



Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits – Swiss evidence

Michael Gerfin^{a,b,*} and Martin Schellhorn^{b,c}

^a*Department of Economics, Universität Bern, Switzerland*

^b*IZA, Bonn, Germany*

^c*GSF, Institute of Health Economics and Health Care Policy, Munich, Germany*

Summary

We evaluate the effect of the size of deductibles in the basic health insurance in Switzerland on the probability of a doctor visit. We employ nonparametric bounding techniques to minimise statistical assumptions. In order to tighten the bounds we consider two further assumptions: mean independence of an instrument and monotone treatment response. Under these two assumption we are able to bound the causal effect of high deductibles compared to low deductibles below zero. We conclude that the difference in health care utilisation is partly due to a reduction of moral hazard effects. Copyright © 2006 John Wiley & Sons, Ltd.

JEL classification: C14; I19

Keywords treatment effects; nonparametric bounds; health insurance

Introduction

Between 1996 and 2004, health care expenditure in Switzerland increased on average by 5.5%/year. The average premium for the compulsory basic health insurance increased by even 6.2%/year in the same period. Currently, the average premium amounts to about 10% of the median of equivalent household income. For many families paying the health care insurance premium has become a major financial burden. Not surprisingly, this development has put health care costs on top of the health and social policy agenda. One instrument thought to reduce the growth in health care expenditure is increased cost sharing because both insurance theory predicts and empirical evidence confirms that there is a negative correlation between the degree of cost sharing and expendi-

ture. The Swiss health insurance law from 1996 provides the possibility of cost sharing in basic health insurance. The insured can choose between five possible deductible levels ranging from a minimum of 230 CHF year to a maximum of 1500 CHF. Choosing higher deductibles is combined with premium reductions.

However, as is well known the negative correlation between the degree of cost sharing and the demand for health care can be due to two reasons: selectivity or moral hazard. If only healthy people who do not go to the doctor anyway choose high deductibles the observed negative correlation is mainly due to selectivity. In this case, the size of the deductible has no impact on health care demand and health care costs. If on the other hand, people become more cost-aware and change their health care demand behaviour there is a causal effect of the deductibles, thus reducing the

*Correspondence to: Department of Economics, Universität Bern, Schanzeneckstrasse 1, PO Box 8573, CH 3001 Bern, Switzerland. E-mail: michael.gerfin@vwi.unibe.ch

moral hazard problem. Hence, in order to assess the cost-reducing potential of variable deductibles it is necessary to isolate the moral hazard (or treatment) effect.

The empirical evidence on the existence of moral hazard effects is mixed. Probably the most famous study is the RAND experiment [1]. In this experiment, people in several US states were randomly assigned to different insurance plans. These plans differed in the degree of cost sharing between patients and insurance company. The main result is that higher cost sharing reduces health care demand (measured by the number of doctor visits). Chiappori *et al.* [2] analyse a natural experiment in France where health insurance companies varied in their reaction to a change in health insurance regulations. They do not find a significant effect of increased cost sharing on health care demand measured by the number of visits to general practitioners. On the other hand, they estimate a significant effect on the number of home visits. Winkelmann [3] analyses the effect of introducing a fee for drug prescriptions in Germany in the year 1997. Using panel data, he estimates a significant reduction of the expected number of doctor visits due to this reform.

The empirical evidence for Switzerland is mixed as well. Gardiol *et al.* [4] and Werblow and Felder [5] estimate a significant negative effect of the size of the deductible on health care costs. On the other hand, Schellhorn [6,7] does not find any significant effects of the deductible on the number of doctor visits. The first two papers are based on data from a large insurance company, each covering only one canton (Vaud and Zurich, respectively), whereas Schellhorn uses data from the Swiss Health Survey, which covers all of Switzerland.

One problem of the existing empirical evidence for Switzerland is the underlying econometrics. The insurance company data used by Werblow and Felder [5] are very informative regarding health care costs. However, there is a lack of variables that may help to solve the selection problem. Not surprisingly, the paper is not very clear about exclusion restrictions. The estimation method is a parametric selection model. Given these problems it is not clear whether the estimated parameters are unbiased estimates of the causal effect of deductibles on health care costs. Gardiol *et al.* [4] use data from the same insurance company, but for a different canton. Their identification strategy is based on the assumption that there should be no moral hazard effect with

respect to hospital expenditure (based on empirical evidence that the price elasticity of hospital care is close to zero). Under this assumption, the estimated effect of the deductible on hospital costs is only due to selection. Given this estimate of the selection effect it is possible to decompose the estimated coefficients for outpatient care into the moral hazard and the selection effect. The estimation method is the parametric two-part model. Hence identification depends crucially both on the assumption regarding hospital expenditure and the correct specification of the statistical model. The papers by Schellhorn employ zero inflated count data regression models with endogenous regressors. In this case, the exclusion restrictions are clearly explained. However, estimating the model requires strong distributional assumptions that may be violated by the data at hand.

The aim of this paper is to analyse the effect of deductibles on the probability of going to the doctor with as little assumptions as possible. We employ the method of nonparametric bounds on treatment effects introduced by Manski [8] and further developed among others by Manski [9], Manski and Pepper [10] and Shaikh and Vytlačil [11]. The bounds give the maximum and minimum value the treatment effect can have given the imposed assumptions. Our primary focus is on the question whether it is possible to exclude zero from the bounds, which would imply that moral hazard reduction plays a role. Imposing two assumptions allows to bound the treatment effect away from zero. Hence our results indicate a negative causal effect of higher deductibles on the probability of going to a doctor.

This paper is organised as follows: the next section gives a brief overview of the Swiss health insurance system. The data used in the empirical analysis are described in the following section. The next section presents the estimation strategy and the penultimate section discusses the estimation results. The last section concludes.

The Swiss health insurance system

Since the reform of the health insurance law in 1996 a basic health insurance is mandatory in Switzerland. This basic health insurance is provided by competing private insurers and covers a widely defined set of medical services. The health insurance premium is heavily regulated. Premiums for adults

can vary between three regions in each canton but are not allowed to be related to risk-factors like age and sex or income. Every individual is insured with a separate contract. In 2002, the insured faced a choice between a minimal deductible of 230 CHF and higher deductibles of 400, 600, 1200 or 1500 CHF are optional. In order to make higher deductibles attractive they are combined with premium reductions. These reductions are regulated to be at most 8, 15, 30 and 40, respectively, of the premium with the minimal deductible. Therefore, the potential savings from choosing a higher deductible vary substantially between cantons and regions. We will use this fact to generate an instrumental variable in our estimation strategy.

The insured have free choice between the insurance companies with open enrolment and can change their insurer and their deductible at the beginning of each calendar year. In 2002, the average monthly insurance premium for the contract with the minimal deductible was 245 CHF and varied between 159 CHF in the canton of Appenzell and 364 CHF in the canton of Geneva. There is a large variance of premiums within cantons as well. For example, the average premium in the canton of Zurich was 249 CHF in 2002; the lowest premium was 170 CHF and the largest premium was 390 CHF for virtually the same insurance. The main reason for this huge spread of premiums is that insurance companies face different risk and cost structures. To reduce the amount of risk selection induced by this type of premium regulation all insurance companies have to participate in a risk-adjustment pool. Depending on the risk structure of their clients insurers pay or receive contributions from this pool. The difference in risk structures across insurance companies implies that premiums are indirectly not independent of risk.

There is a co-payment rate of 10% when costs exceed the chosen deductible, with a ceiling for co-payments of 600 SFr. year irrespective of the chosen deductible. Health insurance premiums and out-of-pocket payments can eat up a substantial part of household income especially in larger families. To ease the financial hardship associated with per capita premiums, government provides means – tested subsidies to low income residents. In 2001, roughly one-third of the insured were subsidised to some extent. Subsidies are paid if premiums exceed a certain percentage of household income.

Compulsory insurance in the fee-for-service sector offers direct access and free choice of

physician for outpatient care (general practitioners and specialists). Except for emergency cases hospitalisation requires referral by a physician. There is no choice of physician in the hospital. However, patients may freely choose among all hospitals which are included on the cantonal eligibility list.

Supplementary insurance covers additional treatments and check-ups, all drugs, extended home care, provides generally higher benefits and up to 100% universal coverage world wide. Most important it provides access to the private ward of all public and private hospitals in a one or two bedroom and free choice of physician in the hospital (medical or assistant medical director), depending on insurance package. A more detailed description of the Swiss health care system can be found e.g. in [12].

Data

The data come from the Swiss Health Survey (SHS) 2002 conducted by the Federal Bureau of Statistics. This cross-section contains a large number of socioeconomic characteristics and information on health status and health care utilisation in the past 12 months. In this paper, we use the subsample of those whose interview took place in the first quarter of 2002. These persons were re-interviewed 6 months after the first interview. This second interview gives information on health care demand since the first interview and detailed information on health care insurance. Persons were asked to provide the name of their insurance company and their chosen deductible. Since each company has to publish their premium structure this allows to compute the precise health insurance premium each person has to pay. It also allows to compute the premiums each person would face at each deductible level. This subsample provides all the relevant information: deductible choice at the beginning of 2002 and health care demand in the first 6 months of 2002. In addition, there are several important variables to control for selection: age, gender, subjective health status, previous demand behaviour and risk factors like weight or smoking.

This subsample of the SHS 2002 contains 3623 observations. After deleting individuals younger than 26 (these face different regulations concerning premiums and deductibles) and individuals in

alternative forms of insurance such as HMO and individuals with missing information on crucial variables the data contains 2714 observations. Table 1 provides descriptive statistics of some central variables separated by treatment status. In the remainder of this paper, we consider three possible treatment states denoted by D , where

$$D = \begin{cases} 0 & \text{if deductible} = 230 \\ 1 & \text{if deductible} \in \{400, 600\} \\ 2 & \text{if deductible} \in \{1200, 1500\} \end{cases}$$

This appears to be a natural combination of the five possible deductible levels. Further separation of the two combined deductible levels would lead to relatively small sample sizes in each state. A little less than half of the individuals have chosen the lowest deductible. Roughly one-third opted for the medium deductible, and the remaining 18% decided to take the largest deductible. Not surprisingly, this group is on average younger and feels more healthy. Men are more likely to opt for the largest deductible. People with chronic conditions have a much smaller probability of choosing the highest deductible. There are also noticeable differences with respect to education, income, previous health care demand, and the

premium at the lowest deductible. This premium is computed using the information on the insurance company and the region of residence. The higher the premium at the lowest deductible, the higher is the saving potential for choosing higher deductibles. As discussed in the previous section there is a large variation in the premium at the lowest deductible.

The outcome variable used in this paper is the indicator variable taking the value one if the person did see a doctor (either general practitioner or specialist) in the past 6 months. In the lowest deductible group 64% had at least one doctor visit, whereas only 40% in the highest deductible group went to the doctor. This is evidence for the well-known negative correlation between the degree of cost sharing and health care utilisation. The same is true if we consider number of doctor visits. People in the lowest deductible group had on average 2.9 doctor visits compared to 1.3 visits of people in the highest deductible group. However, comparing the average number of doctor visits conditional on having seen a doctor reveals a much less pronounced difference across groups. This finding corresponds to the well-known fact that the main choice people have is whether to see a doctor or not. Given they decided on a first

Table 1. Descriptive statistics

Variable	Deductible:230	Deductible:400/600	Deductible:1200/1500
Male	0.39	0.45 ^a	0.57 ^{a,b}
Age	52.76	51.34 ^a	45.63 ^{a,b}
Education: high	0.14	0.19 ^a	0.29 ^{a,b}
Education: average	0.67	0.68	0.64
Education: low	0.19	0.13 ^a	0.07 ^{a,b}
Subjective health: excellent	0.18	0.19	0.22
Subjective health: good	0.65	0.66	0.71 ^{a,b}
Subjective health: not good	0.17	0.16	0.07 ^{a,b}
Swiss	0.83	0.84	0.83
Heavy smoker	0.22	0.22	0.19
Chronic disease	0.33	0.30	0.13 ^{a,b}
Premium (at lowest deductible)	236	245 ^a	247 ^a
Income	2939	3173 ^a	3702 ^{a,b}
Used up deductible in previous year	0.45	0.40 ^a	0.16 ^{a,b}
Doctor visit in 2001: 1 = yes	0.78	0.74 ^a	0.59 ^{a,b}
Doctor visit in 2002: 1 = yes	0.64	0.58 ^a	0.40 ^{a,b}
Number of doctor visits 2002	2.88	2.52	1.32 ^{a,b}
Number of doctor visits 2002 (excluding zeros)	4.50	4.36	3.32 ^a
Number of observations	1255 (46.2%)	950 (35.0%)	509 (18.8%)

Source: Swiss Health Survey 2002, own calculations.

^aSignificantly different compared to mean in Deductible: 230 group.

^bSignificantly different compared to mean in Deductible: 400/600 group.

consultation, further doctor visits are to a large extent out of their control. For this reason we focus on the probability of going to a doctor as outcome variable.

Ordered probit estimates of the determinants of deductible choice clearly indicate that deductible choice is strongly influenced by income, previous health care demand (measured by the number of doctor visits, a dummy for hospital stay and a dummy for exhaustion of deductible in the previous year), the premium in case of the lowest deductible and regional differences. These estimation results are available on request.

The estimation strategy described below requires an instrumental variable. Possible instruments given the ordered probit results could be the regional indicators reflecting premium differences across cantons (as used in [6,7]) or the premium. We use the premium as an instrument in our analysis. The reasoning is that the premium does not affect the costs of going to a doctor. Recall that the premium for basic insurance is independent of risk factors such as age or previous illnesses. Hence we believe that the premium satisfies the weak assumptions required for an instrument as outlined below.

Estimation strategy

For each person i we observe (Y_i, D_i, X_i) , where Y_i is the outcome variable, D_i is the indicator of treatment, $D_i \in \{0, 1, 2\}$ as defined above, and X_i is a vector of covariates.

To simplify notation, the subscript i will be dropped in the following. The observed outcome Y is given by

$$Y = 1[D = t]Y_t, \quad t = 0, 1, 2 \quad (1)$$

where Y_t is the potential outcome if $D = t$ and $1[\cdot]$ is the indicator function equal to one if the expression in brackets is true. Hence each person has three potential outcomes (Y_0, Y_1 , and Y_2) but only the one corresponding to the treatment choice is observed.

The average treatment effect (ATE) is defined as $ATE^{t,s} = E[Y_t|X] - E[Y_s|X] = E[Y_t - Y_s|X]$, $t \neq s$

(2)

where $ATE^{t,s}$ measures the effect of treatment t relative to treatment s . In order to make the notation more compact we leave the conditioning

on X implicit in the following. As is well known, ATE is not identified by the data. This can be easily shown by writing

$$E[Y_t] = E[Y_t|D = t]P[D = t] + E[Y_t|D \neq t]P[D \neq t] \quad (3)$$

The data identify $E[Y_t|D = t]$, $P[D = t]$ and $P[D \neq t]$ but not $E[Y_t|D \neq t]$. The large and growing literature on estimating treatment effects is primarily concerned with solving this identification problem. There is a large variety of identification strategies: conditional independence assumptions, parametric selection models and estimators using instrumental variables. These assumptions can be very strong both in terms of functional form and distributional assumptions. Necessary exclusion restrictions for instrumental variables and selection models are often difficult to justify.

In this paper, we analyse the effect of deductible choice with as few assumptions as possible. We employ the method of nonparametric bounds on treatment effects introduced by Manski [8]. Further developments have been made by Manski [9], Manski and Pepper [10], and Shaik and Vytlačil [11], among others. Applications are still relatively rare; examples include [13–16].

The main idea of bounds can be illustrated as follows. Replace the unobserved $E[Y_t|D \neq t]$ in $E[Y_t] = E[Y_t|D = t]P[D = t] + E[Y_t|D \neq t]P[D \neq t]$ by its bounds Y^l and Y^u , where Y^l is the smallest value Y can take and Y^u is the largest value Y can take. The lower bound for $E[Y_t]$ is then $E[Y_t|D = t]P[D = t] + Y^lP[D \neq t]$ and the corresponding upper bound for $E[Y_t]$ is $E[Y_t|D = t]P[D = t] + Y^uP[D \neq t]$.

If $E[Y_t]$ is a probability we have: $Y^l = 0$ and $Y^u = 1$. Hence the bounds on $E[Y_t]$ are

$$E[Y|D = t]P[D = t] \leq E[Y_t] \leq E[Y|D = t]P[D = t] + P[D \neq t] \quad (4)$$

or compactly

$$B_t^l \leq E[Y_t] \leq B_t^u \quad (4')$$

with B_t^l as lower bound and B_t^u as upper bound on $E[Y_t]$. All terms in (4) are identified.

If $E[Y_t]$ is bounded then ATE is bounded as well. Assume that $t > s$. The lower bound of ATE is the difference between the lower bound of $E[Y_t]$ and the upper bound of $E[Y_s]$ (this is the smallest possible value of ATE), and the upper bound is the difference between the upper bound of $E[Y_t]$ and the lower bound of $E[Y_s]$ (this is the

largest possible value of ATE).

$$(B_t^l - B_s^u) \leq \text{ATE}^{t,s} \leq (B_t^u - B_s^l) \quad (5)$$

These are the ‘worst-case’ bounds of Manski [8]. In practice these bounds are too wide to be useful (and they always contain zero). The estimates in the present context are displayed in Table A1. In order to tighten the bounds we need to make further assumptions. In the following subsections we introduce possible further assumptions and combinations thereof.

Instrumental variable (IV)

Manski [8] analyses the case where an instrument Z is available that satisfies $E[Y_t|Z] = E[Y_t]$, i.e. Z satisfies mean independence. Manski has shown that in this case the bounds on $E[Y_t]$ change to

$$\begin{aligned} & \sup_z \{E[Y|D = t, Z = z]P[D = t|Z = z]\} \\ & \leq E[Y_t] \\ & \leq \inf_z \{E[Y|D = t, Z = z]P[D = t|Z = z] \\ & \quad + P[D \neq t|Z = z]\} \end{aligned} \quad (6)$$

If Z is discrete the supremum is the maximum and the infimum is the minimum over the possible values of Z . If the instrument has some identifying power, i.e. there is variation in $P[D = t|Z = z]$ over Z , the IV bounds on $E[Y_t]$ will be tighter than the worst case bounds. Consequently, the bounds on ATE will be tighter as well.

The mean independence assumption can be tested: it implies that for each treatment $E[Y|D = t, Z = z]$ is constant over Z (see [14]). The test involves estimating the worst case bounds conditional on Z and testing whether a constant value of $E[Y_t]$ is possible given these bounds. A constant value is possible if the maximum of the lower bounds is smaller than the minimum of the upper bounds, i.e. if the IV bounds on $E[Y_t]$ do not cross.

Using the premium at lowest deductible as instrument yields bounds on ATE that are too wide to be helpful. Table A1 displays the estimates. The expectations and probabilities conditional on Z in (6) are estimated by nonparametric kernel regressions. The mean independence assumption for Z cannot be rejected in any of 1000 bootstrap replications of the estimates of the IV bounds on $E[Y_t]$. These results are available on request.

Monotone instrumental variable (MIV)

The mean independence assumption for instrument Z is sometimes hard to justify. Therefore, Manski and Pepper [10] analyse the possibility of weaker assumptions for the instrument and propose MIV. The MIV assumption is that mean response varies weakly monotonically across the distribution of Z . Formally this can be stated as

$$\begin{aligned} z_1 \leq z \leq z_2 & \rightarrow E[Y_t|Z = z_1] \leq E[Y_t|Z = z] \\ & \leq E[Y_t|Z = z_2] \end{aligned} \quad (7)$$

Hence $E[Y_t|Z = z]$ is no smaller than the worst case lower bound on $E[Y_t|Z = z_1]$ and no larger than the worst case upper bound on $E[Y_t|Z = z_2]$. This holds for all $z_1 \leq z$ and all $z_2 \geq z$. There are no other restrictions on $E[Y_t|Z = z]$. Hence the bounds on $E[Y_t|Z = z]$ are

$$\begin{aligned} & \sup_{z_1 \leq z} \{E[Y|D = t, Z = z_1]P[D = t|Z = z_1]\} \\ & \leq E[Y_t|Z = z] \\ & \leq \inf_{z_2 \geq z} \{E[Y|D = t, Z = z_2]P[D = t|Z = z_2] \\ & \quad + P[D \neq t|Z = z_2]\} \end{aligned} \quad (8)$$

Manski and Pepper [10] show that the MIV bound on the marginal mean $E[Y_t]$ can easily be obtained from (8). It is given by

$$\begin{aligned} & \sum_{z \in Z} P[Z = z] \left\{ \sup_{z_1 \leq z} \{E[Y|D = t, Z = z_1] \times \right. \\ & \quad \left. P[D = t|Z = z_1]\} \right\} \\ & \leq E[Y_t] \\ & \leq \sum_{z \in Z} P[Z = z] \left\{ \inf_{z_2 \geq z} \{E[Y|D = t, Z = z_2] \times \right. \\ & \quad \left. P[D = t|Z = z_2] + P[D \neq t|Z = z_2]\} \right\} \end{aligned} \quad (9)$$

The MIV bounds are subsets of the corresponding worst case bounds and supersets of the corresponding IV bounds. The MIV bounds and the worst case bounds are the same if the worst case lower and upper bounds on $E[Y_t|Z = z]$ weakly increase with z . Then the MIV assumption has no identifying power. On the other hand, the MIV bounds and the IV bounds coincide if the worst case lower and upper bounds on $E[Y_t|Z = z]$ weakly decrease with z .

In the present case, the mean independence assumption for the premium at the lowest deductible may be questioned because premiums are not entirely risk independent. Companies with higher premiums have a larger share of high risk clients. Hence the probability of going to a doctor may not be mean independent of the premium but weakly monotonically increasing with the premium. It turns out that the MIV assumption has less identifying power than the IV assumption but still helps to shrink the worst case bound to some extent. The estimates for the MIV bounds can be found in Table A1.

Monotone treatment response

Manski [9] introduced the notion of monotone treatment response (MTR). MTR means that if $t > s$ then $Y_t \geq Y_s$ or $Y_t \leq Y_s$: In our case, MTR implies $Y_t \leq Y_s$, i.e. with increasing deductibles the probability of going to the doctor is not increasing. In other words, assuming MTR implies that we know the sign of the treatment effect. This is a strong assumption but it appears reasonable in our case; the question is whether the treatment effect is zero or negative, which is compatible with the MTR assumption. There are of course applications where it is impossible to make the MTR assumption. For example, the sign of the effect of active labour market programmes is not obvious *a priori*. Another example is the effect of treatment of emergency patients, which is analysed by Bhattacharya *et al.* [16].

Under MTR, the bounds are modified as follows. Note that MTR implies that $E[Y_t] \leq E[Y_s]$. Then

$$E[Y|D \geq t] P[D \geq t] \leq E[Y_t] \\ \leq E[Y|D \leq t] P[D \leq t] + P[D > t] \quad (10)$$

Comparing (10) with (4) shows that under MTR not only observations with $D=t$ are used to estimate the expectations in (7) but all observations are informative. Because MTR implies that $E[Y_t] \leq E[Y_s]$ for $t > s$ all observations with treatment equal to or larger than t can be used to compute the expectation in the lower bound on $E[Y_t]$. Compared to the worst case bounds on $E[Y_t]$, the MTR lower bound is increased by $E[Y|D > t] P[D > t]$. On the other hand, the upper bound on $E[Y_t]$ is reduced by $\{1 - E[Y|D \leq t]\} P[D \leq t]$.

Interestingly, there is a tight link to the bounds introduced by Shaikh and Vytlačil [11] who

assume that both treatment choice and outcome are binary and generated by a nonparametric threshold crossing model. These assumptions are stronger than those imposed by Manski but it is not necessary to know the sign of the treatment effect *a priori*. It can be shown that the Shaikh and Vytlačil bounds are identical to the MTR+IV bounds if the assumed sign of the treatment effect corresponds to the sign identified by the Shaikh and Vytlačil assumptions.

Combinations of assumptions

It is possible to combine the MTR assumption with the IV and the MIV assumption for the instrument Z . This simply involves substituting the worst case bounds in (6) and (9) by the MTR bounds conditional on Z .

We do not consider the monotone treatment selection assumption introduced by Manski and Pepper [10]. There are two main reasons for this. First, given that the main question in this paper is whether there is a moral hazard effect or not, the monotone treatment selection assumption does not buy us anything because it only affects the lower bound (in cases where treatment effects are negative). Second, Chevalier and Lanot [17] have recently shown that the monotone treatment selection assumption (and especially its combination with the MTR assumption) implies strong untestable restrictions on the correlations between treatment choices and outcomes.

Introducing covariates

Allowing for covariates, X changes the analysis in a trivial fashion; the same analysis then holds conditional on X . If X is discrete or can be discretised, the bounds can be estimated within the cells defined by the values of X . If X is continuous, the relevant conditional expectations have to be estimated by nonparametric smoothing estimators such as kernel regressions.

Results

This section presents the main results of the described bounding analysis. A full set of results is presented in Appendix A. Not surprisingly it

turns out that most of the bounding strategies yield bounds that are too wide to be informative. This is especially true for the worst case bounds, the IV and the MIV bounds but also for the MTR bounds. Table 2 displays the estimated bounds on $E[Y_t]$ and on ATE under the MTR + IV and under the MTR + MIV assumptions.

Inspecting the bounds on $E[Y_t]$ reveals that the bounds on $E[Y_0]$ are much tighter than the bounds on $E[Y_2]$. This implies that the assumptions have more identifying power if treatment is the lowest deductible. Imposing MIV instead of IV has also different effects across treatments. Whereas the bounds on $E[Y_2]$ do not differ much, the width of the bounds on $E[Y_0]$ almost doubles. This again reflects that the mean independence assumption has more identifying power for $t=0$.

If one is willing to accept the MTR + IV assumptions, the results indicate a negative treatment effect comparing the high deductible with the low deductible ($ATE^{2,0}$) and for the comparison of the medium deductible with the low deductible ($ATE^{1,0}$). This second effect is barely below zero, however. Taking sampling variation into account by bootstrapped confidence bands for the estimated bounds (given in brackets below the estimates of the bounds) we cannot reject the hypothesis that the upper bound on $ATE^{1,0}$ is zero. The upper bound on $ATE^{2,0}$, on the other hand, is significantly different from zero. This implies that if we accept the MTR + IV assumption, we can say that at least 0.08 of the difference

in the probabilities of going to a doctor is due to a reduced moral hazard effect. Given that the observed difference is 0.24 at least one-third of the difference can be attributed to a reduced moral hazard effect.

Replacing the mean independence by the weaker MIV assumption widens the bounds on the treatment effects considerably. The upper bound on $ATE^{2,0}$ is now only -0.04 , but still significantly different from zero. Hence even under the weaker MIV assumption, we can still say that there is a moral hazard reduction of at least -0.04 .

Table 3 presents results conditional on covariates. The considered covariates are 'No Chronic Condition', gender, and two age groups (below and above 50). Also displayed are the treatment effects under the exogenous selection assumption. Overall, results do not vary much across the covariate values under the MTR + IV assumption and all lie within the estimated confidence band reported in Table 2. However, weakening the mean independence assumption has different effects across subgroups. For some subgroups the upper bound does not change much ("no chronic condition" and "age 50+") and remains significantly smaller than zero. In the other cases, the upper bound is increased substantially by weakening the IV assumption and is not longer significantly different from zero. A more detailed analysis of subgroup-specific effects is left for future research.

Table 2. Estimated bounds on $E[Y_t]$ and ATE under MTR + IV and MTR + MIV, Z =premium

	MTR + IV		MTR + MIV	
	Lower bound	Upper bound	Lower bound	Upper bound
$E[Y_0]$	0.61	0.73	0.59	0.82
$E[Y_1]$	0.33	0.59	0.29	0.66
$E[Y_2]$	0.09	0.52	0.08	0.55
$E[Y_1]-E[Y_0]=ATE^{1,0}$	-0.40	-0.01	-0.53	0
	[-0.45, -0.37]	[-0.05, 0]	[-0.57, -0.50]	
$E[Y_2]-E[Y_1]=ATE^{2,1}$	-0.50	0	-0.58	0
	[-0.54, -0.46]		[-0.64, -0.55]	
$E[Y_2]-E[Y_0]=ATE^{2,0}$	-0.64	-0.08	-0.74	-0.04
	[-0.69, -0.59]	[-0.11, -0.05]	[-0.78, -0.71]	[-0.07, -0.01]

Source: Swiss Health Survey 2002, own calculations.

$ATE^{t,s}$ measures the effect of treatment t relative to treatment s .

Under MTR, the upper bounds cannot be larger than 0 by definition. If the numerical value exceeds 0, the upper bound is set to 0 (see [10] for a discussion).

The figures in brackets are 90% confidence bands for the estimated bounds based on 1000 bootstrap replications.

Table 3. Estimated upper bounds ATE under MTR + IV and MTR + MIV, Z = premium, with Covariates

	Exogenous Selection			MTR + IV, upper bound			MTR + MIV, upper bound		
	ATE ^{0,1}	ATE ^{1,2}	ATE ^{0,2}	ATE ^{0,1}	ATE ^{1,2}	ATE ^{0,2}	ATE ^{0,1}	ATE ^{1,2}	ATE ^{0,2}
No chronic condition	-0.05	-0.13	-0.19	-0.02	0	-0.06	0	0	-0.04
Women	-0.02	-0.20	-0.22	0	0	-0.06	0	0	(-0.03)
Men	-0.11	-0.14	-0.25	0	0	-0.10	0	0	0
Age									
26-50	-0.13	-0.11	-0.18	0	0	-0.10	0	0	(-0.03)
50+	-0.04	-0.22	-0.26	0	0	-0.07	-0.07	0	-0.06

Source: Swiss Health Survey 2002, own calculations.

ATE^{t,s} measures the effect of treatment *t* relative to treatment *s*.

Figures in parentheses are not significantly different from zero (90% level).

Conclusions

There is an ongoing debate in health economics whether cost sharing in health insurance changes the behaviour of patients. The well documented negative correlation between the degree of cost sharing and health care costs can be attributed to changes in behaviour (reduction of moral hazard) or selection (more healthy people select higher cost sharing). Previous empirical evidence for Switzerland, which was obtained based on strong identification assumptions, is mixed. In this paper, we analyse the effect of deductibles with as few statistical assumptions as possible. We employ the nonparametric bounding techniques introduced by Manski [8] and further developed by Manski [9], and Manski and Pepper [10], among others. Treatment is ordered in our application. We consider three possible treatment states: low deductible, medium deductible and high deductible. The outcome variable is the probability of going to the doctor.

Our empirical analysis shows that we need some assumptions to tighten the bounds; the no-assumption bounds are too wide to be useful. We consider two further assumptions: treatment response is mean independent of an instrument, and treatment response is monotone (i.e. we assume the sign of the treatment effect to be known). The mean independence assumption is also weakened by assuming that mean response varies weakly monotonically with the instrument. Under the first two assumptions, we estimate bounds for the treatment effect of the high deductible compared to the low deductible that are below zero. Hence given these two assumptions, we conclude that there is a negative treatment effect (i.e. there is a change in behaviour) of at least -0.08 . Given that

the observed difference is -0.24 at least one-third of the difference in the probabilities of going to a doctor can be attributed to a reduced moral hazard effect. Weakening the mean independence assumption reduces the treatment effect to -0.04 , but it remains different from zero. This finding is in contrast to previous empirical analyses based on the same data.

The research presented in this paper can be extended in several directions. Methodologically, we plan to apply the bounding techniques proposed by Shaikh and Vytlacil [11]. To do so, we need to concentrate on binary treatments (as opposed to three treatments in the present application). Given that there appears to be mainly an effect of the high deductible compared to the low deductible, the reduction to a binary treatment appears to be meaningful. Finally, we can attempt to expand the Shaikh and Vytlacil analysis to ordered treatments.

In addition, one may attempt to point identify the treatment effect with as little assumptions as possible. Nonparametric matching estimates might be a fruitful approach. Alternatively, semiparametric estimators such as the ones proposed by Lewbel [18] or Abadie [19] may be considered.

Acknowledgements

Michael Gerfin thanks the Swiss Science Foundation for the financial support (Grant 610-062887). We thank the Federal Office of Statistics for permission to use the data (Swiss Health Survey) and Robert E. Leu for making the second wave of the data possible. This additional interview was financed by the National Research Program 45, Future of the Social State. Elizabeth Savage and participants of the 14th European Workshop on Health Economics and Econometrics in Dublin,

of the Annual Meeting of the Verein für Socialpolitik in Bonn and of the Econometrics brown bag seminar in Mannheim and two anonymous referees provided helpful comments. Donna Ginther and Libertad Gonzalez helped by sharing some computer programmes. The usual disclaimer applies.

Appendix A

The estimated bounds on ATE under different assumptions is given in Table A1.

Table A1. Estimated bounds on ATE under different assumptions

ATE ^{0,1}		ATE ^{1,2}		ATE ^{0,2}	
B ^l	B ^u	B ^l	B ^u	B ^l	B ^u
		Worst case			
-0.63	0.56	-0.78	0.68	-0.76	0.59
		IV			
-0.49	0.48	-0.74	0.61	-0.64	0.50
		MIV			
-0.60	0.49	-0.75	0.65	-0.74	0.52
		MTR			
-0.56	0	-0.61	0	-0.76	0
		MTR + IV			
-0.40	-0.01	-0.50	0	-0.64	-0.08
		MTR + MIV			
-0.53	0	-0.58	0	-0.74	-0.04

Source: Swiss Health Survey 2002, own calculations.

ATE^{*t*,*s*} measures the effect of treatment *t* relative to treatment *s*.

References

- Manning WG, Newhouse JP, Duan N, Keeler EB, Leibowitz A, Marquis MS. Health insurance and the demand for medical care: evidence from a randomized experiment. *Am Econ Rev* 1987; **77**: 251–277.
- Chiappori P-A, Durand F, Geoffard P-Y. Moral hazard and the demand for physician services: first lessons from a french natural experiment. *Eur Econ Rev* 1998; **42**: 499–511.
- Winkelmann R. Health care reform, the number of doctor visits – an econometric analysis. *J Appl Econ* 2004; **19**: 455–472.
- Gardiol L, Geoffard P, Grandchamp C. *Separating selection and incentive effects: an econometric study of swiss insurance claims data*. Working Paper 2003-27, DELTA, Paris, 2003.
- Werblow A, Felder S. Der Einfluss von freiwilligen Selbstbehalten in der Gesetzlichen Krankenversicherung. Evidenz aus der Schweiz. *Schmollers Jahrb* 2003; **123**: 235–264.
- Schellhorn M. The effect of variable health insurance deductibles on the demand for physician visits. *Health Econ* 2001; **10**: 441–456.
- Schellhorn M. *Wählbare Selbstbehalte in der Krankenversicherung in der Schweiz: Nachfragesteuerung oder Selektion?* Discussion Paper 04/02, Department of Economics, University of Bern, 2004.
- Manski CF. Nonparametric bounds on treatment effects. *Am Econ Rev, Pap Proc* 1990; **80**: 319–323.
- Manski CF. Monotone treatment response. *Econometrica* 1997; **65**: 1311–1334.
- Manski CF, Pepper J. Monotone instrumental variables: with an application to the returns to schooling. *Econometrica* 2000; **68**: 997–1010.
- Shaikh A, Vytlacil E. *Threshold crossing models and bounds on treatment effects: a nonparametric analysis*. NBER Technical Working Paper 307, 2005.
- Leu RE, Schellhorn M. *The evolution of income-related inequalities in health care utilization in Switzerland over time*. Discussion Paper 04/13, Department of Economics, University of Bern, 2004.
- Pepper J. The intergenerational transmission of welfare receipt: a nonparametric bounds analysis. *Rev Econ Stat* 2000; **82**: 472–488.
- Ginther DK. Alternative estimates of the effect of schooling on earnings. *Rev Econ Stat* 2000; **82**: 103–116.
- Gonzalez L. Nonparametric bounds on the returns to language skills. *J Appl Econ* 2005; **20**: 771–795.
- Bhattacharya J, Shaikh A, Vytlacil E. *Treatment effect bounds: an application to Swan-Ganz catheterization*. NBER Working Paper 11263, 2005.
- Chevalier A, Lanot G. Monotonicity and the Roy model. *Manchester Sch* 2004; **72**: 560–567.
- Lewbel A. *Simple estimators for endogenous selection models*. Mimeo, Boston College, 2005.
- Abadie A. Semiparametric instrumental variable estimation of treatment response models. *J Econometrics* 2003; **113**: 231–263.