl double PHU = (`p1'*`fhu')/(`p1'*`fhu'+`p2'*`flu') if entra == 1
predictnl double PLU = (`p2'*`flu')/(`p1'*`fhu'+`p2'*`flu') if entra == 1
quietly generate HU = 0 if PHU < PLU & entra == 1
quietly replace HU = 1 if PHU > PLU & entra == 1
quietly generate LU = 0 if entra == 1 & PLU < PHU
quietly replace LU = 1 if PLU > PHU & entra == 1

**** EFFECT OF EDUCATION ON lOW:HIGH USERS
predictnl double ef = _b[`lu':educ]*`m2' if LU == 1, g(ef)
sum ef if LU == 1
local me = r(mean)
forvalues j= 37/71 {
  quietly sum ef`j' if LU == 1
  matrix U1[`j',1]=r(mean)
}

 matrix W = U1'*V*U1
 local se_me = sqrt(W[1,1])
 local z_me = `me'/`se_me'
 display "equ1:`i'"
 display "equ1:ME `i' ------> "`me'
 display "equ1:SE_ME `i' ---> "`se_me'
 display "equ1:Z_ME `i' ----> "`z_me'

capture drop ef
capture drop ef1-ef35
capture drop ef37-ef71
```

Bibliographic references

Acton, J. P. (1975). "Nonmonetary Factors in the Demand for Medical Services: Some Empirical Evidence." The Journal of Political Economy 83(3): 595-614.

Ai, C. R. and E. C. Norton (2003). "Interaction terms in logit and probit models." Economics Letters 80(1): 123-129.

Aitkin, M. and D. B. Rubin (1985). "Estimation and Hypothesis-Testing in Finite Mixture-Models." Journal of the Royal Statistical Society Series B-Methodological 47(1): 67-75.

Akin, J. S., C. C. Griffin, et al. (1986). "The Demand for Primary Health-Care Services in the Bicol Region of the Philippines." Economic Development and Cultural Change 34(4): 755-782.

Alberto, A. S. (1999). Uma abordagem histórica. O Papel dos Sistemas Privados de Saúde num Sistema em Mudança. Lisboa, Associação Nacional de Sistemas de Saúde: 19-25.

Állvarez, B. and M. A. Delgado (2002). "Goodness-of-fit techniques for count data models: an application to the demand for dental care in Spain." Empirical Economics 27(3): 543-567.

Atella, V., F. Brindisi, et al. (2004). "Determinants of access to physician services in Italy: a latent class seemingly unrelated probit approach." Health Economics 13(7): 657-668.

Bago d'Uva, T. (2005). "Latent class models for use of primary care: evidence from a British panel." Health Economics 14(9): 873-892.

Bago d'Uva, T. (2005). Latent class models for utilisation of health care. HEDG Working paper 05/01. University of York.

Bago d'Uva, T. (2006). "Latent class models for utilisation of health care." Health Economics 15(4): 329-343.

Barros, P. P. (1999). Os sistemas privados de saúde e a reforma do sistema de saúde. O Papel dos Sistemas Privados de Saúde num Sistema em Mudança. Associação Nacional de Sistemas de Saúde. Lisboa, ANSS: 90-115.

Barros, P. P. (2005). Economia da Saúde: Conceitos e Comportamentos. Coimbra,

Almedina.

Barros, P. P., M. P. Machado, et al. (2005). Moral Hazard and the Demand For Health Services: A Matching Estimator Approach. Working paper 05/59, . Departamento de Economia, Universidade Carlos III de Madrid.

Barros, P. P. and P. Olivella (2005). "Waiting Lists and Patient Selection." Journal of Economics & Management Strategy 14(3): 623-646.

Batista, C. (1999). Uma abordagem Quantitativa: Os subsistemas de saúde, Compromisso na eficácia e na qualidade. O Papel dos Sistemas Privados de Saúde num Sistema em Mudança. Associação Nacional de Sistemas de Saúde. Lisboa, ANSS: 26-65.

Becker, G. S. (1965). "A Theory of the Allocation of Time." Economic Journal 75(299): 493-517.

Bentes, M., C. M. Dias, et al. (2004). Health Care Systems in Transition: Portugal. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies.

Bijan, J. B. (2006). "A mixed logit model of health care provider choice: analysis of NSS data for rural India." Health Economics 15(9): 915-932.

Blundell, R. and F. Windmeijer (2000). "Identifying demand for health resources using waiting times information." Health Economics 9(6): 465-474.

Bohning, D. (1995). "A Review of Reliable Maximum-Likelihood Algorithms for Semiparametric Mixture-Models." Journal of Statistical Planning and Inference 47(1-2): 5-28.

Bohning, D. and R. Kuhnert (2006). "Equivalence of Truncated Count Mixture Distributions and Mixtures of Truncated Count Distributions." Biometrics 0(0).

Bolduc, D., G. Lacroix, et al. (1996). "The choice of medical providers in rural Benin: A

comparison of discrete choice models." Journal of Health Economics 15(4): 477-498.

Brannas, K. and G. Rosenqvist (1994). "Semiparametric Estimation of Heterogeneous Count Data Models." European Journal of Operational Research 76(2): 247-258.

Buchmueller, T. C., A. Couffinhal, et al. (2004). "Access to physician services: does supplemental insurance matter? Evidence from France." Health Economics 13(7): 669-687.

Cabral, M. V., P. A. Silva, et al. (2002). Saúde e Doença em Portugal. Lisboa, Imprensa de Ciências Sociais.

Cameron, A. C. and P. K. Trivedi (1986). "Econometric Models Based on Count Data: Comparisons and applications of Some Estimators and Tests." Journal of Applied Econometrics 1: 29-53.

Cameron, A. C. and P. K. Trivedi (1996). Count data models for financial data. Handbook of Statistics. G. S. Maddala and C. R. Rao. 14: 363-391.

Cameron, A. C. and P. K. Trivedi (1998). Regression analysis of count data. Cambridge, UK ; New York, NY, USA, Cambridge University Press.

Cameron, A. C. and P. K. Trivedi (2005). Microeconometrics : methods and applications. New York, NY, Cambridge University Press.

Cameron, A. C., P. K. Trivedi, et al. (1988). "A Microeconometric Model of the Demand for Health-Care and Health-Insurance in Australia." Review of Economic Studies 55(1): 85-106.

Cauley, S. D. (1987). "The Time Price of Medical-Care." Review of Economics and Statistics 69(1): 59-66.

Chernoff, H. (1954). "On the Distribution of the Likelihood Ratio." Annals of Mathematical Statistics 25(3): 573-578.

Chiappori, P. A., F. Durand, et al. (1998). "Moral hazard and the demand for physician services: First lessons from a French natural experiment." European Economic Review 42(3-5): 499-511.

Coffey, R. M. (1983). "The Effect of Time Price on the Demand for Medical-Care Services." Journal of Human Resources 18(3): 407-424.

Colle, A. D. and M. Grossman (1978). "Determinants of Pediatric Care Utilization." Journal of Human Resources 13: 115-158.

Cragg, J. G. (1971). "Some Statistical Models for Limited Dependent Variables with Application to Demand for Durable Goods." Econometrica 39(5): 829-844.

Creel, M. D. and J. B. Loomis (1990). "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California." American Journal of Agricultural Economics 72(2): 434-441.

Cullis, J. G., P. R. Jones, et al. (2000). Waiting lists and medical treatment: Analysis and policies. Handbook of Health Economics. 1: 1201-1249.

Culyer, A. J., E. Vandoorslaer, et al. (1992). "Utilization as a Measure of Equity by Mooney, Hall, Donaldson and Gerard." Journal of Health Economics 11(1): 93-98.

Davidson, R. and J. G. Mackinnon (1993). Estimation and Inference in Econometrics. New York, Oxford University Press.

Deb, P. (2001). "A discrete random effects probit model with application to the demand for preventive care." Health Economics 10(5): 371-383.

Deb, P. (2002). A Discrete Random Effects Probit Model with Application to the Demand for Preventive Care. Econometric Analysis of Health Data. O. O. D. Andrew M. Jones: 189-200.

Deb, P. and A. M. Holmes (2000). "Estimates of use and costs of behavioural health care: A comparison of standard and finite mixture models." Health Economics

9(6): 475-489.

Deb, P. and P. K. Trivedi (1997). "Demand for medical care by the elderly: A finite mixture approach." Journal of Applied Econometrics 12(3): 313-336.

Deb, P. and P. K. Trivedi (2002). "The structure of demand for health care: latent class versus two-part models." Journal of Health Economics 21(4): 601-625.

Deb, P. and P. K. Trivedi (2006). "Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization." Econometrics Journal 9(2): 307-331.

Dempster, A. P., N. M. Laird, et al. (1977). "Maximum Likelihood from Incomplete Data Via Em Algorithm." Journal of the Royal Statistical Society Series B-Methodological 39(1): 1-38.

Direcção Geral da Saúde. (2005). "Guia da Utente do Serviço Nacional de Saúde." Retrieved May 2005, 2005, from http://www.dgsaude.pt/default.aspx?cn=55065715AAAAAAAAAAAAAAAA.

Dixon, A. and E. Mossialos (2000). "Has the Portuguese NHS Achieved its Objectives of Equity and Efficiency?" International Social Security Review 53(4): 49-78.

Dor, A., P. Gertler, et al. (1987). Non-Price Rationing and the Choice of Medical-Care Providers in Rural Cote-Divoire. Journal of Health Economics. 6: 291-304.

Duan, N., W. G. Manning, et al. (1983). "A comparison of alternative models for the demand for medical care." Journal of business and Statistics 1: 115-126.

Englin, J. and J. S. Shonkwiler (1995). "Estimating Social-Welfare Using Count Data Models - an Application to Long-Run Recreation Demand under Conditions of Endogenous Stratification and Truncation." Review of Economics and Statistics 77(1): 104-112.

Ferreira, P., V. Raposo, et al. (2005). A Voz dos Utilizadores dos Centros de Saúde.

Lisboa, Ministério da Saúde, Instituto da Qualidade em Saúde.

Ferreira, P. F. (1999). A Voz dos Doentes: Satisfação com a Medicina Geral e Familiar. Instrumentos para a melhoria continua da qualidade, Direcção Geral da Saúde.

Ferreira, P. L., A. Luz, et al. (2001). "Determinantes da satisfação dos utentes dos cuidados primários: O caso de Lisboa e Vale do Tejo." Revista Portuguesa de Saúde Publica Volume temático 2: 53-61.

Geil, P., A. Million, et al. (1997). "Economic incentives and hospitalization in Germany." Journal of Applied Econometrics 12(3): 295-311.

Gerdtham, U. G. (1997). "Equity in health care utilization: Further tests based on hurdle models and Swedish micro data." Health Economics 6(3): 303-319.

Gerdtham, U. G. and P. K. Trivedi (2001). Equity in Swedish health care reconsidered: New results based on the finite mixture model. Health Economics. 10: 565-572.

Gertler, P., L. Locay, et al. (1987). Are User Fees Regressive - the Welfare Implications of Health-Care Financing Proposals in Peru. Journal of Econometrics. 36: 67-88.

Goddard, J. A. and M. Tavakoli (1998). Referral rates and waiting lists: Some empirical evidence. Health Economics. 7: 545-549.

Gould, W. and W. Sribney (2003). Maximum Likelihood Estimation with Stata, Stata Press

Gourieroux, C., A. Monfort, et al. (1984). "Pseudo Maximum-Likelihood Methods - Applications to Poisson Models." Econometrica 52(3): 701-720.

Gourieroux, C. and M. Visser (1997). "A count data model with unobserved heterogeneity." Journal of Econometrics 79(2): 247-268.

Gouveia, M. (1999). Os sistemas privados de saúde no financiamento. O Papel dos Sistemas Privados de Saúde num Sistema em Mudança. Lisboa, Associação Nacional de Sistemas de Saúde: 131-148.

Gravelle, H., M. Dusheiko, et al. (2002). The demand for elective surgery in a public system: time and money prices in the UK National Health Service. Journal of Health Economics. 21: 423-449.

Greene, W. H. (2003). Econometric analysis. Upper Saddle River, N.J., [Great Britain], Prentice Hall.

Gritz, R. M. (1993). "The Impact of Training on the Frequency and Duration of Employment." Journal of Econometrics 57(1-3): 21-51.

Grogger, J. T. and R. T. Carson (1991). Models for Truncated Counts. Journal of Applied Econometrics. 6: 225-238.

Grol, R., M. Wensing, et al. (2000). Patients in Europe evaluate general practice care: an international comparison. British Journal of General Practice. 50: 882-887.

Grootendorst, P. V. (1995). "A Comparison of Alternative Models of Prescription Drug Utilization." Health Economics 4(3): 183-198.

Grootendorst, P. V. (2002). A Comparison of Alternative Models of Prescription Drug Utilization. Econometric Analysis of Health Data. O. O. D. Andrew M. Jones: 71-86.

Grossman, M. (1972). "Concept of Health Capital and Demand for Health." Journal of Political Economy 80(2): 223-225.

Gurmu, S. (1991). "Tests for Detecting Overdispersion in the Positive Poisson Regression-Model." Journal of Business & Economic Statistics 9(2): 215-222.

Gurmu, S. (1997). "Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization." Journal of Applied Econometrics 12(3): 225-242.

Gurmu, S., P. Rilstone, et al. (1999). Semiparametric estimation of count regression models. Journal of Econometrics. 88: 123-150.

Gurmu, S. and P. K. Trivedi (1992). Overdispersion Tests for Truncated Poisson Regression-Models. Journal of Econometrics. 54: 347-370.

Gurmu, S. and P. K. Trivedi (1996). "Excess zeros in count models for recreational trips." Journal of Business & Economic Statistics 14(4): 469-477.

Hakkinen, U., G. Rosenqvist, et al. (1996). "Economic depression and the use of physician services in Finland." Health Economics 5(5): 421-434.

Harmon, C. and B. Nolan (2001). "Health insurance and health services utilization in Ireland." Health Economics 10(2): 135-145.

Hauenstein, E. J., S. Petterson, et al. (2006). "Rurality, gender, and mental health treatment." Family & Community Health 29(3): 169-185.

Hausman, J., B. Hall, et al. (1981). Econometric Models for count data with an application to the Patents-R&D relationship, NBER.

Heckman, J. and B. Singer (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric-Models for Duration Data. Econometrica. 52: 271-320.

Holly, A., L. Gardiol, et al. (1998). "An econometric model of health care utilization and health insurance in Switzerland." European Economic Review 42(3-5): 513-522.

Instituto Nacional de Estatística (1999). Indicadores Urbanos do Continente. Lisboa, Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2001). Censos 2001 : resultados provisórios : XIV recenseamento geral da população. Lisboa, Instituto Nacional de Estatística.

James, L. F., C. E. Preibe, et al. (2001). "Consistent estimation of mixture complexity." Annals of Statistics 29(5): 1281-1296.

Janssen, R. (1992). "Time Prices and the Demand for Gp-Services." Social Science &

Medicine 34(7): 725-733.

Jimenez-Martin, S., J. M. Labeaga, et al. (2002). "Latent class versus two-part models in the demand for physician services across the European Union." Health Economics 11(4): 301-321.

Jiménez-Martín, S., J. M. Lebeaga, et al. (2002). Latent Class versus Two-Part Models in the Demand for Physician Services Across the European Union. Econometric Analysis of Health Data. O. O. D. Andrew M. Jones: 101-116.

Jones, A. M. (2000). Health econometrics. Handbook of Health Economics. A. J. Culyer and J. P. Newhouse. New York, Elsevier. 1: 265-344.

Jones, A. M., X. Koolman, et al. (2004). The impact of supplementary private health insurance on the use of specialists in European countries.

Jones, A. M. and O. O'Donnell (2002). Econometric analysis of health data. Chichester, Wiley.

Jones, A. M., N. Rice, et al. (2006). The Dynamics of Health. Elgar Companion to Health Economics. A. M. Jones, Edward Elgar Publishing.

Keeler, E. B., W. G. Manning, et al. (1988). "The Demand for Episodes of Mental-Health Services." Journal of Health Economics 7(4): 369-392.

Lindsay, B. G. (1995). Mixture models : theory, geometry, and applications. Hayward, Calif. Alexandria, Va., Institute of Mathematical Statistics ; American Statistical Association.

Lindsay, B. G. and M. L. Lesperance (1995). "A Review of Semiparametric Mixture-Models." Journal of Statistical Planning and Inference 47(1-2): 29-39.

Lindsay, C. M. and B. Feigenbaum (1984). "Rationing by Waiting Lists." American Economic Review 74(3): 404-417.

Lourenço, Ó. (2003). Os utentes dos subsistemas de saúde vs os utentes do SNS: A

utilização de cuidados de saúde privados e especializados. 8th Meeting of Portuguese Association of Health Economics.

Lourenço, Ó. D. and P. L. Ferreira (2005). "Utilization of public health centres in Portugal: effect of time costs and other determinants. Finite mixture models applied to truncated samples." Health Economics 14(9): 939-953.

Manning, W. G. and M. S. Marquis (1996). "Health insurance: The tradeoff between risk pooling and moral hazard." Journal of Health Economics 15(5): 609-639.

Manning, W. G., J. P. Newhouse, et al. (1987). "Health-Insurance and the Demand for Medical-Care - Evidence from a Randomized Experiment." American Economic Review 77(3): 251-277.

Martin, S. and P. C. Smith (1999). "Rationing by waiting lists: an empirical investigation." Journal of Public Economics 71(1): 141-164.

McLachlan, G. J. and D. Peel (2000). Finite mixture models. New York, Wiley.

Ministério da Saúde - Instituto Nacional de Saúde (1999). INS 1998/1999. Continente. Dados Gerais., INSA - Instituto Nacional de Saúde.

Mocan, H. N., E. Tekin, et al. (2004). The demand for medical care in urban China. World Development. 32: 289-304.

Mooney, G., J. Hall, et al. (1991). "Utilization as a Measure of Equity - Weighing Heat." Journal of Health Economics 10(4): 475-480.

Mossialos, E. and S. M. S. Thomson (2002). "Voluntary health insurance in the European Union: A critical assessment." International Journal of Health Services 32(1): 19-88.

Mullahy, J. (1986). "Specification and Testing of Some Modified Count Data Models." Journal of Econometrics 33(3): 341-365.

Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models.

Journal of Applied Econometrics. 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(3): 247-281.

Mwabu, G., M. Ainsworth, et al. (1993). "Quality of Medical-Care and Choice of Medical-Treatment in Kenya - an Empirical-Analysis." Journal of Human Resources 28(4): 838-862.

Nagin, D. S. and K. C. Land (1993). "Age, Criminal Careers, and Population Heterogeneity - Specification and Estimation of a Nonparametric, Mixed Poisson Model." Criminology 31(3): 327-362.

National Social Security. (2005). "Web page. https://www.seg-social.pt." Retrieved 05 May 2005.

Newhouse, J. P., C. E. Phelps, et al. (1980). "On Having Your Cake and Eating It Too - Econometric Problems in Estimating the Demand for Health-Services." Journal of Econometrics 13(3): 365-390.

Observatório Português dos Sistemas de Saúde (2003). SAÚDE: que rupturas? Lisboa, Escola Nacional de Saúde Pública.

Observatório Português dos Sistemas de Saúde (2004). Incertezas … gestão da mudança na saúde. Lisboa, Escola Nacional de Saúde Pública.

Observatório Português dos Sistemas de Saúde (2005). Novo Serviço Público da Saúde. Novos desafios. Lisboa, Escola Nacional de Saúde Pública.

OECD (2004). OECD Economic Surveys, Portugal, OECD (ed.). Paris, OECD.

OECD. (2004). "OECD GLOSSARY OF STATISTICAL TERMS." Retrieved February 2004, 2004.

OECD (2006). OECD health data 2006. Organization for the Economic Co-operation and Development, Paris.

Oehlert, G. W. (1992). "A Note on the Delta Method." American Statistician 46(1): 27-29.

Oliveira, M. D. (2001). Voluntary health insurance in Portugal: a study for the European Commission. London, London School of Economics and Political Science.

Oliveira, M. D. (2004). Modelling demand and supply influences on utilization: a flow demand model to predict hospital utilization at the small area level. Applied Economics. 36: 2237-2251.

Oliveira, M. D. and G. Bevan (2003). Measuring geographic inequities in the Portuguese health care system: an estimation of hospital care needs. Health Policy. 66: 277-293.

Oliveira, M. D., Gouveia Pinto, C. (2005). "Health care reform in Portugal: an evaluation of the NHS experience." Health Economics 14(S1): S203-S220.

Pereira, I. (2004). Medical Care Demand and Health Systems - Portuguese Case. U. A. d. Barcelona. Barcelona.

Pereira, J. (1990). "Equity Objectives in Portuguese Health-Policy." Social Science & Medicine 31(1): 91-94.

Pereira, J. (1995). Equity, health and health care: an economic study with reference to Portugal. Department of Economics and Related Studies of the University of York. York, University of York: 341 p : il.

Phelps, C. E. and J. P. Newhouse (1974). "Coinsurance, The Price of Time, and the Demand for Medical Services." The Review of Economic Statistics 56(3): 334-342.

Pohlmeier, W. and V. Ulrich (1995). An Econometric-Model of the 2-Part Decision-Making Process in the Demand for Health-Care. Journal of Human Resources. 30: 339-361.

Portugal (1992). Decreto-Lei n.o 54/92, Ministério da Saúde. n.o 54/1992: 7101-7103.

Propper, C. (1990). Contingent Valuation of Time Spent on Nhs Waiting Lists. Economic Journal. 100: 193-199.

Propper, C. (1995). The Disutility of Time Spent on the United-Kingdom National-Health-Service Waiting Lists. Journal of Human Resources. 30: 677-700.

Rodriguez, M. and A. Stoyanova (2004). "The effect of private insurance access on the choice of GP/specialist and public/private provider in Spain." Health Economics 13(7): 689-703.

Ruiz, F., L. Amaya, et al. (2007). "Progressive segmented health insurance: Colombian health reform and access to health services." Health Economics 16(1): 3-18.

Santana, P. (1996). Utilização de cuidados hospitalares: uma abordagem da geografia da saúde. As reformas dos sistemas de saúde. C. G. Pinto, F. Ramos and J. Pereira. Lisboa, APES.

Santana, P. (1999). Geografia das Desigualdades em Saúde e Estado de Saúde. Estudos em Homenagem a Augusto Mantas. P. P. Barros and J. A. Simões. Lisboa, APES.

Santos-Silva, J. M. C. (1997). "Unobservables in count data models for on-site samples." Economics Letters 54(3): 217-220.

Santos-Silva, J. M. C. (2001). "A score test for non-nested hypotheses with applications to discrete data models." Journal of Applied Econometrics 16(5): 577-597.

Santos-Silva, J. M. C. (2003). "A note on the estimation of mixture models under endogenous sampling." Econometrics Journal 36: 46-52.

Santos-Silva, J. M. C. and F. Windmeijer (2001). Two-part multiple spell models for health care demand. Journal of Econometrics. 104: 67-89.

Sarma, S. and W. Simpson (2006). "A microeconometric analysis of Canadian health care utilization." Health Economics 15(3): 219-239.

Schellhorn, M. (2001). "The effect of variable health insurance deductibles on the demand for physician visits." Health Economics 10(5): 441-456.

Shaw, D. (1988). "On-Site Samples Regression - Problems of Non-Negative Integers, Truncation, and Endogenous Stratification." Journal of Econometrics 37(2): 211-223.

Sin, C. Y. and H. White (1996). Information criteria for selecting possibly misspecified parametric models. Journal of Econometrics. 71: 207-225.

van Doorslaer, E. and A. M. Jones (2004). Income-related inequality in health and health care in the European Union. Health Economics. 13: 605-608.

van Doorslaer, E. and X. Koolman (2004). Explaining the differences in income-related health inequalities across European countries. Health Economics. 13: 609-628.

van Doorslaer, E., X. Koolman, et al. (2004). "Explaining income-related inequalities in doctor utilisation in Europe." Health Economics 13(7): 629-647.

van Doorslaer, E., C. Masseria, et al. (2006). "Inequalities in access to medical care by income in developed countries." Canadian Medical Association Journal 174(2): 177-183.

van Doorslaer, E. K. A., C. Masseria, et al. (2004). Income-related Inequality in the Use of Medical Care in 21 OECD Countries, Organisation for Economic Co-operation and Development.

Van Ourti, T. (2004). "Measuring horizontal inequity in Belgian health care using a Gaussian random effects two part count data model." Health Economics 13(7): 705-724.

Vella, F. (1998). "Estimating models with sample selection bias: A survey." Journal of Human Resources 33(1): 127-169.

Vera-Hernandez, A. M. (1999). "Duplicate coverage and demand for health care. The

case of Catalonia." Health Economics 8(7): 579-598.

Vuong, Q. H. (1989). "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." Econometrica 57(2): 307-333.

Wagstaff, A. (1986). "The Demand for Health - Some New Empirical-Evidence." Journal of Health Economics 5(3): 195-233.

Wagstaff, A. and E. van Doorslaer (2000). Equity in health care finance and delivery. Handbook of Health Economics. A. J. Culyer and J. P. Newhouse. New York, Elsevier. 1: 1803-1862.

Wagstaff, A. and E. van Doorslaer (2000). "Measuring and testing for inequity in the delivery of health care." Journal of Human Resources 35(4): 716-733.

Wagstaff, A., E. Vandoorslaer, et al. (1991). "On the Measurement of Horizontal Inequity in the Delivery of Health-Care." Journal of Health Economics 10(2): 169-205.

Wang, P. M., I. M. Cockburn, et al. (1998). "Analysis of patent data - A mixed-Poisson-regression-model approach." Journal of Business & Economic Statistics 16(1): 27-41.

Wedel, M., W. S. Desarbo, et al. (1993). "A Latent Class Poisson Regression-Model for Heterogeneous Count Data." Journal of Applied Econometrics 8(4): 397-411.

White, H. (1982). "Maximum-Likelihood Estimation of Mis-Specified Models." Econometrica 50(1): 1-25.

Windmeijer, F. A. G. and J. M. C. S. Silva (1997). Endogeneity in count data models: An application to demand for health care. Journal of Applied Econometrics. 12: 281-294.

Winkelmann, R. (2003). Econometric analysis of count data. Berlin ; New York, Springer.

Winkelmann, R. (2004). "Health care reform and the number of doctor visits - An econometric analysis." Journal of Applied Econometrics 19(4): 455-472.

Wooldridge, J. M. (1997). Quasi-Likelihood methods for count data. Handbook of apllied econometrics, Vol. II: Microeconomics. M. Pesaran and P. Schmidt, Blackwell Publishers. 2.

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. Cambridge, Mass., MIT Press.

Yawn, B. P., A. G. Mainous, et al. (2001). "Do rural and urban children have comparable asthma care utilization?" Journal of Rural Health 17(1): 32-39.

Yen, S. T., C. H. Tang, et al. (2001). Demand for traditional medicine in Taiwan: A mixed Gaussian-Poisson model approach. Health Economics. 10: 221-232.

Zhang, P., G. Tao, et al. (2000). "Utilization of preventive medical services in the United States: A comparison between rural and urban populations." Journal of Rural Health 16(4): 349-356.

Zweifel, P. (1981). Supplier-Induced Demand in a Model of Physician Behavior. Health, Economics, and Health Economics. J. Van der Gaag and M. Perlman, North-Holland: 245-267.