# Laboratoire d'Economie de Dauphine

WP n°1/2020

Document de travail

**Supplementary Health Insurance and medical prices: an inflationary spiral ?**

Mathilde Péron

Brigitte Dormont

Pôle Laboratoire d'Economie et de Gestion des Organisations de Santé (LEGOS)
Place du Maréchal de Lattre de Tassigny 75775 Paris Cedex 16
Tél (33) 01 44 05 44 46 Fax (33) 01 44 05 40 67
Site : www.legos.daupine.fr

# Supplementary Health Insurance and medical prices: an inflationary spiral?

**Mathilde Péron**
University of York

**Brigitte Dormont**
PSL, Université Paris Dauphine

January 27, 2020

Acknowledgments: We gratefully acknowledge financial support from the Health Chair; a joint initiative by PSL, Université Paris-Dauphine, ENSAE, MGEN and ISTYA under the aegis of the Fondation du Risque (FDR)

E-mail: mathilde.peron@york.ac.uk (Péron); brigitte.dormont@dauphine.fr (Dormont)

**Abstract**

We investigate the existence of an inflationary spiral on medical prices due to supplementary health insurance (SHI), focusing on the demand for specialists who balance bill their patients, i.e. charge them more than the regulated fee. We ask three questions. First, is balance billing consumption a motive to buy SHI? Second, does SHI have a positive impact on the use of physicians who charge balance billing? Third, is the behavioral reaction to balance billing coverage correlated with self selection into SHI? We consider a structural model that links demand for balance billing, decision to take out balance billing coverage and the reaction to better coverage. From this model, we derive marginal treatment effects to estimate the causal impact of SHI coverage on balance billing consumption on a French database of 58,519 individuals observed in 2012. We are able to identify adverse selection, to estimate heterogeneous moral hazard, driven by both observed and unobserved characteristics, and to test for selection on return. We find that SHI can feed an inflationary spiral on medical prices. Specifically, we find that observable characteristics such as specialists availability are associated with a higher use of balance billing, are also determinants of balance billing coverage. Coupled with moral hazard, this form of adverse selection is likely to encourage the rise in prices. Furthermore, we show that selection on return reinforces the inflationary spiral. Individuals with observed and unobserved characteristics that make them more likely to subscribe to SHI are also those who exhibit stronger moral hazard, i. e. a larger increase in balance billing per visit. The role of income is particularly interesting. Without coverage, the poor consume less balance billing than the rich but increase their consumption more sharply once covered. They are also more likely to take out balance billing coverage. (JEL: I13; I18; C23)

Keywords: Health insurance, selection, moral hazard, marginal treatment effects, balance billing.

---

## 1. Introduction

In most European countries, individuals have the option to buy private coverage that supplements social health insurance. On the grounds that social coverage is quite generous, the quantitative importance of supplementary health insurance remains rather limited. In 2015, the coverage of health expenditure by voluntary private insurance was 0.3% in Norway, 3% in Germany, 5% in the United Kingdom, 7% in the Netherlands, 7.7% in Switzerland and 14% in France (OECD 2015). However, considering the growing pressures in favor of reducing public expenditures, an expansion of supplementary health insurance is expected.

Supplementary health insurance is definitely not a neutral part of healthcare systems, just sitting alongside social health insurance. It interferes with its performance regarding efficiency and equity in access to care. Understanding the nature and extent of these interferences is crucial for the regulators, in order to appropriately design the scope of social health insurance. Our purpose is to provide evidence on the influence of supplementary health insurance on medical prices. Precisely, the aim of this paper is to investigate the following question for France: is there an inflationary spiral regarding doctors' fees, due to the coverage of balance billing by supplementary health insurance?

The issue of physician payments is emblematic of negative interferences between social insurance and private supplementary insurance. Indeed, social health insurance coverage is only effective if the regulator can control medical prices. Social health insurance generally sets regulated fees that are used as reference prices to calculate reimbursements. For cost containment purposes, the regulator cannot allow fees to increase as much as demanded by doctors. In this context, many countries authorize some doctors to charge balance billing, that is to charge more than the regulated fee. For policy makers, balance billing has the advantage of permitting an increase in physicians' earnings with no additional burden on social health insurance.

We observe in many countries that supplementary health insurance is then used to provide coverage of balance billing and compensate the deterioration of social coverage (Sagan and Thomson 2016). Such a compromise raises concerns about the efficiency of the health care system since supplementary coverage can favor demand for expensive physicians who can increase their fees in turn. This might encourage demand for more supplementary coverage and hence fuel an inflationary spiral.

To limit harmful consequences on medical prices, regulators have introduced direct restrictions, for instance with a cap on balance-billing, as in the USA for Medicare (McKnight 2007), or taxes to discourage too generous coverage of balance billing by private insurers, as in France in 2014.

To generate an inflationary spiral, supplementary health insurance must be jointly involved in two mechanisms: (i) it should increase demand for doctors who charge balance billing and (ii) access to doctors who charge balance billing should be a motive for the purchase of supplementary health insurance. In a previous paper we have provided empirical evidence for mechanism (i), showing that in France an increase in supplementary coverage causes a rise in the demand for specialists who charge high fees (Dormont and Péron 2016). To prove the existence of an inflationary spiral, however, we must also show that balance billing triggers the purchase of supplementary insurance (mechanism (ii)). For that purpose, we need to go beyond a simple estimation of the demand for insurance. Indeed, the fact that individuals self-select into the voluntary purchase of supplementary insurance entails two selection effects, referred hereafter as "classical adverse selection" and "selection on return". Classical adverse selection is linked to heterogeneity as regards demand for balance billing. In that case, some individuals consume more balance billing than others and are also more likely to buy insurance in order to reduce the financial risk associated with their health care expenditures. Selection on return appears when there is heterogeneity as regards the behavioral response to health insurance. If so, some individuals might be more prone to buy insurance because they expect a particular increase in their use of balance billing due to better coverage.

In the literature on insurance, selection on return is also called “selection on moral hazard”: individuals who are responding more to insurance coverage are also more likely to enroll Einav et al. (2013). Equally, the econometric literature on treatment effect refers to “essential heterogeneity” when individuals who benefit more from a treatment are also more likely to take it. Estimating a model with selection on return is methodologically challenging. Indeed, the estimated relationship between demand for supplementary coverage and balance billing consumption is influenced by endogenous selection and heterogeneity in reaction to coverage. Nevertheless, we think it is a key mechanism, nested at the heart of the inflationary spiral we want to examine. In the literature that evaluates the impact of health insurance on health care use, results based on randomization such as the RAND Health Insurance Experiment (Manning et al. 1987; Newhouse 1993), or quasi-natural experiments such as the Oregon Experiment (Finkelstein et al. 2012), are usually considered as a gold standard. These evaluations rely on the estimation of reduced forms, with methods designed to eliminate selection biases. In our case, insurance is voluntary, and we are precisely interested in the relation between self-selection (decision to buy better coverage), and reaction to better coverage. For this reason, we use a structural approach where selection mechanisms are explicitly specified and estimated.

We set down a structural model that links the demand for balance billing, the decision to take out more coverage for balance billing and the behavioral response to better coverage. Precisely, we consider a Roy model, where two different equations explain the use of balance billing, depending on whether the individual is covered for it or not. Our model specifies the choice to purchase coverage for balance billing and allows for possible correlation between the decision to buy better coverage and the impact of better coverage on consumption (essential heterogeneity). Following Heckman and Vytlacil (2001) and Heckman et al. (2006), we estimate marginal treatment effects (MTE) estimators, which capture the impact of a treatment likely to vary within a population in correlation with observed and unobserved characteristics, when individuals select themselves into treatment. This allows us to examine the decision to buy coverage for balance billing and to identify classical adverse selection, as well as selection on return. Our database stems from administrative data provided by a French supplementary insurer, the Mutuelle Générale de l'Education Nationale (MGEN). We use data which provide for 58,519 individuals information on health care claims and reimbursements for social and supplementary health insurance coverage in 2012. We are able to observe two groups of individuals: a control group (53,456) that is never covered for balance billing and a treatment group (5,063) that was not covered for balance billing in 2010, decided to buy better coverage in 2011 and thus have benefitted from balance billing coverage over 2012.

We consider three questions. First, is balance billing consumption a motive to buy supplementary health insurance? Second, does supplementary coverage have a positive impact on the use of physicians who charge balance billing? Third, is the behavioral reaction to balance billing coverage correlated with self selection into supplementary health insurance?

Our findings show that supplementary insurance can feed an inflationary spiral on medical prices: balance billing is a determinant of the purchase of supplementary health insurance and supplementary coverage has a positive effect on balance billing consumption. In addition, this inflationary spiral is reinforced by selection on return. We find that reaction to better coverage (moral hazard) is significantly heterogeneous, and that stronger reaction to coverage is generally associated with a higher propensity to take out supplementary insurance. This is apparent through the estimated impact of some observable variables, and through our marginal treatment effects estimates which refer to unobservable heterogeneity. Interestingly, we find cases where there is no classical adverse selection but only selection on return. The role of income is critical in this. Low income individuals appear to consume less balance billing than others when they are not covered. However, they react more strongly to supplementary health insurance than the rich and are also more likely to demand supplementary coverage.

This paper contributes to the rather seldom literature on the influence of supplementary health insurance on the performance of healthcare systems (Stabile and Townsend 2014), by drawing attention on the impact on medical prices of a supplementary coverage which is subscribed on a voluntary basis. Our paper also adds results to the emerging empirical literature that considers heterogeneity in treatment effects. In particular, we contribute to the parsimonious literature providing evidence on selection on return. As concerns health insurance, we know only two papers examining selection on return (Einav et al. 2013; Kowalski 2018).

In our case, providing evidence on selection on return is of major importance to disentangle the mechanisms underlying the inflationist spiral due to supplementary coverage. Selection on return reinforces demand for expensive physicians by the first subscribers, triggering the inflationist impact of supplementary insurance. To our knowledge, our paper is the first one to use the Roy model in a structural approach as developed by Heckman and Vytlacil (2005) to understand demand for, and reaction to, health care coverage. This model is particularly relevant in our context, where supplementary insurance is voluntary, as it is the case in many European countries. We further contribute to the debate on the choice between reduced form versus structural approach in empirical work, by illustrating how both approaches can be fruitful to answer a specific research question (Einav and Finkelstein 2018). On the one hand, it was important to eliminate any selection effect in a first contribution by using a reduced form approach to evaluate the causal impact of supplementary coverage on the use of expensive physicians (Dormont and Péron 2016). On the other hand, using a structural approach, as we do in this paper, is also crucial to demonstrate the existence of an inflationary spiral. Indeed, for this purpose, we must understand self-selection behaviors in supplementary coverage and estimate selection on return effects. Finally, one of our results, though not central to our problematic, contributes to a very controversial literature on the nature of moral hazard (Pauly 1968; Nyman 1999). Is moral hazard only a substitution effect due to the fact that insurance coverage reduces the relative price of healthcare, or does it also entail an income effect? Our estimates show that low income people are more likely to purchase supplementary coverage and have a stronger reaction to it. Assuming that preferences are homogeneous across individuals, this can be seen as an evidence that income effects are at stake in reaction to coverage and that access to care motive is present in the purchase of supplementary insurance.

The paper is organized as follows. Section 2 summarizes the economic and econometric literature on selection on returns. Section 3 gives insights on the regulatory context of balance billing in France and discusses potential sources of heterogeneity in the response to balance billing coverage. We describe our data and empirical strategy in section 4 before going through our empirical specification in more details in section 5. Finally, we present and discuss our results in section 6. Section 7 concludes.

## 2. Literature review

Empirical contributions that aim to estimate the causal effect of insurance on healthcare use acknowledge that there is heterogeneity in the demand for healthcare and control for classical adverse selection (Cameron et al. 1988; Coulson et al. 1995; Holly et al. 1998; Vera-Hernández 1999; Schellhorn 2001; Buchmueller and Couffinhal 2004; Jones et al. 2006). In this literature, the response to health insurance is often assumed to be homogeneous across individuals and moral hazard is estimated through a single parameter associated with the price elasticity of demand for healthcare. In this framework, studies based on randomization such as the RAND Health Insurance Experiment (Manning et al. 1987; Newhouse 1993), or quasi-natural experiments (Chiappori et al. 1998; Finkelstein et al. 2012) are elegant solutions to eliminate selection bias from the estimation of the impact of insurance on care use. But randomization is not necessarily of interest when insurance is voluntary. Because these analyses remove the endogenous choice component from the equation, they are not able to estimate a potential selection on moral hazard and predict the impact of a voluntary insurance on healthcare consumption. The question of selection on moral hazard has been addressed empirically by Einav et al. (2013). They use individual-level panel data from an American firm where employees can choose among different level of coverage. They find heterogeneity on moral hazard together with selection on moral hazard: individuals who buy more comprehensive coverage exhibit greater moral hazard.

In the econometric literature, selection on moral hazard is more generally known as selection on return or essential heterogeneity. Assuming that there is individual heterogeneity in treatment effects, essential heterogeneity arises when individuals decide to take the treatment in relation with their expected response to the treatment. Heckman and Vytlacil (2007) show that in the presence of essential heterogeneity, instrumental variable (IV) methods, which are frequently used to control for endogenous selection, do not estimate an average treatment effect (ATE), nor a treatment effect on treated. Indeed, IV methods only estimate a local average treatment effect (LATE), specific to individuals who would react to the shock induced by the instrument. In the presence of essential heterogeneity, this local effect cannot be extended to the average population. Another consequence is that different instruments are likely to give different estimates of the treatment effect because they rely on compliers with different reactions to the treatment. Beyond the objective to estimate unbiased causal effects, we can question the relevance of estimating an ATE in a context where individuals can decide to participate or not in the treatment. Indeed, in this case, we pay more attention to the treatment effect of those who are more likely to take the treatment rather than to the average effect on the whole population. MTE estimators have been developed to capture the impact of a treatment likely to vary within a population in correlation with observed and unobserved characteristics, in a setting where individuals select themselves into treatment. First defined by Bjorklund and Moffitt (1987), MTE have been comprehensively described by Heckman and Vytlacil (2001) and Heckman et al. (2006). Empirically, MTE have been used to capture returns in education (Carneiro et al. 2011), breast cancer treatment effects (Basu et al. 2007), the effect of family size on children's outcome (Brinch et al. 2017) or the marginal returns of universal childcare (Cornelissen et al. 2017). Recently, Kowalski (2018) uses MTE in an experimental framework to assess the external validity of the Oregon health insurance experiment.

## 3. Why should reactions to balance billing coverage be heterogeneous?

In this section we briefly present the regulatory context around balance billing in France before discussing several potential sources of heterogeneity of response to balance billing coverage.

### *3.1. The French regulation of ambulatory care*

Individuals can take out supplementary health insurance (SHI) to enhance their coverage and limit out-of-pocket expenditure, either voluntary in the individual market or through their employer. For ambulatory care, the National Health Insurance (NHI) sets a regulated price and reimburses only a fraction of it (70% of the regulated fee for a visit to a specialist). On top of NHI copayments, patients may also have to pay balance billing. In France, ambulatory care is mostly provided by self-employed physicians paid on a fee-for-service basis and patients have the choice to visit two types of physicians. 'Sector 1' (S1) physicians are mandated to charge the NHI regulated fee (€23 in 2012 for a routine visit) whereas 'sector 2' (S2) physicians are allowed to balance bill, i.e. charge more than the regulated price. Access to S2 has been closed to most GPs since 1990, so most of them are in S1: 87% in 2012. Hence, balance billing concerns mostly specialists. National figures show a steady increase of the total amount of balance billing for specialists in France, exhibiting a striking factor 2.6 increase between 2000 (1.02 billion euros) and 2016 (2.69 billion euros) (Figure 1). Balance billing have contributed for about a third of the 66% increase in total fees (regulated plus balance billing) for all specialists between 2000 and 2016[1]. At the time of our study, in 2012, the average proportion of specialists operating in S2 amounts to 42% and balance billing adds about 35% to their annual earnings. This proportion varies strongly across specialties. For instance, the proportion of specialists operating in S2 is 19% for cardiologists, 73% for surgeons and 53% for ophthalmologists.

### *3.2. Sources of sensitivity to balance billing coverage*

Patients are free to choose to visit a specialist of sector 1 or 2. Because of balance billing, a visit to a S2 specialist is more expensive than a visit to a S1 specialist, but the relative price between sector 2 and 1 can be reduced by the coverage offered by a supplementary health insurance (SHI). In France, almost 95% of the population is covered by a SHI contract and all SHI contracts cover NHI copayments (30% for a

---

1. French national statistics on specialists' fees: SNIR and AMOS from 2015

consultation). Still, there are important differences between SHI contracts in terms of balance billing coverage: in polls, only 48.5% of SHI policyholders state that they are well covered against balance billing (Célant et al. 2014). Therefore, besides the magnitude of the balance billing charged by the physician, the relative price of a S2 specialist also depends on the extent of balance billing coverage provided by SHI. One reason for heterogeneous reactions to coverage is simply that its extent varies across individuals.

For a given coverage, i.e. for a given relative price of S2 visit, the demand for S2 specialist will depend on the degree of substitutability between care services provided by S2 versus S1 specialists. Specialists of the two sectors are supposed to provide the same medical service. However, they are not perfect substitutes because of differences in geographical location, in waiting time and in perceived quality of care. This will induce heterogeneity between individuals in the elasticity of the demand for S2 medical services with respect to their relative price. Time and monetary costs of access to care, as well as waiting time, are determined by the local availability of S1 and S2 specialists.

In France, the number of specialists in S1 and S2 for 100,000 people varies dramatically between regions and so does the share of consultations of S2 specialists. Geographical patterns displayed in Figure 2 show a correlation between a limited availability of S1 specialists and use of S2 specialists. In Paris and its surroundings for instance, there are few S1 specialists, and this induces large waiting times for an appointment and transportation time costs to get to the doctor's office. Depending on their preferences, patients might be willing to pay balance billing to avoid waiting time and travel. This creates differences between individuals regarding the substitutability between S1 and S2 specialists, resulting in heterogeneity in their reaction to balance billing coverage and their propensity to buy coverage[2].

Another source of heterogeneity is the perception of the quality of care provided by S2 physicians. In France, S1 and S2 specialists are supposed to provide the same medical service and there is no public information released on the quality of care provided by doctors. Yet, because access to S2 is restricted to physicians who have been practicing in a qualifying hospital setting, their consultations can be associated by patients with a better quality. This belief can be reinforced by their higher price, which acts as a signal in a context of asymmetrical information. Beliefs, which are unobserved, are likely to be heterogeneous: they can explain both heterogeneous response to a better coverage and decision to take out SHI, resulting in selection on return.

A last source of heterogeneity in reactions to insurance coverage refers to income effects in access to healthcare rather than substitution effects based on the relative price of S2 and S1 visits. Classical models of health insurance (Friedman and Savage 1948; Pauly 1968) see moral hazard as a pure price effect. Contributions by Nyman (1999, 2003) suggest that better coverage also creates an income effect which releases the budget constraint and gives patients access to care that they could not afford without insurance. In this perspective, better coverage could give low income individuals access to S2 consultations that they could not purchase otherwise. This can be a motive to buy coverage of balance billing. If this is true, we should observe that low income people are more likely to buy coverage for balance

---

2. Of course, there is a two-way causality: S2 specialists are more likely to choose their location in areas where people have high incomes. This is of no consequence on our reasoning.

billing, and should react more to an improvement in coverage than rich people. Our data and specification give us an opportunity to implement an empirical test of this interpretation of moral hazard.

## 4. Data and empirical strategy

We use a data set from a French supplementary insurer (MGEN) which is a not-for-profit insurer who provides mandatory basic health insurance for teachers and Ministry of education's employees. MGEN also supplies supplementary health insurance in the form of a unique[3] contract (SHI-basic) which offers a minimal supplementary coverage: it covers only copayments and not balance billing. People can subscribe to this SHI-basic on a voluntary basis, or take out another SHI. Our data stem from administrative MGEN data: they provide, for each policyholder, detailed information about their medical bills and reimbursements for basic health insurance and for supplementary insurance when the individual is a SHI-basic subscriber. Very often in the empirical literature on SHI, administrative claims do not include information on total healthcare consumption. The great advantage of our data is that all medical services that are used and all fees that are charged, including balance billing, are recorded for every individual, whether or not he or she is a SHI subscriber. This is due to the fact that MGEN manages the NHI reimbursements for all teachers.

3. This is true for our observational period. From 2016 onwards, MGEN started to supply a choice between different contracts for SHI.

We do not observe the coverage of balance billing for people who subscribed to another SHI than SHI-basic. However, MGEN used to send a questionnaire to people who switched to another SHI. This allows us to know, for people who have terminated a SHI-basic contract, if they have subscribed to another SHI. For this reason, we selected, for year 2012, a sample of subscribers of SHI-basic and of subscribers of another SHI, who were in 2010 subscribers of SHI-basic and have terminated their contract in 2011. In this case, we know that their new coverage will be at least equal and probably better than before, because SHI-basic coverage on balance billing is zero. We name this new contract "SHI-plus".

Our original sample was composed of 91,629 subscribers of SHI-basic and 8,249 subscribers of SHI-plus. We excluded individuals who live outside continental France as well as the top 1% of care users in 2012. Because in France balance billing concerns mostly specialists, our analysis focuses on the impact of coverage of balance billing on the use of specialists. As a result, we only keep individuals who have at least one visit to a specialist in 2012, with or without balance billing. Our final sample includes 58,519 individuals: 53,456 subscribers of SHI-basic and 5,063 subscribers of SHI-plus, observed in 2012, who have visited a specialist at least once in 2012.

### *4.1. Empirical strategy*

To investigate whether there is an inflationary spiral due to the coverage of balance billing by SHI, we consider a structural model that links demand for balance billing, decision to take out balance billing coverage and the reaction to better coverage. Precisely, we specify a Generalized Roy model, on a cross section of individuals observed in 2012. It is a switching regression model that explains together the decision to take out coverage for balance billing (SHI-plus), and the demand for consultations with balance billing when the individual is – or is not – covered for balance billing. From this model, we derive the marginal treatment effect for someone who is at the margin, i.e. who is indifferent between being covered for balance billing or not. We are able to estimate heterogeneous moral hazard, driven by both observed and unobserved characteristics, such as income and preferences, and to test for selection on return.

Our empirical strategy requires the use of an instrument to explain the decision to terminate SHI-basic contract in order to take out a SHI-plus. A valid instrument must be correlated to the decision to quit SHI-basic and be uncorrelated to the consumption of balance billing. The decision to retire in 2011 for people younger than 55 years-old, that we used in Dormont and Péron (2016), is a reliable instrument. The age threshold refers to a specific right for teachers and civil servants who raised three or more children to retire before 55. This right has been revoked in January 2012, creating an important incentive for individuals meeting the criteria to retire in 2011. Figure 3 shows the number of individuals in our sample who retired in 2010, 2011 and 2012. One can observe that the number of individuals retiring before 55 increased notably in 2011 compared to 2010 and dropped to almost 0 in 2012. Furthermore the decision to retire is linked to the decision to quit SHI-basic. In our sample, 368 individuals decided to retire before 55 in 2011 and half of them quit SHI-basic the same year. Indeed, SHI-basic premiums raise from 2.97% of wages before retirement to 3.56% of pensions after. We argue that this retirement policy change creates an exogeneous shock that gives individuals incentives to terminate their SHI-basic contract for a SHI-plus, but has no reason to drive their balance billing consumption. We do not observe eligibility (the number of children raised is not available) so we use the decision to retire before 55. Of course compliers are people who retired before 55 because of the rules until 2011 and hold SHI-plus. However they might have chosen to retire for health reasons, which would question our instrument exogeneity. We checked that holding SHI-plus in connection with our instrument is not associated with a higher use of GP consultations, nor of drug consumption, ruling out the existence of a health shock for our compliers (results available on request).

### *4.2. Basic features of the data*

Our data provide, for each individual in 2012, the number of visits to a specialist, including the number of visits to S2 specialists who charge balance billing, as well as the total amount of balance billing. One variable of interest is the average balance billing per consultation, which indicates the intensity of use of balance billing. Its magnitude is influenced by the proportion of visits to S2 specialists and by the average balance billing charged by these S2 specialists. Our second variable of interest is the total number of visits to (S1 and S2) specialists. If there are difficulties in accessing S1 specialists due to waiting times or geographical location, higher fees charged by S2 specialists might yield a reduction in the use of specialist consultations. In total, we are able to distinguish two dimensions in the demand for specialist consultations: quantity and quality. Whilst quantity refers to the number of visits, quality is reflected by the choice between S1 and S2 specialists and the average balance billing per visit to a S2 specialist.

On the grounds that our data do not provide the fee for each consultation, we compute for each individual an annual average of balance billing per visit. We are able, however, to control for the individual's needs regarding medical specialties. This is important because, the availability of S1 and S2 specialists varies dramatically from a specialty to another in France. Gynecologists, ophthalmologists, surgeons and ENT specialists[4] charge balance billing in a larger proportion than their colleagues. As a matter of fact, patients' choice to visit a S2 is likely to be far more constrained when they need to visit one of these specialties. We therefore use a dummy variable called "expensive physicians" which equals 1 when the individual visited one of these specialists at least once in 2012.

Our information on individual characteristics include gender, age, income and health status, all measured in 2012. The impact of age is modelled with three age groups: 20-40, 40-60 and over 60 years old. Our income variable is based on individuals' wage or pension used by MGEN to compute SHI-basic premiums. The dummy variable "Chronic disease", which equals 1 if individuals have at least one chronic disease, is used as an indicator of health status. To measure the availability of S1 or S2 specialists, we use the "specialist : population ratios" (SPR) provided by national statistics in 2012. The SPR is the number of specialists either in S1 or in S2 per 100,000 inhabitants in each region[5]. We use three categories for the availability of S1 specialists: "low availability" includes the bottom third of regions in terms of number of S1 specialists per 100,000 inhabitants ($[20, 41[$), "medium

4. Ear, Nose and Throat specialists

5. Our data are at the French (*département* level.

availability" includes the second third ($[41, 52[$), "high availability" includes the top third ($[52, 56]$). We define two categories for S2 specialists: "low availability" includes the bottom third of regions in terms of number of S2 specialists per 100,000 inhabitants ($[2, 15[$; "high availability" includes the middle and top thirds ($[15, 29]$).

Table 1 displays the characteristics of the 58,519 individuals from our final sample: there is a high proportion of women (72.5%), the average age is close to 58 years, the average income amounts to €2,500 and 22% have a chronic disease. In comparison, the average wage is in France equal to €2,157[6] and 19.5%[7] of people have a chronic disease. These characteristics derive from the fact that (i) MGEN covers teachers and civil servants who have a certain education level and are mostly women; (ii) we have restricted our sample to those who visited a specialist at least once in 2012. Compared to SHI-basic holders, SHI-plus holders are on average 12 years younger, count more women (82% vs 72.5%) and less individuals with chronic disease (9.4% vs 22%). To sum up, those who decided to quit SHI-basic are on average younger and healthier. This is a common result in the literature on switching behavior: in the US (Buchmueller and Feldstein 1997; Strombom et al. 2002), Switzerland (Dormont et al. 2009) or in the Netherlands (Duijmelinck and van de Ven 2016), switchers are invariably younger and also tend to be healthier. We discuss the motivations to subscribe to SHI-plus further in the paper.

---

6. Average net mensual wage in 2012; source: INSEE.

7. Source: ESPS survey.

Table 2 displays statistics on the annual number of visits to a specialist and amount of balance billing in 2010 and 2012 for SHI-basic holders and future SHI-plus holders (who are covered by SHI-basic in 2010 and SHI-plus in 2012)[8]. Of course, in 2012, SHI-plus holders are likely to enjoy better coverage for balance billing than SHI-basic holders. Whilst the total number of visits is not significantly different between SHI-basic and SHI-plus holders, the latter consume significantly more balance billing, both in quantity (The share of S2 visits is 51% for SHI-plus holders vs 43% for SHI-basic holders) and price (€26.1 of balance billing per S2 visit vs €24.2). Consequently, SHI-plus holders' average consumption of balance billing per visit amounts to €13.7 in 2012, which is about 30% higher than for SHI-basic holders.

These differences might reflect adverse selection, as well as moral hazard and, if there is heterogeneity in moral hazard, possible selection on return. Actually, our data design enables us to observe the use of balance billing by SHI-plus subscribers in 2010, before they take out better coverage. In 2010, all individuals in our sample, including future SHI-plus, are SHI-basic holders, hence equally not covered for balance billing. Table 2 shows that in 2010 the future SHI-plus holders, who will buy better coverage the next year, consumed more balance billing than those who will keep their SHI-basic contract. This reveals classical adverse selection: those who ask for better coverage used to consume more balance billing, even when they were not covered for it.

8. This comparison is not possible for all the 58,519 individuals observed in 2012 since only 43,612 of them used at least a specialist visit in 2010.

## 5. Empirical specification

We consider a Generalized Roy model to derive and estimate MTE. The estimates will allow us to identify different sources of heterogeneity in the treatment impact.

### *5.1. Model*

We consider the following Generalized Roy model that specifies the two potential levels of balance billing consumption $(Y_0, Y_1)$ which are observed if the individual takes out respectively SHI-plus (treated) $(D = 1)$ or SHI-basic (not treated) $(D = 0)$.

$$Y = DY_1 + (1 - D)Y_0 \tag{1}$$

$$Y_1 = X\beta_1 + U_1 \tag{2}$$

$$Y_0 = X\beta_0 + U_0 \tag{3}$$

$$D^* = \gamma Z - V = \gamma_1 X + \gamma_2 EarlyRetiree - V \tag{4}$$

$$D = \begin{cases} 1 & \text{if } D^* > 0 \\ 0 & \text{if } D^* \leq 0 \end{cases} \tag{5}$$

$D$ is equal to 1 if the individual chooses to take out SHI-plus in 2011. In 2012, people covered by SHI-plus benefit from balance billing coverage whilst SHI-basic enrollees (those who stayed) do not. Taking out SHI-plus has an impact on unobserved heterogeneity (from $U_0$ to $U_1$) and on the effect of covariates $X$ (from

$\beta_0$ to $\beta_1$). In the general case, it is assumed that $U_0$, $U_1$ and $V$ are independent of $Z$, conditional on $X$. In addition, the probability of treatment is a non-trivial function of $Z$, conditional on $X: \; Pr(D|X = x, Z = z) \neq Pr(D|X = x)$ (Basu et al. 2007).

The choice equations (4) and (5) explain the individual's decision to take out another SHI to enjoy better coverage (SHI-plus) than the one provided by SHI-basic. The choice of taking out SHI-plus is modeled as a function of observables $Z$ and unobservables $V$; and linked to the observed outcome $Y$ through a latent variable $D^*$. This enables us to understand coverage choices' determinants and provides the propensity scores that are used to identify MTE.

In equations (2) to (4) $X$ is a vector of covariates which includes individuals' gender, age, income and whether they suffer from a chronic disease. It includes also local availability of specialists of S1 (not allowed to charge balance billing) and S2 (allowed to charge balance billing) and for the individual's needs as regards medical specialty (the proportion of S2 specialists is particularly high for ophthalmologists, gynaecologists and ENT).

In equation (4), $EarlyRetiree$ is our excluded instrument: the decision to retire before 55 years old is correlated with the decision to subscribe to SHI-plus, but not with the consumption of balance billing. $V$ is an unobservable random variable corresponding to the individual reluctance to buy SHI-plus. It is linked with unobservable individual characteristics such as disutility of administrative switching costs, minus utility of coverage for given risk level, i.e. belief that S2 doctors provide better quality of care, waiting time aversion and risk aversion.

The propensity score $P(Z)$ is the probability of receiving treatment conditional on $Z$:

$$P(Z) \equiv Pr(D = 1|Z = z) = Pr(V < \gamma Z|Z = z) = F_V(\gamma Z)$$

where $F_V$ is the cumulative distribution function of $V$, hence a monotonic and absolutely continuous function. An individual chooses to take out SHI-plus if the latent variable $D^*$ is positive:

$$D = 1 \Leftrightarrow D^* > 0 \Leftrightarrow \gamma Z > V \Leftrightarrow F_V(\gamma Z) > F_V(V) \Leftrightarrow P(Z) > F_V(V)$$

Defining $U_D = F_V(V)$, the condition to be treated is that the propensity score is greater than $U_D : P(Z) > U_D$. Without a loss of generality we can assume that $U_D$ is a uniformly distributed random variable between 0 and 1. In this case the $p^{th}$ quantile of $U_D$ is $p$ and different values of $U_D$ correspond to different quantiles of $V$.

The propensity score must be interpreted as the incentive to take out SHI-plus, for given covariates $Z$. Conversely, $U_D$ can be seen as the individual idiosyncratic disutility of switching to SHI-plus. Conditionally on her characteristics $z$, which provide a propensity score $p$, an individual will ultimately take out SHI-plus if her disutility $u_D$ is lower than $p$ (and be indifferent if $u_D = p$). The econometrician observes variables $Z$ but not the realizations $u_D$. However, given that values of $U_D$ are quantiles of $V$, it is possible to compare $P(Z)$ and $U_D$ on the same interval $[0, 1]$ on the horizontal axis of a diagram that plots the idiosyncratic disutility against the treatment effect (Figure 4).

### 5.2. Marginal Treatment Effects

In our model, the decision to take out SHI-plus and its impact on balance billing consumption vary across individuals. MTE capture the treatment effect $(Y_1 - Y_0)$ for the 'marginal individual' who is indifferent between being treated or not, conditional on her observed characteristics $X = x$. By definition, the marginal individual has a propensity score equal to her disutility of taking the treatment: $U_D = p$.

$$MTE \equiv E(Y_1 - Y_0|X = x, U_D = p) \tag{6}$$

We follow the method suggested by Heckman and Vytlacil (2001) and generally known as Local IV. This method identifies MTE as the derivative of the conditional expectation of the outcome $E(Y|X = x, Z = z)$, with respect to the propensity score $P(Z)$. First, note that

$$E(Y|X = x, Z = z) = E\{Y|X = x, P(Z) = p\} \tag{7}$$

The observed outcome can be written as:

$$E\{Y|X = x, P(Z) = p\} = E(Y_0|X = x) + E(Y_1 - Y_0|X = x, D = 1)p \tag{8}$$

$$= E(Y_0|X = x) \tag{9}$$

$$+\int_0^p E(Y_1 - Y_0|X = x, U_D = u_D)du_D \tag{10}$$

As a consequence,

$$\frac{\partial E\{Y|X=x,P(Z)=p\}}{\partial p}=E(Y_1-Y_0|X=x,U_D=p) \tag{11}$$

Expression (11) shows how the derivative of $E(Y|X=x,Z=z)$ identifies MTE, i.e the expected treatment effect conditional on $X$ and $U_D$. Note that a high value of $P(Z)=p$ identifies MTE at a high value of $U_D=u_D$. In other words, the fact that individuals with a high propensity score are indifferent between taking out SHI-plus or staying with SHI-basic implies that they have a very high idiosyncratic disutility of switching towards SHI-plus ($u_D$). Therefore, MTE with high $p$ values identify returns for individuals who are less likely to take out SHI-plus. Conversely, MTE with low values of $p$ identify returns for individuals prone to take out SHI-plus.

Ideally, a continuous instrument with sufficient variation conditional within all $X=x$ would allow for a fully non-parametric estimation of the MTE, conditional on X, and would produce a separate MTE curve for each value of $X$. However, our instrument $EarlyRetiree$ is a binary instrument (1 if individuals retired before 55 in 2011, 0 otherwise) and in this case, further assumptions are required to identify MTE (Cornelissen et al. 2016). A first assumption is to condition the outcome on $X$ in a parametric linear way and model potential outcomes as $Y_1=x\beta_1+U_1$ and $Y_0=x\beta_0+U_0$ and the selection equation as $D_i^*=\gamma Z-V$. A second assumption is to assume that the shape of the MTE curve is independent of observed characteristics $X$. Only the intercept of the MTE curve is allowed to vary with $X$. This is implied by the full independence assumption $(X,Z)\perp\!\!\!\perp(U_0,U_1,V_1)$, which is stronger than the conditional independence assumption $Z\perp\!\!\!\perp(U_0,U_1,V_1)|X$ necessary for fully non-parametric approach. The linear separability and the full independence assumptions imply that the MTE is additively separable into an observed and unobserved component. Combining (8) with the linear expressions of $Y_1$ and $Y_0$, one obtains:

$$E\{Y|X=x,P(Z)=p\} = x\beta_0 + x(\beta_1-\beta_0)p + K(p) \ , \tag{12}$$

$$\text{with } K(p) = E\{U_0|P(Z)=p\} + E\{U_1-U_0|P(Z)=p\}p \tag{13}$$

$K(p)$ is a nonlinear function of the propensity score. It serves here as a control function, as defined by Heckman and Robb (1985). It takes into account the fact that the difference between the outcome and the specification on the right-hand side is a function of $p$. Hence, a regression applied on (12) consistently estimates parameters $(\beta_0, \beta_1)$. Note that the fact that $K(p)$ does not depend on $X$ reflects the assumption that the slope of the MTE curve does not depend on observable characteristics.

The MTE are computed as the partial derivative of the conditional expectation of $Y$ with respect to $P(Z)$ :

$$\frac{\partial E\{Y|X=x,P(Z)=p\}}{\partial p} = x(\beta_1-\beta_0) + \frac{\partial K(p)}{\partial p} \tag{14}$$

Writing the control function $K(p)$ as a polynomial in $p$ , equation (12) becomes:

$$E\{Y|X=x,P(Z)=p\} = x\beta_0 + \{x(\beta_1-\beta_0)\}p + \sum_{i=1}^{\vartheta} \varphi_i p^i \tag{15}$$

As mentioned above, implementing the local IV method with a binary instrument implies that identification relies crucially on the full independence assumption $(X,Z) \perp\!\!\!\perp (U_0,U_1,V_1)$. Alternatively, Brinch et al. (2017) propose a separate approach.

They assume additive separability between the observed and unobserved component in the expected outcomes conditional on $U_D = u$. The conditional expectations of $Y_1$ and $Y_0$ are then estimated separately and the MTE is derived from $MTE(x, u) = E(Y_1|X = x, U_D = u) - E(Y_0|X = x, U_D = u)$. One advantage of this method is to allow for fully non-parametric estimation of the MTE with a binary instrument, as used in Kowalski (2018). A linear MTE can also be identified based on the binary instrument only, without relying on the functional form of the covariates. We use the Stata program made available by Andresen et al. (2018) to check that our results are robust to the separate approach.

### 5.3. Estimation

Estimating MTE by the local IV method requires three steps. In a first stage we estimate the propensity score for each individual, $\hat{P}(z) = Pr(\gamma Z > V|Z = z) = p$. The propensity score is fitted by a probit model[9]. As noted is section 5.1, $Z$ includes covariates $X$ and our excluded instrument $EarlyRetiree$. We then determine the common support, i.e. the values of $\hat{P}(z) = p$ for which we have positive frequencies of individuals who decided to take out SHI-plus ($D = 1$) and of individuals who remained SHI-basic enrollees ($D = 0$).

---

9. The results are robust to the use of a Logit model. It is preferable not to consider a linear probability model because it does not allows to constrain the range of $\hat{P}(z)$ to be $(0, 1)$, see Brave et al. (2014).

In a second step, we perform OLS on equation (15), assuming that the function $K(p)$ is a polynomial of degree 3:

$$y = x\beta_0 + \{x(\beta_1 - \beta_0)\}p + \varphi_1 p + \varphi_2 p^2 + \varphi_3 p^3 \tag{16}$$

$y$ is the log-transformation of our variables of interest: the average balance billing per visit $BB/Q$ and the number of visits to a specialist $Q$. We also present in the online appendix the estimates of the log-transformations of the components of $BB/Q$, i.e. the share of S2 visits $Q2/Q$, and the average balance billing per S2 visit $BB/Q2$. As for the choice equation, $x$ is a vector of covariates which includes individuals' gender, age, income, chronic disease, local availability of S1 and S2 specialists, and individual's needs regarding ophthalmologists, gynaecologists and ENT ("expensive physicians"). Subscript 1 (respectively, 0) refers to SHI-plus enrollees (respectively to SHI-basic enrollees). SHI-plus enrollees are treated, i.e benefit from balance billing coverage, but this is not the case for SHI-basic enrollees. According to the Roy model, when an individual chooses to switch from SHI-basic to SHI-plus, her behavior switches from $Y_0 = X\beta_0 + U_0$ to $Y_1 = X\beta_1 + U_1$.

In a third step, the parametric estimator of MTE is computed for given values $x$ as

$$MTE\{x, p\} = x(\beta_1 - \beta_0) + \varphi_1 + \varphi_2 p + \varphi_3 p^2 \tag{17}$$

In our setting, MTE capture the effect of having better balance billing coverage for the individual 'at the margin', who is indifferent between subscribing to SHI-plus or remaining enrolled in SHI-basic ($U_D = p$).

Alternatively, we compute a semiparametric estimator of MTE by running a local polynomial regression of $\widetilde{y}$ on $p$ with:

$$\widetilde{y} = y - x\widehat{\beta}_0 - \{x(\widehat{\beta_1 - \beta_0})\}p. \tag{18}$$

Note that the semiparametric approach differs only in the estimation of the unobserved component $K(p)$. Furthermore, the semiparametric estimator can only be estimated on the common support of the propensity score. Precisely, the common support assumption requires that there exist positive frequencies of $\hat{P}(z)$ for individuals that receive $(D = 1)$ and do not receive $(D = 0)$ the treatment. It is worth noting that, although a parametric estimator of MTE can be estimated on the whole range $[0, 1]$, its precision also crucially depends on the common support (Brave et al. 2014). Therefore, our interpretation of the results will be limited to the common support.

To run the estimations, we use the Stata command ***margte*** (Brave et al. 2014) with a polynomial of degree 3 to estimate the parameters of the MTE. We use an epanechikov kernel function in the semiparametric estimation. Standard errors are computed using bootstrap (50 reps). Parametric and semiparametric MTE are computed at mean values of $x$ as in equations (19) and (20):

$$MTE\{\overline{x}, p\} = \overline{x}(\beta_1 - \beta_0) + \varphi_1 + \varphi_2 p + \varphi_3 p^2 \tag{19}$$

$$MTE\{\overline{x}, p\} = \overline{x}(\beta_1 - \beta_0) + \frac{\widehat{\partial K(p)}}{\partial p} \tag{20}$$

We are also able to compute an "empirical average treatment effect", conditional on $X$. This empirical ATE is constructed as a weighted average of the semiparametric

MTE by integrating over $U_D$ on the common support: $ATE \equiv E(Y_1 - Y_0|X = x)$. We do so for all the covariates at their mean as well as for different levels of income and S1 and S2 availability.

### *5.4. Interpretation of the estimates*

Our empirical specification allows for a detailed analysis as regards the impacts of observable characteristics:

- $\beta_0$ captures the impacts of observed individual characteristics on the use of balance billing, without coverage for it;
- $\gamma$ captures the impact of observed characteristics on the decision to switch;
- $(\beta_1 - \beta_0)$ measures the change in the impact of observed characteristics on balance billing consumption which is due to the treatment (coverage for balance billing).

Note that in our model the fact that the impacts of regressors can be modified by better coverage is a source of heterogeneity in moral hazard that comes in addition to the heterogeneity linked to unobserved characteristics. Suppose that $(\beta_1 - \beta_0) < 0$ for income. This would mean that low-income people react more strongly to insurance.

We also compare the signs of the estimates obtained for $\beta_0$, $\gamma$ and $(\beta_1 - \beta_0)$ to identify the situations of classical adverse selection (relationship between $\beta_0$ and $\gamma$) and the situations of selection on return (relationship between $(\beta_1 - \beta_0)$ and $\gamma$).

As regards essential heterogeneity, Heckman et al. (2006) propose a simple test to explore the assumption of a variable treatment effect due to unobservable characteristics. The joint significance of the polynomial coefficients $\varphi_1, \varphi_2, \varphi_3$ in equation (16) reveals the presence of essential heterogeneity. Indeed, the signs of $\varphi_2$ and $\varphi_3$ determine the slope of the curve that characterizes the relationship between the treatment effect and the value of $U_D$. Precisely, $\varphi_1 = \varphi_2 = \varphi_3 = 0$ would mean that the treatment effect does not vary with unobservable characteristics, i.e. that there is no evidence of essential heterogeneity. On the contrary, depending on the values of $\varphi_2$ and $\varphi_3$, one can find that individuals with a low (or high) disutility to switch benefit more (or less) from better balance billing coverage.

Figure 5 displays our common support. Since it is not defined for all values of $U_D$ between 0 and 1, we are only able to compute an ATE on a limited support ("empirical ATE"). Note also that although parametric MTE are estimated on $[0, 1]$, their precision strongly decreases for $U_D > 0.35$ which makes the value of MTE difficult to interpret for higher values of $U_D$. So, in any case, we restrict our analysis of MTE on the values corresponding to the common support.

## 6. Results

Results of the main estimations are presented in Tables 3 to 5. Table 3 displays the effects of observable individual characteristics on balance billing per visit ($log(BB/Q)$) without coverage, propensity to switch and moral hazard. Table 4 presents the same set of results for the number of visits to a specialist ($log(Q)$). Table 5 gives the estimates of the empirical ATE and the parameters related to essential heterogeneity (here, selection on return due to unobserved individual characteristics). Significance tests enable us to test for the existence of selection on return due to unobserved heterogeneity. Figures 6 and 7 display respectively semiparametric MTE over $U_D$ for balance billing per visit and number of visits, evaluated at mean values of observable characteristics $X$. In the appendix, Tables A.1-A.3 display the same set of results for the log-transformations of the components of $BB/Q$, i.e. the share of S2 visits $Q2/Q$, and the average balance billing per S2 visit $BB/Q2$.

### *6.1. Evidence of classical adverse selection*

Classical adverse selection is captured through the estimates of parameters $\gamma$ and $\beta_0$: the relation between them can show whether patients with a higher balance billing consumption without coverage are more likely to take out coverage for balance billing.

The estimates of $\beta_0$ (first column of Table 3) show that the main determinants of the amount of balance billing per visit paid by patient who are not covered for it are income, medical needs and local availability of S1 and S2 specialists. The average amount of balance billing per visit significantly increases with income: a 10% increase in income drives up balance billing per visit by 5.3%. Individuals aged of 60 years old and more, those who suffer from a chronic disease or visit gynaecologists, ophthalmologists or ENT specialists consume also more balance billing than others. The availability of S1 and S2 specialists has also a strong impact on the amount of balance billing paid by patients. Indeed, balance billing use is 18% higher for patients living in areas where the number of S1 specialists is low and 56% higher for those who lived in areas where S2 specialists are numerous. Otherwise, we do not find unambiguous evidence of limitation in access to specialists in regions with low availability of S1 specialist. Without balance billing coverage, in regions with low availability of S1, we find no significantly lower use of specialist care, as measured by the number of visits to a specialist ($Q$). Whereas there is evidence of a significantly lower use in regions with medium availability of S1 (Table 4).

The effects of observed individual characteristics on the probability of subscribing to SHI-plus are captured through the coefficients $\gamma$ in the first step of the estimation (second columns of Tables 3 and 4 - we reproduce the same estimates in both tables). We find that in our sample, young and healthy (with no chronic disease) individuals are more likely to quit SHI-basic. Low income individuals are more likely to take out SHI-plus than high income. Individuals who live in regions where there are few S1 specialists or a lot of S2 specialists are also more likely to take out SHI-plus.

To sum up, the low availability of S1, high availability of S2 specialists and the type of specialists that individuals visit are a source of classical adverse selection. The similar signs of $\gamma$ (propensity to take out SHI-plus) and $\beta_0$ (balance billing use without coverage) for these variables show that they explain both a higher use of balance billing and a propensity to switch to better coverage.

### *6.2. Evidence of selection on return*

Selection on return is captured through the estimates of the MTE and of parameters $\gamma$ and $(\beta_1 - \beta_0)$. For observable characteristics, relationships between $\gamma$ and $(\beta_1 - \beta_0)$ can show whether patients with a stronger reaction to balance billing coverage are more likely to take out coverage for it (second and third columns of Tables 3 and 4). If the estimated MTE vary in connection with the reluctance to buy SHI-plus $U_D$, there is also selection on return linked to unobserved heterogeneity. In Table 5 we present the patterns of estimated MTEs for our two variables of interest, as well as the resulting empirical ATE.

First of all, we find a significant empirical ATE regarding the use of balance billing (+321%[10], Table5 - column 1), whilst this is not the case for the number of visits to a specialist (column 2). In other words, there is, on average, moral hazard regarding the composition and price of specialists care, but not regarding the quantity of care.

Turning to the heterogeneity of our estimated moral hazard, we first examine the impacts of observable characteristics. We find that better coverage induces significant changes ($\beta_1 - \beta_0$) in the impact of regressors (Table 3, column 3). There is heterogeneous moral hazard linked to observed characteristics: the reaction to better coverage is significantly heterogeneous between different levels of income and availability of S1 specialists. More precisely, the effect of insurance on balance billing consumption is significantly decreasing with income: the poor react more to insurance than the rich. They increase more strongly their use of balance billing per visit. This results on a higher empirical ATE for the bottom decile of the income distribution (+380%, other covariates at their mean) than for the top decile (+267%). Moral hazard on the use of balance billing per visit is 156% higher in areas where the availability of S1 specialists is low and 178% higher in areas where the availability of S2 specialist is high. Hence, we estimate the empirical ATE for individuals living in regions with low S1 and high S2 availability at +444%. On the contrary the empirical ATE in regions with High S1 and Low S2 is not significant. When looking at the detailed impacts on the two components of $BB/Q$, we find that in both cases, the sharper increase in balance billing use is driven by both a

10. In our sample, the average balance billing per visit for SHI-basic holders is €10.5 (sd=13.4), see Table 2.

stronger increase in the share of S2 visits and the average amount of balance billing per S2 visits (Tables A.1 & A.2 in the appendix).

Selection on return appears clearly as regards income. Indeed, the estimated values of $\beta_0$, $\gamma$ and $(\beta_1 - \beta_0)$ for income (first row of Table 3) show that low income people use less balance billing than others when they are not covered for it (no classical adverse selection) but react strongly to coverage (moral hazard) and are more likely to buy SHI-plus coverage (selection on return). In other words, they buy coverage to be able to afford balance billing. As concerns the impacts of the availability of S1 and S2 specialists, selection on return comes is addition to classical adverse selection. Individuals living in regions with few S1 and many S2 specialists benefit more than others from SHI-plus and are also more likely to switch to better coverage.

Are individual unobserved preferences also responsible for heterogeneity in moral hazard and is reaction to coverage related to the decision to buy it? A simple test of joint significance on the terms of the propensity score polynomial shows that we have to reject the hypothesis of a homogeneous treatment effect. Hence, the parametric estimates of the MTE vary with $U_D$ and the signs the of $p$ and $p^2$ give us the form of the MTE function depending on $U_D$ (Table 5). We find similar results with semiparametric estimates[11] The semiparametric MTE on balance billing per visit $(BB/Q)$ is decreasing in $U_D$ (Table 5 and Figure 6). This shows selection on return: individuals who are more likely to take out better coverage have a stronger reaction to health insurance because of unobserved characteristics. Note that this

11. Recall that because the common support is relatively restricted (Figure 5), roughly for $p$ included in $[0.02, 0.35]$, we cannot interpret the MTE results for $U_D > 0.35$.

is also true for both components of balance billing use, i.e. the share of S2 visits and for balance billing per S2 visits (Table A.3 and Figure A.1, in the appendix). Remarkably, we find the contrary for the number of visits to a specialist: the MTE is increasing in $U_D$ and hardly significant, except for values of $U_D$ lying between 0.2 and 0.3. Only those who are the less prone to take out better coverage show moral hazard in the number of visits to a specialist. These semiparametric MTE are computed at the mean of all covariates $X$. Figure 8 illustrates the shift of the MTE curve for different levels of our covariates, in this instance income (bottom and top decile) and specialists availability (low S1 and high S2 v High S1 and low S2)[12].

### *6.3. The inflationary spiral*

We find evidence of classical adverse selection where observable characteristics associated with a higher use of balance billing are also determinants of the purchase of balance billing coverage. Coupled with moral hazard, this is likely to encourage the rise in prices. Furthermore, we argue that selection on return reinforces the inflationary spiral. We find evidence of heterogeneous moral hazard, both on observable and unobservable characteristics and that stronger reaction to coverage is generally associated with a higher propensity to take out balance billing coverage.

Specifically, we show that supply side characteristics play a significant role in the inflationary spiral. Individuals living in regions with few S1 specialists show both classical adverse selection and selection on moral hazard, which explains their high motivation to switch. Finding that low income people react more to an improvement

12. This is a direct implication of the linear separability and full independence assumptions needed to estimate MTE with a binary instrument (see Section 5.2).

in coverage is also particularly interesting. Assuming that all individuals have the same marginal rate of substitution between medical services and consumption of other goods, such a result can be seen as an empirical evidence of Nyman's interpretation of moral hazard (Nyman 1999, 2003). Poor people would react more to coverage than others because better coverage not only changes the relative price of consultations with balance billing, but also releases their budget constraint.

The evidence of selection on return on the use of balance billing but not on the number of visits proves to be an exciting finding as well. To interpret this result, we need to go back to the model specified in equation (4). $U_D$ corresponds to quantiles of $V$. For a given propensity score, the decision to take out better SHI depends on the value of $V$ $(Z\gamma > V)$. The lower $V$, the higher the probability of choosing SHI-plus. $V$ can be linked with unobserved individual characteristics such as disutility ($V_1$) of administrative switching costs, belief ($V_2$) that sector 2 doctors provide better quality of care, or risk aversion ($V_3$). Assuming for simplicity that risk aversion is homogeneous across individuals, the decision depends on $V_1 - V_2$: SHI-plus subscription is restrained by the disutility of switching costs ($V_1$) but encouraged by faith in better quality ($V_2$). Following this interpretation, individuals who are more prone to switch for better SHI are those with stronger faith in the quality of care provided in sector 2[13].

13. In our specification, $Z$ is by definition uncorrelated with $V$, $U_1$ and $U_0$, while $V$ can be correlated with the unobserved components, $U_1$ and $U_0$, in the demand for balance billing or for consultations. While there is only one $V$ driving the decision to switch, $U_1$ and $U_0$ are different for each of our dependent variables.

Our findings give empirical support for such a story: the highest impact of better coverage on balance billing consumption is observed for the first switchers. For the first decile of $U_D$ (i.e. of $V$), they increase their balance billing per visit by 704% (Table 5). Then MTE decrease for higher values of $U_D$ and become non significant for values between 0.2 and 0.3 (Figure 6). Similar results are found for the share of S2 visits and the average amount of balance billing per S2 visits, which are all variables measuring the use of S2 consultations. The reverse is found for the number of visits to a specialist. For this variable, MTE are increasing with $U_D$ but are generally non significant in the semiparametric estimation. In any case, they are not significant for low values of $U_D$. These individuals do not believe that S2 specialists provide better quality of care, or do not value it. Hence, the disutility of administrative costs delays their decision to take out SHI-plus. Furthermore, the improvement in coverage has no impact on their use of S2 specialists. If any significant impact, it is only on the number of visits without distinction between sectors. Obviously, this interpretation is based on a story on the 'content' of the unobserved components of the decision to subscribe to SHI-plus. Nevertheless, the contrast between the decreasing profiles of MTE regarding balance billing use and the increasing or flat profile of MTE regarding the use of specialist consultations provide a strong support to our econometric approach. In any case, our results are consistent with the expected effect of heterogeneous beliefs in the quality of S2 specialists.

### *6.4. Robustness checks*

In our sample, we observe that compliers, i.e. individuals who respond to the exogenous change in retirement rules and switch towards SHI-plus, are only women aged between 38 and 55 years-old. Note that the age boundaries are consistent with the specific rules of early retirement (agents need at least 15 years of experience, and after 55 other rules apply). Nevertheless, one may be concerned that the characteristics of our compliers are driving the results. To check their robustness we restrict the sample, and therefore the control group, to women aged between 35 and 60. As for the results presented above, we use the local IV method and estimate parametric MTE. In the case of a restricted sample, the best specification seems to be a linear MTE. However the results remain very similar in sign, magnitude and significance (Table A.4, column 2 in the appendix). The poor still show a higher impact of insurance on balance billing consumption and we find evidence of heterogeneity on moral hazard and selection on return.

We also test the robustness of our results to the separate approach developed by Brinch et al. (2017) (see subsection 5.2). We use the command *mfete* made available by Andresen et al. (2018) and estimate semiparametric MTE, with a polynomial of degree 2. Table A.4 in the appendix shows that estimates obtained from the separate approach are consistent with our results based on the local IV method. We find evidence of heterogeneity and selection on moral hazard, only the negative impact of income on moral hazard appears to be smaller, although still significant.

## 7. Conclusion

When insurance is voluntary, some individuals may buy insurance because they expect an increase in their consumption due to better coverage. Defined as 'selection on moral hazard' by Einav et al. (2013), this phenomenon is likely to play a preponderant role in a context of supplementary health insurance, where subscription is voluntary.

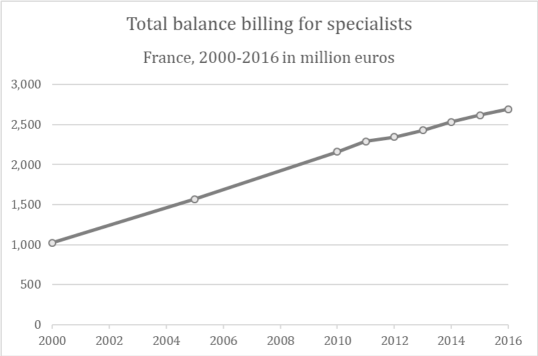
In this paper we investigate the relationships between healthcare use, decision to take out supplementary health insurance and response to better coverage. We use a model that specifies individual heterogeneity in demand for healthcare and in moral hazard. We focus on the demand for specialists who balance bill their patients, i.e. charge them more than the regulated fee set by NHI. Indeed, the demand for specialists who balance bill relies on preferences and beliefs in quality of care. Individuals are likely to be heterogeneous in their preferences and beliefs, while these unobserved characteristics both drive demand for care and decision to take out SHI, resulting in selection on return. In the econometric literature, selection on moral hazard is generally known as 'essential heterogeneity'. Marginal treatment effects estimators have been developed to capture the impact of a treatment likely to vary across individuals. We use MTE to estimate the causal effect of SHI coverage on balance billing consumption on a French database of 58,519 individuals observed in 2012.

We find that supplementary insurance can feed an inflationary spiral regarding medical prices. Specifically, we find evidence of individual heterogeneity in the response to better coverage and of selection on return. Individuals with unobserved characteristics that make them more likely to subscribe to comprehensive SHI are also those who exhibit stronger moral hazard, i. e. a larger increase in balance billing per consultation. Regarding the influence of observed characteristics, we also find that individuals' income is a determinant of balance billing consumption and influences the behavioral response to better coverage. Without coverage, the poor consume less balance billing than the rich but increase their consumption more sharply once covered. They are also more likely to take out comprehensive coverage.

In a context where SHI is voluntary, the inflationary impact of SHI coverage on balance billing might be worsened by selection on return. Our policy conclusions as regards the role of income are of different nature. The negative effect of income on the demand for balance billing consultations coupled with its positive effect on moral hazard provides evidence that insurance plays an important role in terms of access to care for low-income individuals.

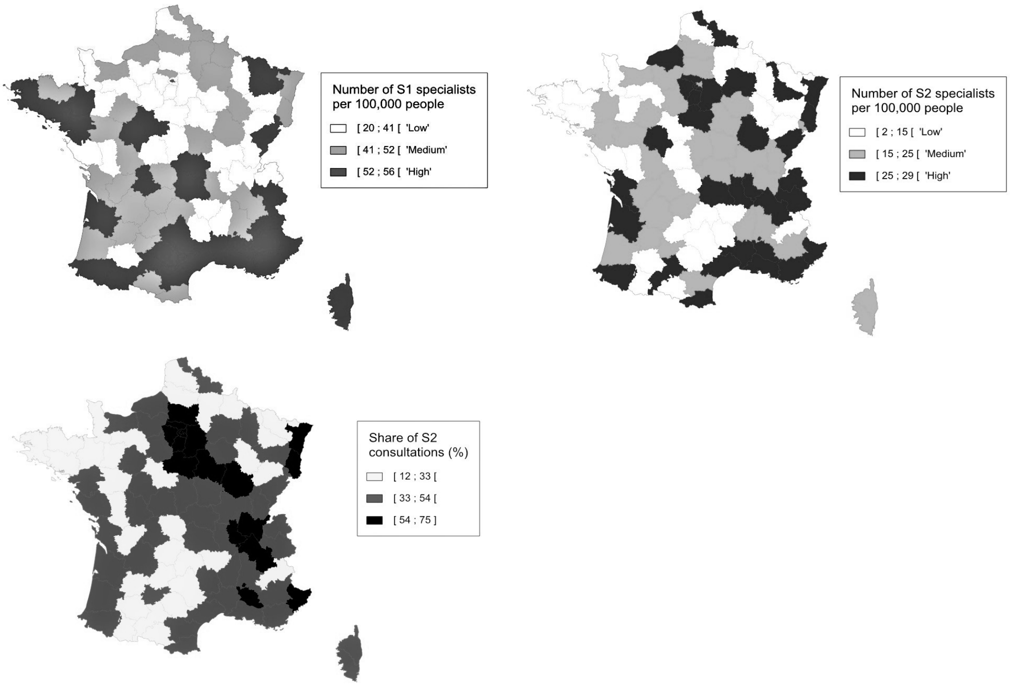
## Tables and Figures

FIGURE 1. A steady increase of the total amount of balance billing in France between 2000 and 2016[a]




a. Data source: SNIR and AMOS (from 2015)

FIGURE 2. Regional variations in sector 1 (S1) and sector 2 (S2) specialists availability[a], share of sector 2 visits $(Q2/Q)^{b}$ in 2010




a. Data source: SNIR data

b. Data source: authors' dataset (MGEN sample, N=58,336)

FIGURE 3. Frequency of individuals retiring in 2010, 2011 and 2012, by age

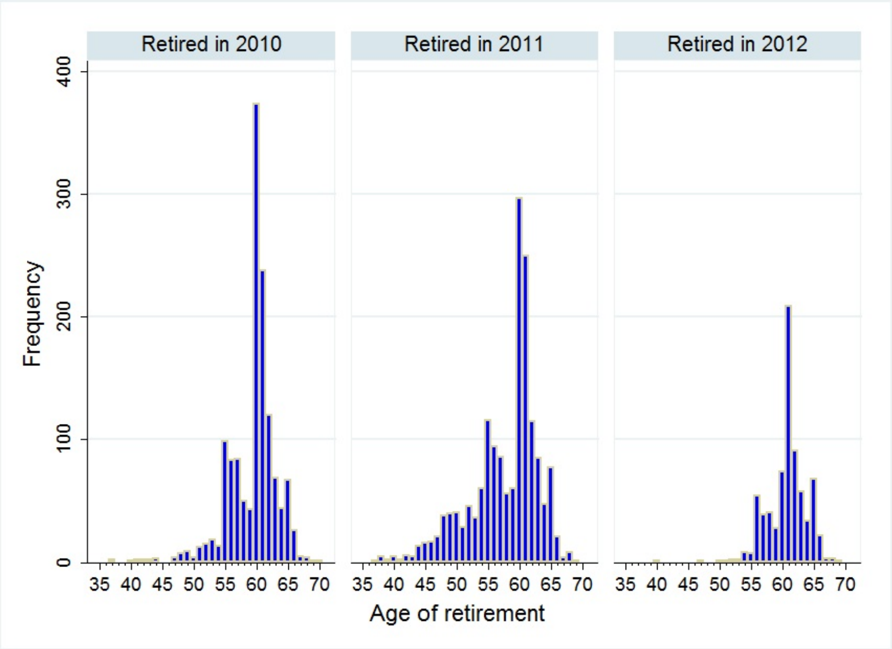


FIGURE 4. Treatment choice for given propensity score $P(Z)$ and values of disutility $U_D$

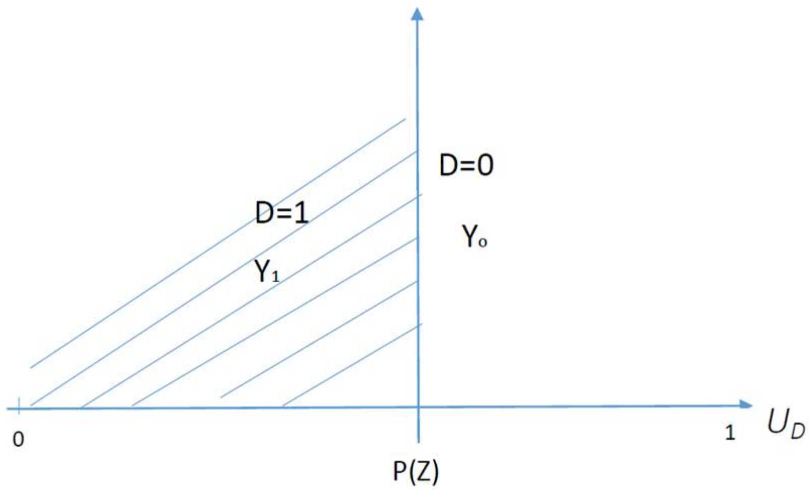


Table 1. Descriptive statistics on individual characteristics of SHI-basic and SHI-plus holders

| | N | Women % | Age mean (sd) | Income mean (sd) | Chronic Disease % |
|---|---|---|---|---|---|
| SHI-basic holders | 53,456 | 72.5 | 57.7 (15.2) | 2,499 (764) | 22 |
| SHI-plus holders | 5,063 | 82*** | 45.2*** (13.3) | 2,406*** (712) | 9.4*** |

*** Significantly different from SHI-basic holders, at 1%.

Sample: 58,519 individuals with at least one specialist consultation in 2012.

Table 2. Descriptive statistics on visits to a specialist and use of balance billing

| | Number of visits[a] $Q$ mean (sd) | Share of S2 visits[a] $Q2/Q$ mean (sd) | BB per visit[a] $BB/Q$ mean (sd) | BB per S2 visit[b] $BB/Q2$ mean (sd) |
|---|---|---|---|---|
| **In 2010** | | | | |
| SHI-basic | 3.6 (4.6) | 44% (0.43) | 10.4 (12.6) | 22.8 (11.6) |
| Future SHI-plus | 3.7 (4.6) | 52%*** (0.43) | 13.2*** (13.6) | 24.7*** (11.6) |
| **In 2012** | | | | |
| SHI-basic | 3.3 (3.4) | 43% (0.43) | 10.5 (13.4) | 24.2 (11.7) |
| SHI-plus | 3.3 (2.3) | 51%*** (0.43) | 13.7*** (14.3) | 26.1*** (12.2) |

*** Significantly different from SHI-basic holders, at 1%.

a. Sample: 43,612 individuals with at least one specialist consultation in 2010 and 2012.

b. Sample: 26,557 individuals with at least one S2 specialist consultation in 2010 and 2012.

Figure 5. Common support

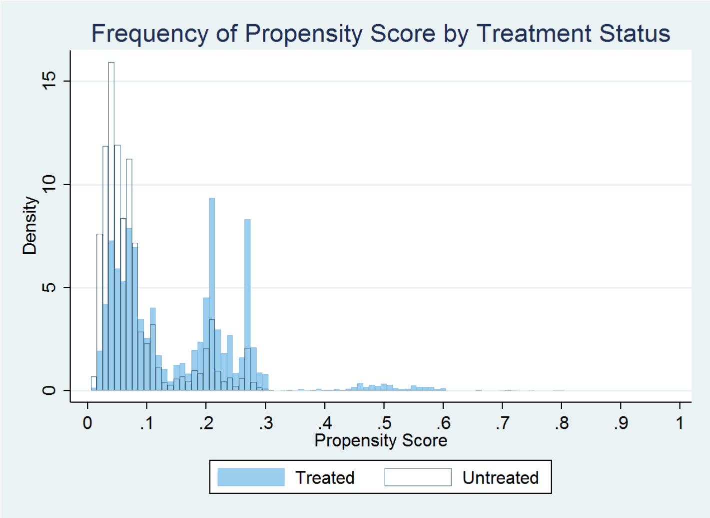


TABLE 3. Main estimates: balance billing per visit to a specialist ($log(BB/Q)$)

| | [1] Balance billing without coverage[a] $\beta_0$[c] | [2] Propensity to switch to SHI-plus[b] $\gamma$[d] | [3] Moral hazard on balance billing [a] $(\beta_1 - \beta_0)$[c] |
|---|---|---|---|
| Log(income) | 0.53*** (0.04) | -0.14*** (0.03) | -1.27*** (0.30) |
| Women | 0.00 (0.03) | 0.08*** (0.02) | -0.68*** (0.26) |
| 20-40 yo. | -0.11 (0.19) | 0.60*** (0.02) | -0.75 (0.58) |
| 40-60 yo. | ref. | ref. | ref. |
| 60+ yo. | 0.17*** (0.06) | -0.29*** (0.02) | 4.35*** (1.06) |
| Chronic disease | 0.17*** (0.05) | -0.20*** (0.02) | -0.49 (0.38) |
| Expensive physicians | 0.81*** (0.03) | 0.09*** (0.02) | 0.15 (0.22) |
| | | | |
| *Sector 1 availability* | | | |
| High | ref. | ref. | ref. |
| Medium | -0.11*** (0.02) | 0.01 (0.02) | -0.01 (0.21) |
| Low | 0.18*** (0.06) | 0.21*** (0.02) | 1.56*** (0.26) |
| *Sector 2 availability* | | | |
| Low | ref. | ref. | ref. |
| High | 0.56*** (0.06) | 0.24*** (0.02) | 1.78*** (0.35) |
| | | | |
| *Excluded instrument* | | | |
| Early retirees | .. | 1.36*** (0.07) | .. |
| | | | |
| N | 58,519 | 58,519 | 58,519 |

*Significant at 10%; **significant at 5%; ***significant at 1%.

Bootstrap standard errors in brackets.

a. The dependent variable is a log transformation of the average balance billing per visit to a specialist ($log(BB/Q)$).

b. The dependent variable equals 1 if individuals switched to SHI-plus, 0 otherwise.

c. Coefficients from Local IV estimation - see equation (16).

d. Coefficients from Probit estimation.

TABLE 4. Main estimates: number of visits to a specialist ($log(Q)$)

| | [1] Number of visits without BB coverage[a] $\beta_0$[c] | [2] Propensity to switch to SHI-plus[b] $\gamma$[d] | [3] Moral hazard on the number of visits[a] $(\beta_1 - \beta_0)$[c] |
|---|---|---|---|
| Log(income) | 0.01 (0.03) | -0.14*** (0.03) | -0.08 (0.18) |
| Women | 0.10*** (0.02) | 0.08*** (0.02) | 0.89*** (0.15) |
| 20-40 yo. | 0.25** (0.12) | 0.60*** (0.02) | 0.16 (0.38) |
| 40-60 yo. | ref. | ref. | ref. |
| 60+ yo. | -0.21*** (0.04) | -0.29*** (0.02) | 3.01*** (0.65) |
| Chronic disease | 0.15*** (0.03) | -0.20*** (0.02) | 1.19*** (0.23) |
| Expensive physicians | 0.16*** (0.02) | 0.09*** (0.02) | -0.02 (0.16) |
| *Sector 1 availability* | | | |
| High | ref. | ref. | ref. |
| Medium | -0.13*** (0.01) | 0.01 (0.02) | 0.25** (0.12) |
| Low | -0.03 (0.03) | 0.21*** (0.02) | 0.07 (0.13) |
| *Sector 2 availability* | | | |
| Low | ref. | ref. | ref. |
| High | 0.12*** (0.03) | 0.24*** (0.03) | 0.39* (0.18) |
| *Excluded instrument* | | | |
| Early retirees | .. | 1.36*** (0.07) | .. |
| N | 58,519 | 58,519 | 58,519 |

*Significant at 10%; **significant at 5%; ***significant at 1%.
Bootstrap standard errors in brackets.
a. The dependent variable is a log transformation of the number of visits to a specialist ($log(Q)$).
b. The dependent variable equals 1 if individuals switched to SHI-plus, 0 otherwise.
c. Coefficients from Local IV estimation - see equation (16).
d. Coefficients from Probit estimation.

TABLE 5. Evidence of heterogeneity in the reaction to better coverage: empirical ATE, parametric MTE and semiparametric MTE

| | Balance billing per visit $log(BB/Q)$ | Number of visits $log(Q)$ |
|---|---|---|
| **Empirical ATE**[a] | | |
| Covariates at their mean | **3.21**** (1.49) | **0.72** (0.55) |
| Income - bottom 10% | 3.80*** (1.39) | 0.76 (0.60) |
| Income - top 10% | 2.67** (1.29) | 0.68 (0.53) |
| Low S1 and High S2 | 4.44*** (1.10) | 0.75 (0.47) |
| High S1 and Low S2 | 1.10 (1.35) | 0.28 (0.62) |
| **Parametric MTE**[b] | | |
| Polynomial coef. $p$ | -27.20*** (8.02) | 16.20*** (5.21) |
| Polynomial coef. $p^2$ | 21.89*** (6.82) | -14.17*** (4.99) |
| Joint test (p-value) | 0.000 | 0.002 |
| **Semiparametric MTE**[c] | | |
| **MTE at p=0.1** | **7.04*** ** | **-0.97** |
| lower bound | 3.76 | -2.56 |
| upper bound | 10.33 | 0.62 |
| **MTE at p=0.2** | **2.23**** | **1.15**** |
| lower bound | 0.49 | 0.05 |
| upper bound | 3.96 | 2.24 |
| **MTE at p=0.3** | **0.15** | **2.58*** |
| lower bound | -0.10 | -0.04 |
| upper bound | 0.39 | 5.21 |
| N | 58,519 | 58,519 |

*Significant at 10%; **significant at 5%; ***significant at 1%.

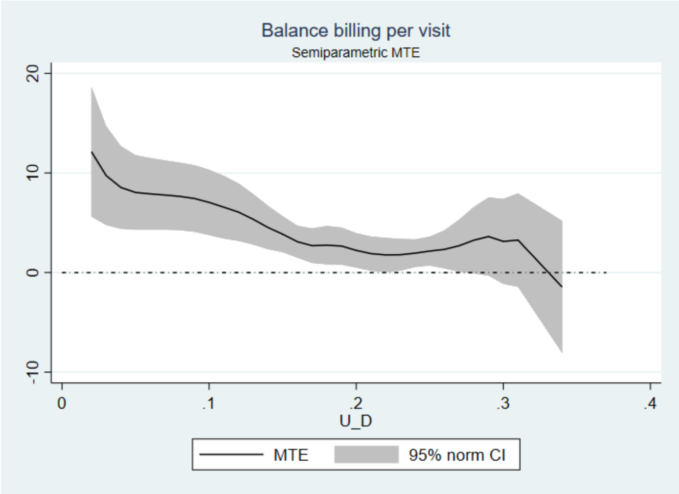
Bootstrap standard errors in brackets.

a. Computed by *margte* command (Brave et al. 2014) on the common support only.

b. $MTE\{x,p\} = x(\beta_1 - \beta_0) + \varphi_1 + \varphi_2 p + \varphi_3 p^2$, covariates at their mean.

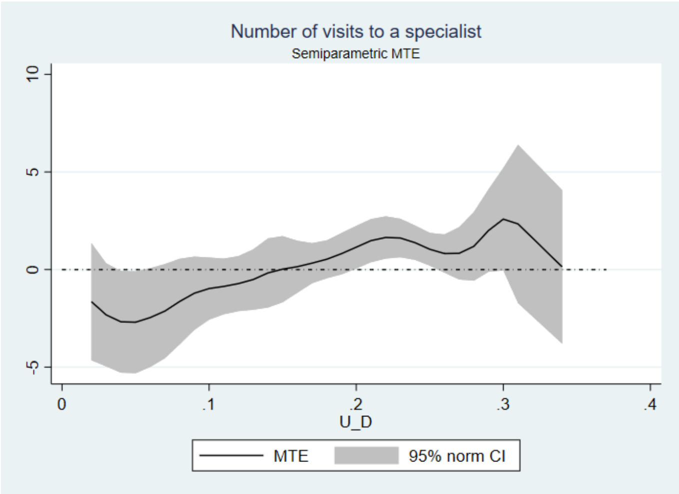
c. Covariates at their mean.

FIGURE 6. Balance billing per visit ($log(BB/Q)$): MTE, semiparametric estimates



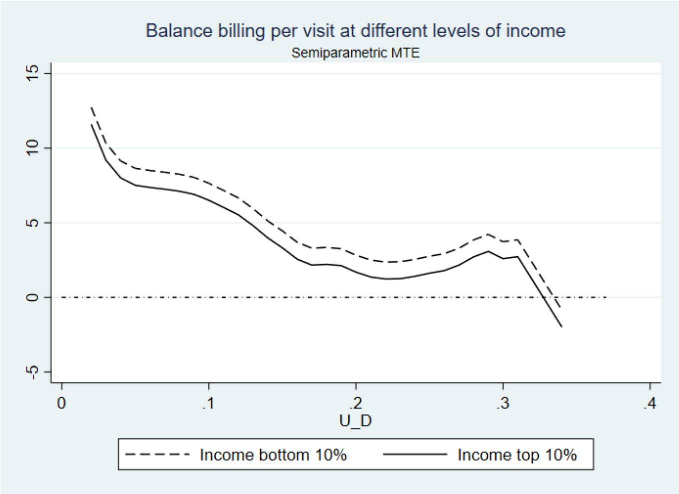

All covariates at their mean.

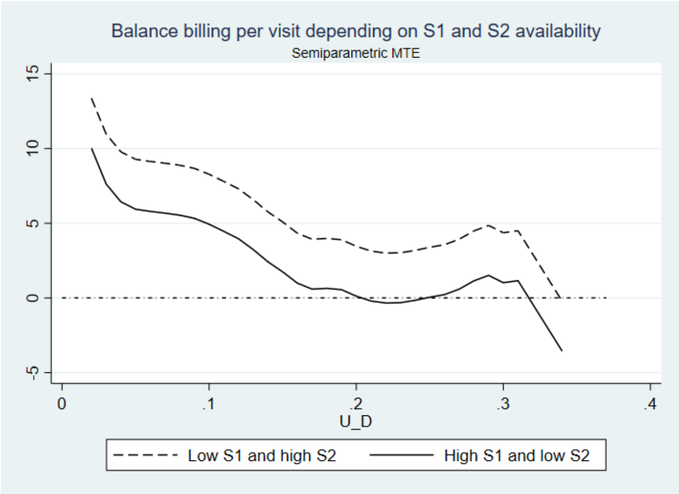
FIGURE 7. Number of visits to a specialist ($log(Q)$): MTE, semiparametric estimates




All covariates at their mean.

FIGURE 8. Balance billing per visit ($log(BB/Q)$): MTE at different levels of covariates




All other covariates at their mean.

## Appendix

Table A.1. Main estimates: share of sector 2 (S2) visits ($log(Q2/Q)$)

| | [1] Share of S2 visits without BB coverage[a] $\beta_0$[c] | [2] Propensity to switch to SHI-plus[b] $\gamma$[d] | [3] Moral hazard on share of S2 visits[a] $(\beta_1 - \beta_0)$[c] |
|---|---|---|---|
| Log(income) | 0.09*** (0.01) | -0.14*** (0.03) | -0.19*** (0.07) |
| Women | -0.01 (0.01) | 0.08*** (0.02) | -0.19*** (0.06) |
| 20-40 yo. | 0.00 (0.03) | 0.60*** (0.02) | -0.15 (0.10) |
| 40-60 yo. | ref. | ref. | ref. |
| 60+ yo. | 0.02* (0.01) | -0.29*** (0.02) | 0.69*** (0.19) |
| Chronic disease | 0.01 (0.01) | -0.20*** (0.02) | -0.12 (0.07) |
| Expensive physicians | 0.14*** (0.01) | 0.09*** (0.02) | -0.01 (0.05) |
| *Sector 1 availability* | | | |
| High | ref. | ref. | ref. |
| Medium | -0.01 (0.00) | 0.01 (0.02) | -0.03 (0.04) |
| Low | 0.05*** (0.01) | 0.21*** (0.02) | 0.22*** (0.05) |
| *Sector 2 availability* | | | |
| Low | ref. | ref. | ref. |
| High | 0.12*** (0.01) | 0.24*** (0.03) | 0.27*** (0.06) |
| *Excluded instrument* | | | |
| Early retirees | .. | 1.36*** (0.07) | .. |
| N | 58,519 | 58,519 | 58,519 |

*Significant at 10%; **significant at 5%; ***significant at 1%.

Bootstrap standard errors in brackets.

a. The dependent variable is a log transformation of the share of S2 visits.

b. The dependent variable equals 1 if individuals switched to SHI-plus, 0 otherwise.

c. Coefficients from Local IV estimation - see equation (16).

d. Coefficients from Probit estimation.

TABLE A.2. Main estimates: balance billing per S2 visit ($log(BB/Q2)$)

| | [1] BB per S2 visit without BB coverage[a] $\beta_0$[c] | [2] Propensity to switch to SHI-plus[b] $\gamma$[d] | [3] Moral hazard on BB per S2 visit[a] $(\beta_1 - \beta_0)$[c] |
|---|---|---|---|
| Log(income) | 0.26*** (0.03) | -0.16*** (0.03) | -0.74*** (0.18) |
| Women | 0.05** (0.02) | 0.08*** (0.03) | -0.28* (0.15) |
| 20-40 yo. | -0.24* (0.09) | 0.57*** (0.02) | 0.11 (0.29) |
| 40-60 yo. | ref. | ref. | ref. |
| 60+ yo. | 0.16*** (0.03) | -0.29*** (0.03) | 0.47 (0.53) |
| Chronic disease | 0.13*** (0.03) | -0.21*** (0.03) | -0.09 (0.15) |
| Expensive physicians | 0.16*** (0.02) | 0.08*** (0.03) | 0.43** (0.16) |
| *Sector 1 availability* | | | |
| High | ref. | ref. | ref. |
| Medium | -0.13*** (0.01) | 0.02 (0.03) | 0.17 (0.13) |
| Low | -0.02 (0.03) | 0.22*** (0.02) | 0.69*** (0.16) |
| *Sector 2 availability* | | | |
| Low | ref. | ref. | ref. |
| High | 0.13*** (0.03) | 0.22*** (0.04) | 0.76*** (0.18) |
| *Excluded instrument* | | | |
| Early retirees | .. | 1.37*** (0.09) | .. |
| N | 33,332 | 33,332 | 33,332 |

*Significant at 10%; **significant at 5%; ***significant at 1%.

Bootstrap standard errors in brackets.

a. The dependent variable is a log transformation of the average BB per S2 visit.

b. The dependent variable equals 1 if individuals switched to SHI-plus, 0 otherwise.

c. Coefficients from Local IV estimation - see equation (16).

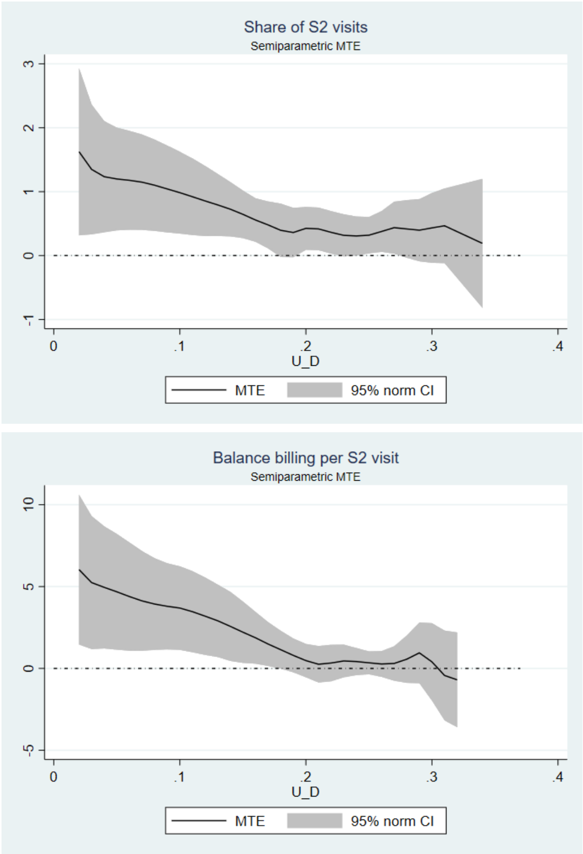
d. Coefficients from Probit estimation.

TABLE A.3. Evidence of essential heterogeneity: empirical ATE, parametric MTE, semiparametric MTE

| | Share of S2 visits $log(Q2/Q)$ | BB per S2 visit $log(BB/Q2)$ |
|---|---|---|
| **Empirical ATE**[ac] | 0.56*** (0.21) | 0.63 (0.73) |
| **Parametric MTE**[bc] | | |
| Polynomial coef. $p$ | -3.74** (1.60) | -16.93*** (4.13) |
| Polynomial coef. $p^2$ | 3.05** (1.50) | 12.12*** (3.76) |
| Joint test (p-value) | 0.007 | 0.000 |
| **Semiparametric MTE**[c] | | |
| **MTE at p=0.1** | **0.98**** | **3.69**** |
| lower bound | 0.34 | 1.13 |
| upper bound | 1.62 | 6.24 |
| **MTE at p=0.2** | **0.43**** | **0.48** |
| lower bound | 0.09 | -0.54 |
| upper bound | 0.76 | 1.50 |
| **MTE at p=0.3** | **0.43** | **0.40** |
| lower bound | -0.11 | -1.97 |
| upper bound | 0.98 | 2.77 |
| N | 58,519 | 33,332 |

*Significant at 10%; **significant at 5%; ***significant at 1%.
Bootstrap standard errors in brackets.
a. Computed by *margte* command (Brave et al. 2014) on the common support only.
b. $MTE\{x,p\} = x(\beta_1 - \beta_0) + \varphi_1 + \varphi_2 p + \varphi_3 p^2$
c. Covariates at their mean.

FIGURE A.1. Share of S2 visits ($log(Q2/Q)$) and balance billing per S2 visit ($log(BB/Q2)$): MTE, semiparametric estimates




Covariates at their mean.

Table A.4. Robustness checks: separate approach and restricted sample; balance billing per visit ($log(BB/Q)$)

| | [1] Separate approach[a] | | [2] Restricted sample[b] | |
|---|---|---|---|---|
| | $\beta_0$ | $(\beta_1 - \beta_0)$ | $\beta_0$ | $(\beta_1 - \beta_0)$ |
| Log(income) | 0.37*** (0.02) | -0.13* (0.07) | 0.52*** (0.09) | -1.53*** (0.53) |
| Women | 0.00 (0.01) | -0.03 (0.06) | .. | .. |
| Age | 0.01*** (0.00) | -0.01** (0.00) | 0.01** (0.00) | -0.03* (0.02) |
| Chronic disease | -0.02 (0.02) | -0.04 (0.07) | 0.03 (0.06) | 0.00 (0.47) |
| Expensive physicians | 0.86*** (0.01) | 0.10* (0.05) | 0.78*** (0.04) | 0.24 (0.41) |
| *Sector 1 availability* | | | | |
| High | ref. | ref. | ref. | ref. |
| Medium | -0.10*** (0.01) | 0.00 (0.05) | 0.00 (0.05) | -0.39 (0.48) |
| Low | 0.42*** (0.02) | 0.18*** (0.05) | 0.49*** (0.05) | 0.27 (0.39) |
| *Sector 2 availability* | | | | |
| Low | ref. | ref. | ref. | ref. |
| High | 0.79*** (0.02) | 0.15* (0.08) | 0.71*** (0.06) | 1.38*** (0.42) |
| Polynomial coefficients of MTE | $p$ | $p^2$ | $p$ | |
| | -13.8*** (4.78) | 14.33* (7.59) | -3.24** (1.67) | |
| Joint test of significance (p-value) | 0.0037 | | 0.0069 | |
| N | 58,519 | 58,519 | 20,251 | 20,251 |

*Significant at 10%; **significant at 5%; ***significant at 1%.

Bootstrap standard errors in brackets.

The dependent variable is a log transformation of the average balance billing per visit to a specialist.

a. Semiparametric MTE of degree 2 - run with *mtefe* command (Andresen et al. 2018).

b. Women aged 35-60 - local IV, parametric MTE of degree 1.

## References

Andresen, Martin Eckhoff et al. (2018). "Exploring marginal treatment effects: Flexible estimation using Stata." *Stata Journal*, 18(1), 118–158.

Basu, Anirban, James J Heckman, Salvador Navarro-Lozano, and Sergio Urzua (2007). "Use of instrumental variables in the presence of heterogeneity and self-selection: An application to treatments of breast cancer patients." *Health economics*, 16(11), 1133–1157.

Bjorklund, Anders and Robert Moffitt (1987). "The estimation of wage gains and welfare gains in self-selection models." *The Review of Economics and Statistics*, pp. 42–49.

Brave, Scott, Thomas Walstrum, et al. (2014). "Estimating marginal treatment effects using parametric and semiparametric methods." *Stata Journal*, 14(1), 191–217.

Brinch, Christian N, Magne Mogstad, and Matthew Wiswall (2017). "Beyond LATE with a discrete instrument." *Journal of Political Economy*, 125(4), 985–1039.

Buchmueller, Thomas C and Agnes Couffinhal (2004). "Private health insurance in France."

Buchmueller, Thomas C and Paul J Feldstein (1997). "The effect of price on switching among health plans." *Journal of Health Economics*, 16(2), 231–247.

Cameron, A Colin, Pravin K Trivedi, Frank Milne, and John Piggott (1988). "A microeconometric model of the demand for health care and health insurance in Australia." *The Review of economic studies*, 55(1), 85–106.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil (2011). "Estimating Marginal Returns to Education." *American Economic Review*, 101, 2754–2781.

Célant, Nicolas, Stéphanie Guillaume, and Thierry Rochereau (2014). "Enquête sur la santé et la protection sociale 2012." *Les rapports de l'Irdes*, (556).

Chiappori, Pierre-André, Franck Durand, and Pierre-Yves Geoffard (1998). "Moral hazard and the demand for physician services: First lessons from a French natural experiment." *European economic review*, 42(3), 499–511.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg (2016). "From LATE to MTE: Alternative methods for the evaluation of policy interventions." *Labour Economics*, 41, 47–60.

Cornelissen, Thomas, Christian Dustmann, Uta Schönberg, and Anna Raute (2017). "Who benefits from universal child care?: Estimating marginal returns to early child care attendance." *Journal of Political Economy*.

Coulson, N Edward, Joseph V Terza, Cheryl A Neslusan, and Bruce C Stuart (1995). "Estimating the moral-hazard effect of supplemental medical insurance in the demand for prescription drugs by the elderly." *The American Economic Review*, 85(2), 122–126.

Dormont, B., P.Y. Geoffard, and K. Lamiraud (2009). "The influence of supplementary health insurance on switching behaviour: evidence from Swiss data." *Health Economics*, 18(11), 1339–1356.

Dormont, Brigitte and Mathilde Péron (2016). "Does health insurance encourage the rise in medical prices? A test on balance billing in France." *Health economics*, 25(9), 1073–1089.

Duijmelinck, Danielle MID and Wynand PMM van de Ven (2016). "Switching rates in health insurance markets decrease with age: empirical evidence and policy implications from the Netherlands." *Health Economics, Policy and Law*, 11(02), 141–159.

Einav, Liran and Amy Finkelstein (2018). "Moral hazard in health insurance: What we know and how we know it." *Journal of the European Economic Association*, 16(4), 957–982.

Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen (2013). "Selection on Moral Hazard in Health Insurance." *The American Economic Review*, 103(1), 178–219.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group (2012). "The Oregon health insurance experiment: evidence from the first year." *The Quarterly journal of economics*, 127(3), 1057–1106.

Friedman, Milton and Leonard J Savage (1948). "The utility analysis of choices involving risk." *The journal of political economy*, pp. 279–304.

Heckman, James J and Richard Robb (1985). "Alternative methods for evaluating the impact of interventions: An overview." *Journal of econometrics*, 30(1-2), 239–267.

Heckman, James J, Sergio Urzua, and Edward Vytlacil (2006). "Understanding instrumental variables in models with essential heterogeneity." *The Review of Economics and Statistics*, 88(3), 389–432.

Heckman, James J and Edward Vytlacil (2001). "Policy-relevant treatment effects." *The American Economic Review*, 91(2), 107–111.

Heckman, James J and Edward Vytlacil (2005). "Structural equations, treatment effects, and econometric policy evaluation 1." *Econometrica*, 73(3), 669–738.

Heckman, James J and Edward J Vytlacil (2007). "Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments." *Handbook of econometrics*, 6, 4875–5143.

Holly, Alberto, Lucien Gardiol, Gianfranco Domenighetti, and Brigitte Bisig (1998). "An econometric model of health care utilization and health insurance in Switzerland." *European economic review*, 42(3), 513–522.

Jones, Andrew M, Xander Koolman, and Eddy Van Doorslaer (2006). "The impact of having supplementary private health insurance on the use of specialists." *Annales d'Economie et de Statistique*, pp. 251–275.

Kowalski, Amanda E (2018). "Extrapolation using Selection and Moral Hazard Heterogeneity from within the Oregon Health Insurance Experiment." Tech. rep., National Bureau of Economic Research.

Manning, Willard G, Joseph P Newhouse, Naihua Duan, Emmett B Keeler, and Arleen Leibowitz (1987). "Health insurance and the demand for medical care: evidence from a randomized experiment." *The American economic review*, pp. 251–277.

McKnight, Robin (2007). "Medicare balance billing restrictions: Impacts on physicians and beneficiaries." *Journal of health economics*, 26(2), 326–341.

Newhouse, Joseph P (1993). *Free for all?: lessons from the RAND health insurance experiment*. Harvard University Press.

Nyman, John A (1999). "The value of health insurance: the access motive." *Journal of health economics*, 18(2), 141–152.

Nyman, John A (2003). *The theory of demand for health insurance*. Stanford University Press.

OECD (2015). *Health at a Glance 2015*. URL https://www.oecd-ilibrary.org/content/publication/health_glance-2015-en.

Pauly, Mark V (1968). "The economics of moral hazard: comment." *The American Economic Review*, pp. 531–537.

Sagan, Anna and Sarah Thomson (2016). *Voluntary Health Insurance in Europe: Role and Regulation: Health Policy Series*. 43, World Health Organization.

Schellhorn, Martin (2001). "The effect of variable health insurance deductibles on the demand for physician visits." *Health Economics*, 10(5), 441–456.

Stabile, Mark and M Townsend (2014). "Supplementary private health insurance in national health insurance systems." *In Encyclopedia of Health Economics, Volume 3, edited by Anthony J. Culyer*, p. 362–65.

Strombom, Bruce A, Thomas C Buchmueller, and Paul J Feldstein (2002). "Switching costs, price sensitivity and health plan choice." *Journal of Health economics*, 21(1), 89–116.

Vera-Hernández, Ángel Marcos (1999). "Duplicate coverage and demand for health care. The case of Catalonia." *Health economics*, 8(7), 579–598.