THIRTEENTH EUROPEAN WORKSHOP ON ECONOMETRICS AND HEALTH ECONOMICS

The impact of non-monetary factors on the primary care utilization in Portugal. Finite mixture models applied to on-site and truncated samples

Óscar Domingos Lourenço & Pedro Lopes Ferreira

Venafro, Italy

22-25 September 2004

The impact of non-monetary factors on the Primary Care utilization in Portugal: Finite mixture models applied to on-site and truncated samples

Óscar D. Lourenço [a,*], Pedro L. Ferreira [a]

School of Economics, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal

Draft version

Abstract

In the Portuguese National Health Service (NHS) patients have to pay a co-payment of 2€ to visit a GP in the health centres. Therefore, the monetary price associated to each visit is low and, with a high probability, is not a factor that affects the utilization of consultations in health centres. On the other hand, in any health system in which the monetary cost to consume medical care is very low other kind of costs can emerge as determinants of medical care utilization. The Portuguese NHS suffers from several time-related inefficiencies and so, the non-monetary form of co-payment is a non negligible reality. With data in our database we have concluded that the average waiting time to visit a GP is approximately 9 days. Moreover, the average waiting time in the waiting room for a consultation is approximately 1 hour. Therefore, this study aims at analysing the impact of non-monetary factors on the utilization of public GPs. This study can be useful for policy making, as well as for econometric reasons.

In the other hand, sometimes the empirical researcher faces non-random samples. So, modelling based on the assumption that we have a simple random sample can be inappropriate and misleading. In this research we face this same situation. Our data resulted from the application of two endogenous sampling schemes: an sample collected on-site and a truncated sample. Therefore each sampling scheme generates a selected sample. Thus, to make valid inference, adequate econometric modelling has to be used. To model our dependent variable, number of visits, and to take into account the unobserved heterogeneity, we relied on a semi-parametric specification through the use of finite mixture models.

The data were obtained from the Europep questionnaire, a standardized questionnaire designed to measure primary patient satisfaction in European countries.

Keywords: endogenous sample, finite mixture model, primary care utilization

* Corresponding author: Tel./fax: + 351 230 790 507
E-mail address: osl@fe.uc.pt

1. Introduction

In the Portuguese National Health Service (NHS), as in other European NHS, the entry point to the health system is the Health Centre (HC) where the General Practitioners (GP) are generally paid on a salary basis. The visit to a public GP costs (monetary cost) approximately 2€ to the patient. However, some patients with special needs, (e.g., elderly, children under 14 years old, pregnant women and patients with some chronic diseases) are exempt from that payment. Therefore, the monetary price to visit a GP is small and, with a high probability, it is not a factor that affects the demand of health centres consultations.

In a health system where the monetary cost to consume medical care is very low, at least at the moment of consumption, other kind of costs can emerge as determinants of medical care utilization. Medical care is a kind of service that requires time to be consumed (Folland, 2000). Therefore, as people tend to value their time, it is reasonable to assume that time comes out as an important factor to which people are sensible when making decisions regarding the utilization of medical care. It should be noted that the importance of the time as a determinant of the demand for activities requiring time to be consumed was already suggested by Becker (1965). Subsequently, Acton (1976) studied this issue empirically applying to the health care field and arguing that, in a setting where the monetary price of health care is null or near null, a mechanism involving time is quite likely to assume the task of demand shifter since consumption of health care usually entails a payment in travel time and waiting time.

These non-monetary costs, or time related costs, may then be seen as a form of co-payment, and medical care demand/utilization has to be sensitive to them. This is an assumption for this paper.

As it is referred in several publications, the Portuguese NHS suffers from several inefficiencies (Bentes et al, 2004) with consequences in the importance of the non-monetary form of co-payment. For example, Cabral et al. (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. Our data led us to the conclusion that the average waiting time to visit a GP is approximately 9 days. Moreover, the average waiting time in the waiting room for a consultation is approximately 1 hour. These non-monetary or time related costs, are a source of

disutility to the patients (Cullis and Jones, 2000; Folland, 2000; Serrano-del-Rosal et al., 2004).

Even considering the large waiting time imposed by the NHS to patients to access to a public GP, nothing is known about how these non-monetary factors impact on the actual use of primary health care visits. Therefore, the knowledge about the reaction of the patients to the non-monetary related costs to see a GP is very important in every health system and, especially in the Portuguese NHS where the monetary effort that a patient makes to see a doctor is very low and almost null relatively to the non-monetary cost involved in the full utilization of health care.

The first major argument to focus on how the Portuguese patient's decisions about the utilization of GPs are affected by the non-monetary factors - time needed to have an appointment and to wait for the doctor inside the health centre - are the changes proposed by the current government to better the organization of the Portuguese primary health care system. In a recent public interview the Minister of Health has announced that 2004 would be devoted to the full reorganization of the primary health care system. In the same interview, he also stated that, among other things, this reorganization would aim at reducing the waiting time to get a medical appointment and to increase the number of patients seen per hour by a GP. According to the health minister, the figure should increase from 3.8 to 4 patients assisted in one hour, in average.

Meanwhile, little is known about the factors that influence the utilization of primary health care facilities in the Portuguese NHS. We lack knowledge about the response of the health centre's users to the non-monetary costs, especially the time costs, i.e., travel time, waiting time inside the health centre, and waiting time to get an appointment. This information would be crucial to assess the effects of policies aiming at improving the accessibility of primary health care. The success of such policies depends on the reaction of the patients to changes in the time-price.

Therefore, one major goal of this paper is to measure the elasticity of primary health care utilization relatively to the time-price, and the elasticity of waiting time to get an appointment with the physician.

Additionally, this study will allows us to better understand the importance of other determinants of health centre utilization, e.g., age, gender, education, health status as well as the impact of other socio-economics variables. This knowledge can be important to improve the financing system used to fund the primary health care at a regional level.

In the Portuguese primary health care organization each GP, although paid on a salary basis, is responsible for a given number of patients varying from 1000 to 2000 patients per GP. It is obvious that the size of the list, *per se*, is unimportant because its relevance varies upon the age and gender distribution as well as the burden of the disease; what is important to know is the expected amount of work that the doctor has to do with the patients in his/her list. Therefore, this paper can also be useful to build equivalent lists of patients in terms of amount of work. Moreover, in the area of policy making and health economics, this study may be important as a contributor for the assessment of the equity at a regional level. In conclusion, the results obtained in this study can be very important because they can reveal relevant information to improve the organization of the Portuguese Primary Health care System.

This study can still be important on the statistical econometric grounds. As far as we know, there are very little econometric models estimations on health taking into account the characteristics of an on-site sampling. Even in the international literature, is rare a description of models using truncated samples applied to health count data. Moreover, our econometric specification uses finite mixture models, a very recent econometric specification applied to data in the health field (Deb and Trivedi, 2002).

To reach our goal to analyse the main factors that influence the individual behaviour in the decisions to visit a GP, we need to have some form of measuring the utilization of medical care at a HC. In this study, our medical care utilization indicator is the number of visits that the individual made to the GP in the twelve months prior to the filling of the survey. This generates a variable that is non-negative and integer. The econometric modelling of variables with these characteristics requires the use of count data models. However, due to the sampling scheme (our first sample was obtained choosing patients randomly on-site), we had to use modified count data models. To carry on with this investigation we have estimated finite mixture models, applied to samples obtained on-site. Our methods heavily rely on the econometric specification presented by Santos-Silva (2003).

This paper contains 4 sections. In Section 2 we set up the methodological apparatus to respond to our research question. Section 3 describes the data source for the empirical results, presenting also some summary statistics. Still in this same section, we will briefly describe some aspects of the Portuguese Health System, and the mode of HC functioning. Empirical results are reported in section 4 with a discussion of the results.

2. Econometric specification and estimation

The behaviour of people in the health sector, more specifically, regarding the utilization of health care facilities, has been studied using continuous measures, such as individual expenditures, or discrete measures, such as the number of medical acts (e.g. visits to the physician) in a given period of time (Deb and Trivedi, 2002). In recent times the use of discrete measures has been more common than the use of continuous measures. Probably, the great use of discrete measures of health care utilization to analyse individual health care use is due to the greater availability and accuracy of this kind of discrete data relative to their continuous counterparts. We may refer Cameron et al. (1988), Barros (1999), Deb and Trivedi (1997, 2002), Gerdtham (1997), Gerdtham and Trivedi (2001), Hernandéz (1999), Martin et al. (2004), Yen et al. (2001), among many others, as authors who have used discrete measures to explain individual medical care utilization. All of these papers used standard count data modeling to explain the count dependent variable and estimated the econometric models using data obtained by sampling at random from the underlying population of interest.

The assumption of a random sample from the actual population is not always realistic (Wooldridge, 2002). In many projects that analyze economic data, the available data should be seen as a selected sample and not as a simple random sample. This is due to the sample design or to the behaviour of the individuals being sampled. So, it is crucial that sampling issues don't be ignored in empirical work. Modelling based on the assumption that we have a simple random sample can be inappropriate and misleading.

However, it is when the selection mechanism depends on the endogenous variable (in this case, the sampling scheme is referred to as endogenous sampling) that it is fundamental to account for the non-random nature of the sample, because under endogenous sampling methods the data do not represent appropriately the population of interest, and in order to make valid inference correct econometric specifications should be used (Santos-Silva, 2003; Wooldridge, 2002).

A truncated sample is a well know example of an endogenous sample. Notice that in a truncated sample the units included in the sample are chosen on the basis of the value of the dependent variable. Creel and Loomis (1990), Grogger and Carson (1991) and Gurmu and Trivedi (1992) made some theoretical and empirical research on the econometrics applied to truncated samples. They treated the case when the dependent variable is a count.

Both authors argued that one sampling scheme that generates a truncated sample is when the sampling units are selected on-site. Inclusion in the sample is only possible when the individual participates in the activity of interest (Gurmu and Trivedi, 1992).

Nevertheless, with such sampling method, on-site sampling, it is easy to see that the probability of inclusion in the sample depends on the frequency of utilization of the "site" (Englin and Shonkwiler, 1995 Santos-Silva, 1997; Winkelmann, 2000). Thus, as more regular users are more likely to be included in the sample, the empirical distribution will have more weight in the right tail, relative to the "true" distribution of the actual population. As a result, samples collect on-site, being also examples of endogenous samples, although, are of a different nature than the truncated samples. According to Shaw (1988), an on-site sample generates data with 3 characteristics that the researcher has to be aware of: Non-negative variables integers, truncation and endogenous stratification. This last characteristic makes on-site samples different from truncated samples. Hence, the probability models used to analyze truncated samples are different from the probability models that should be applied to data collected on-site.

In this paper we measure medical care utilization as the number of visits to the health centre in the year prior to the filling of the survey. However, due to the particular characteristics of our data resulting from the endogenous sampling scheme, we cannot straightforwardly apply count data models; the parameter estimates would be biased and inconsistent (Santos-Silva, 1997; Shaw, 1988; Wooldridge, 2002).

Using Santos-Silva's (2003) framework, if $f(y_i|x_i)$ is the density function of the i^{th} person in the underlying population, then the density of the same person in the sample (on-site or truncated), $f_S(yi|xi)$ is given by,

$$f_s(y_i \mid x_i) = \frac{f(y_i \mid x_i)}{h(y_i, x_i)} \tag{1}$$

where, in the case of sampling on-site,

$$h(y_i, x_i) = \frac{E(y_i \mid x_i)}{y_i} \tag{2}$$

and, in the case of truncated, at zero, sample,

$$h(y_i \mid x_i) = 1 - F(0 \mid x_i) \quad if \; y_i > 0 \tag{3}$$

where $F(0 \mid x_i)$ is the distribution function, conditional on x_i, corresponding to the probability function f, evaluated at value zero

Combining equations 1 and 2, we obtain the probability model that corrects the problems generated for on-site samples, and is given by

$$f_s(y_i \mid x_i) = \frac{y_i f(y_i \mid x_i)}{E(y_i \mid x_i)} \tag{4}$$

This equation 4 was firstly developed by Shaw (1988) and, subsequently, Santos-Silva (1997) arrived at the same expression using endogenous stratified samples ideas. Englin and Shonkwiler (1995 applied this model to an on-site sample to study the demand of recreational sites. In their paper, they assumed that the underlying population was distributed according to, on one hand, a Poisson distribution and on the other hand, a negative binomial distribution.

On the other hand, using equations 1 and 3, the probability model to correct for the problems created by truncated at zero samples is,

$$f_s(y_i \mid x_i) = \frac{f(y_i \mid x_i)}{1 - F(0 \mid x_i)} \qquad y_i > 0 \tag{5}$$

where expression (5) is the well known probability function of a truncated, at zero, random variable, as is presented in most count data econometrics textbooks (Cameron and Trivedi, 1998; Grogger and Carson, 1991; Gurmu and Trivedi, 1992; Winkelmann, 2000; Wolldridge, 2002).

In the regression analysis of non-negative integer valued dependent variables it is frequent to assume that the conditional probability function of the dependent variable is Poisson with exponential mean function $\lambda_i = \exp(x_i'b)$, where x_i is a column vector of regressors, and b is a vector containing the model parameters to be estimated. The exponential specification for the conditional mean ensures that λ_i is non-negative, as is required by the Poisson probability model.

In the general framework of count data modelling, i.e., count data models applied to random samples, a common criticism that is made to the use of the Poisson probability model, usually the starting point to analyse count data, is the conditional mean-variance equality feature that is imposed by the model, property that is known as equidispersion (Cameron and Trivedi, 1988; Jones, 2000). This conditional mean variance equality is often violated in real world as overdispersion, a situation where the conditional mean is lower than the conditional variance, is the

more common feature of empirical data (Jones, 2000). As Mullahy (1997) argues, what is causing overdispersion is the presence unobserved heterogeneity in the data.

This specification error, equidispersion, continues to be relevant in the case of endogenous sampling, both for on-site sampling (Englin and Shonkwiler, 1995; Santos-Silva, 1997), as well as for truncated samples (Grogger and Carson, 1991). To get consistent estimates with truncated and on-site data, the researcher has to correctly specify the first two moments of the distribution of the dependent variable (Grogger and Carson, 1991; Santos-Silva, 1997).

Thus, to estimate the parameters of the population of interest we have to explicitly take into account the unobserved heterogeneity. One way to proceed is to specify the probability model of the dependent variable as a mixture models.

The formal specification of a mixture model is easily done, but estimation of this kind of models is more complex because this estimation requires the distribution of the unobserved random variable which represents the unobserved heterogeneity, ε ,to be specified; justify one explicit density is difficult. Usually, the independence of the individual unobserved heterogeneity and the regressors are assumed (Santos-Silva, 2003). The unobserved heterogeneity ε can be assumed to be a continuous or discrete random variable. If the error term is assumed to be represented by a discrete random variable, we are using a semi-parametric model (Deb and Trivedi, 2002; Heckman and Singer, 1984). It consists in approximating the density of the error term using a discrete distribution with a finite number of support points. Deb and Trivedi (1997) pioneered the utilization of finite mixture models to explain the utilization of medical care, measured as a count. Afterwards, Martin et al. (2004) and Gerdtham, and Trivedi, (2001) applied this same type of models to health data. The application of finite mixture models to endogenous samples raises a number of important methodological issues that the researcher should be aware.

Let us now briefly explain how the unobserved heterogeneity is integrated in the assumed probability model of the dependent variable through the use of finite mixture models. Assume that the unobserved heterogeneity, ε , has a discrete probability distribution with m, support points of support, $\omega_1, \omega_2, ..., \omega_m$ and associated probability masses p_1, p_2, ..., p_m, with $\sum_{k=1}^{m} p_k = 1$ and $p_k \geq 0, k = 1,...m$. Further, assume that the distribution of the dependent variable, y_i, conditional on a

set of regressors and on ω_k is $f_k(y_i \mid x_i, \theta_k)$, where θ_k is a parameter vector to be estimated.

Then, the probability distribution of y_i, conditional on x_i, is

$$f(y_i \mid x_i, \theta) = \sum_{k=1}^{m} p_k f_k(y_i \mid x_i, \theta_k) \tag{6}$$

where $\theta = [\theta_1' \ \cdots \ \theta_k' \ \ p_1 \ \cdots \ p_{k-1}]$ represents a vector of parameters to be estimated. The mixing probabilities p_k, k =1...m are estimated together with all other model parameters. In a finite mixture model the random variable y_i can be viewed as being postulated as being drawn from a population that is an additive mixture of m distinct subpopulations (Deb and Trivedi, 2002).

Finite mixture models can have a very interesting interpretation if we relate the finite mixture model with the Latent Class Analysis (Wedel et al., 1993). In the latent class, each observation may be considered as a member of a specific group (a latent class). Under this framework, it is assumed the existence of m (unknown) latent classes or group of individuals. It is also hypothesized that each group has its own distribution. Group k has $f_k(y_i \mid x_i, \theta_k)$ as its representing probability function. From the construction, the a priori probability of an individual belongs to group k is p_k, that is, $P(y_i \in Group\, k) = p_k$.

To specify finite mixture models to endogenous samples, for example, truncated and on-site samples, care must be taken concerning the assumptions about the distribution of the unobserved heterogeneity. As Santos-Silva (2003) shows, model specification can be performed assuming the distribution of heterogeneity in the actual population or in the population induced by the sampling scheme, being the choice between the two alternatives not innocuous, as we will show now, relying heavily in the paper written by Santos-Silva, (2003).

As we above referred, in the case of truncated samples, the probability function of the sample is $f_s(y_i \mid x_i) = \dfrac{f(y_i \mid x_i)}{1 - F(0 \mid x_i)} \quad y_i > 0$. Assume now that the correct model for the dependent variable in the actual population, is $f(y_i \mid x_i, \theta) = \sum_{k=1}^{m} p_k f_k(y_i \mid x_i, \theta_k)$, i.e., a finite mixture model. We are now taking assumptions about the error term in the population of interest.

Under this assumption, the denominator in (5) is $1-F(0\mid x_i)=1-\sum_{k=1}^{m} p_k F_k(0\mid x_k,\theta_k)$, thus,

$$f_s(y_i\mid x_i)=\frac{\sum_{k=1}^{m} p_k f_k(y_i\mid x_i,\theta_k)}{1-\sum_{k=1}^{m} p_k F_k(0\mid \theta_k)}=\sum_{k=1}^{m}\frac{f_k(y_i\mid x_i,\theta_k)}{\left(1-\sum_{k=1}^{m} p_k F_k(0\mid \theta_k)\right)}p_k \tag{7}$$

$$=\sum_{k=1}^{m}\frac{f_k(y_i\mid x_i,\theta_k)*(1-F_k(0)\mid \theta_k)}{\left(1-\sum_{k=1}^{m} p_k F_k(0\mid \theta_k)\right)*(1-F_k(0\mid \theta_k))}p_k=\sum_{k=1}^{m} f_{s,k}(y_i\mid x_i,\theta_k)\hat{p}_k \tag{8}$$

$$f_{s,k}(y_i\mid x_i,\theta_k)=\frac{f_k(y_i\mid x_i,\theta_k)}{(1-F_k(0\mid x_i,\theta_k))} \tag{9}$$

and

$$\hat{p}_k=\frac{(1-F_k(0\mid\theta_k))}{\left(1-\sum_{k=1}^{m} p_k F_k(0\mid \theta_k)\right)}p_k \tag{10}$$

So, finite mixture models applied to truncated samples can be specified in two ways, namely (7) and (8). Equation (7) assumes that the unobserved heterogeneity is present in the population of interest, while equation (8) makes assumptions about the distribution of the unobserved heterogeneity in the population induced by the sampling scheme (Santos-Silva, 2003). This same author argues that the researcher should use (7) when the population of interest is the actual population. In contrast, specification (8) is the appropriate when the research intends to study the population induced by the sampling scheme.

This reasoning also applies to on-site samples, by replacing $1-F(0\mid x_i)$ with $\frac{E(y_i\mid x_i)}{y_i}$ in (5). Under the assumption that the probability model of the dependent variable in the actual population is a mixture model, $f(y_i\mid x_i,\theta)=\sum_{k=1}^{m} p_k f_k(y_i\mid x_i,\theta_k)$, thus $\frac{E(y_i\mid x_i)}{y_i}=\frac{1}{y_i}\sum_{k=1}^{m} p_k E(y_i\mid x_i,\theta_k)$, where $E(y_i\mid x_i,\theta_k)$ is the expected value of $f_k(y_i\mid x_i,\theta_k)$.

Given that the unobserved heterogeneity can be specified in the actual population, and the model (7) applies (or in the population induced by the sampling scheme, and the model (8) applies), we have to ask which finite mixture

specification is the most appropriate to analyse our data? The response depends on the population that we are interested in, and on the assumptions that we do about the behaviour of the individuals excluded from the sample due to the sampling mechanism.

If we believe that the probability model that describes the behaviour of the individuals excluded from the sample due to the sampling mechanism (individuals with zero visits) is different from the model that describes the behaviour that have a positive count, i.e., the utilization of the health centre is described by a Hurdle model (hurdle models are described in detail in Cameron and Trivedi, (1998), Gurmu (1997), Mullahy (1986), Pohlmeier and Ulrich (1995)) then, we can only use specification (8) and make inference to the population induced by the sampling mechanism. On the contrary, if we assume that the utilisation of the health centre of both types of individuals, zeroes and positives, is described by the same stochastic model, then the choice between (7) and (8) depends only on the objectives of the study.

In this study, we are interested in the actual population, so we will primarily use specification (7), that is, we assume that the unobserved heterogeneity is distributed in the actual population.

To estimate finite mixture models, we have to identify the number of component distributions of the mixture, that is, the value of m. However, the empirical estimation of m is an unanswered question; so, it is common to estimate models with a given m and then use LR tests, when possible, or information criteria (AIC and BIC) to select between alternative models (Deb and Trivedi, 2000).

The maximum likelihood estimation of a finite mixture model can be very thorny. The likelihood function of finite mixture models may have several local maxima hence there is always the chance to get convergence to one of these local extreme solutions which does not correspond to the maximum likelihood estimator (McLachlan and Peel, 2000). So, it is important to ensure that the algorithm converges to the global maximum, and this task is not easy to accomplish.

We estimated all models using the STATA ML command. For solving the unconstrained maximization problem, STATA uses the Broyden-Fletcher-Goldfarb-Shanno, (bfgs) algorithm. To "almost" guarantee that the algorithm converges to a global maximum, we did as Deb and Trivedi (2002) and estimated each model using different starting values. In most cases, we got the same estimates, regardless of the

starting values. Therefore, although we can not be a hundred per cent sure, there is a large probability that we have got the global maximum likelihood estimator.

Other widely recommended method for estimating this type of models is the Expected Maximization algorithm (EM algorithm). This procedure is described in McLachlan and Peel, (2000), Wedel et. al (1993). We didn't estimate our model using this procedure because it may be slow to converge (Deb and Trivedi, 1997).

All the models were estimated using (pseudo) maximum likelihood, that is to say, the standard errors of the coefficients were estimated using the robust variance formula (Deb and Trivedi, 1997). This is a sound procedure due to the possibility of a misspecification of the unknown density. This estimation was performed without imposing constraints on the parameters of the composing distributions.

Let is assume that our dependent variable y_i (number of visits to the health centre in the year prior to application of the questionnaire) has is distributed according to the finite mixture model $f(y_i \mid x_i, \theta) = \sum_{k=1}^{m} p_k f_k(y_i \mid x_i, \theta_k)$. Further, let us assume that the m component distributions of the mixture are Poisson, like

$$f_k(y_i \mid x_i, \theta_k) = \frac{\exp(-\lambda_{ik})\lambda_{ik}^{y_i}}{y_i!} \tag{11}$$

Therefore, if we want to make inference to the population of health centre users, and not only to the population of individuals that visited the health centre in September 2003, we should use specification 7, and assume that the unobserved heterogeneity is present in the population of interest.

In this case, for the truncated sample, the probability model used is

$$f_s(y_i \mid x_i, \theta) = \frac{y_i \sum_{k=1}^{m} p_k f_k(y_i \mid x_i, \theta_k)}{1 - \sum_{k=1}^{m} p_k F_k(0 \mid x_i, \theta_k)} \tag{12}$$

If f_k is Poisson distributed with mean parameter $\lambda_{ik} = \exp(\alpha_k + x_i' b_k)$ (x_i is a column vector of independent variables without a column of ones), then $F_k(0 \mid x_i, \theta_k) = \exp(-\lambda_{ik})$ and (12) becomes,

$$f(y_i \mid x_i, \theta_k) = \frac{\sum_{k=1}^{m} p_k \exp(-\lambda_{ik})\lambda_{ik}^{y_i}}{y_i!\left(1 - \sum_{k=1}^{m} p_k \exp(-\lambda_{ik})\right)} \tag{13}$$

For on-site samples, the probability model can be written as,

$$f_s(y_i \mid x_i, \theta) = \frac{y_i \sum_{k=1}^{m} p_k f_k(y_i \mid x_i, \theta_k)}{\sum_{k=1}^{m} p_k E(y_i \mid x_i, \theta_k)} \tag{14}$$

Assuming that $f(y_i \mid x_i, \theta_k)$ is distributed according to a Poisson model with location parameter $\lambda_k = \exp(\alpha_k + x_i' b_k)$, then $f_s(y_i \mid x_i)$ is given by

$$f_s(y_i \mid x_i) = \frac{1}{(y-1)!} \frac{\sum_{k=1}^{m} p_k e^{-\lambda_k} \lambda_k^{\,y}}{\sum_{k=1}^{m} p_k \lambda_k} \tag{15}$$

In this paper we estimated several finite mixture models based on (7) and (8) for both truncated data and data collected sampling individuals on-site. For instance, using (7), we estimated 4 models,

- Poisson truncated ()
- Finite mixture, with 2 component Poisson distributions with mean parameter $\lambda_{ik} = \exp(\alpha_k + x_i' b)$. That is, the two component distributions are different only because the constant term of the mean parameter
- Finite mixture, with 2 component Poisson distributions with mean parameter $\lambda_{ik} = \exp(\alpha_k + x_i' b_k)$
- Finite mixture, with 3 component Poisson distributions with mean parameter $\lambda_{ik} = \exp(\alpha_k + x_i' b)$

Some of these models are nested, so we are able to use Likelihood ratio tests to choose between them. For example, the simple truncated Poisson is nested, both, in the finite mixture, with 2 component distributions (Poisson), with mean parameter $\lambda_{ik} = \exp(\alpha_k + x_i' b)$ and in the finite mixture, with 2 component distributions (Poisson), with mean parameter $\lambda_{ik} = \exp(\alpha_k + x_i' b_k)$.

To choose between non-nested models we've used two traditional model selection criteria, specifically, the AIC - Akaike Information Criteria (-2log L + 2K) and BIC - Bayesian Information Criteria (-2log L+K*ln(N)), where ln L is the maximized log-likelihood, K the number of parameters

3. Institutional setting, variables and data

In Portugal, primary health care is mainly provided through public health centres covering, each of them an average of 28000 people and employing in total 30000 professionals, 25% GPs (Bentes et al., 2004). However, the system is a mix of both public and private health service providers.

The questionnaire used to obtain the satisfaction scores from the health centre's users was based on the Europep questionnaire, created by an international task force on patient evaluation of general practice care (Grol et al., 2000), financed by the European Union, initially involving researchers from 7 countries (Denmark, Germany, The Netherland, Norway, Portugal, Sweden and United Kingdom). Lately, other countries entered in 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.

It encompasses 23 outcome questions grouped into five major dimensions: (1) Patient-doctor interaction; (2) Medical care; (3) Information and support; (4) Continuity and cooperation; and (5) Organization of services. Other questions complete the underlying conceptual model of this questionnaire. They may be grouped into three different dimensions: (6) Consultation, accessibility and appointment; (7) Professionals; and (8) Health centre environment and services provided. The questionnaire also included socio-demographic questions, about patients' health status and the number of visits.

The Portuguese version of the Europep was validated in a nationwide sample (Ferreira, 1999) 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 centres (Ferreira et al, 2001). A sample of 3964 answers from users from 86 health centres was collected.

In this research we use two samples obtained during two consecutive years: 2001/2002 (S0102) and 2003/2004 (S0304). In the first round we obtained an on-site sample of 4714 answers from a universe of 194 health centres (54% of all health centres). The second round covered all health centres from the Portugal mainland and a population of almost 68000 users.

The questionnaires of the first round were directly administered by the GP, having some of then not fully followed the premises given. In the second round the authors followed a new strategy. After obtaining the list of all users during the first

six months of 2003 a random sample, proportionally to the distribution of age and gender within each health centre, was built. The questionnaires were directly sent by mail to user's residencies and we have received a total sample of 12000 answers. In this research we used the first 2563 questionnaires returned.

In every round the questionnaire were filled by patients and sent back in a prepaid envelope. The anonymity was, in this way, granted. All the answers were optically read through a scanner by using Teleform software.

In Table 1 we present the list of the dependent variables. Unfortunately the survey does not provide any information about the income of the respondents. However, we can use education as a proxy to income.

Insert Table 1 about here

The covariates used in our study are those usually used in other studies on the determinants of health care utilization.

Table 2 presents the list of the dependent variables, as well as their definitions.

4. Results

In this section we present the results obtained in both on-site and truncated samples. For each one, we begin by the descriptive data of the sample and we analyze the results obtained from econometric models.

On site sample

Although the entire sample was formed by 4714 health centres users, our final workable database had only 3181 usable records. The large majority of the deleted records corresponded to individuals that missed to report the necessary information about important variables, generating a missing value. A couple of records were, in addition, deleted because the patient report strange values on some variables. For instance, some individuals reported more than 8 hours waiting in the health centre

The empirical distribution of dependent variable, (visits), the number of visits in the last year, is shown in Figure 1. The sample mean for the dependent variable is 5.52 visits.

Insert Table 2 about here

Analysing this empirical distribution we were concern with a data measurement problem that can, in some way, contaminate our estimation results, and therefore our conclusions. Recall that the period over which we register the visits to the health centre is 12 months, a long period, that can induced response errors, mainly due to the lack of memory. This fact can be observed in the previous graphic. Above 4 visits, we see some kind of digit preference in the response to the question about the number of consultations. Note the percentage of people that answered 10, 11 and 12. So, care must be taken in the analysis of the results.

Table 2 presents the descriptive statistics of the independent variables (on-site sampling).

Insert table 3 about here

Despite the fact that we have estimated more than a few alternative finite mixture specifications that correct for the on-site sampling potential bias, we will present the results of the model that assumes that the discrete unobserved heterogeneity is represented by a discrete probability function with 3 points of support. The unobserved heterogeneity is assumed to be distributed in the actual population, according to specification (7). We choose this specification on the basis of LR tests and information criteria, as it was referred at the end of section 2.

Table 3 reports the estimates, robust standard errors, p-value and statistical significance levels for the parameters of the 3 points of support mixture model estimated, where the component distributions are assumed to be Poisson. The interpretation of the figures shown in the table is as follows: a negative coefficient means that as the dependent variable increases the mean utilization decreases, and vice-versa.

Insert table 4 about here

Next we will describe the impact of each variable on the mean utilization of the primary health care centres.

Socio-demographic variables

Male individuals tend to use the health care provided by health centres less than the female (-0.01***). Individuals with higher education have a tendency to go less often to the health centre; contrarily lower educated people are more willing to receive health care in the health centre. This result maybe a consequence of higher income, i.e., people with more purchase power can more frequently use the private substitutes, in this case, the private health sector. Or are more willing to have special access to hospital (emergency and ambulatory) care. In this specification, we assumed that age has a quadratic effect. The coefficient of the quadratic term has a near zero value, and the other term has a negative sign.

Health status variables

The health status variables, both objective (chronic disease) and self-assessed, have the expected impact on the dependent variable, even taken into account other characteristics as gender and age. People who perceive their health status as excellent, go less regularly to the health centre. The opposite happens with people rating their health status as bad. Regarding the chronic disease for all of them, people with a chronic disease go more often to the consultation.

Price-Time cost

For this group of variables, we have considered the time (days) waiting for a consultation as well as the time (minutes) waiting in the waiting room of the health centre. Both failed to be statistically significant. This means that, at least for the population of health centre's users, they do not respond to price-time. This may be explained because Portuguese patients are, in general, protected by welfare state, meaning that few of them do not feel an immediate impact (e.g., in day-off salary) of being absent from work so many hours.

How to reach the heath centre

All the 4 variables (on foot, car, public transportation and others) that we hypothesized to have an impact in the mean utilization of the health centre failed to be statistically significant. These variables may also be irrelevant for patients. Even when they have to pay by themselves to go to a health centre (e.g., there is no public transportation available and they have to go by taxi), their need and the lack

of citizenship experience force them to do whatever has to be done to maximize their expectations, i.e., to be seen by a doctor.

Characteristics of the health centre

Under this group we included location of the health centre (in a costal region, or not) and the perception of adequacy of the time spent the doctor during the interview. To begin this analysis, patients living in a costal region are less willing to use the health centres. These regions are those, in Portugal, where people have a better access to other health care providers, namely, emergency care and private providers, thus more alternatives to the health centres.

The enough time variable shows that patients who perceive not having spent enough time with their GP are less willing to go to the health centre. Dissatisfaction may be a reason for this behaviour. This result is interesting and can be used to forecast the effect of an intended reduction in duration of the visits. For a substantial reduction on the time each patient spend with his/her GP, it is likely that the number of users unsatisfied with the visit's duration will increase. Therefore, the demand of public GP would be reduced. So, we foresee that the effect of an increase in the number of visits per hour will decrease the demand for public health care visits.

Truncated sample

The empirical distribution of dependent variable for the truncated sample, (VISITS), is presented in table 5. The sample mean for the dependent variable is 5.6 visits.

Insert table 5 about here

The total sample was formed by 2957 health centres users, but, as happened in the case of the other sample, our final workable database has only 1618 utilizable records. The deleted records corresponded to individuals that missed to report the necessary information about important variables, consequently generating a missing value.

It is interesting to analyse this empirical distribution comparatively with the distribution presented in Table 2.

As was argued in Section 2, in the on-site samples, the high counts would be overrepresented. This can be evidenced in Figure 1.

Insert Figure 1 about here

Replicating what has been done for the on-site sample, we have estimated several alternative finite mixture specifications that correct for truncation. We will present the results of the model that assumes that the discrete unobserved heterogeneity is represented by a discrete probability function with 2 points of support. The unobserved heterogeneity is assumed to be distributed in the actual population, according to specification (7). We choose this specification on the basis of LR tests. We also chose this specification in order to interpret the model in terms of latent classes.

Table 6 reports the estimates, robust standard errors, p-value and statistical significance levels for the parameters of the 2 points of support mixture model estimated, where the component distributions are assumed to be Poisson, and with location parameter described by a different set of parameters. The interpretation of the figures shown in the table is as follows: a negative coefficient means that as the dependent variable increases the mean utilization decreases, and vice-versa.

Insert table 6 about here

Assuming the representativeness of our sample, the mixture model suggested that the population be split into two sub-populations. The first one contains about 84% of individuals and corresponds to the sub-sample of users who visit less often the health centre. In average, individuals from this (latent) sub population make 3.98 visits a year. The remaining 16% of individuals are more frequent users with an average of visits of about 13.49 per year. The population estimated mean of visits after the split was 5,59, being the original empirical mean equal 5.62.

The following analysis will be split into the two latent classes previously obtained.

Less frequent users

For individuals in this class, the average number of visits to the health centre is dependent upon gender, education, place of living, self-assessed health, time to reach the health centre, difficult to book an appointment and satisfaction. In fact, male, high educated and healthier individuals, living in the coast and spending less time to reach the health centre tend to visit it less often. These individuals also have, in general their consultation booked and are less satisfied with the time they spend with their GP as well as with the location of the health centre.

More frequent users

For individuals from this class, the average number of visits to the health centre is dependent upon their occupation, sickness, health status and satisfaction. In fact, individuals with lower social strata, sicker and with bad self assessed health are more willing to visit their GP in the health centre. These individuals are more satisfies with the location of the health centre.

It is worth to emphasize that the only variables that are statistically significant across both latent classes are the logarithm of the total time to reach the health centre and the satisfaction with its location. The coefficient of the first variable, measuring the elasticity of average utilization relatively to total time to reach the health centre, is greater than zero in both latent classes. This means that, as total time to reach the health centre increases, the average utilization also increases, which seems to contradict the theory that, in systems where the financial cost to consume the service is low, the time becomes more relevant as a cost .

This apparent contradiction may be explained because the places associated to a more difficult access to health centres (e.g., places in rural areas) are also locations where there are few alternatives. Therefore their inhabitants are "forced" to go more often to the health centres.

Regarding the offer of care, measured by the number of inhabitants per doctor and per GP, table 6 also shows significant values for the more frequent users. However the magnitude of the coefficients is economically insignificant.

After estimating our model, we were able to compute the posterior probability that individual i belongs to the class k. With these figures we could assign each individual to a latent class. We followed McLachlan and Peel (2000). Table 7 presents the descritives of the variables for each latent class. As we can see from it, the

latent classes have almost identical characteristics. What distinguish them the most is the number of chronic disease. In fact, users from latent class 1 (less frequent users) have less chronic diseases then individuals from latent class 2 (more frequent users). At the same time, the percentage of individual who have lower self assessed health is greater in this latter class.

Insert table 7 about here

5. Bibliographic references

Acton, J.P., 1976 Non-Monetary factors in the demand for medical services: Some empirical evidence. The Journal of Political Economy 83, 595-614.

Barros, P.P., 1999. Os sistemas privados de saúde e a reforma do sistema de saúde, in: Associação Nacional de Sistemas de Saúde (Eds), O papel dos sistemas privados de saúde num sistema em mudança, pp. 90-115.

Becker, G.,1965. A theory of the allocation of time, Economics Journal 75, 493-517.

Bentes M., Dias C.M., Sakellarides C., Bankauskaite, V., 2004. Health care systems in transition: Portugal. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies, 2004.

Cabral, M.V., Silva, P.A., Mendes, H., 2002. Saúde e Doença em Portugal, Imprensa de Ciências Sociais, Lisboa.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Econometric Society Monograph 30, Cambridge University Press, New York.

Cameron, A.C., Trivedi, P.K., Milne, F., Piggot, J., 1988. A Microeconometric model of the demand for health care and health insurance in Australia. Review of Economic Studies 55, 85-106

Creel, M.D., Loomis, J.B., 1990. Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California, American Journal Agricultural Economics 72, 434-441.

Cullis, J., Jones, P., 2000. Waiting lists and medical treatment: analisys and policies,in: A. J. Culyer and J.P. Newhouse (Eds), Handbook of Health Economics, volume 1, Elsevier Science, pp. 1201-1249.

Deb, P., Trivedi, P.K., 1997. Demand for medical care by the elderly in the United States: a finite mixture approach. Journal of Applied Econometrics 12, 313-336.

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

Englin, J., Shonkwiler, J.S., 1995. Estimating social welfare using count data models: an application to long-run recreation demand under conditions of endogenous stratification and truncation. The review of Economics and Statistics 77, 104-12.

Ferreira, P.L., 1999. A voz dos doentes: satisfação com a medicina geral e familiar, in: Direcção-Geral da Saúde (Eds), Instrumentos para a melhoria contínua da qualidade.

Ferreira, P.L, Luz A, Valente S, Raposo V, Godinho P, Felício ED., 2001. Determinantes da satisfação dos utentes dos cuidados primários: o caso de Lisboa e Vale do Tejo. Revista Portuguesa de Saúde Pública, Volume Temático 2: 53-61.

Folland, S., Goodman, A., Stano, M., 1997. Economics of Health and Health Care, Prentice-Hall.

Gerdtham, U.D.,1997. Equity in health care utilization: further evidence based on hurdle models and Swedish micro data. Health Economics 6; 303-319.

Gerdtham, U.D., Trivedi, P., 2001. Equity in Swedish Health Care Reconsidered: New results Based on the Finite Mixture Model. Health economic Letters.

Grogger, J.T., Carson, R.T., 1991. Models for truncated counts. Journal of Applied Econometrics 6, 225-238.

Grol, R., Wensing, M., Mainz, J., Jung, H.P., Ferreira, P., Hearnshaw, H., Hjortdahl, P., Olesen, F., Reis, S., Ribacke, M., et al., 2000. Patients in Europe evaluate general practice care: an international comparision. British Journal of General Practice 50:882–7.

Gurmu, S., 1997. Semi-Parametric estimation of hurdle regression models with an application to medicaid utilization, Journal of Applied Econometrics 12, 225-242.

Gurmu S., Trivedi P., 1992. Overdispersion tests for truncated Poisson models. Journal of Econometrics 54: 347-370.

Heckman, J.J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data, Econometrica 52, 271-320.

Hernández, A.M., 1999. Duplicate coverage and demand for health care. The case of catalonia, Health Economics 8, 579-598.

Jones, A., 2000. Health Econometrics, in: A.J. Culyer, J.P. Newhouse (Eds), Handbook of Health Economics, vol 1, Elsevier Science, pp. 264-343.

Martin, S.J., Labeaga, J.M., Granado, M.M., 2004. An empirical analyses of the demand for physician services across the European Union, European Journal of Health Economics 5: 150-165.

Mansky, S., 1995. *Identification problems in the social sciences*, Harvard University Press, London.

McLachlan, G., Peel, D., 2000. Finite Mixture Models, New York: John Wiley & Sons.

Mullahy, J., 1986. Specification and testing of some modified count data models, Journal of Econometrics 33, 341-365.

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

Pohlmeier, W., Ulrich, V., 1995. An econometric Model of the two-part decision process in the demand for health, Journal of Human resources 30, 339-361.

Serrano-del-Rosal, R., Vera-Toscano, V., Ateca-Amestoy, V., 2004. The disutility of waiting time: evidence from the Public Primary Health Care Service in Andalucía. Working Paper IESA 06-04, Instituto de Estudios Sociales de Andalucía, Córdoba.

Santos-Silva, J.M.C., 1997. Unobservable in count data models for on-site samples, Economics Letters 54, 217-220.

Santos-Silva, J.M.C., 2003. A note on the estimation of mixture models under endogenous sampling, Econometrics Journal 6, 46-52.

Shaw, D., 1988. On-site's samples regression. Problems of non-negative integers, truncation and endogenous stratification. Journal of Econometrics 37, 211-33.

Wedel, M., Desarbo, W., Bult, J., Ramaswami, V., 1993. A latent class Poisson regression model for heterogeneous count data, Journal of Applied Econometrics 8, 397-411.

Winkelmann, R., 2000. Econometric Analysis of Count Data.Berlin: Springer.

Wooldridge, J., 2002. Econometric analysis of cross section and panel data. Cambridge, Mass.: MIT Press.

Yen, S.T., Tang, C.H., Su, S.J., 2001. Demand for traditional medicine in Taiwan: A mixed Gaussian-Poisson model approach, Health Economics 10, 221-232.

Table 1 - Definitions of dependent variable and explanatory variables

Variable Name	Used in regression
Visits	Number of visits to the GP in the previous year
Gender - Two dummy variables. Female is the excluded category	male
Age - in tenths of years, and its square	Age and sqage
Education - Three dummy variables representing the maximum level of formal education of the individual: Educ1 (low education), Educ2 (medium education) and Educ3 (high education). High education is the excluded category	Educ1, Educ2
Occupation - Three dummy variables representing the profession of the individual: Prof1 (high level profession), Prof2 (medium level profession), Prof3 (low level profession)	Prof1, Prof2
Marital Status - Two dummy variables representing the marital status of the individual: Live alone (widow or divorced), married.	Live_alone, married
Living region - One dummy variable: individual lives in a costal region	coastal
Health status - Five dummy variables representing long run and short run health status. Long run health: Bad health (Self assesses Health - SAH- is bad), good health (SAH is good) and Excellent health (SAH is excellent). Excellent health is the excluded category. Number of chronicle diseases (max = 10. Short run health: Sick (individual feels sick in the days prior to the GP visit and, urgent (sudden disease)	Bad_h, good_h, sick and urgent, n_chronic
Time cost - Total Time spent in the travel to the health centre and total time spent in the health centre, both in minutes. We took the logarithms of both variables	Log_travel_time, Log_hc_time
Access - Three dummy variables representing the mean used to travel to the health centre in the last visit. On foot (individual went to the HC on foot), Car (car used), Pub_transp (public transportation used). Dific_access (had difficulties in accessing the health centre, booked_visit (last visit booked in advance).	On_foot, car, pub_transp, dific_access, booked visit
Satisfaction - Two dummy variable: Not enough time (individual felt that the GP doesn't give him enough time) and HC_location (individual satisfied with the location of the health centre).	Not enough time, HC_location
Health supply (macro Level) - Two variables representing the number of doctors at regional level: number of inhabitants_per_doctor (all physicians included) and number of inhabitants per GP	Inhabitants_per_doctor, Inhabitants_per_GP
Interaction variables - Inactive is a dummy variable representing the occupational (active or not) status of the individual. The re	Inactive* Log_hc_time, age* Log_hc_time

Table 2 - Distribution of the variable VISITS (on-site sample)

visits	Freq.	Percent
1	214	6.73
2	470	14.78
3	491	15.44
4	471	14.81
5	321	10.09
6	378	11.88
7	126	3.96
8	165	5.19
9	41	1.29
10	194	6.1
11	11	0.35
12	165	5.19
13	4	0.13
>=14	130	4.08

Table 3 - Descriptive statistics for the on-site sample

Variable	Mean	Std. Dev.	Min	Max
male	0.33	0.47	0	1
age	4.82	1.73	1.8	9.9
sqage	26.21	17.32	3.24	98.01
Live alone	0.12	0.32	0	1
Married	0.73	0.44	0	1
Educ1	0.40	0.49	0	1
Educ2	0.44	0.50	0	1
Coastal	0.35	0.48	0	1
Sick	0.35	0.48	0	1
urgent	0.13	0.34	0	1
bad_h	0.11	0.31	0	1
good_h	0.79	0.41	0	1
num_chronic	1.81	1.69	0	10
Log_hc_time	4.41	0.77	0.69	6.17
On_foot	0.31	0.46	0	1
car	0.49	0.50	0	1
Pub_transp	0.12	0.32	0	1
Not enough time	0.10	0.30	0	1
booked visit	0.90	0.30	0	1
age* Log_hc_time	21.22	8.42	2.63	54.15

Table 4 - Parameter estimates of the finite mixture model (Model 2) with 3 Poisson component distributions with means with equal slopes and different intercepts. Robust standard errors.

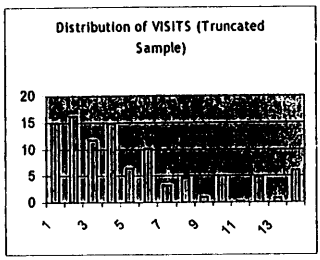
numcons	Coef.	Robust Std. Err.	z	P>z
male	-0.099	0.066	-1.50	0.134
age	-0.227	0.120	-1.89	0.058
sqage	0.020	0.006	3.23	0.001
Live alone	0.185	0.083	2.22	0.026
Married	0.209	0.066	3.16	0.002
Educ1	0.335	0.067	5.02	0.000
Educ2	0.194	0.066	2.95	0.003
Coastal	-0.102	0.070	-1.46	0.144
Sick	-0.058	0.089	-0.65	0.514
urgent	-0.211	0.042	-4.96	0.000
bad_h	0.539	0.089	6.08	0.000
good_h	0.229	0.071	3.21	0.001
num_chronic	0.088	0.024	3.70	0.000
Log_hc_time	-0.097	0.079	-1.23	0.219
On_foot	0.025	0.367	0.07	0.946
car	-0.018	0.368	-0.05	0.960
Pub_transp	0.073	0.406	0.18	0.857
Not enough time	-0.169	0.085	-2.00	0.045
booked visit	-0.022	0.059	-0.37	0.711
age* Log_hc_time	0.010	0.016	0.64	0.525
const1_				
_cons	2.912	0.967	3.01	0.003
const2_				
_cons	2.062	0.979	2.11	0.035
const3_				
_cons	1.111	0.886	1.25	0.210
p1_				
_cons	0.006	0.007	0.91	0.361
p2_				
_cons	0.141	0.071	1.97	0.048

Table 5 - Distribution of the variable VISITS (Truncated sample)

Visits	Freq.	Percent
1	242	14.96
2	261	16.13
3	189	11.68
4	238	14.71
5	104	6.43
6	165	10.2
7	51	3.15
8	76	4.7
9	16	0.99
10	83	5.13
11	4	0.25
12	77	4.76
13	12	0.74
>=14	100	6.16

)

Figure 1 - Comparison of the distribution of visits from both samples



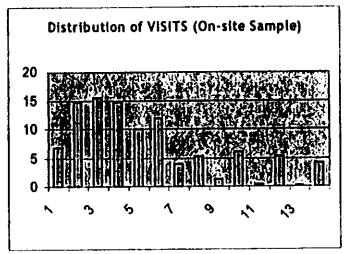


Figure 6 - Truncated model for both latent classes

	Class1 - less frequent P_1=0.84				Class 2 - more frequent P_1=0.16			
Variables	Coef.	Stdv.	P	sig	Coef.	Stdv.	P	sig
Male	-0,317	0,050	-6,38	0,000	-0,120	0,088	-1,36	0,174
Age	0,072	0,115	0,63	0,528	0,056	0,194	0,29	0,774
Educ1	-0,003	0,008	-0,42	0,673	0,005	0,016	0,30	0,761
Educ2	0,327	0,100	3,27	0,001	0,074	0,234	0,32	0,751
Prof1	0,199	0,090	2,22	0,026	-0,180	0,126	-1,43	0,153
Prof2	-0,055	0,076	-0,73	0,465	-0,245	0,111	-2,22	0,027
Live alone	-0,067	0,075	-0,89	0,371	-0,179	0,102	-1,75	0,080
Coastal regions	0,198	0,105	1,90	0,058	0,297	0,217	1,37	0,170
Sick	-0,180	0,058	-3,10	0,002	-0,002	0,093	-0,02	0,982
urgent	0,081	0,052	1,56	0,119	-0,255	0,092	-2,78	0,005
Bad_hs	0,012	0,088	0,14	0,887	-0,318	0,124	-2,57	0,010
good_hs	0,383	0,111	3,44	0,001	0,386	0,163	2,37	0,018
n_chronic s	0,014	0,095	0,15	0,884	0,069	0,135	0,51	0,610
Log_travel_time	0,063	0,018	3,53	0,000	0,019	0,039	0,49	0,626
Log_hc_time	0,105	0,041	2,58	0,010	0,186	0,072	2,57	0,010
on foot	-0,043	0,110	-0,39	0,693	0,086	0,176	0,49	0,624
Car	-0,004	0,097	-0,04	0,971	-0,323	0,213	-1,51	0,130
Pub_transp	0,003	0,098	0,03	0,978	-0,453	0,214	-2,12	0,034
Dific_access	0,032	0,111	0,28	0,776	-0,087	0,241	-0,36	0,718
booked visit	0,129	0,067	1,93	0,054	0,044	0,260	0,17	0,867
Satisf w/ GP time	-0,241	0,095	-2,54	0,011	-0,213	0,258	-0,83	0,409
Satisf w/ loc hc	-0,151	0,068	-2,22	0,027	0,151	0,104	1,45	0,146
Inhabitants per doctor	-0,411	0,073	-5,65	0,000	2,677	0,227	11,77	0,000
Inhabitants per GP	1,526	0,536	2,85	0,004	0,278	0,785	0,35	0,723

Table 7 - Descriptive of both latent classes of the truncated sample

	Class1 - less frequent P_1=0.84				Class 2 - more frequent P_1=0.16			
Variables	**Mean**	**Stdv.**	**Min**	**Max**	**Mean**	**Stdv.**	**Min**	**Max**
Male	0,33	0,47	0	1	0,32	0,474	0	1
Age	5,53	1,85	0,9	9,6	5,5	1,90	1,2	9,3
Educ1	0,60	0,49	0	1	0,58	0,49	0	1
Educ2	0,29	0,45	0	1	0,30	0,46	0	1
Prof1	0,14	0,34	0	1	0,15	0,35	0	1
Prof2	0,15	0,36	0	1	0,15	0,36	0	1
Live alone	0,17	0,37	0	1	0,21	0,41	0	1
Coastal regions	0,71	0,45	0	1	0,67	0,47	0	1
Sick	0,40	0,49	0	1	0,43	0,491	0	1
urgent	0,11	0,31	0	1	0,13	0,33	0	1
Bad_hs	0,17	0,38	0	1	0,21	0,41	0	1
good_hs	0,75	0,43	0	1	0,74	0,44	0	1
n_chronic s	1,971	1,48	0	8	2,35	1,53	0	7
Log_travel_time	2,62	0,76	0	5,30	2,65	0,76	0,69	5,01
Log_hc_time	4,63	0,73	1,79	6,17	4,63	0,75	2,30	6,17
on foot	0,38	0,48	0	1	0,36	0,48	0	1
Car	0,43	0,49	0	1	0,48	0,50	0	1
Pub_transp	0,11	0,32	0	1	0,09	0,28	0	1
Dific_access	0,12	0,33	0	1	0,09	0,29	0	1
booked visit	0,91	0,28	0	1	0,90	0,30	0	1
Satisf w/ GP time	0,22	0,41	0	1	0,22	0,41	0	1
Satisf w/ loc hc	0,89	0,31	0	1	0,89	0,31	0	1
Inhabitants per doctor	581	400	106	1727	561	365	106	1727
Inhabitants per GP	1383	206	1019	1973	1387	206	1019	1973

Series Organisers: Andrew Jones (University of York) and Owen O'Donnell (University of Macedonia)
Series Secretary: Liz Hudson (University of York)
Local Organisers: Vincenzo Atella and Franco Peracchi (Tor Vergata University, Rome)
Local Secretary: Maria Laura Cantarelli (Tor Vergata University, Rome)

Funding

We are grateful for funding for the workshop series from:

- University of York

and for additional local funding from:

- Abbott Italy