DISCUSSION PAPER SERIES

No. 3021

HEALTH CARE REFORM AND THE NUMBER OF DOCTOR VISITS – AN ECONOMETRIC ANALYSIS

Rainer Winkelmann

LABOUR ECONOMICS AND PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP3021.asp**

ISSN 0265-8003

HEALTH CARE REFORM AND THE NUMBER OF DOCTOR VISITS – AN ECONOMETRIC ANALYSIS

Rainer Winkelmann, Universität Zürich and CEPR

Discussion Paper No. 3021
October 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3021

October 2001

ABSTRACT

Health Care Reform and the Number of Doctor Visits – An Econometric Analysis*

This Paper evaluates the German health care reform of 1997, using the individual number of doctor visits as an outcome measure. A new econometric model, the Probit-Poisson-log-normal model with correlated errors, describes the data better than existing count data models. Moreover, it has an attractive structural interpretation, as it allows the reforms to have a different effect at different parts of the distribution. The overall effect of the reform was a 10% reduction in the number of doctor visits. The effect was much larger in the lower part of the distribution than in the upper part.

JEL Classification: C25, I11 and I18
Keywords: co-payment, count data, moral hazard and Probit-Poisson-log-normal model

Rainer Winkelmann
Statistisches Seminar
Universität Zürich
Raemistr. 62
8001 Zurich
SWITZERLAND
Tel: (41 1) 634 2301
Fax: (41 1) 634 4996
Email: winkelmann@sts.unizh.ch

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=116561

* I am grateful to Joao Santos-Silva as well as seminar participants at IZA, the Universities of Heidelberg, München, Gottingen, Zurich, Dortmund and Kiel and the IZA-CEPR European Summer Symposium in Labour Economics 2001 for helpful comments.

Submitted 20 July 2001

1 Introduction

Expenditures for health services make up a substantial portion of total GDP in all OECD countries. For most countries, health expenditures as a share of total GDP have trended upward over the last years and decades. In Germany, for example, the share increased from 8.4 percent in 1980 to 10.5 percent in 1996 (Breyer and Zweifel, 1999). The most commonly cited reasons for this increase are the expanding technological possibility in the health service sector as well as the ageing of the population, coupled in many countries with a large public health service where the incentive structures do not promote economic use of the resources.

One such country with a large publicly funded health sector is Germany. There have been regular attempts to reform the health care system in order to reduce cost. The purpose of this paper is to evaluate the success of a major reform that took place in 1997. In that reform, the co-payments for prescription drugs were raised by up to 200 percent. In addition, a modified budget system imposed upper limits for reimbursements of physicians by the state insurance.

The contributions of this paper are twofold. First, it provides an answer to the substantive question whether or not the health care reform of 1997 has been a success, using as outcome measure the individual number of visits to a doctor. Second, it derives a new econometric model for the number of doctor visits, the Probit-Poisson-log-normal Model with correlated errors. This model describes the data better than existing count data models. Moreover, it has an attractive structural interpretation, as it allows the reforms to have a different effect at different parts of the distribution. The overall effect turns out to be quite substantial, a 10 percent reduction in the number of doctor visits. The effect is much larger in the lower part of the distribution (for the choice between having no visit or at least one visit) than in the upper part of the distribution (the number of visits given at least one visit).

2 The German Health Care Reform of 1997

More than 90 percent of the German population obtains health insurance through the federal social insurance system that is mostly financed by mandatory payroll deductions. For dependent employees, the premium is proportional to earnings (up to a ceiling), and coverage automatically extends to (non-working) spouse and dependent children. Special membership arrangements exist for other groups, such as the unemployed or students. Once a person is part of the system, the coverage is uniform. In particular, the insurance pays for the cost of doctor visits, hospital stays, and prescription drugs. However, the full treatment costs are not always reimbursed as there is a requirement for a co-payment in many cases.

The focus of this paper is on co-payments for prescription drugs. Such co-payments were increased substantially on July 1, 1997, by a fixed amount of DM 6 relative to a year earlier. Since the absolute amount of the co-payment is a function of the package size, after the reform DM 9 for small, DM 11 for medium and DM 13 for large sizes, the relative effect of the 1997 reform was largest for small sizes, where it amounted to a 200 percent increase. Social considerations resulted in a number of exemptions (co-insured children, low-income households with family gross income under DM 1700/DM2350, maximum cumulative annual co-payments limited to 2 percent of annual gross income; 1 percent for the chronically sick).

The change in co-payments was the most radical element of the 1997 reform. It was reinforced by a number of additional measures that extended previously existing regulations such as an exclusion list (*Negativliste*) defining drugs not covered at all by social insurance, price ceilings related to the availability of generics, as well as a binding overall annual budget for drugs. A further cost saving element of the 1997 reform targeted directly the provision of physician's services. A quarterly budget was introduced at the surgery level as the product of an average cost per patient times the number of patients, with allowances made for emergency treatments. In no way was the budget linked to the actual condition of a patient. By the same token the budget was made fully transferable among patients, in anticipation of the fact that the costs average out at the level of the individual physician. Foreshadowing a later discussion of this point, one might expect that such a budget, while possibly reducing the intensity of treatment chosen by the doctor, might also increase the number of proposed re-appointments. This is so because a re-appointment (for a below average cost treatment) scheduled for a later quarter will actually increase the overall budget and allow for cross-subsidization of above average treatment cost for other patients.

The combination of these different measures, it was hoped, would contain health care expenditures, or better, its rate of increase. By definition, an increased co-payment has a direct fiscal effect, reducing the share of the cost covered by the insurer. For instance, the patient pays the full amount for all drugs with price below the co-payment. Equally important, though, it was hoped that the increased out-of-pocket expenses would raise the awareness of the "customer" and lead to a change in attitudes, reducing what has been perceived as avoidable and excessive use of prescription drugs. Co-payments may increase the incentive to act responsibly and reduce the moral hazard problem.

The following empirical analysis deals with the second aspect. It does so by focusing on the effect of the reform on the number of visits to the doctor at the individual level. This approach is chosen mainly because direct information on the use of prescription drugs is not available. In addition, there are good reasons why the increased co-payments could have changed the demand for doctors visits (in addition to other effects of the praxis budget, if any). The demand for prescription drugs and the demand for doctor visits are closely related, and they might be complements indeed.

Figure 1 clarifies the idea. The 1997 reform increased the out-of-pocket expenses for prescription drugs. To obtain a prescription, one has to see a doctor, the doctor has to fill out a prescription, and one has to go to the pharmacy. Several responses to the price increase are possible, including influencing the doctor to prescribe a larger package size, or not seeing a doctor. Both behavioral changes would reduce the number of visits to a doctor. Alternatively, one might still see a doctor in order to seek advice on non-prescription or self-treatment, and one might not comply with the prescription and just not buy the drug. In either case, the number of visits would tend to be unaffected by the increased co-payment. If there is a combination of the two effects, the number of visits will go down, and it is an empirical question to quantify the size of the overall effect.

Finally, it is worth noting that the 1997 reform enjoyed only a short lifespan. A new coalition government led by the social democrats emerged from general elections held in 1998. The partial repeal of the 1997 reform was one of the first items on the political agenda, and a new law lowered the co-payments by between DM 1 and DM 3, effective January 1, 1999. From an econometric point of view, this second reform is a fortuitous occurrence, as it introduces an additional source of variation in the health environment that can be used to identify the individuals' responses.

3 A previous study

The consequences of the German health care reform of 1997 on demand for health services were assessed previously by Lauterbach, Gandjour and Schnell (2000). The study was based on data collected in October - December 1998 in Cologne among visitors to pharmacies. To be in the sample, one had to be covered by the social insurance, be aged 18 or older, suffer from an acute or chronical sickness, and not be exempted from the co-payment. 10.000 questionnaires were distributed and 695 returned.

The Cologne study included a number of different outcome measures. I concentrate here on the number of visits to a doctor. Those who responded to the survey reported on average 9.2 doctor visits over the previous 12 months. 80.2 percent of all respondents said that the health care reform had no effect on the number fo visits. 8.6 percent reported that they had renounced once, while 11.2 percent said that they had renounced more than one visit due to the reform. Based on this information, Lauterbach et al. estimate a reduction of consultations by 4.5 percent. Thus, the effect of the policy change is economically substantial.

But how robust is this result? The study has a number of shortcomings that may affect the conclusions. The sample size is small and the response rate is very low, raising the issue of response bias. More importantly, the sampling design induces an overrepresentation of heavy users. This is an example for a so-called "on-site", or endogenous sample (see Santos-Silva, 1997, for a clear discussion of this issue). Presence at a pharmacy is highly correlated with an immediate previous doctor visit. Hence, the inclusion in the sample depends on the outcome of the dependent variable, and the results cannot be representative for the population at large. Occasional users of health care services are underrepresented, and non-users are excluded a-priori.

There are two possible responses to this problem. The first would consists in using appropriate econometric techniques to correct for the endogenous sampling, effectively inferring from the distributional form of observations conditional on visits the probability of being observed at all. Of course, this approach requires that the same model applies to those observed in the sample and those not observed (the "users" and the "non-users"), an assumption that can be questioned in the present context. Therefore, if one wants to estimate the effect of the reform in the overall population, one needs a random sample of the entire population, such as is provided for instance by the German Socio-Economic Panel (GSOEP).

Details of this annual household survey are given in the next section. It offers a number of advantages in addition to the representativeness of the sample. In particular, it gives independent measurements of the number of doctor visits before and after reform, from where the change can be computed. This is likely to yield a more accurate estimate than a retrospective self-assessment of the direction of response to reform as considered in the above study. Finally, the GSOEP contains a rich set of other socio-economic characteristics that can be used as control variables, and the individual number of doctor visits over time can be modeled directly using count data models.

4 Data

The GSOEP was initiated in 1984. The latest available release includes data for 1999. For the purpose of this study, I select a period of five years centered around the year of the reform, i.e., 1995 - 1999. The GSOEP has a few variables relating to the usage of health service. One of them is the number of visits to a doctor during the previous 3 months. In some earlier years, this question was asked separately for visits to a general practitioner and visits to a specialist, separately by field. However, only the aggregate count is available in the 1995-1999 waves. Note that visits to a dentist are included in the definition.

I use observations on men and women aged 20-60 from Sample A, i.e., persons associated with non-Guestworker-households in the original sample for West Germany. Privately insured individuals (about 6 percent of the sample) are excluded. Accounting for observations with missing values on any of the dependent or independent variables, the final sample comprises 32837 observations.

The basic empirical strategy, as detailed in the next section, is to pool the data over the five years and estimate the effects of the reforms by comparing the expected number of visits in 1998 and 1996 ceteris paribus, i.e., for an individual with given characteristics. The years 1998 and 1996 are chosen, since the reform took place in the middle of 1997. Thus, depending on the interview month, some 1997 observations fell before the reforms, and some after. Another argument for using the longer time span is a reduced risk of biases due to timing issues. For instance, people might have developed an "extra-demand" for doctor visits just prior to the reform in anticipation of the upcoming changes.

The models that will be estimated in the following sections all include a systematic component (linear predictor) of the type

$$x_{it}'\beta = \beta_0 + \beta_1 \text{age}_{it} + \beta_2 \text{age}_{it}^2 + \beta_3 \text{years of education}_{it} + \beta_4 \text{married}_{it} + \beta_5 \text{household size}_{it} + \beta_6 \text{active sport}_{it} + \beta_7 \text{good health}_{it} + \beta_8 \text{bad health}_{it} + \beta_9 \text{self employed}_{it} + \beta_{10} \text{full-time employed}_{it} + \beta_{11} \text{part-time employed}_{it} + \beta_{12} \text{unemployed}_{it} + \beta_{13} \text{equivalent income}_{it} + \beta_{96} (\text{year} = 1996)_{it} + \beta_{97} (\text{year} = 1997)_{it} + \beta_{98} (\text{year} = 1998)_{it} + \beta_{99} (\text{year} = 1999)_{it}$$

The reference year is 1995. In addition, there are three dummies for the season of the interview (winter, fall, spring). The linear predictor will be embedded in various alternative count data models, starting with the Poisson model. It is assumed that the reform effect is the same for all groups of the population or, alternatively, that the interest lies in the average response. One could allow for heterogeneous responses by estimating the model for subgroups, or including interactive terms.

There are three general channels through which these variables can affect the demand for doctor visits. The first is the underlying health status, the second the budget constraint, and the third the preference formation. The health status is poorly measured in the GSOEP. In particular, no details of current medical conditions are known. A time-consistent measure of health over 1995-1999 is subjective self-assessment in response to a question: "How good do you perceive your own health at current?", with responses "very good", "good", "fair", "poor", and "very poor". The two best responses are classified as "good health", the two lowest responses as "bad health", with fair health being the reference group. An other proxy for health is the age polynomial. Finally, engaging in "active sports" (defined as a frequency of weekly or higher) is seen as a further proxy for good health. Clearly, these are only crude measures of health, and one may want to account for the possibility of additional unobserved heterogeneity to capture any remaining health aspects, as well as other unobserved influences.

The budget constraint is determined by income and prices. Income is measured through household equivalent income, where the OECD scale has been applied (weight of one for the first person, 0.7 for the second person, and 0.5 for each additional person). Income is expressed in 1995 values using the CPI deflator published by the Sachverstän-digenrat (German Council of Economic Advisors). The main price variables are the opportunity costs of a visit to a doctor which, in turn depend on education levels and employment status. The influence of insurance status cannot be modeled in any meaningful way. The number of uninsured persons in Germany is too small to be empirically relevant, and privately insured persons are excluded from the analysis, mainly because no systematic information on the nature of the insurance contract is available.

Several of the variables affect more than one aspect at a time. Age, for instance, matters for health, opportunity cost (through the effect of experience on earnings) as well as potentially preferences. Similarly, education is an important factor in determining the optimal investment in health capital (Grossman, 1972). It is not the goal of this paper to disentangle these various transmission channels. Rather, the focus lies on the year dummies, whereas the other right hand sight variables serve as controls for any effects of these variable on the changes in visits over time in order to estimate the adjusted, or ceteris paribus, effect of the reform.

5 Econometric models

The standard probability distribution for count data is the Poisson distribution:

$$P(y_i|\lambda_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \tag{1}$$

where

$$\mathrm{E}(y_i|\lambda_i) = \mathrm{Var}(y_i|\lambda_i) = \lambda_i$$

In a regression model, we assume that the population is heterogeneous with covariates x_i, and λ_i is specified as $\lambda_i = \exp(x_i'\beta)$ where $i = 1, \ldots, N$ indexes observations in the sample. Let $y = (y_1, \ldots, y_N)'$ and $x = (x_1, \ldots, x_N)'$. Under random sampling

$$P(y|x) = \exp\left[-\sum_{i=1}^{N}\exp(x_i'\beta)\right]\prod_{i=1}^{N}\frac{[\exp(x_i'\beta)]^{y_i}}{y_i!} \tag{2}$$

and estimation of the parameters by maximum likelihood is straightforward.

Furthermore, the reform effect given by the expected change of doctor visits can be computed as follows:

$$\begin{aligned}\Delta\%_{(98,96)} &= \left[\frac{\mathrm{E}(y_{i,98}|x)}{\mathrm{E}(y_{i,96}|x)} - 1\right] \times 100 \\ &= [\exp(\beta_{98} - \beta_{96}) - 1] \times 100 \end{aligned} \tag{3}$$

This estimate serves as a benchmark. Clearly, the simple Poisson model can be criticized on a number of grounds. To begin with, it does not allow for unobserved heterogeneity. Alternative models, such as the negative binomial model or the Poisson-log-normal model provide potentially more efficient estimators (see e.g., Winkelmann, 2000). Secondly, it ignores the panel structure of the data. There are up to five observations for a given persons. The presence of an individual specific heterogeneity term will invalidate the assumption of independent sampling. One can incorporate this into estimation by using a random effects panel model. Depending on the assumptions, such models again can be of a negative binomial or a Poisson-log-normal variety. Alternatively, one could suspect dependence between the individual effects and the covariates. In this case, a fixed effects Poisson model would be preferred.

The focus of this paper is somewhat different, however. First, quasi-likelihood results imply that the Poisson maximum likelihood estimator remains consistent even if unobserved heterogeneity, be it individual specific or individual and period specific, is present. In this sense, the estimator is robust and the panel structure does not invalidate the estimator (although the standard errors need to be corrected). Secondly, fixed effects models are not considered either. Two considerations have led to this restriction. First, the reform effect is independent of any individual specific error by construction. Second, the main methodological interest of this paper is to evaluate a number of alternative "structural models" for which fixed effects estimators simply do not exist at this stage. It is to such structural models that the attention of the following subsection now turns.

5.1 Structural models

The defining element of the Poisson regression model is the single index log-linear conditional expectation function. More general models can be thought of, and each of them has interesting structural interpretations. Historically, the most important model is the hurdle model (Mullahy, 1986). The hurdle model combines a binary model for the decision of use with a truncated (1+) count data model for the extent of use given use. Define $d_i = 1$ if a person does **not** see a doctor in a given period, i.e., $d_i = 1 - \min(1, y)$. The probability function is then given by

$$f(y_i) = f_{1i}^{d_i} \left[(1 - f_{1i}) f_T(y_i|y_i > 0)\right]^{1-d_i} \tag{4}$$

where $f_{1i} = P(d_i = 1)$, $f_T(y_i|y_i > 0) = f_2(y_i)/[1 - f_{2i}(0)]$, and independence between hurdle and positive part is assumed. Estimation is simple, since the log-likelihood factors into two parts

$$\ln L = \sum_i d_i \ln f_{1i} + (1 - d_i) \ln(1 - f_{1i}) + \sum_{d_i=0} \ln f_2(y_i) - \ln(1 - f_{2i}(0))$$

To close the model, one needs to specify f_1 and f_2. Choices for the hurdle f_1 include:

- $\exp(-\exp(x_i'\gamma))$ (Poisson)
- $[\alpha/(\exp(x_i'\gamma) + \alpha)]^\alpha$ (negative binomial)
- $\exp(x_i'\gamma)/(1 + \exp(x_i'\gamma))$ (Logit)
- $\Phi(x_i'\gamma)$ (Probit)

Choices for f_2 include:

- Poisson
- Negative Binomial
- Poisson log-normal

The first two hurdle expressions for f_1 possess the advantage that, if combined with the same distribution for f_2, the hurdle model nests the simple model. The Probit assumption has the advantage that it can be easily generalized to a model with correlated hurdle, as shown below.

The hurdle model has been popular in the health literature, partially because it can be given a structural interpretation. It describes the dual decision structure of the demand process, with the contact decision made independently by the person, and the treatment and referral decisions (co)influenced by the physician. Rightfully, Deb and Trivedi (1999, p.2) note that

> "In modeling the usage of medical services, the two-part model (i.e. hurdle model, my insertion) has served as a methodological cornerstone of empirical analysis."

However, Deb and Trivedi (1999) go on to criticize this conventional approach. They point out an incongruence between model and data situation: medical consultations are measured per period and not per illness episode. Moreover, healthy individuals consult physicians as well. In a similar spirit, Santos Silva and Windmeijer (2000) note that several illness episodes are possible (i.e., one cannot identify a single binary contact decision). One rather should model episodes and contacts per episode jointly within the framework of stopped sum distributions.

As alternative candidate to the hurdle model, Deb and Trivedi (1999) advocate a finite mixture model in order to discriminate between frequent and less frequent users. Such a model can, for instance, capture unobserved differences with respect to the long-run state of health that affect the constant as well as the slope coefficients. For instance, let

$$f(y_i|\theta) = \sum_{j=1}^{s} \pi_j f_j(y_i|\theta_j) \tag{5}$$

where f_j is a Poisson- or Negbin distribution. For $s = 2$, the model has the same number of parameters as the hurdle model, and the two can be compared directly. Deb and Trivedi (1999, p. 1) conclude from an application to health data that "there is compelling evidence in favor of a latent class model."

Santos Silva and Windmeijer (2000) by contrast formulate a model of the form

$$Y = R_1 + R_2 + \ldots + R_S = \sum_{i=j}^{S} R_j$$

where Y is the total number of visits, R is the number of contacts per episode, and S is the number of episodes. If $S = 0, 1, 2, \ldots$ is Poisson distributed, and $R_j = 1, 2, \ldots$ are identically logarithmically distributed, all independently, with means

$$\mathrm{E}(S_i) = \exp(x_i'\beta)\,, \qquad \mathrm{E}(R_{ij}) = \frac{\exp(x_i'\gamma)}{\ln[1 + \exp(x_i'\gamma)]}$$

then one can show that Y is negative binomial distributed with

$$f(y_i|x_i) = \frac{\Gamma\left(y_i + \frac{\exp(x_i'\beta + x_i'\gamma)}{\ln[1+\exp(x_i'\gamma)]}\right)\exp(-\exp(x_i'\beta))}{\Gamma(y_i+1)\Gamma\left(\frac{\exp(x_i'\beta + x_i'\gamma)}{\ln[1+\exp(x_i'\gamma)]}\right)(1+\exp(-x_i'\gamma))^{y_i}} \tag{6}$$

and

$$\mathrm{E}(y_i|x_i) = \frac{\exp(x_i'\beta + x_i'\gamma)}{\ln[1 + \exp(x_i'\gamma)]}\,.$$

5.2 A new hurdle model

Do these criticism of the hurdle model mean that we should abandon the model and switch to the finite mixture or multi-episode alternatives, investigated by Deb and Trivedi (1999) and Santos Silva and Windmeijer (2000), respectively? The answer to this question depends on one's philosophical stance. If one insists that the model must be able to separately identify contact and frequency decisions, then the hurdle model is inappropriate indeed. If, however, one sees all models as mere approximations to some underlying "true" model, then the question becomes rather which one of several approximation is the best, a question that cannot be decided in general but depends on the specific empirical application.

Two general points are pertinent, though. First, one should be interested in a parsimonious representation of the data in any case. In this respect, hurdle, 2-group finite mixture, and multi episode models fare all equally well, as they all roughly double the initial number of regression parameters, unless one is willing to impose some arbitrary prior assumptions. Second, the three proposed structural model have in common that they allow different responses to covariate changes in different parts of the distribution. This is an interesting feature, and a great advantage over the simple Poisson model. Its relevance for policy is detailed below.

The conventional hurdle model has one striking feature that may be responsible for making it an inferior approximation in health demand applications. The standard model assumes conditional independence between the hurdle step and the distribution model for the positives. What happens to the hurdle model, if this particular assumption is relaxed? We have already seen before that there are many possible formulations for a hurdle model, with at least four different "plausible" models for the hurdle. To state the case as clearly as possible, I follow the approach that minimizes numerical difficulties. In particular, I combine a probit model for the hurdle with a truncated Poisson-log-normal model for the positives. Let z_i be a latent indicator variable such that

$$z_i = x_i'\gamma + \varepsilon_i$$

and

$$y_i = 0 \text{ iff } z_i \geq 0 \ .$$

Moreover, for the positive part of the distribution

$$y_i | y_i > 0 \sim \text{truncated Poisson}(\lambda_i)$$

where

$$\lambda_i = \exp(x_i'\beta + u_i)$$

The model is completed by the assumption that ε_i and u_i are bivariate normal distributed with mean 0 and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}$$

The structure of this model is thus very similar to the class of models with *endogenous selectivity* described in detail in Winkelmann (2000).

To derive the log-likelihood function, note first that

$$\varepsilon_i | u_i \sim N(\rho u_i/\sigma, 1 - \rho^2)$$

and

$$
\begin{aligned}
P(y_i = 0|u_i) &= P(\varepsilon_i \geq -x_i'\gamma|u_i) \\
&= \Phi\left(\frac{x_i'\gamma + \rho u_i/\sigma}{\sqrt{1-\rho^2}}\right) \\
&= \Phi_i^*(u_i)
\end{aligned}
$$

Thus one obtains

$$
f(y_i|u_i) = \Phi_i^*(u_i)^{d_i} \times \left[(1-\Phi_i^*(u_i))\frac{\exp(-\lambda_i(u_i))(\lambda_i(u_i))^{y_i}}{[1-\exp(-\lambda_i(u_i))]y_i!}\right]^{1-d_i} \tag{7}
$$

and

$$
f(y_i) = \int_{-\infty}^{\infty} f(y_i|u_i)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{u_i}{\sigma}\right)^2}\mathrm{d}u_i \tag{8}
$$

The likelihood can be evaluated using Gauss-Hermite integration. The correlation should be negative (unobserved factors). If P(no use) is high, then E(y|use) is low.

5.3 Reform effect in the different models

The ultimate goal of this paper is the evaluation of the reform effect, namely the ceteris paribus reduction in the expected number of doctor visits between 1996 and 1998. The appropriate formula for the Poisson model was already given in (3). Identical computations apply for the Negbin and Poisson-log-normal models. Due to the log-linearity of the conditional expectation function, the proportional effect is independent of the values taken by other independent variables.

For the structural models, the reform effect is more complex. In the hurdle model, the overall effect can be decomposed into an effect for the hurdle and an effect for positive counts. These two effects can complement or counteract each other. Similarly, in the finite mixture model the reform may impact differently on the two groups. Finally, in the multi-episode model, separate effects are identified for the number of spells and the number of referrals.

Formally, the computations in the three models are as follows:

1. Hurdle model

$$\begin{aligned}\frac{\mathrm{E}(y_{98})}{\mathrm{E}(y_{96})}-1 &= \frac{P(y_{98}>0)}{P(y_{96}>0)}\frac{\mathrm{E}(y_{98}|y_{98}>0)}{\mathrm{E}(y_{96}|y_{96}>0)}-1\\ &= (1+\Delta_{P(Y>0)})(1+\Delta_{E(Y|Y>0)})-1\end{aligned}$$

2. Finite mixture model

$$\frac{\mathrm{E}(y_{98}|\text{group}=j)}{\mathrm{E}(y_{96}|\text{group}=j)}-1=\exp(\beta_{98}^{j}-\beta_{96}^{j})-1\,, \qquad j=1,2$$

3. Multi-episode model

$$\begin{aligned}\frac{\mathrm{E}(y_{98})}{\mathrm{E}(y_{96})}-1 &= \frac{\mathrm{E}(S_{98})}{\mathrm{E}(S_{96})}\frac{\mathrm{E}(R_{98})}{\mathrm{E}(R_{96})}-1\\ &= (1+\Delta_{\mathrm{E}(S)})(1+\Delta_{\mathrm{E}(R)})-1\end{aligned}$$

Except for the finite mixture model, the estimated effects will depend on the realized values of the other independent variables. The computations in the following section evaluate these effects at the sample means of the variables.

6 Results

Table 1 gives the sample means for the variables involved in the analysis. The average number of doctor visits per quarter declined from 2.66 to 2.35 between 1996 and 1998.

This is a more than 11 percent reduction in the number of quarterly visits. There was a 1 percent decline between 1995 and 1996, and a 2 percent increase between 1998 and 1999. Thus, the large drop in the number of visits clearly coincides with the timing of the reform. Also, the 1999 “counter” reform went in hand with an increase in the number of visits, again consistent with the hypothesis of a behavioral effect.

Throughout the sample period, there is a large fraction of non-users. The proportion is highest in 1998, when it reaches 37 percent of the sample, a 4.4 percentage point increase over the pre-reform year 1996. A simple Poisson distribution with parameter equal to the sample mean would predict a much lower proportion of non-users, e.g., 9.5 percent in 1998. Although this comparison does not take into account the variation generated by the regressors, it suggests the presence of extra-zeros on the data. Finally note that the annualized number of doctor visits in 1998 is remarkably close to the statistic reported in the survey by Lauterbach et al. (2000).

The average age increases by less than a year. This is a reflection of the fact that the panel is not balanced. One reason for this is that young people enter and old people leave due to restriction to those aged 20-60 at a given point in time, in addition to attrition and non-response. The proportion of the sample unemployed (based on the ILO definition) captures the state of the business cycle. Indeed, it closely traces the movements of the official unemployment rates (see e.g. Sachverständigenrat 2000) that peaked for former West Germany at 11 percent in 1997.

Finally, Table 1 also informs about the other health related variables used in the analysis. Interestingly, the statistics indicate a general improvement in the health status of the population between 1996 and 1998. The proportion of people in active sports increased from 25 to 31 percent, although these averages are very volatile. A steadier trend is observed for the self-reported health condition. The proportion of people reporting good health increased from 56 to 60 percent, while the proportion of people reporting poor health decreased from 14 to 13 percent. These trends are important for two reasons. Firstly, improvements in the perceived health might be able to explain part of the reduction in the number of doctor visits, and one should control for that in order to isolate the reform effect. Secondly, they provide some evidence against the possibility that the reforms, while being successful in containing costs, actually worsened the general health status. Of course, this is only a very crude measure, and more research would be needed to study the long-term consequences of expenditure reductions in the health sector on public health. This is beyond the scope of the current analysis.

The estimates for the basic Poisson model are displayed in Table 2. The first column shows the coefficients and the second the incidence ratio, the ratio of the expected values for a unit increase in the independent variable which is a constant equal to the exponential of the coefficient in the exponential Poisson regression. Many of the results are common in the literature: men have less doctor visits than women, employment reduces the demand for visits, as does household size. The health indicators have the largest effect among all variables. Interestingly, controlling for health, engaging actively in sports increases the number of visits. Most importantly, though, the coefficients on the year dummies indicate that the expected number of visits fell by 9.9 percent between 1996 and 1998, a statistically significant reduction.

While these results are interesting in their own right, a major purpose of this paper is to extent and test the econometric methodology for the analysis of counts. To this end, six additional models were estimated on the same data: Negbin, Poisson-log-normal, Hurdle-Negbin, finite mixture Negbin with two components, Multi-episode model, and Probit-Poisson-log-normal with correlated errors. The following discussion of the models is guided by the following questions: Is the result found in the base Poisson model robust with respect to model choice? Does the Dep and Trivedi (1999) conclusion of the superiority of the finite mixture model over the hurdle model hold up? And to what extent can one uncover asymmetries in the responses to the reform in different parts of the distribution (rather than focusing on the mean effect only)?

There are several ways to select between the models. Some of them are nested (such as the Poisson and the Negbin model), most of them are not (such as the finite mixture, the hurdle Negbin and the multi-episodes models). Table 3 shows the log-likelihood values of the different models. Likelihood ratio tests clearly reject the Poisson model against the alternative models with unobserved heterogeneity. To pick the best model among all seven, a comparison of the simple likelihoods is a first indicator. From this, one can compute for instance the average probability that the model has generated the data, denoted here as S.

However, both the log-likelihood and the S statistics do not account for the fact that the number of parameters differ across the estimated models. Hence, the Akaike Information Criterion (AIC) is included as well. The overall result stays the same, regardless of what model selection criterion one chooses. The new model with probit hurdle and log-normal unobserved heterogeneity, allowing for correlation between the two, offers a substantial improvement over all other models. One should also point out that the results corroborate the Deb and Trivedi (1999) result that the finite mixture Negbin model outperforms the hurdle Negbin model. Thus their result has to be interpreted as evidence against the particular hurdle parameterization, and not against hurdle models in general.

To analyze the particular relationship between the two hurdle models and the finite mixture model more formally, one can use the Vuong (1989) test for model selection among non-nested models. This test does not require any of the two models be correctly specified, but rather picks the model that is closer to the true distribution. The test is directional and symmetric. Under the null-hypothesis that the two models are equivalent, the test statistic

$$\frac{\sum_{i=1}^{N} \log f(y_i|x_i) - \log g(y_i|x_i)}{\sqrt{\sum_{i=1}^{N} \left(\log f(y_i|x_i) - \log g(y_i|x_i)\right)^2}}$$

is standard normal distributed. Note that the numerator is nothing else than the log of the likelihood ratio. To implement the test, one chooses a critical values c from the standard normal distribution. If the value of the statistic is greater than c, one rejects the null hypothesis of equivalence against the alternative that model f is better than model g. If the test statistic is smaller than $-c$, g is better than f. The test statistic for the probit Poisson log normal against the finite mixture model is 5.7, and against the hurdle Negbin model 14.7. Hence, the null hypothesis is rejected in both cases in favor of the new model. The Deb and Trivedi comparison leads to a test statistic of 6.4, leading to a selection of the finite mixture model over the hurdle Negbin.

Table 4 reports some of the parameter estimates for the probit Poisson-log-normal model. (The full set of estimation results for all models is available on request). The first columns gives the coefficients for the hurdle, the second for the positive part. Due to the parameterization of the model (the hurdle is parameterized for the event of no visit), the coefficients should normally be of opposite sign, implying that the effect of a variable on the probability of use and on the extent of use, given use, go in the same direction. However, this is a property that is decided by the data, and not imposed a-priori. While the "sign-test" in Table 4 indeed reveals no anomalies, one still obtains the interesting result that the variable "Active sport" has a significant impact on the probability of non-use, but is insignificant for the positives. Such asymmetries are most interesting with regard to the reform effect, as discussed below. First, however, it is opportune to point out that the correlation parameter ρ, while negative, is insignificant. Hence, the initial conjecture that the poor comparative performance of the hurdle model was caused by the independence assumption, though plausible a-priori, is not supported by the evidence. The probit Poisson-log-normal model works very well, albeit for reasons other than the suspected ones. It must be the specific distributional assumptions (probit rather than Negbin hurdle; Poisson-log-normal rather than Negbin for positives) that contribute to the better performance.

Finally, we turn to the main question associated with the application, the size of the overall reform effect, measured by the percentage reduction in the expected number of doctor visits, ceteris paribus. Those changes are listed for the various models in Table 5. The estimates for the base model, with or without unobserved heterogeneity, are all in the same range, varying between 9.9 to 10.4 percent. These estimates are substantially above those of the Lauterbach et al. study who reported a decline of 4.5 percent. How can these two findings be reconciled? It is possible that the differences have to do with the low response rate in their survey, or the way the question was posed that differs from the GSOEP approach. The analysis of this paper suggests, however, a more fundamental reason, namely the fact that the Cologne study sampled individuals on-site and thus overrepresented heavy users. If it is the case that heavy users have a lower demand elasticity than occasional users, the two findings can be reconciled.

The structural models estimated above can exactly deal with this question of different elasticities in different parts of the distribution. Table 5 confirms that such a differential effect is present indeed. This is most obvious from the probit-Poisson-log-normal hurdle estimates. The reduction is greatest at the left margin of the distribution: the probability of being a user (for at least one time) decreased by an estimated 6.7 percent between 1996 and 1998, whereas the expected number of visits, conditional on use, decreased only by an estimated 2.6 percent. Compare this to the alternative of a single Poisson-log-normal model without hurdle. In this case, the implied changes are -3.0 percent for $P(Y > 0)$ and -6.1 percent for $E(Y|Y > 0)$, respectively. Hence, the evidence clearly suggests an excess sensitivity at the left end of the distribution.

This important result is confirmed by the other two structural models, although quantitative details differ. The finite mixture model separates the population into two groups. Two-thirds of the population belong to a low user group with a mean number of quarterly visits of 1.6, and one third belongs to a high user group with a mean number of 3 visits per quarter. Consistent with the above argument, the low user group shows the larger response to the reform with a 13 percent reduction. Similarly, in the multi-episode model, the effect on the number of spells is much greater than the effect on the number of referrals (which actually are estimated to have slightly increase by 1.3 percent). In each case, the two effects add up to a combined effect in the vicinity of a 10 percent reduction in the number of visits between 1996 and 1998.

7 Discussion

Is the effect uncovered by this analysis really causal? Identification is through variation in time. Thus, it is assumed that other things didn't change as well, beyond the individual socio-economic characteristics controlled for in the regression. It is hard to imagine what these other things should have been. It is unlikely that the underlying unobserved health status varied substantially between the two years beyond the controls, or that a health epidemic of major proportion hit in 1996 but was absent in 1998.

Even if one accepts the interpretation that the health care reform of 1997 was causally responsible for reducing the subsequent number of visits, it is still an additional step to attribute this effect to the increased co-payments. The reforms consisted of several measures, the demand side policy of increased co-payments being one of them. A 1997 change that affected the supply side was the introduction of a "Praxis Budget", which intended to affect the physician's treatment choices. While this system did provide incentives for reduced treatment intensity, it also made it beneficial for a physician to see a patient at least once each quarter. Thus, supply side arguments are in contradiction with the observed evidence of a the large reduction in first time visits, combined with a small effect on repeated visits.

Certainly, future work needs to pursue these issues further. Such work can build on the methodological insights of this paper. When studying the effects of reforms on demand on the number of doctor visits, hurdle-, or two-part-models should be given serious consideration. This paper has extended previous approaches by developing a model that allows for correlation between the zero-hurdle and the positive part of the distribution. The results showed that the reforms affected the hurdle step much more than the positive part of the distribution. To the extent that the positive part represents the subpopulation of the seriously or chronically ill, whereas the left end of the distribution represents the healthy, this might have been an intended consequence of the reforms.

References

Breyer, F. and P. Zweifel (1999) *Gesundheitsökonomie*, Springer: Heidelberg.

Deb, P. and P.K. Trivedi (1999) The structure of demand for health care: latent class versus two-part models,

Grossman, M. (1972) On the concept of health capital and the demand for health, *Journal of Political Economy* 80, 223-255.

Lauterbach, K.W., A. Gandjour, and G. Schnell (2000) Zuzahlungen bei Arzneimitteln, Powerpoint presentation at http://www.medizin.uni-koeln.de/kai/igmg/stellungnahme/zuzahlungen/index.htm, version 25.2.2000.

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

Sachverständigenrat (2000) *Chancen auf einen höheren Wachstumspfad, Jahresgutachten 2000/01* Metzler-Poeschel, Stuttgart.

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

Santos-Silva, J.M.C. and F. Windmeijer (1999) Two-part multiple spell models for health care demand. Institute for Fiscal Studies Working Paper No. W99/02.

Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica* 57(2), 307-333.

Winkelmann, R. (2000): Econometric Analysis of Count Data, 3rd ed. Springer: Heidelberg, New York.

Table 1. Sample means of doctor visits and selected socio-demographic characteristics, 1995-1999.

Year	1995	1996	1997	1998	1999
Number of doctor visits[1]	2.687	2.657	2.553	2.353	2.391
(relative change in %)		(-1.1)	(-3.9)	(-7.8)	(+1.6)
No doctor visits (0/1)	0.348	0.328	0.352	0.372	0.346
Age	38.08	38.20	38.47	38.73	38.92
Unemployed (0/1)	0.085	0.084	0.092	0.085	0.075
Active sport[2] (0/1)	0.295	0.247	0.262	0.307	0.266
Good health[3] (0/1)	0.568	0.562	0.581	0.595	0.580
Bad health[3] (0/1)	0.145	0.138	0.134	0.127	0.129
Observations	6790	6555	6480	6781	6231

Source: German Socio-Economic Panel (N = 32837).

[1] During the three months prior to the interview

[2] At least once per week.

[3] How do you assess your current health status: "very good / good" or "bad / very bad"

Table 2. Poisson Results (N = 32837)

	coefficients	incidence ratio
Age $\times 10^{-1}$	-0.106	0.900
	(0.066)	(0.059)
Age^2 $\times 10^{-3}$	0.158	1.171
	(0.080)	(0.094)
Male	-0.209	0.812
	(0.021)	(0.017)
Education 10-1	-0.058	0.944
	(0.037)	(0.035)
Married	0.081	1.084
	(0.022)	(0.024)
Householdsize	-0.052	0.949
	(0.008)	(0.007)
Active sport	0.047	1.048
	(0.019)	(0.020)
Good health	-0.611	0.543
	(0.019)	(0.010)
Bad health	0.813	2.255
	(0.023)	(0.051)
Fulltime employed	-0.238	0.789
	(0.025)	(0.020)
Parttime employed	-0.253	0.776
	(0.030)	(0.023)
Unemployed	-0.164	0.849
	(0.032)	(0.027)
Social assistance	0.086	1.090
	(0.044)	(0.048)
Log(income)	0.093	1.098
	(0.023)	(0.025)
Year = 1996	0.001	1.001
	(0.027)	(0.027)
Year = 1997	-0.030	0.970
	(0.027)	(0.026)
Year = 1998	-0.105	0.900
	(0.027)	(0.024)
Year = 1999	-0.099	0.906
	(0.027)	(0.025)

Source: GSOEP, own calculations. Model includes a constant and three quarterly dummies. Robust standard errors in parentheses.

Table 3. Model Selection

	Log-Likelihood	Parameter	AIC	S
Poisson	-86,566.18	22	173,176.36	7.16
Unobserved heterogeneity:				
Negbin	-64,611.55	23	129,269.10	13.97
Poisson-Log-Normal	-64,202.78	23	128,451.56	14.15
Hurdle models:				
Hurdle Negbin	-64,281.70	45	128,653.40	14.11
Probit-Poisson-log-normal model	-63,870.59	46	127,833.18	14.29
Finite mixture model:				
2-Components Negbin	-64,020.05	47	128,134.10	14.23
Multi episodes model:				
Poisson-logarithmic	-64,246.58	44	128,581.16	14.13

AIC = - 2 ln L + 2 K, S = exp(ln L/N) 100 , N = 32,837.

Table 4. Probit-Poisson-log-normal model with correlated errors (N = 32837)

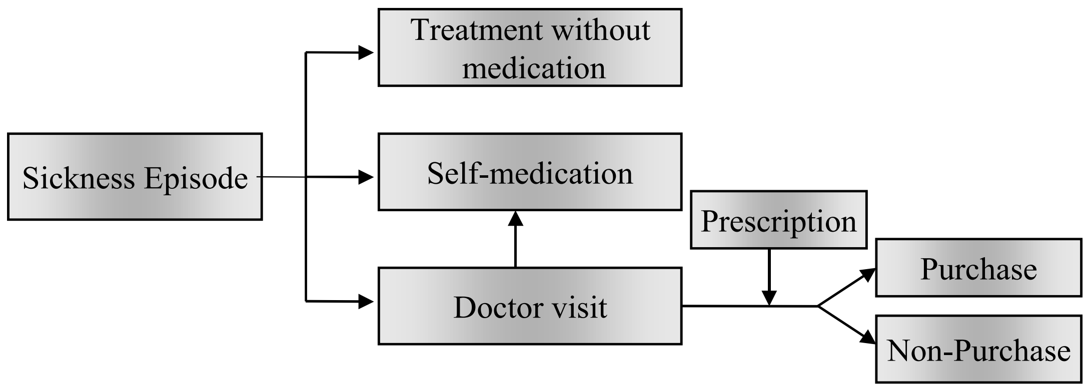
	0/1+	1+
Male	0.403	-0.117
	(0.017)	(0.020)
Active sport	-0.143	0.008
	(0.017)	(0.018)
Good health	0.459	-0.465
	(0.017)	(0.026)
Bad health	-0.573	0.658
	(0.029)	(0.024)
Year = 1996	-0.067	-0.010
	(0.023)	(0.023)
Year = 1997	-0.002	-0.019
	(0.023)	(0.023)
Year = 1998	0.059	-0.042
	(0.023)	(0.024)
Year = 1999	-0.009	-0.075
	(0.024)	(0.023)
σ^2	0.801	
	(0.008)	
ρ	-0.043	
	(0.078)	
Log-Likelihood	-63,870.59	

Further Variables: Age, Years of schooling, married, household size, employment status, income, quarterly dummies and a constant.

Table 5. Evaluation of the reform effect

	$\Delta\%(96{,}98)$
Poisson model	-9.9
Negbin	-8.9
Poisson-log-normal	-10.4
Two-components Negbin	
Group 1 ($p_1 = 0.663$, $\mu_1 = 1.59$)	-12.9
Group 2 ($p_2 = 0.337$, $\mu_2 = 3.01$)	-4.9
Total	-10.2
Probit-Poisson-log-normal	
Hurdle $P(Y > 0)$	-6.7
Positives $E(Y\|Y > 0)$	-2.6
Total	-9.1
Poisson-logarithmic(multi episodes)	
Spells	-10.2
Referrals	+1.3
Total	-9.1

Figure 1: Prescription Drugs and the Demand for Doctor Visits



Source: Lauterbach et al. (2000)