ESTIMATING UTILISATION OF HEALTH CARE: A grouped data regression approach

Autora: *Mabel Amaya Amaya* [(a)]

P.T. N.º 5/01

(a) Corresponding author:
Mabel Amaya
Health Economics Research Unit
University of Aberdeen
Polwarth Building
Forresterhill
Aberdeen AB25 2ZD
Phone: +44(0) 1224 681818
Fax: +44(0) 1224 662994
E-mail: mma@heru.abdn.ac.uk

ÍNDEX

ABSTRACT

In an era in which the share of the public budget taken by health care rises, it becomes increasingly important to determine the basic forces influencing the demand for medical services and the decision process underlying it. In this paper, redefining health care consumption in terms of utilisation, we explore how alternatives estimation models perform for a cross section of data from the British Household Panel Survey. Grouped data regression is shown to be the most appropriate modelling approach due to the nature of the dependent variable. We find the familiar result the frequency of physician visit is clearly responsive to both socio-economic characteristics and need, proxied by morbidity.

Keywords: demand of health care, medical care utilisation, grouped data, ordered probit.

JEL classification: I18, C21, C25

1. INTRODUCTION AND OVERVIEW

As the share of the public budget taken by health care rises, it becomes increasingly important to determine those factors influencing the demand for medical services and the decision process underlying it. However, it is well established in the health economics literature that the usual microeconomic approach, i.e. the neoclassical framework, needs to be altered to cope with the peculiarities of the health care sector. Therefore, in empirical studies, health care consumption has usually been redefined in terms of utilisation, allowing proper acknowledgement of the influence of the supplier, who becomes instrumental in specifying the consumption pattern of the consumer, as well as of the role of health insurance. This approach will be adopted in this paper.

Using a cross section of data from the British Household Panel Survey, the purpose is to search, by means of econometric modelling, the main determinants of health care utilisation in Britain. We explore how alternative estimation models perform in our dataset, concluding that, rather than an ordered probit model, a grouped data regression approach turns out to be empirically more adequate. Although using a different modelling approach, our results are very much in line with those found in previous empirical work in health care demand.

The rest of the paper is structured as follows. In section 2 we review previous empirical literature in the field. Section 3 is devoted to specifying the statistical model we use. In section 4 we briefly describe the dataset and the measurement of variables employed in the empirical analysis. Section 5 presents estimation results for the alternative models tried. We end in Section 6 with some conclusions together with suggestions for further research.

2. STATISTICAL MODEL AND ESTIMATION METHODS

The review of the theoretical analyses of medical care utilisation suggests two traditions: consumer theory approach (Grossman, 1972) and principal-agent set-up (Zweifel, 1981). The first approach assumes the patient is the sole actor determining the demand for medical services while the general principle in the tradition of Zweifel (1981) is that decisions are based on a mixture of own and expert information, through which suppliers can manipulate demand. Consequently, the econometric modelling approach in each case has to be different. This paper can be said in the tradition of Grossman.

The statistical model for health care utilisation should take account of a special feature of the BHPS data – medical use is recorded as a categorical variable with 5 response categories which represents the number of consultations 1 = 0, 2 = 1 to 2, 3 = 3 to 6, 4 = 6 to 10, 5 = >10-. An appropriate specification for similarly recoded responses where the categories represent choices is the multinomial probit or logit model. However, the application of such models here would fail to account for the ordinal nature of the dependent variable. According to Greene (2000), ordinary regression analysis would err in the opposite direction, however. That is, linear regression would treat the difference between two adjacent response codes as the same, irrespective of where they lie in the distribution of responses, whereas in the fact they are only a ranking.

The most common way to deal with ordered response data is to use an ordered qualitative response model, usually either the ordered probit model or the ordered logit model. The key feature of ordered qualitative response models is that all the choices depend on a single index function

$$y_i = \boldsymbol{b}' x_i + u_i \quad (i = 1,2,\ldots,n) \tag{2.1}$$

where y_i is the underlying response variable (number of visits to GP), x_i is a set of explanatory variables (including socio-economic and health status measures) and u_i is the residual error.

The variable of theoretical interest y_i^* (use of medical services) is unobservable but we know which of the M categories it belongs to. We observe

$$\begin{aligned} y_i &= 0 \text{ if } y_i^* \le m_1 \\ &= 1 \text{ if } m_1 < y_i^* \le m_2 \\ &\vdots \\ &= M \text{ if } m_{M-1} \le y_i^*, \end{aligned} \tag{2.2}$$

which is a form of censoring.

The probability that an observation falls in one of the categories is given by

$$\begin{aligned} &\Pr ob(y_i = M) = F(\boldsymbol{b}' x_i) \\ &\Pr ob(y_i = M-1) = F(\boldsymbol{b}' x_i + m_1) - F(\boldsymbol{b}' x_i) \\ &\Pr ob(y_i = M-2) = F(\boldsymbol{b}' x_i + m_1 + m_2) - F(\boldsymbol{b}' x_i + m_1) \end{aligned} \tag{2.3}$$

and so forth.

$m_0, m_1, \ldots m_M$ denote $M+1$ real numbers (thresholds), with $m_0 = -\infty, m_M = +\infty$ and $m_1, m_2, \ldots m_{M-1} > 0$, to be estimated with β.

For the distribution function F, we can use the logistic or the cumulative normal defining the model as ordered logit or probit respectively. Differences in practice should be negligible both distributions are very close to each other except for the tails.

For the latter case and imposing the identifying restrictions $m_1 = 0, s = 1$, the probability of observing a particular value of y is

$$\Pr(y_{ik} = k - 1) = \Phi[m_k - x_i b] - \Phi[m_{k-1} - x_i b] \tag{2.4}$$

where Φ is the cumulative standard normal.

The likelihood function for the model is

$$L = \prod_{i=1}^{N} \prod_{k=1}^{M} \left[\Phi\left(m_k - b' x_i\right) - \Phi\left(m_{k-1} - b' x_i\right)\right]^{Y_{ik}} \tag{2.5}$$

and, with independent observations, the log-likelihood takes the form

$$L^* = \log L = \sum_{i=1}^{N} \sum_{k=1}^{M} Y_{ik} \log\left[\Phi\left(m_k - b' x_i\right) - \Phi\left(m_{k-1} - b' x_i\right)\right] \tag{2.6}$$

For the maximization of the log.likelihood, McKelvey and Zavoina (1975) reported good results for the convergence of the Newton-Raphson method.

Marginal effects of changes in the covariates on the cell probabilities $\frac{\partial P(y = j)}{\partial x}$ are not the coefficients but these multiplied by the change in the probability distribution

$$\left[f(m_{j-1} - xb) - f(m_j - xb)\right] b \,. \tag{2.7}$$

However, in the general case, from the knowledge of the coefficients, one can only infer the marginal effects for the first and the last categories, Prob (y = 0) and Prob (y = J). What happens in the middle depends on the two densities. According to Greene (2000), one must be very careful in interpreting the coefficients of this model. Indeed, without a fair amount of extra calculations, it is quite unclear how the coefficients in the ordered probit should be interpreted[1].

Furthermore, when, as it is the case, the ranges of the underlying variable to which each category refers are known, a variant of the ordered probit, the so-called grouped data regression, appears more adequate. Theoretically, this approach is more efficient than an alternative ordered probit approach since

[1] This point seems uniformly to be overlooked in the literature. Authors routinely report coefficients and *t* ratios, occasionally with some comments about significant effects, but rarely suggest upon what or in what direction they are exerted.

the estimation procedure (maximum likelihood) utilises information on the scale of Y^* (provided by the thresholds values) to produce an estimate of σ, rather than requiring that this be normalised to one (Horowitz, 1994). With data from the British General Household Survey for 1978-1990, Sutton and Godfrey (1995) use this alternative to the ordered probit to estimate a model in which socio-economic characteristics, along with health-related attitudes and behaviour predict levels of drinking.

In this paper we examine how these two alternative models, ordered probit and grouped data regression perform on our data. For completeness, we also estimate ordinary least squares and ordered logit models.

We test linear and logarithmic functional forms. Following Sutton and Godfrey (1995), for the ordered response models, logarithmic forms are generated by using logged values of the threshold terms and taking logs of the continuous regressors.

In comparing the models we use a RESET test[2] as a general check for the validity of the estimated coefficients. Although there are alternative tests for misspecification of limited dependent variable models[3], leading to inconsistency of maximum likelihood estimators, the evidence provided by Horowitz (1994) and others suggest the RESET is a fairly reliable and convenient general check. Further, Godfrey et al.(1988) has indicated that using only the squared term provides an effective check. Thus we use this approach at the 1% level to discriminate between models.

For those models not passing the RESET test, thus apparently misspecified, we further test for heteroskedasticity in the errors. In the case of ordinary least-squares models we employ the general test suggested by Cook and Weisberg (1983). For the ordered qualitative response models we use a RESET-type test assuming the heteroskedasticity is a function of an individual's age.

3. DATA AND VARIABLE SPECIFICATION

3.1. The dataset

The data source is the first wave of the British Household Panel Survey (BHPS) conducted between 1st September and 30th April 1991 by the Institute

[2] Ramsey (1969)

[3] See, e.g. Maddala (1995), Pagan and Vella (1989) and Glewwe (1997). See also Vuong (1989) for a more general discussion on likelihood ratio tests for model selection.

for Social and Economic Research at the University of Essex "… *to further our understanding of social and economic change at the individual and household level in Britain*" (User manual)

The BHPS was designed as an annual survey of each adult (+ 16) member of a nationally representative sample of around 5,500 households, making a total 13,840 individual interviews in the first wave. Although measures have been taken in an attempt to maintain the sample broadly representative of the population of Britain, only using the first wave (1991) data can guarantee the sample is truly nationally representative.

For the ease of estimation, we use a subset of the 8167 Original Sample Members who completed the questionnaire giving valid responses for the variables we use in our estimation. After excluding individuals with missing values on variables of interest we obtain a working sample of 7881 individuals, consisting in 4080 males and 3801 females.

3.2. Measurement of Variables

The dependent variable is the frequency of physician visits. Individuals in the survey were asked to approximately state the number of times they had talked to or visited a General Practitioner or family doctor about their own health during the year before the interview. The question specifically asks respondents to exclude any visits to a hospital. Responses are coded one to five depending on the interval the number of visits falls into: 1 if none; 2 if one or two; 3 if visits were between three and five; 4 if between six and ten and 5 if more than ten. The nature of the variable dictates the modelling approach to empirical estimation.

As it is common use, the explanatory variables are categorised according to whether they reflect socio-economic or health status (need) factors. Following Windmeijer and Santos Silva (1997), no price variables are directly included in the model, since every individual in the UK, even those who have taken out a private insurance, is covered by the National Health Service (NHS), which is paid by payroll National Insurance (unavoidable) contributions. Further, information on private insurance taken is not available in the dataset[4].

- *Socio-economic factors*

Following Pohlmeier and Ulrich (1994), I include monthly household income deflated to per person level (EHHINC). This standardisation is derived by

[4] Nevertheless, it seems likely the non-availability of this information leads to only minor misspecification, given that private insurance offer more choice at specialist level but less so at the GP (see Besley et al. 1996)

using the household equivalence scale before housing costs contained in the survey. Rather than capturing income effect, this variable is likely to reflect opportunity costs (for example, in terms of foregone time costs).

I further use three dummies for current labour force status: self-employed (JOBST1); in-paid employed (JOBST2) or other, such as on maternity leave or long term-sick or disabled (JOBST4). Reference category is the unemployed. A dummy accounting for employment status (JOBFT=1 if currently working full-time, zero otherwise) is also included. According to Phelps et al (1974) and Cauley (1987) the argument is that individuals who are employed may incur a larger "time price" of going to the physician than the unemployed. Moreover the two dummy variables separating out the employed control for potential differences in "time price" for individuals who are self-employed and those in-paid employment. Following Gerdtham (1997) I include a 0-1 dummy for individuals working full-time (JOBFT=1 if individual works fulltime, zero otherwise). This variable is defined for both self and in-paid employed people and, again, it will also reflect different time costs for people employed full-time and those who are not so.

I also include covariates on education specified as 0-1 dummy variables for individuals with higher qualification (HDEGREE)[5] and no qualification (NOQUAL). The reference category is individuals with qualification other than higher[6]. The *a priori* expectation of education, which eventually correlates with medical knowledge, is ambiguous[7].

As with Colle and Grossman (1978), we introduce three 0-1 dummy variables for ethnic group the respondent belongs to: BLACK =1 if the individual is black, either Caribbean, African or other; INDIAN if he/she is indian, pakistani or bangladeshi and OTHER if the individual belongs to some other ethnic group. Reference category is white/european individuals. These covariates control for differences in health care utilisation between white/European and memberships of a different ethnic group that are not due to differences in the other independent variables.

According to Feldstein (1979), single persons generally use more health care. To test this proposition, I also include three dummy variables for different marital statuses: child under 16 years old (CHILD<16=1); married or living as a couple

---

[5] Including individuals with a first or higher degree and those with a higher national certificate/diploma or teaching qualification.

[6] This variable summarise, amongst others, the following categories in the questionnaire: nursing qualification, GCE A levels, GCE O levels or equivalent, clerical o commercial qualifications; CSE 2-5 or Scottish Standard Grade ; recognised trade apprenticeship and any other qualifications.

[7] See Grossman (1972); Muurinen (1982) and Wagstaff (1986).

(MARRIED = 1) and divorced, separated or widowed (DIVSEPWI = 1). Reference category is single people. For Feldstein's claim to be data consistent, the coefficient of the last two covariates should be found significantly positive whereas children under sixteen should not behave different from single people.

Finally, following Deb and Trivedi (1997), in order to control for behavioural differences across geographical regions, country of origin is included as two 0-1 dummies for individuals coming from countries other than England (SCOT and WALES). Reference category is English people.

- *Health status measures*

As a proxy for long-term status, I use a self-assessed health question which asks the individual to rate their health on average over the last twelve months relative to someone of their own age. This variable is coded as excellent, good, fair, poor and very poor. Following Contoyannis and Rice (2000) I created three dummy variables (SAHEX, SAHGOOD and SAHPOOR) equal to one if an individual has excellent, good or fair or worse health. The reference category is individuals reporting good health. Under fairly weak assumptions, Grossman (1972) and Phelps (1973) show that the quantity of medical care demanded will rise as health falls hence it is expected that the coefficient on the excellent variable will be negative and the one on the fair or worse variable positive.

I also use a composite measure (GHQ1) on the Likert scale derived from the results of the reduced version of the Goldberg's General Health Questionnaire (GHQ) consisting of twelve questions concerning the subjective general well being of the respondent. For the same reason as stated above, we expect the coefficient on this variable to be positive.

Further I include a third measure of subjective well-being, a measure of the self-perceived energy compared with people of the same age. This is entered as a dummy variable (ZESTLESS) and equals one if the individual finds himself less energetic than people of his same age and zero if he/she states to be more energetic or about the same.

Gender and age are included in the model, which, according to Gerdtham (1997 p.306) "… *may capture imperfect measures of morbidity for individuals of different ages or sex (e.g. obstetric charges")*. Gender is represented by a 0-1 dummy for males, thus females are the reference category. As it is common use, I use a quadratic parameterisation of age on utilisation.

I also include number of children in the household by age using three dummies for age groups: 3-4 years (CHILD4); 5 to 11 years (CHILD511) and 12 to 18 years (CHLD1218). The reference category is pre-school children (up to and including two years). According to Colle and Grossman (1995), an increase in the number of children in a family lowers the quantity of care demanded. They also claim visits to physicians fall with age. Therefore we expect a negative coefficient for all the three dummies.

Following Pohlmeier and Ulrich (1994), to capture the incidence of an illness in the previous year we include a dummy variable (HOSPYES) taking value one whenever an individual was hospitalised during that period.

Further, we use two dummy variables to indicate whether the individual is a current smoker or not. Reference category is non-smokers. Since smoking has been proved to seriously damage health, a positive sign is expected for the coefficient of this variable.

Our last set of covariates, following Windmeijer and Santos Silva (1997), relate to short term health. First we create two dummy variables referring to whether health limits in any way daily activities compared to most people of the same age of the respondent (LIMDAY=1 if health limits in some way daily activities) or not (NOLIMDAY=1 if health does not limit daily activities at all). Following Cameron et al (1988) and Gerdtham (1997), for further account for need (=morbidity), another health status variable measuring chronic conditions of the respondent is included. Individuals were asked whether they had health problems or not at all. Respondents reporting health problems were asked a further question identifying the type of problem. I create two 0-1 dummies; the first (HLPROB) for individuals reporting health problems (any of the twelve types included in the questionnaire) and the other for those reporting no health problems at all (HLPRB0). Further, I also create two dummy variables for whether health limits type or amount of work (LIMWORK) or not (LIMNOT). We include LIMDAY, HLPROB and LIMWORK thus we standardise the dummies on “non-chronic conditions”. Obviously it is expected a positive sign for the coefficient of all these covariates since it is likely people suffering from limiting conditions visit the physician more often.

Sample statistics for all the variables are shown in Table 1

5. ANALYSIS OF RESULTS

Table 2 presents the alternative estimation results for number of physician visits.

Table 1: Sample descriptive statistics

Variable	No. of observations	Mean	Std. Deviation	Minimum	Maximum
NVISIT	7864	2.2913	1.1271	1	5
SOCIO-ECONOMIC CHARACTERISTICS					
MALE	7881	0.5177	0.4997	0	1
AGE	7881	44.5752	17.9335	16	97
AGE2	7881	2308.52	1756.25	256	9409
CHILD<16	7881	0.0013	0.0356	0	1
MARRIED	7881	0.6705	0.4701	0	1
DIVSEPWI	7881	0.1349	0.3416	0	1
HDEGREE	7881	0.0836	0.2768	0	1
NO QUAL	7881	0.4012	0.4902	0	1
JOBST1	7881	0.0902	0.2865	0	1
JOBST2	7881	0.6169	0.4862	0	1
JOBST4	7881	0.2303	0.4210	0	1
JOBFT	7881	0.1608	0.3673	0	1
SCOT	7881	0.0940	0.2919	0	1
WALES	7881	0.0485	0.2148	0	1
EHHINC	7881	16403.07	10669.29	109.0909	119060.3
BLACK	7881	0.0114	0.1063	0	1
INDIAN	7880	0.0122	0.1097	0	1
OTHRACE	7880	0.0076	0.0869	0	1
CHILD4	7881	0.1062	0.3081	0	1
CHILD511	7881	0.1622	0.3686	0	1
CHILD1218	7881	0.1630	0.3694	0	1
HEALTH STATUS MEASURES					
SAHEX	7875	0.3062	0.4609	0	1
SAHPOOR	7881	0.2304	0.4211	0	1
ZESTLESS	7831	0.1190	0.3238	0	1
LIMWORK	7875	0.1374	0.3443	0	1
HOSPYES	7879	0.0943	0.2923	0	1
SMOKES	7881	0.2924	0.4549	0	1
GHQ1	7881	10.4716	4.6218	0	36
LIMDAY	7877	0.1051	0.3067	0	1

Table 2: Alternative estimates for the frequency of physician visits

Variable	(1) OLS (linear)	(2) OLS (log)	(3) Ordered probit (linear)	(4) Ordered probit (log)	(5) Ordered logit (linear)	(6) Ordered logit (log)	(7) Grouped data (linear)	(8) Grouped data (log)
Constant	2.408991 (25.505)	.7894975 (8.441)	-	-	-	-	2.768473 (10.536)	.7462005 (3.683)
SOCIO-ECONOMIC CHARACTERISTICS								
MALE	-.294997 (-12.764)	-.1362459 (-12.941)	-.3514345 (-12.882)	-.3514788 (-12.884)	-.6070664 (-12.888)	-.6073172 (12.893)	-.7039396 (-10.970)	-.2936229 (-12.899)
AGE	-.022004 (-5.390)	-.0111264 (-5.971)	-.0277927 (-5.813)	-.0274438 (-5.728)	-.0465427 (-5.608)	-.0459151 (-5.519)	-.0448337 (-3.943)	-.0223229 (-5.565)
AGE2	.0002119 (4.835)	.0001079 (5.390)	.0002672 (5.221)	.000263 (5.126)	.0004511 (5.057)	.004436 (4.958)	.0004241 (3.469)	.0002137 (4.973)
CHILD<16	-.2414606 (-.820)	-.0941811 (-.702)	-.251915 (-.719)	-.2502454 (-.714)	-.4222486 (.466)	-.4195594 (.469)	-.6378122 (-.790)	-.2259205 (-.736)
MARRIED	.1258832 (3.647)	.0621452 (3.952)	.1578735 (3.873)	.1587247 (3.893)	.266288 (3.800)	.2678393 (3.822)	.2739488 (2.864)	.12951 (3.789)
DIVSEPWID	.0786565 (1.747)	.0341654 (1.665)	.0915479 (1.727)	.0917905 (1.731)	.1474123 (1.616)	.1477465 (1.619)	.218071 (1.743)	.074191 (1.669)
HDEGREE	.0050288 (.125)	.0149004 (.819)	.0189818 (.398)	.0251591 (.533)	.0372372 (.465)	.0477707 (.603)	-.0541693 (-.486)	.0133165 (0.337)
NO QUAL	-.0168698 (-.647)	-.0154691 (-1.296))	-.0272161 (-.884)	-.0300452 (-.972)	-.0645932 (-1.219)	-.0695632 (-1.307)	.0062062 (.086)	-.0213265 (-.823)
JOBST1	-.1589645 (-2.807)	-.0694631 (-2.692)	-.1875999 (-2.793)	-.1835442 (-2.732)	-.3316965 (-2.872)	-.3243998 (-2.808)	-.3847035 (-2.450)	-.1553739 (-2.758)
JOBST2	-.1050839 (-2.259)	-.0394692 (-1.840)	-.1113106 (-2.030)	-.1069847 (-1.927)	-.1955941 (-2.070)	-.1877663 (-1.963)	-.290698 (-2.249)	-.0955409 (-2.052)
JOBST4	-.0147308 (-.253)	-.0115089 (-.434)	-.0214595 (-.314)	-.0209056 (-.306)	-.0670182 (-0.567)	-.065671 (-.554)	-.0031915 (-.020)	-.0151281 (-.264)
JBFT	.0030085 (.095)	.009333 (.650)	.0188896 (.511)	.0172591 (.467)	.0191166 (.303)	.0162477 (.258)	-.0560744 (-.642)	.0095667 (.309)
SCOT	.0997868 (2.753)	.0331693 (2.008)	.1049127 (2.462)	.1041046 (2.442)	.1826927 (2.478)	.181405 (2.460)	.3284532 (3.259)	.0940952 (2.634)
WALES	.0189805 (.389)	-.0031213 (-.140)	.0083177 (.144)	.0068871 (.119)	.0098488 (.099)	.0071473 (0.072)	.1281261 (.944)	.0120938 (.250)
EHHINC	7.73e-07 (.675)	.0028246 (.302)	1.00e-06 (.741)	.0027519 (.114)	1.73e-06 (.748)	0.00413 (.100)	-7.17e-07 (.226)	.0016757 (-.083)
BLACK	-.0210316 (.210)	-.0011038 (-.024)	.0163982 (.139)	.0150591 (.127)	.0370969 (.180)	.03406 (.165)	.1243126 (.446)	.0207472 (.209)
INDIAN	-.0174131 (-.181)	-.0002287 (-0.005)	-.0000168 (0.000)	-.0018759 (-0.017)	.0243296 (.128)	.21147 (.111)	-.0975584 (–.367)	-.0035162 (-0.037)
OTHER	-.0494931 (.411)	.0328647 (.599)	.070117 (.496)	.0690117 (.488)	.1272379 (.524)	.1251947 (.516)	.0276267 (.083)	.0456688 (.386)

Table 2 (cont.): Alternative estimates for the frequency of physician visits

Variable	(1) OLS (linear)	(2) OLS (log)	(3) Ordered probit (linear)	(4) Ordered probit (log)	(5) Ordered logit (linear)	(6) Ordered logit (log)	(7) Grouped data (linear)	(8) Grouped data (log)
CHILD4	.1744852 (4.586)	.063195 (3.640)	.186884 (4.17)	.1832097 (4.082)	.3230775 (4.187)	.316278 (4.097)	.5367724 (5.087)	.1597675 (.000)
CHILD511	-.0635489 (-2.042)	-.0275552 (-1.940)	-.0712614 (-1.936)	-.0748878 (-2.031)	-.1297881 (-2.063)	-.1361288 (-2.161)	-.1743645 (-2.025)	-.0631592 (-2.044)
CHLD1218	-.0531996 (-1.743)	.-.0218384 (-1.569)	-.0614261 (-1.698)	-.0637491 (-1.761)	-.0986832 (-1.604)	-1.027372 (-1.668)	-.1473185 (-1.746)	-.0561182 (-1.851)
HEALTH STATUS MEASURES								
SAHEX	-.2649206 (-10.566)	-.1391688 (-12.189)	-.3565324 (-11.849)	-.3556328 (-11.824)	-.6090675 (-11.862)	-.607421 (-11.835)	-.4818523 (-6.976)	-.2884545 (-11.493)
SAHPOOR	.5712384 (18.884)	.2213925 (16.032)	.5803748 (16.479)	.5799447 (16.466)	1.022601 (16.527)	1.021854 (16.513)	1.679858 (19.821)	.5077197 (17.376)
ZESTLESS	.1690748 (4.618)	.0565195 (3.389)	.1671732 (3.930)	.1674344 (3.936)	.2939771 (3.979)	.2944324 (3.985)	.599409 (5.837)	.1504905 (4.222)
LIMWORK	.2852889 (6.955)	.0985402 (5.273)	.280047 (5.929)	.2791225 (5.909)	.5247351 (6.318)	.5229375 (6.296)	.947644 (8.217)	.2493456 (6.302)
HOSPYES	.6681339 (18.107)	.2565399 (15.264)	.6999815 (16.236)	.701206 (16.266)	1.194027 (15.735)	1.195652 (15.756)	2.052297 (19.647)	.6086637 (16.975)
SMOKES	-.085815 (-3.581)	-.0433006 (-3.963)	-.1085445 (-3.834)	-.1095103 (-3.865)	-.197121 (-4.051)	-.1988928 (-4.084)	-.1825549 (-2.746)	-.0904452 (-3.809)
GHQ1	.0133577 (5.464)	.0054029 (4.852)	.0150085 (5.232)	.014986 (5.224)	. 0259994 (5.240)	.0259632 (5.233)	.0390564 (5.720)	.129123 (5.372)
LIMDAY	.1227005 (2.632)	.0302566 (.154)	.0991948 (1.848)	.0999613 (1.863)	.1775984 (1.887)	.1787774 (1.899)	-.5437793 (0.000)	.099006 (2.199)
HLPROB	.3603756 (15.082)	-.1788895 (16.434)	.4516759 (15.920)	.4517189 (15.921)	.7691919 (15.751)	.7695113 (15.755)	.7468865 (11.314)	.3696505 (15.675)
Sigma. 95% Conf. Interval	-	-	-	-	-	-	2.4714 2.5582	.8352 .8709
No. observat.	7777	7777	7777	7777	7777	7777	7777	7777
No. iterations	-	-	4	4	5	5	4	5
-Log-L	-	-	9591.2585	9591.5268	9605.6153	9605.8898	15978.074	11437.046
R-squared	(adj) 0.3265	(adj) 0.42108	(pseudo) 0.1284	(pseudo) 0.1284	(pseudo) 0.1271	(pseudo) 0.1271	N/A	N/A
Overall sig. Prob > c^2	F(30,7746 126.67 = . 0000	F(30,7746 105.89 = . 0000	c^2(30) = 2827.05 = .0000	c^2(30) = 2826.51 = .0000	c^2(30) = 2798.34 = .0000	c^2(30) = 2797.79 = .0000	c^2(30) = 3083.11 = .0000	c^2(30) = 2949.25 = .0000
RESET2 Prob > c^2(1)	t (7745) = 1.364 = .173	t (7745)= -2.023 = .043	c^2(1) = 1.49573 = 0. 221	c^2(1) = 1.42086 = 0.233	c^2(1) = .58217 = 0. 446	c^2(1) = .53436 = 0.465	c^2(1) = 48.678 = 0. 0000	c^2(1) = 0.0471 =.829
Hetero Prob > c^2(1)	-	c^2(1) = 0.12 = 0.7269	-	-	-	-	c^2(1)= 5.01 = 0. 0252	-

Notes:

1. The baseline individual in all models is male; single; has qualification but not higher degree; with children under 2 years old, is unemployed, belongs to white/European ethnic group; lives in England, reports good health, feels about the same or more energetic that people of him same age, is non-smoker, declares to have no health problems and health does limit either daily activities or type/amount of work. Also he was not hospital inpatient during the year before the interview.

2. Coefficients reported for all models are the coefficients on x in the index function $x\beta$. t- ratios are in parentheses. Critical values for t-student distribution are 1.65; 1.96 and 2.576 at the significance levels 10%, 5% and 1% respectively.

3. For the ordered probit and logit models, the cut-off points are estimated as

	Ord. Probit (lin)	Ord. Probit(log)	Ord. Logit (lin)	Ord. Logit (log)
μ_1 =	-1.0376	-1.0201	-1.7695	-1.7454
μ_2 =	.2408	0.2576	0.3854	0.4093
μ_3 =	1.0956	1.1123	1.8735	1.8973
μ_4 =	1.6994	1.7162	2.9806	3.0044

Interestingly we find that most of the specifications cannot be rejected as appropriate models of the determinants of health care utilisation according to the RESET test at the 1% level. Only for the grouped data regression in the linear form is the RESET statistic significant. However, there is no evidence of heteroskedasticity in the errors, based on the test used.

In general the estimated coefficients do exhibit the expected signs except for those corresponding to the covariates on education. However these effects appear not to be significant.

Since the grouped data regression approach is theoretically more appealing and the estimated effects quite similar[8], the following discussion concentrates on the results of the estimated coefficients for this model in the logarithmic form. Moreover, as we get a negative estimate for the first of the ancillary parameters (m_1 in Table 4) , following Maddala (1983), we can assume there is some specification error in both ordered probit and logit which is not picked up by the general check used.

As stated above, most of the estimated coefficients have the expected sign. The influences of personal characteristics are found to be significantly different from zero. We find familiar results: men are less likely to use medical care than women and there is a strong convex relationship between the number of consultations and age[9].

---

[8] [8] We should bear in mind that the estimates of the parameters are not directly comparable. Since the logistic distribution has a variance $\sigma^2/3$, the estimated coefficients obtained from the logit model have to be multiplied by $\sqrt{3}/\sigma$ to be comparable to the estimates obtained from the probit model. Amemiya (1981) suggests that the logit estimates be multiplied by 1/1.6 = 0.625 produces a closer approximation between the logistic and the distribution function of the standard normal.

[9] Cameron et al. (1986) also find an inverted bell shaped age pattern for doctor consultations using an instrumental variable estimator.

We do find enough evidence favouring Feldstein's proposition (1979). The effect of not being single is estimated positive and it is clearly significant. Thus single people appear to visit the GP less often than married or divorced/widowed whereas there are not significant behavioural differences between the former and children under 16 years old.

The estimated relationship between age of children in the household and number of GP visits is not monotonic. With respect to the reference category (children under two years), probability of visiting a physician is first increasing for children up to four years old; then it decreases for dependent children between five and eleven years. For children over twelve years old we also find an increased probability to visit compared with children less than two years but the effect is only significant at the 10% significance level.

Membership of other than white/european ethnic group appears not to affect number of GP consultations. For all of the three dummies included the coefficients are not significant at any traditional significance level. Thus, there is enough evidence against a potential selectivity process with non-white people more likely to visit "public care" sites (hospital, emergency rooms, hospital outpatient departments and public clinics not connected with hospitals) or, alternatively, against a lower latent health status for races other than white.

Wagstaff's (1986, p.216) argument that people with a higher education can improve their health more efficiently and therefore would contact a GP less often cannot be ruled out in these data. Although individuals having a higher qualification are estimated to be significantly more likely to visit the GP than those with other type of qualification, this effect is clearly not significant. Moreover, despite non-qualified people appearing to be less likely to visit the GP, the effect is clearly not significant.

Health care utilisation appears quite unresponsive to changes in the level of income. A small positive and not significant coefficient is found for this covariate, thus, conditional on the other variables in the model, the hypothesised negative effect of income (meaning greater opportunity costs the higher the level of income) are not borne out in these data.

The dummies for self-employment status (JOBST1) and in-paid employed (JOBST2) have a negative significant effect, which is consistent with *a priori* expectations. Both people are estimated to be less likely to visit a doctor than the unemployed, with the effect being greater for those in self-employment. However we do not find significant differences between the unemployed and people on maternal leave, retired or other job situations. Interestingly we find a positive but not significant effect of full-time employment compared to part-time. Therefore plausible greater stress related illnesses for individuals in full-time employment leading to higher frequency of GP visits is ruled out in our sample.

The estimated effect of the dummies for non-English people is positive and significant for Scottish people and negative but not significant for Welsh. The different behaviour for Scottish people, estimated to be more likely to visit a GP than English, may be explained by looking at Public Heath statistics. Scotland shows higher than the average mortality rates from many diseases such as lung cancer or coronary heart disease hence there are good reasons to support this finding here. Although the determinants of these regional differences in health behaviour are beyond the scope of this thesis, it appears to be an interesting subject to explore in future research.

Not surprisingly, the frequency of physician visits is clearly responsive to need, proxied by morbidity. The estimated effects pertaining to the dummies for self-assessed health (SAHEX = 1 and SAHPOOR = 1) are significant at the 1% level. Individuals reporting excellent health are less likely to visit the doctor than those claiming that their health is good whereas reporting own health is just fair or worse has a positive effect in the probability of visiting a GP. We also find a significant positive effect for those individuals reporting their energy is less than people of their own age. Also, as expected, the higher the score in the General Health Questionnaire the higher the probability of a GP visit, although the effect shown is not very large.

It is not surprising either that people who were seriously ill in the previous year, thus hospitalised, require more treatment from general practitioner. This may be indicative of low latent, unobserved health or, alternatively, may be due to post-hospitalization on-going care provided through their general practitioner (such as follow-up for prescriptions, monitoring of health status, … etc).

Unexpectedly we find a significant negative effect of being a smoker on the number of GP consultations. A plausible explanation of this finding is that smokers know that this habit seriously damage their health and that they will be told to stop by physician. Thus they tend to avoid visiting the doctor for minor illness spells.

Consistent with previous studies individuals whose health limits type or amount of work are found more likely to visit the doctor than those without work limiting conditions. Also the probability of a GP visit by people reporting chronic diseases is estimated to be significantly higher than the probability of doing so by people declaring to have no health problems at all. Moreover, as expected, the effect of limitations in daily activities due to health is found to be significant and positive.

Although using a different modelling approach, in general, our results for the full sample are in line with those found for the UK by Windmeijer and Santos Silva (1997), for Australia by Cameron et al. (1986) and for Germany by Pohlmeier and Ulrich (1994).

6. CONCLUSIONS

The basic question we have tried to solve in this paper is: *What basic forces influence the utilisation of health care services?* Finding the answer may be crucial to improve the health system in general and particularly for resource allocation, to make it more equitable and adjusted to need. Moreover, answer can help to understand causes underlying the spectacular increase of health care public expenditure over the last thirty years.

Therefore, in this paper, as with other empirical studies, we redefine health care consumption in terms of utilisation, allowing proper acknowledgement of the influence of the supplier as well as of the role of health insurance. We make use of General Practitioner services dependable on socio-economic factors, such as age, income and education, along with proxied health status measures

Using a cross section of data from the British Household Panel Survey, we explored how alternatives estimation models perform. Grouped data regression is shown to be the most appropriate modelling approach due to the limited nature of the dependent variable.

The estimated regression coefficients using the latter functional form are compatible with previous studies in health care utilisation. We find familiar results. Personal characteristics, such as gender, age or region of origin, significantly influence the use of medical care and, not surprisingly, the frequency of physician visit is clearly responsive to need, proxied by morbidity.

However, some difficulties of interpretation remain. We have not controlled for the potential endogeneity of included explanatory variables, for example, variables measuring morbidity. Hence, our estimates may be contaminated by biases as found in previous cross-sectional analyses.

Further, the nature of the available data has dictated the modelling approach used. Possible future research might employ panel data instead of a cross-section which would allow analysing how individuals and household experience change in their socio-economic environment and how they respond to such changes and how conditions, life events, behaviour and values are linked with each other dynamically over time.

Moreover, more sophisticated econometric models[10] could be used such as hurdle or finite mixture models although this requires count data, i.e. data on number for consultations, instead of intervals for the dependent variable.

[10] See Jones (1998) for a review

Ackowledgements

The data used in this thesis were made available through The Data Archive. The data were originally collected by the ESRC Research Centre on Micro-Social Change at the University of Essex (UK). Financial support from the Spanish Ministry of Education is gratefully ackowledged. The author wish to specially thank Nigel Rice from the Centre for Health Economics (CHE), University of York (UK), for valuable comments which helped to improve earlier drafts of this paper. Helpful comments were also obtained from Javier Salinas Jiménez, Francisco Pedraja Chaparro and Julián Ramajo Hernández from the Economics Department at the University of Extremadura (Spain). I would also like to acknowledge Peter C. Smith, John Wildman, Andrew Jones and Roberto León González from the University of York (UK) for assistance at various stages of the project. I alone bear all responsibility for errors of omission and co-mission in the analyses or interpretations presented here.

REFERENCES

Amemiya, T. (1981), "Qualitative Response Model: A Survey", *Journal of Economic Literature*, p.1488.

Besley, T, J. Hall and I. Preston (1996), "Private health insurance and the state of the NHS", The Institute for Fiscal Studies, Commentary No. 52.

BHPS, User's manual, Vol. A

Cameron, A.C. and F.A.G. Windmeijer (1996), "R^2 measures for count data regression models with application to health care utilisation", *Journal of Business and Economic Statistics*, 14:pp. 209-220.

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(1), pp. 29-53

Cauley, S.D., (1997) "The time price of medical care", *The Review of Economics and Statistics*, 13: pp.56-66.

Colle, A.D and M. Grossman, (1978), "Determinants of pediatric care utilisation", *The Journal of Human Resources* , XIII, Supplement, pp. 115-151.

Contoyannis, P and N. Rice (2000), "The impact of health on wages: Evidence from thee British Household Panel Survey". Unpublished

Cook, R.D. and S. Weisberg, (1983), "Diagnostics for heteroskedasticity in regression", *Biometrika* 70: .1-10.

Cullis, J. and P. West (1979), "The economics of health: an introduction", Martin Robertson. Oxford.

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

Feldstein, P.J. (1979), "Health Economics", New York, Wiley.

Gerdtham, ULF.G. (1997), "Equity in health care utilisation: further tests based on hurdle models and Swedish Micro Data", *Health Economics*, vol.6: 303-319.

Glewwe, P. (1997), "A test of normality assumption in the ordered probit model", *Econometrics reviews*, 16: pp.1-19.

Godfrey, L.G., McAleer, M. and MacKenzie, C.R., (1988) "Variable addition and Lagrange multiplier tests for linear and logarithmic regression models". *Review of Economics and Statistics*, 70: 492-503.

Greene, W.H, (2000),“Econometric Analysis”, Forth edition, Prentice Hall International.

Grossman, M. (1972),“The demand for health: a theoretical and empirical investigation*”, National Bureau of Economic Research*. New York.

Horowitz, J.L. (1994) “Bootstrap-based critical values for the information matrix test*”, Journal of Econometrics*, 61 (2), pp. 395-411.

Jones, A (1998) “Health econometrics”, University of York. Preliminary draft prepared for Newhouse, J.P and A.J. Culyer, “Handbook of Health Economics”, North Holland

Maddala, G.S.

(1983), “Limited -Dependent and QualitativeVariables in Econometrics”, Econometric Society Monographs. Cambridge University Press.

(1995), “ Specification test in limited dependent variables models”, pp. 1-49. Blackwell. Oxford. In G.S. Maddala, P.C.B. Phillips and T.N. Srinivasan (eds.), “Advances in econometrics and quantitative economics: essays in honour of Professor C.R.Rao”

McKelvey, R.D. and W. Zavoina (1975), “An statistical model for the analysis ordinal level dependent variables”, *Journal of Mathematical Sociology*, 4, pp.103-120.

Muurinen, J.M. (1982), “The demand for health: a generalised Grossman model”, *Journal of Health Economics*. Vol.1, N1, pp.5-28

Pagan, A. And F. Vella (1989), “Diagnostic tests for models based on individual data: a survey*”, Journal of Applied Econometrics*, 4: S29-S59.

Phelps, C.E. and J.P. Newhouse (1974), “Coinsurance, the price of time and the demand for medical services” *Review of Economics and Statistics*, 56, pp. 334-347.

Pohlmeier, W. And V.Ulrich (1994), “An econometric model of the two-part decision-making process in the demand for health care”, *Journal of Human Resources*, XXX, 2, pp.339-359.

Pratt, J.W (1981), “Concavity of the log-likelihood”, Journal of Mathematical Sociology 4: 103-120.

Ramsey, J.B. (1969), “Tests for specification errors in classical linear squares regression analysis”. *Journal of the Royal Statistical Society*, Series B, 31: pp.350-371.

Sutton, M. and C. Godfrey (1995), “A grouped data regression approach to estimating economic and social influences on individual drinking behaviour”. *Health Economics* 4. pp. 237-247..

Vuong, Q.H. (1989), " Likelihood ratio tests for model selection and non-nested hypotheses", *Econometrica*, Vol. 57, No 2, pp. 307-333.

Wasgtaff, A.(1986),"The demand for health: some new empirical evidence", *Journal of Health Economics*, 5, pp. 195-233.

Windmeijer, F.A.G. and J.M.A. Santos Silva, (1997),"Endogeneity in count data models: an application for Health Care demand", *Journal of Applied Econometrics*, vol. 12, pp. 281-294.

Zweifel, P. (1981), "Supplied-Induced-Demand in a model of physician behaviour" in Van der Gaag, J. and M. Perlman (eds.) "Health, Economics and Health Economics", Amsterdam, North Holland, pp. 245-267.

NORMAS DE PUBLICACIÓN DE PAPELES DE TRABAJO DEL INSTITUTO DE ESTUDIOS FISCALES

Esta colección de *Papeles de Trabajo* tiene como objetivo ofrecer un vehículo de expresión a todas aquellas personas interasadas en los temas de Economía Pública. Las normas para la presentación y selección de originales son las siguientes:

1. Todos los originales que se presenten estarán sometidos a evaluación y podrán ser directamente aceptados para su publicación, aceptados sujetos a revisión, o rechazados.

2. Los trabajos deberán enviarse por duplicado a la Subdirección de Estudios Tributarios. Instituto de Estudios Fiscales. Avda. Cardenal Herrera Oria, 378. 28035 Madrid.

3. La extensión máxima de texto escrito, incluidos apéndices y referencias bibliográficas será de 7000 palabras.

4. Los originales deberán presentarse mecanografiados a doble espacio. En la primera página deberá aparecer el título del trabajo, el nombre del autor(es) y la institución a la que pertenece, así como su dirección postal y electrónica. Además, en la primera página aparecerá también un abstract de no más de 125 palabras, los códigos JEL y las palabras clave.

5. Los epígrafes irán numerados secuencialmente siguiendo la numeración arábiga. Las notas al texto irán numeradas correlativamente y aparecerán al pie de la correspondiente página. Las fórmulas matemáticas se numerarán secuencialmente ajustadas al margen derecho de las mismas. La bibliografía aparecerá al final del trabajo, bajo la inscripción "Referencias" por orden alfabético de autores y, en cada una, ajustándose al siguiente orden: autor(es), año de publicación (distinguiendo a, b, c si hay varias correspondientes al mismo autor(es) y año), título del artículo o libro, título de la revista en cursiva, número de la revista y páginas.

6. En caso de que aparezcan tablas y gráficos, éstos podrán incorporarse directamente al texto o, alternativamente, presentarse todos juntos y debidamente numerados al final del trabajo, antes de la bibliografía.

7. En cualquier caso, se deberá adjuntar un disquete con el trabajo en formato word. Siempre que el documento presente tablas y/o gráficos, éstos deberán aparecer en ficheros independientes. Asimismo, en caso de que los gráficos procedan de tablas creadas en excel, estas deberán incorporarse en el disquete debidamente identificadas.

Junto al original del Papel de Trabajo se entregará también un resumen de un máximo de dos folios que contenga las principales implicaciones de política económica que se deriven de la investigación realizada.

PUBLISHING GUIDELINES OF WORKING PAPERS AT THE INSTITUTE FOR FISCAL STUDIES

This serie of *Papeles de Trabajo* (working papers) aims to provide those having an interest in Public Economics with a vehicle to publicize their ideas. The rules governing submission and selection of papers are the following:

1. The manuscripts submitted will all be assessed and may be directly accepted for publication, accepted with subjections for revision or rejected.

2. The papers shall be sent in duplicate to Subdirección General de Estudios Tributarios (The Deputy Direction of Tax Studies), Instituto de Estudios Fiscales (Institute for Fiscal Studies), Avenida del Cardenal Herrera Oria, nº 378, Madrid 28035.

3. The maximum length of the text including appendices and bibliography will be no more than 7000 words.

4. The originals should be double spaced. The first page of the manuscript should contain the following information: (1) the title; (2) the name and the institutional affiliation of the author(s); (3) an abstract of no more than 125 words; (4) JEL codes and keywords; (5) the postal and e-mail address of the corresponding author.

5. Sections will be numbered in sequence with arabic numerals. Footnotes will be numbered correlatively and will appear at the foot of the corresponding page. Mathematical formulae will be numbered on the right margin of the page in sequence. Bibliographical references will appear at the end of the paper under the heading "References" in alphabetical order of authors. Each reference will have to include in this order the following terms of references: author(s), publishing date (with an a, b or c in case there are several references to the same author(s) and year), title of the article or book, name of the journal in italics, number of the issue and pages.

6. If tables and graphs are necessary, they may be included directly in the text or alternatively presented altogether and duly numbered at the end of the paper, before the bibliography.

7. In any case, a floppy disk will be enclosed in Word format. Whenever the document provides tables and/or graphs, they must be contained in separate files. Furthermore, if graphs are drawn from tables within the Excell package, these must be included in the floppy disk and duly identified.

Together with the original copy of the working paper a brief two-page summary highlighting the main policy implications derived from the research is also requested.

ÚLTIMOS PAPELES DE TRABAJO EDITADOS POR EL INSTITUTO DE ESTUDIOS FISCALES

2000

1/00 Crédito fiscal a la inversión en el impuesto de sociedades y neutralidad impositiva: Más evidencia para un viejo debate.
Autor: Desiderio Romero Jordán.
Páginas: 40.

2/00 Estudio del consumo familiar de bienes y servicios públicos a partir de la encuesta de presupuestos familiares.
Autores: Ernesto Carrilllo y Manuel Tamayo.
Páginas: 40.

3/00 Evidencia empírica de la convergencia real.
Autores: Lorenzo Escot y Miguel Ángel Galindo.
Páginas: 58.

Nueva Época

4/00 The effects of human capital depreciation on experience-earnings profiles: Evidence salaried spanish men.
Autores: M. Arrazola, J. de Hevia, M. Risueño y J. F. Sanz.
Páginas: 24.

5/00 Las ayudas fiscales a la adquisición de inmuebles residenciales en la nueva Ley del IRPF: Un análisis comparado a través del concepto de coste de uso.
Autor: José Félix Sanz Sanz.
Páginas: 44.

6/00 Las medidas fiscales de estímulo del ahorro contenidas en el Real Decreto-Ley 3/2000: análisis de sus efectos a través del tipo marginal efectivo.
Autores: José Manuel González Páramo y Nuria Badenes Pla.
Páginas: 28

7/00 Análisis de las ganancias de bienestar asociadas a los efectos de la Reforma del IRPF sobre la oferta laboral de la familia española.
Autores: Juan Prieto Rodríguez y Santiago Álvarez García.
Páginas 32.

8/00 Un marco para la discusión de los efectos de la política impositiva sobre los precios y el *stock* de vivienda.
Autor: Miguel-Ángel López García.
Páginas 36.

9/00 Descomposición de los efectos redistributivos de la Reforma del IRPF.
Autores: Jorge Onrubia Fernández y María del Carmen Rodado Ruiz.
Páginas 24.

10/00 Aspectos teóricos de la convergencia real, integración y política fiscal.
Autores: Lorenzo Escot y Miguel-Ángel Galindo.
Páginas 28.

2001

1/01 Notas sobre desagregación temporal de series económicas.
Autor: Enrique M. Quilis.
Páginas 38.

2/01 Estimación y comparación de tasas de rendimiento de la educación en España.
Autores: M. Arrazola, J. de Hevia, M. Risueño, J.F. Sanz.
Páginas 28.

3/01 Doble imposición, “efecto clientela” y aversión al riesgo.
Autores: Antonio Bustos Gisbert y Francisco Pedraja Chaparro.
Páginas 34.

4/01 Non-Institutional Federalism in Spain.
Autor: Joan Rosselló Villalonga.
Páginas 32.