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

Rainer WINKELMANN*

IZA Bonn, und CEPR, London

Abstract

The paper evaluates the German health care reform of 1997, using the individual number of doctor visits as 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 percent 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 Klassifikation: I11, I18, C25

Keywords: co-payment, moral hazard, count data, correlated probit hurdle

* I am grateful to Participants at Seminars at the Universities of Heidelberg, München, and Göttingen, as well as the IZA Brown Bag Seminar for helpful comments. Address for correspondance: IZA Bonn, PO Box 7240, D-53072 Bonn, Germany; winkelmann@iza.org.

1 Introduction

Expenditures for health services make up a substantial portion of total GDP in all OECD countries. For most countries, the share of health expenditures in total GDP has trended upward over the last years. In Germany, for example, it increased from 8.7 percent in 1990 to 10.8 percent in 1996. 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 case-based upper limits for reimbursements of physicians by the state insurance.

The contributions of this paper are twofold. First, it provides answers 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

About 90 percenty of the German population obtain their health insurance through the social insurance system that is finance through mandatory payroll deductions. While the premium is proportional to earnings (up to some ceiling), the coverage is the same for everyone. In particular, coverage includes the cost for prescription drugs. The market for such drugs is highly regulated market. It includes in Germany a much wider set of pharmaceuticals than, say, in the U.S.

Co-payments existed for both prescription drugs and hospital stays already before 1997. However, co-payments for prescription drugs were substantially increased 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 (children, low-income households with familiy 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 explicit aim of the 1997 health care reform was a reduction of the health care costs, or better, its rate of increase. By definition, an increased co-payment has a direct fiscal effect, reducing, as it does, the amount of the cost covered by the insurer. For instance, the patient pays the full amount for all drugs whith 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 focussing on the effect of the reform on the number of visits to the doctor, at the individual level. There are a number of justification for this approach. The first is practical. Individual level data on the number of doctor visits before and after the 1997 reform are publicly available, as part of the German Socio-Economic Panel. But there are also well founded conceptual reasons, why this outcome variable is relevant. In particular, the law maker might have had a reduction of expenditures through a reduced number of doctor visits in mind as well. Such a view is not unfounded, as 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 effect, the number of visits will go down, and it is an empirical question to quantify the size of the overall effect.

3 A previous study

The consequences of the German health care reform of 1997 on demand for health services were assessed in a previous study 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-payments. 10.000 questionnaires were distributed, the response rate was 6.95 percent.

The study included a number of different outcome measures. I will 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, there seem to be an economically substantial effect of the policy change. But how robust is this result? The study has a number of shortcomings that may affect the conclusions. The sample size is very small and the response rate is very low, raising issues of self-selection. More importantly, the sampling design intruduces an extreme selection towards 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 situation. The first would consists in using appropriate econometric techniques to account for the sampling scheme, effectively inferring from the distributional form of observations conditional on visits the probability of being observed in the first place. Of course, this procedure 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 might be highly questionable 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, as is privided for instance by the German Socio-Economic Panel (GSOEP).

Details of this annual household survey are given in the next section. In my view, 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 might give 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 modelled directly using count data models.

4 Data

The GSOEP was started 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 asked separately for visits to a G.P. and visits to a specialist, separately by field. However, in the 1995-1999 waves, only the aggregate count is available. Note that visits to a dentist are included in this 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. After dropping observations with missing values on any of the dependent or independent variables, the final sample comprises 36363 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 reforms occured 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 that people may need some time to adjust their behaviour.

The models that will be estimated in the following sections are 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). $x_{it}'\beta$ is a linear predictor that will be embedded in various alternative count data models, starting with the Poisson model. In some sense, this is a first explorative exercise. In particular, at this stage it is assumed that the reform effect is the same for all groups of the population. This assumption is unrealistic. For instance, for employed persons with high income, the co-payment is only a very small part of the cost of a prescription, the major part being the opportunity cost of time. Accodingly, the relative increase in the overall cost of prescriptions after the reform will be relatively modest, or even neglegible, for this group of people. One could allow for such effects by estimating the model for subgroups, or including interactive terms. The above pooled model gives thus the aggregate effect for the reforms averaged over the potential subgroups.

There are three general channels through which these variables affect the demand for doctor visits. These is the underlying health status, the budet constraint, and the preferences. The health status is poorly measured in the GSOEP. In particular, no details of 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 reponses "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 would want to account in any analysis for the possibility of aditional unobserved heterogeneity to capture the remaining health aspects, among others.

The budget constraint is determined by income and prices. Income is here measured through household equivalent income, where the OECD method 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 Sachverständigenrat. The main price variables are the opportunity costs of a visit to a doctor which, in turn depend on education levels and employment status. Indirectly, the insurance status enters here as well. Among the small fraction of privately insured persons in Germany, the self-employed make up the largest group. This is captured by the dummy variable. Unfortunately, there is no meaningful way to explore the influence of insurance status in further detail. For example, not all self-employed are privately insured, they can join the public scheme as well. Secondly, private insurance contracts come in all kind of varieties. Many contracts copy the provisions of the public provider, offering lower rates for singles (in particular young and males) but in general higher rates for families (in single earner households the public insurance covers all family members at no additional cost). One can also mention at this place that the number of uninsured persons in Germany is too small to be empirically relevant.

Finally, several of the variables affect more than one aspect at a time. Age, for instance, affects health as well as 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 effects. Rather, the focus lies entirely on the year dummies, the other right hand sight variables serving as controls to avoid the interfering effects of these variable on the changes in visits over time.

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

$$\mathrm{E}(y_i|\lambda_i) = \mathrm{Var}(y_i|\lambda_i) = \lambda_i \tag{2}$$

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 $i, i = 1, \ldots, N$ indexes observations in the sample. Let $y = (y_1, \ldots, y_N)'$ and $x = (x_1, \ldots, x_N)'$. Under random sampling

$$P(y|x) = \exp\left[-\sum_{i=1}^{N}\exp(x_i'\beta)\right]\prod_{i=1}^{N}\frac{[\exp(x_i'\beta)]^{y_i}}{y_i!} \tag{3}$$

In the current application, there are up to five observations for a given persons. One can incorporate this information into estimation by using e.g. a random effects panel model, or simply adjust the covariance matrix to allow for intra-cluster variation. Since efficient estimation is not the focus of this paper, such corrections are not undertaken at this preliminary stage.

The maximum likelihood estimator (MLE) of β is obtained by maximizing the log-likelihood function

$$\ln L(\beta|y) = \sum_{i=1}^{N}(-\exp(x_i'\beta) + y_i \ln x_i'\beta - \ln y_i!) \tag{4}$$

Maximization of (4), although requiring a numerical algorithm, does not pose great difficulties, as the log-likelihood is globally concave. The approximate distribution of the MLE for β can be derived as

$$\hat{\beta}_{ML} \overset{\text{app}}{\sim} N(\beta, [-\mathrm{E}H(\beta)]^{-1}) \tag{5}$$

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

$$\begin{aligned}\Delta\%_{(98,96)} &= \left[\frac{\mathrm{E}(y_{i,98}|x)}{\mathrm{E}(y_{i,96}|x)} - 1\right] \times 100 \\ &= [\exp(\beta_{98} - \beta_{96}) - 1] \times 100 \end{aligned} \tag{6}$$

5.1 Unobserved heterogeneity

The importance of unobserved heterogeneity was mentioned before. Write $\tilde{\lambda}_i = \lambda_i \nu_i$ where ν_i is a heterogeneity term with mean $\mathrm{E}(\nu_i|x_i) = 1$ and variance $\mathrm{Var}(\nu_i) = \sigma^2$.

The Poisson parameter is now a random variable, and estimation by maximum likelihood requires marginalization. For example, assume that ν_i has gamma density

$$g(\nu_i) = \frac{\beta^\alpha}{\Gamma(\alpha)}\nu^{\alpha-1}e^{-\nu_i\beta}$$

where $\alpha = \beta$. It follows that $\sigma_\nu^2 = \alpha^{-1}$. Moreover

$$\begin{aligned} f(y_i|x_i) &= \int_0^\infty f(y_i|x_i,\nu_i)g(\nu_i)\mathrm{d}\nu_i \\ &= \int_0^\infty \frac{e^{-\lambda_i\nu_i}(\lambda_i\nu_i)^{y_i}}{y_i!}\frac{\alpha^\alpha}{\Gamma(\alpha)}\nu_i^{\alpha-1}e^{-\nu_i\alpha}\mathrm{d}\nu_i \\ &= \frac{\Gamma(\alpha+y_i)}{\Gamma(\alpha)\Gamma(y_i+1)}\left(\frac{\alpha}{\lambda_i+\alpha}\right)^\alpha\left(\frac{\lambda_i}{\lambda_i+\alpha}\right)^{y_i} \end{aligned}$$

and

$$\mathrm{E}(y_i|x_i) = \lambda_i\,, \qquad \mathrm{Var}(y_i|x_i) = \lambda_i + \sigma^2\lambda_i^2$$

This is the well known negative binomial model (henceforth: Negbin). Relative to the Poisson model the variance of the dependent variable has increased - it is now a quadratic function of the conditional mean. This property is also referred to as "overdispersion". It can be shown that in this model, the Poisson estimator for β is still consistent, but the estimated standard errors using the conventional formula are downward biased. Under the gamma assumption, the Negbin model is asymptotically efficient.

While the assumption is convenient, it is also plausible? If one rewrites the heterogeneity term ν_i as $\nu_i = \exp(\varepsilon_i)$, it follows that $\tilde{\lambda}_i = \exp(x_i'\beta + \varepsilon_i)$. Thus, ε_i captures the combined effect of any omitted variables in the linear predictor. If these additive effects are independent, one can rely on a central limit theorem and reasonably assume that ε_i is approximately normally, and ν_i approximately log-normally distributed. Therefore

$$f(y_i|x_i) = \int_{-\infty}^\infty f(y_i|x_i,\varepsilon_i)g(\varepsilon_i)\mathrm{d}\varepsilon_i$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{\exp(-\exp(x_i'\beta+\varepsilon_i))\exp(x_i'\beta+\varepsilon_i)^{y_i}}{y_i!}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2}\mathrm{d}\varepsilon_i \\
&= \int_{-\infty}^{\infty} \frac{\exp(-\exp(x_i'\beta+\xi_i\sqrt{2}\sigma))\exp(x_i'\beta+\xi_i\sqrt{2}\sigma)^{y_i}}{\sqrt{\pi}y_i!}e^{-\xi_i^2}\mathrm{d}\xi_i
\end{aligned}
$$

This integral can be computed numerically using Gauss-Hermite approximation.

5.2 "Structural" models

The extensions of the Poisson model, while having superior statistical properties when correctly specified, are unlikely to offer much new insights. The main element is, as in the Poisson model, a single index log-linear conditional expectation function. Yet, more general models can be thought of, and each of them has interesting structural interpretations. Historically the most important and widely used departure from the simple log-linear structure is the hurdle model.

In this case, the demand decision is cast as a two-stage process

a) user yes/no

b) if user, then extent of use

This 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 referal decisions (co)influenced by the physician. Rightfully, Deb und Trivedi (1999) note that

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

Technically, the hurdle model combines a binary model with a truncated (1+) count data model. 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

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

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.

Recently, the hurdle model has come under criticism from a number of sides. Deb und Trivedi (1999) point out that there is 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) point out that several illness episodes are possible (i.e., one cannot identify a single binary contact decision). One rather should model jointly model episodes and contacts per episode 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 demanders. 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

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

where f_j is a Poisson- or Negbin distributed. If one mixes two component densities, the model has the same number of parameters as the hurdle model, and the two can be compared directly. Deb und Trivedi (1999) conclude from an application to health data that

> "The analysis ... provides strong evidence in favor of a latent class model as compared to a two-part 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

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

and

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

5.3 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? I think that such a conclusion would be premature. In support of this view, I present a new hurdle model that overcomes a limitation of the original approach, namely the independence assumption between hurdle and positive outcomes. In particular, let

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

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

$y_i | y_i > 0 \sim$ truncated Poisson(λ_i)

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

$$\varepsilon_i, u_i \sim \text{BVN}(0, \Sigma)$$

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

It follows 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

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

and

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

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.

5.4 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 (6). The same computation applies 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 the 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 on the hurdle and an effect on positive counts. These effects may be complementing or counteracting each other. Similarly, in the finite mixture model the reform may have acted 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

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

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

The results reported in this part are based on a previous set of estimates that covered the years 1988, 1995-1998 and a smaller set of explanatory variables. The newer results including the data for 1999 will be added later. There is no change in the substantive conclusions, although the inclusion of the 1999 observation is of substantial interest, as it can inform about the effects of the further reform that took effect on 1. January 1999. This later reform, initiated by the social-democratic government elected in mid-1998, reduced the co-payments by between DM 1 and DM 3, not exactly reverting to the pre-1997 levels but at least undoing some of the increase. Thus, by symmetry and under the assumption that the same mechanisms are at work, one would expect the number of doctor visits to increase again.

Table 1 gives the descriptive statistics for the earlier sample. The average number of doctor visits declined over time, from 2.8 per quarter in 1988 to 2.3 per quarter in 1998. However, the decrease was most pronounced between 1996 and 1998, where it amounted to about 0.13 per year. Also note that the annualized number of doctor visits in 1998 is remarkably close to the statistic found in the survey by Lauterbach et al. (2000). A substantial fraction of the population has zero doctor visits, with an increase of 3.7 percentage points between 1996 and 1998. These descriptive statistics give already an indication that the response elasticity at the 0/1+ margin might exceed the elasticity in the positive part. The other explanatory variables for the most part do not change much over time. The most notable exception are the health related variables that indicate an improvement in the general health status of the population between 1996 and 1998. The proportion of people in active sports increased from 25.6 to 30 percent. The proportion of people reporting good health increased from 56.4 to 58.7 percent, while the proportion of people reporting poor health decreased from 13.6 to 12.9 percent. Thus, it is very important to control for these factors through regression analysis to single out the contribution of the reforms, as opposed to improved health, on the decline in the number of doctor visits between 1996 and 1998.

Following the discussion in the previous section, as total of seven distinct models were estimated: Poisson, Negbin, Poisson-log-normal, Hurdle-Negbin, finite mixture Negbin with two components, Multi-episode model, and Probit-Poisson-log-normal with correlated errors. In principle, one could estimate the Hurdle and finite mixture models based on Poisson distributions as well. For simplicity, these models are excluded from consideration as the evidence against the Poisson assumption of equal conditional variance and mean is overwhelming anyway.

There are several ways to select between the models. Some of them are nested (such as the Poisson and the Negbin model), some of them are non-nested (such as the finite mixture and the hurdle Negbin model). Again for simplicity I adopt a crude model selection strategy, comparing unpenalized or penalized log likelihood values of the different models evaluated at the MLE. The results of are listed in Table 2. The Poisson model is clearly rejected against the models with unobserved heterogeneity. Among those, the Poisson-log-normal model is superior to the Negbin model, a common finding whenever this comparison is explicitly made (Winkelmann, 2000). Among the structural models, the multi-episode model performs worst. The findings confirm the Deb and Trivedi (1999) result that the finite mixture Negbin model outperforms the hurdle model. However, this does not necessarily mean that the hurdle model per-se is inappropriate. Indeed, 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, both in terms of AIC and BIC, and therefore is the preferred model overall.

Table 3 reports the parameter estimates for this model (The coefficients from the other models are 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 effects go in the same direction whenever the coefficients are of opposite sign. For example, being full-time employed increases the probability of having no visit, and reduces the expected number of visits, given that one visits at least once. In some cases, the effect goes in the opposite direction. For example, engaging in active sport reduces the probability of no visit, but reduces also the expected number of positive visits. Finally, we also note that the correlation coefficient is negative as expected, and significantly different from zero.

Next, consider 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 4. The estimates for the base model, with or without unobserved heterogeneity, are all in the same range, varying between 10.5 to 10.9 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 appraoch. 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 this hypothesis is correct, i.e., heavy users might have a lower response than occasional users, this should show up in the estimates of the structural models, that allow for differential responses in different parts of the distribution.

Table 4 confirms that such a differential effect is present indeed. This is most obvious from the hurdle estimates, regardless of whether the Negbin version or the probit-Poisson-log-normal version is considered. In each case, the reduction is greatest at the left margin of the distribution: the probability of being a user (for at least one time) decreased by 6 percent, whereas the expected number of visits, conditional on use, decreased only by 4.3 percent. This latter number is again remarkably similar to the Cologne estimate, and the resulting inference seems therefore robust, although the Cologne study obviously is incomplete, as it ignores the effect of the reform on the occasional user. The GSOEP based analysis provides a more complete picture.

One should also note that the different relative responses in the different parts of the distribution is not a statistical artefact, although one has to be careful in constructing the right benchmark. For instance, it is not the case that a x-percent reduction in the expected number of visits translates into an equally large relative reduction for specific conditional expectations (such as $\mathrm{E}(Y|Y > 0)$) or probabilities (such as $P(Y > 0)$). The proper benchmark is a count data model without hurdle, from which the two relevant quantities, the relative changes in $\mathrm{E}(Y|Y > 0)$ and $P(Y > 0)$ can be computed. For instance, based on the Negbin results, these are reductions of 7.5 and 3.3 percent, respectively. Thus, without allowing the flexibility of a hurdle model, one actually enforces the hurdle effect to be **less** important than the effect on positives, whereas the evidence from the unrestricted model suggests the opposite order of magnitude.

Finally, the results for the finite mixture and the multi-episode model confirm the qualitative conclusions, although quantitative details differ. The finite mixture model separates the population into two groups of roughly equal size, a high user group with a mean number of quarterly visits of 3.3, and a low user group with a mean number of 1.6 visits. Consistent with the above argument, the low user group shows the larger response to the reform. Similary, in the multi-episode model, the effect on the number of spells is much greater than the effect on the number of referrals. In each case, the two effects add up to an estimate of the 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? Or could it be that the timing of the reforms and the subsequent drop in visits is a pure coincidence? Identification is through variation in time. Thus, one needs to assume that other things didn't change as well, beyond the individual socio-economic characteristics controlled for in the regression. In a sense, it is hard to imagine what these other things should have been. It is unlikely that the underlying health status varied substantially between the two years beyond the controls. It is also unlikely that a health epidemic of major proportion hit in 1996 and 1997, but was absent in 1998.

Of course, even if one believes that the health care reform of 1997 was causally responsible for reducing the subsequent number of visits, it is still one additional step to say that the increase in co-payments caused the reduction in visits. Clearly, the reforms consisted of a bundle of measures, and the increased co-payments were only one of them. Another change of 1997 was the introduction of a "Praxis Budget". But this system actually made it profitable for a physician to see a patient once, but only once each quarter. Thus, supply side arguments cannot explain the large reduction in first time visits, combined with a smaller effect on repeated visits.

In theory, one can imagine better experiments. Ideally, one would think of a differences-in-differences estimator: use a group of people for which the reforms did not change anything, such as the uninsured or privately insured persons, as control group. But this approach has many problems as well. Some of them were mentioned before, including data limitations. In particular, only very few people are uninsured, and little is known about the nature of the private insurance contracts, whether they include co-payments or not, or whether any provisions changed following the changes to the public provisions (as is the case in many contracts. More fundamentally, the selection into insurance status for people who have a choice is highly endogenous, and very likely correlated with other unobserved factors that affect the number of doctor visits as well (adverse selection).

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

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.

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.

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