ABSTRACT

ZHENG, YAN. Demand Side Economics of Health Care Provision Under a Single Payer System: The Case of Croatia. (Under the direction of Tomislav Vukina.)

This dissertation consists of three chapters, connected via same data, examining asymmetric information problems in different cohorts.

In the first chapter, we use a structural approach to separately estimate moral hazard and adverse selection effects in health care utilization using hospital invoices data. Our model explicitly accounts for the heterogeneity in the opportunity cost of time for hospital visits which raises the total cost of hospital visits and dampens the moral hazard effect. A measure of moral hazard is derived as the difference between the observed and the counterfactual health care consumption. In the population of patients with non life-threatening diagnoses, our results indicate statistically significant and economically meaningful moral hazard. We also test for the presence of adverse selection by investigating whether patients with different health status sort themselves into different health insurance plans. Adverse selection is confirmed in the data because patients with estimated worse health tend to buy the insurance coverage and patients with estimated better health choose not to buy the insurance coverage.

The second chapter uses truncated count model and simulated maximum likelihood estimation technique to estimate gender differences in moral hazard in health care insurance. We use the dataset which consists of invoices for all outpatient services from a regional hospital in Croatia. Our theoretical model predicts that higher risk aversion is associated with smaller moral hazard effect. Given empirical evidence which seem to be supportive of the fact that women are more risk-averse then men, we hypothesize that moral hazard effect due to health insurance should be lower in women than in men. However our empirical results show that gender differences in moral hazard due to health insurance are statistically insignificant.

The third chapter uses fuzzy regression discontinuity design to estimate the moral hazard effect in health care consumption in the population of young adults. We use invoices data for outpatient hospital services from a regional hospital in Croatia. The estimation is complicated by the fact that the data set consists only of users of medical services. To address this issue we use a modified version of instrumental variables approach and found an 89% reduction in the number of hospital visits for individuals who lost insurance coverage when crossing the 18th birthday threshold.
Demand Side Economics of Health Care Provision Under a Single Payer System: The Case of Croatia

by
Yan Zheng

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Economics
Raleigh, North Carolina
2015

APPROVED BY:

______________________________  ________________________________
Xiaoyong Zheng                  Barry Goodwin

______________________________  ________________________________
Melinda Morrill                 Tomislav Vukina
Chair of Advisory Committee
DEDICATION

This dissertation is dedicated to my mother Liqun. Thank you for your unconditional love, support, and sacrifice throughout my life.
BIOGRAPHY

Yan was born and raised in a small town in the Liaoning Province of China. After graduating with high school diploma in 2005, she attended Central University of Finance and Economics for her undergraduate degree in Beijing. She was chosen to be a member of an elite class for Mathematical Economics and Finance learning from the first semester. She was an active member at the student union and awarded with student fellowship. She graduated with Bachelors of Science in Mathematics and Bachelors in Art in Economics in May 2009.

In August 2009, she came to US and pursued her Doctoral degree in Department of Economics at North Carolina State University in Raleigh, North Carolina. She was heavily involved in graduate life. She worked as a research assistant and teaching assistant for the department of Economics. She interned at Blue Cross Blue Shield Association in 2014 for the sake of her dissertation research. She obtained her Ph. D in May 2015, Her doctoral research, under the direction of Dr. Tomislav Vukina, focused on the demand side economics of health care provision under a single payer system. After graduation, she will pursue an academic career in a Chinese University.
ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family.

First and foremost, I would like to express my deepest gratitude to my advisor Dr. Tomislav Vukina for his guidance, understanding, patience, and most importantly, his friendship during my graduate studies at North Carolina State University. His mentorship was paramount in providing a well rounded experience consistent with my long-term career goals. He encouraged me to not only grow as an economist but also as an independent thinker, a good writer. Without him I would not have achieved my goals for this dissertation. A special thanks for his countless hours of reading, revising and editing my writings. I would also like to sincerely thank Dr. Xiaoyong Zheng for his assistance and guidance in my estimation and providing me with the foundation of programming skills. I would not be able to accomplish my dissertation without him generously supporting me with his expertise and precious time. I also want to thank the members of my dissertation committee, Dr. Melinda Sandler Morrill, Dr. Barry Goodwin for your valuable and insightful comments about my research and academic career path.

There is a special group of graduate students friends whom without their support and understanding in my life I would not have made it in the past few years. Thank you Yiyang Yin, Xia He, Lu Jin, Jie Gong, Huijie Yu, Tsung-Yu Yang, Jian Wen, Zhen Wang, Jieyuan Zhao, Xiaomei Chen, Haozhi Wei, Muyi Yang, Moonhee Cho, Juan He, Sui Wu for making my six years here wonderful.

Church has played an important role in my life in Raleigh, and I sincerely thank those whom I connected with at Living Water Christian Church and Raleigh Chinese Christian Church. I thank my godfather Robert Ching and godmother Laura Ching, for welcoming me when I first arrived at US, helping me get through the first few weeks, and teaching me about their faith. I acknowledge other friends in Christ include Mingxin Wang, Ming Yan for taking care of me like I were their child.

I would like to thank my family members. My mother Liqun Wang, without her I would not have made it this far in life. She has been there for me every step of the way, has always loved me unconditionally, and has aided me through all of my tough decisions. Without her full love, support and sacrifice, I never would have realized my full potential. I thank her everyday of my life. I would also like to express my deepest gratitude to my grandparents, Yin Tao and Jiamin Wang for their support, encouragement, quiet patience and unwavering love. They nurtured and loved me since I was little. I gained the ability to be strong and tackle challenges growing up around them. I also want to thank my uncle Shaofeng Wang, Shaohua Wang, my aunt Lihua Wang. They were always supporting me and encouraging me with their best wishes, being there
for me whenever I needed them. Also, I thank my father Baocheng Zheng, my grandparents,
for the sweet memories I share with them and the time I spent with them.
# TABLE OF CONTENTS

LIST OF TABLES .............................................. vii

LIST OF FIGURES ........................................ viii

### Chapter 1 Estimating Asymmetric Information Effects in Health Care Accounting for the Opportunity Cost of Time .......................... 1

1.1 Introduction ............................................. 1
1.2 Institutional Framework and Data Description ......................... 3
1.3 The model ............................................... 6
1.4 Estimation ............................................... 8
    1.4.1 Income Prediction ................................. 8
    1.4.2 Model Identification ............................. 9
    1.4.3 Moral Hazard .................................... 11
    1.4.4 Adverse Selection ............................... 12
1.5 Empirical Results ..................................... 13
1.6 Conclusion ............................................. 16

### Chapter 2 Risk Aversion, Moral Hazard and Gender Differences in Health Care Utilization .......................... 22

2.1 Introduction ............................................. 22
2.2 Theoretical Framework ................................ 25
2.3 Gender Difference in Risk Aversion .......................... 27
2.4 Institutional Framework and Data Description ......................... 28
2.5 Estimation ............................................... 29
2.6 Empirical Results ..................................... 32
2.7 Conclusion ............................................. 33

### Chapter 3 Using Age-Based Insurance Eligibility Criterion to Estimate Moral Hazard in Medical Care Consumption .................. 40

3.1 Introduction ............................................. 40
3.2 Institutional Framework and Data Description ......................... 42
3.3 Regression Discontinuity Design ................................ 44
    3.3.1 Instrumental Variable Approach .................. 45
3.4 Estimation and Results .................................. 48
3.5 Conclusion ............................................. 52

### Chapter 4 General Conclusion ................................ 59

References .................................................. 61
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Summary Statistics by Insurance Category</td>
<td>18</td>
</tr>
<tr>
<td>Table 1.2</td>
<td>Summary Statistics: CHBS Data</td>
<td>18</td>
</tr>
<tr>
<td>Table 1.3</td>
<td>Regression Results for Income in the Survey Data</td>
<td>19</td>
</tr>
<tr>
<td>Table 1.4</td>
<td>Summary Statistics: Predicted County Income and Premia</td>
<td>19</td>
</tr>
<tr>
<td>Table 1.5</td>
<td>Estimates of Model Parameters</td>
<td>19</td>
</tr>
<tr>
<td>Table 1.6</td>
<td>Counterfactual Effects of Moral Hazard</td>
<td>20</td>
</tr>
<tr>
<td>Table 1.7</td>
<td>Testing for Adverse Selection: Logistic Regression of Insurance Choices</td>
<td>20</td>
</tr>
<tr>
<td>Table 1.8</td>
<td>Summary Statistics: Predicted County Income and Premia</td>
<td>19</td>
</tr>
<tr>
<td>Table 1.9</td>
<td>Estimates of Model Parameters</td>
<td>19</td>
</tr>
<tr>
<td>Table 1.10</td>
<td>Counterfactual Effects of Moral Hazard</td>
<td>20</td>
</tr>
<tr>
<td>Table 1.11</td>
<td>Testing for Adverse Selection: Logistic Regression of Insurance Choices</td>
<td>20</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>Summary of Literature about Gender Difference in Risk Aversion</td>
<td>35</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Summary Statistics by Insurance Category and Gender</td>
<td>37</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Summary Statistics by Insurance Category and Employment Status</td>
<td>37</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Top 10 Diagnoses Subject to Co-payment by Gender</td>
<td>38</td>
</tr>
<tr>
<td>Table 2.5</td>
<td>Probit Regression: Marginal Effects of Independent Variables to the Probabi-</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>lity of Purchasing Insurance</td>
<td></td>
</tr>
<tr>
<td>Table 2.6</td>
<td>Maximum Likelihood Estimation: Hospital Visits</td>
<td>39</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Summary Statistics: One-year Window Around the Age-18</td>
<td>54</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Seemingly Unrelated Regression Results: One-Year Window</td>
<td>54</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>OLS Regression Results: One-Year Window</td>
<td>55</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Percent Change in Uninsured (First Stage) and Total Visits (Reduced Form)</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>and the Moral Hazard (IV): One-Year Window</td>
<td></td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Summary Statistics: Two-year Window Around the Age-18</td>
<td>55</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Seemingly Unrelated Regression Results: Two-Year Window</td>
<td>56</td>
</tr>
<tr>
<td>Table 3.7</td>
<td>OLS Regression Results: Two-Year Window</td>
<td>56</td>
</tr>
<tr>
<td>Table 3.8</td>
<td>Percent Change in Uninsured (First Stage) and Total Visits (Reduced Form)</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>and the Moral Hazard (IV): Two-Year Window</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1.1  Empirical CDF of θ for two groups (Enlarged) . . . . . . . . . . . . . . . . . 21
Figure 3.1  Uninsured visits surrounding patients’ 18th birthday (two-year window) . . . 57
Figure 3.2  Hospital visits surrounding patients’s 18th birthday (two-year window) . . . 58
Chapter 1

Estimating Asymmetric Information Effects in Health Care Accounting for the Opportunity Cost of Time

1.1 Introduction

The empirical literature dealing with the estimation of moral hazard and adverse selection effects in health care utilization is quite large. For example, Manning et al. (1987) used a randomized experiment and found that a catastrophic insurance plan reduces expenditures 31 percent relative to zero out-of-pocket price, indicating a large moral hazard effect. Using the British Household Panel Survey, Olivella and Vera-Hernandez (2013) found that adverse selection is present in the private health insurance market. Individuals who purchase private health insurance (PHI) have a higher probability of both hospitalisation and visiting their general practitioner than individuals who receive PHI as a fringe benefit from employers. Fewer studies succeeded in disentangling these two effects. For example, Wolfe and Goddeeris (1991) estimated adverse selection as the effect of lag health status or lag health expenditure on current insurance decisions in a longitudinal study of Medigap insurance. They found a nontrivial decrease in moral hazard after taking into account the adverse selection. Liu, Nestic and Vukina (2012) successfully disentangled moral hazard from adverse selection, taking advantage of the unique features of health care system in Croatia. Using matching estimators, they discovered favorable selection effect for patients in the 20-30 year cohort, adverse selection for patients in older age cohorts and significant moral hazard for all age cohorts.

Structural estimation of adverse selection and moral hazard has also been used in the health insurance literature. Earlier work was based on constructing health care demand, using price elasticities as the measure of moral hazard, and the link between insurance and health care
demand as the measure of adverse selection. Cameron et al. (1988) derived closed form demand functions for health insurance and health care for a risk-averse consumer under uncertainty. Using data from 1977-1978 Australian Health Survey, they found a statistical dependence between error terms of health insurance and health care demand equations which suggest the presence of adverse selection. They also found a significant price effect in health care consumption which identifies the moral hazard as an important determinant of the overall health care utilization. Cardon and Hendel (2001) integrated health insurance and health care demand using 1987 National Medical Expenditure Survey data. Their objective was to estimate whether consumers’ private information is an important link between insurance and health care demand. They did not find any evidence of adverse selection but found that the gap in expenditure between insured and uninsured can be attributed to observable demographic differences and to price sensitivity. They concluded that this price elasticity to coinsurance rate is an evidence of moral hazard. Vera-Hernandez (2003) introduced a new measure of moral hazard based on the conditional correlation between contractible (treatment cost) and non-contractible (penalty health shocks) variables. If this correlation is zero, the relationship between the health shocks and treatment costs would be deterministic and then there would be no moral hazard. Using RAND Health Insurance Experiment data and simulated maximum likelihood estimator, the author uncovered the structural model parameters. The results reject the nonexistence of moral hazard at 95% confidence level with about two-thirds of the variance of residual health penalty shock explained by the cost and one third left unexplained, a measure of moral hazard. Gardiol, Geoffard and Grandchamp (2005) developed a structural model of joint demand for health insurance and health care. They took advantage of the design feature of the Swiss insurance system where consumers choose their coverage from a menu of insurance plans ranked by the size of their deductibles. They used insurance claims data and found evidence of both selection and moral hazard effects. Their results indicate that 75% of the correlation between insurance coverage and health care expenditures may be attributed to selection and 25% to ex post moral hazard. The most closely related to our paper is Bajari et al. (2014). They proposed a two-step semiparametric estimation strategy to disentangle the adverse selection and moral hazard effects in medical care. They used a unique claims data from a large self-insured employer in the U.S. In the theoretical model, they introduced the latent health status parameter into the utility function and used the GMM approach to estimate it. Moral hazard is then estimated as the difference between health care consumption of people with the insurance and their counterfactual consumption with no coverage. Adverse selection is examined by comparing health status distributions of people in different insurance plan groups. The fact that people with worse latent health status sort themselves into insurance plan with higher coverage is an evidence of adverse selection.

In this paper we use the structural approach to estimate the effects of moral hazard and
adverse selection relying on the hospital invoices data for non life-threatening diagnoses. The data covers a four-month period in 2009 of all outpatient services provided by a small regional hospital in Croatia. Croatia has a government controlled health care system with a single payer insurance fund. Our main contribution is to extend the model by Bajari et al. (2014) to account for the heterogeneity in the opportunity cost of time for hospital visits over and above the direct health care cost. We found that the opportunity cost of time for hospital visits is large and statistically significant. On average, it amounts to 42% of the incurred medical expenses compared to the average out-of-pocket co-payment of only about 30%. We also found a counterfactual evidence of moral hazard. If one takes away the insurance from people with coverage, they would decrease their health care consumption by 32 HRK (8%). If you give the insurance to people without coverage, they would increase their health consumption by 27 HRK (9%). The presence of adverse selection was investigated by comparing the empirical distributions of the estimated latent health status across groups of patients with different insurance types relying on the Kolmogorov-Smirnov test. The results indicate the evidence of adverse selection in the sense that patients without the insurance are relatively healthier than patients who bought the coverage.

1.2 Institutional Framework and Data Description

The main part of the data for this paper comes from Liu, Nestic and Vukina (2012). The original data set consists of all invoices for all outpatient services from a regional hospital in Croatia during the period from March 1 to June 30, 2009. The health care system in Croatia is dominated by a single public health insurance fund: the Croatian Institute for Health Insurance (HZZO). The HZZO offers two types of insurance: the compulsory insurance and the supplemental insurance. The compulsory insurance’s coverage is universal and it is funded by a 15% payroll tax whereas the supplemental insurance can be either bought or is extended automatically free of charge to certain categories of citizens. The Croatian public health care system as provided by the HZZO is very similar to the Medicare system in the U.S. The main difference is that the Croatian compulsory insurance insures all citizens whereas Medicare provides limited public health insurance for the 65 years and older citizens. The individuals covered by Medicare can choose to purchase Medigap to cover the gaps in Medicare such as co-pays, deductibles and uncovered expenses (e.g. prescription drugs, prolonged hospital stays, etc.). Medigap in the U.S. system plays the same role as the supplemental insurance for Croatian citizens (for details see Liu, Nestic and Vukina, 2012).

The full coverage services afforded by the compulsory insurance include: full health care for children under the age of 18, health care of women related to pregnancy and child birth, preventive and curative health care related to infectious diseases, mandatory vaccinations and
immunizations, hospital care for all chronic psychiatric patients, complete treatment of all
cancers, dialysis, organ transplants, emergency room interventions, house-calls and at-home
treatment of patients and prescription drugs from the HZZO basic list. All other health services
are subject to a system of co-payments. The insured are required to pay certain percentage of
the full price of medical care, for example: laboratory, radiological and other diagnostics at
the primary health care level (15.00 HRK), specialists’ visits and all out-patient services except
physical therapy and rehabilitation (25.00 HRK), specialists’ diagnostics not at the primary care
level (50 HRK), orthopedic and prosthetic devices (50 HRK), out-patient and at-home physical
therapy and rehabilitation (25 HRK per day), in-patient care (100 HRK per day), primary care
including family physician, gynecologist and dentist (15 HRK), etc. The largest out-of-pocket
cost-share amount that a person can pay amounts to 3,000.00 HRK per one invoice.¹

Supplemental insurance is a voluntary insurance that can be acquired by a person 18 years
of age or older, having compulsory insurance, by signing a contract with HZZO. A person having
the supplemental insurance policy is entitled to full waiver of all medical expense co-payments
listed above. The premiums for supplemental insurance are determined as follows: 50.00 HRK
per month for a retired person with the monthly pension of less than 5,108.00 HRK; 80.00
HRK per month for a retired person with pension higher than 5,108.00 HRK; 80.00 HRK per
month for an active person with a monthly net income less than 5,108.00 HRK; 130.00 HRK
per month for an active person with net income in excess of 5,108.00 HRK; 80.00 HRK per
month for all family members and dependents.

An interesting feature of the supplemental insurance program is that certain categories of
people are exempt from paying the supplemental insurance; in other words, they are entitled to
it automatically for free. The list of exemptions is quite long. The top five largest categories are
poor people (59.17%), single pensioners (14.11%), people with 80% physical disability (10.86%),
blood donors (3.48%) and war veterans (2.96%). Therefore, with respect to the supplemental
insurance coverage we distinguish three categories of patients: those that bought the insurance
(Bought group), those that received the exact same insurance for free (Free group) and those
that do not have the insurance (No group). The analysis in the rest of the paper involves only the
comparison of health care consumption for diagnoses subject to the system of co-payments, i.e.
all other diagnoses covered in full by the universal compulsory insurance program are ignored.
Hence when we refer to patients with no insurance, we mean patients with no supplemental
insurance, i.e. those that are only covered by the universal (compulsory) program.

The data set consists of 105,646 observations. Each observation reflects the invoice for
one hospital visit. For the purpose of this estimation, patients who visited the hospital only
because of the illnesses that are fully covered by the compulsory insurance are excluded. We

¹Listed co-payments were valid for 2009. The exchange rate for the local currency, Croatian Kuna (HRK), as
of June 20, 2009 was 1USD=5.19 HRK.
also delete patients that are younger than 18 because there is no variation in the type of insurance coverage for this group of people, i.e. they are all entitled to supplemental insurance for free. The data contains the following set of variables: a numeric code for the type of hospital service provided, compulsory health insurance number, supplemental insurance number (if the patient has one), period covered by the supplemental insurance, numeric code for categories entitled to supplemental insurance free of charge, eligibility category for compulsory insurance ($k_1$–employed, $k_2$–farmers, $k_3$–pensioners, $k_4$–unemployed, $k_5$–living on social welfare, $k_6$–self-employed and other), cost of hospital service, part of the cost covered by compulsory insurance, part of cost covered by supplemental insurance, part of cost covered by participation (co-payment), date of birth and sex of the patient. We measure health care utilization using total cost per patient during the four-month period covered by the data. The working data set consists of 70,851 invoices for 22,903 patients. The summary statistics of the working data set is displayed in Table 1.1. The No group has smaller cost, younger population and larger percentage of men relative to the Bought and the Free group. Each patient without the supplemental insurance (No group) paid on average 77.4 HRK in out-of-pocket co-payments during the 4-month period which amounts to less than 1% of their income during the same time period. For patients in the other two groups, the co-payment is zero because they are fully covered by the supplemental insurance.

In addition to invoice data set, we also use the Croatian Household Budget Survey (CHBS) for 2009. The CHBS data set is used for the purpose of forecasting the income variable for patients in the invoices data set. The dataset contains information on individual’s age, gender, eligibility category for compulsory insurance, supplemental insurance status, household income, income per capita (household income divided by the number of household members), individual income by source (employment, self-employment, unemployment benefits, pensions) and county of residence. The income variable used is defined as the larger amount between the household income per capita and the sum of individual income from all sources. There are a total of 8,269 individuals in the CHBS data set; 1,430 of them are under the age of 18 and have been dropped. Among the remaining 6,839 observations, 305 are the residents of the county which coincides with the region from which our hospital draws its patients.\footnote{The actual name of the county is suppressed for confidentiality reason because revealing the name of the county would automatically reveal the name of the hospital.} The summary statistics for the 6,839 observations sub-sample, broken down by active versus retired population, is displayed in Table 1.2. It is interesting to note that there are people with zero income and also that the difference in mean income between the active and retired citizens is not very large. Active people earn on average only HRK 2,593 (US$ 500) annually more than the retired people.
1.3 The model

The model relies on the assumption of a rational economic agent who maximizes his utility function by choosing his optimal health care services $m$ and consumption of composite commodity $c$ subject to a budget constraint. The utility function is additive in aggregate consumption and health care, each in the constant relative risk aversion (CRRA) functional form:

$$U(c_i, m_i; \theta_i, \gamma) = (1 - \theta_i)^{\gamma_1} c_i^{1 - \gamma_1} + \theta_i m_i^{1 - \gamma_2}$$  \hspace{1cm} (1.1)

where $\gamma_1$ and $\gamma_2$ are risk-aversion coefficients for aggregate and health care consumption. The larger the coefficients, the more risk averse the person with respect to variation in two types of consumption. The utility also depends on latent health status parameter $\theta \in [0, 1]$, which is known to the agent but unobservable by the insurance company. It can be interpreted as the importance weight an agent places on health care and aggregate consumption. In case of bad health, $\theta$ is close to one, and the agent would gain relatively more utility from health care consumption and less utility from other consumption.

A consumer’s budget constraint requires that his expenditure on aggregate consumption and health care must not be greater than his income minus the insurance premium:

$$c_i + m_i \beta_{ij} \leq y_i - p_j,$$  \hspace{1cm} (1.2)

where $y$ is income, $j$ denotes insurance status (no, bought, free), $p_j$ is the insurance premium and $\beta_{ij}$ is the co-payment rate which has two parts:

$$\beta_{ij} = \alpha_{ij} + \epsilon_i.$$  \hspace{1cm} (1.3)

In Equation (1.3), $\alpha_{ij}$ denotes the percentage of the cost that patients have to pay out of pocket and it depends on the insurance coverage. If the patient has no supplemental insurance, $\alpha_{ij} \in (0, 1)$; if the patient has the supplemental insurance either by buying it or being entitled to it for free, $\alpha_{ij} = 0$. Idiosyncratic cost $\epsilon_i > 0$ is the opportunity cost of time measured as a percentage of the cost of hospital services. It does not depend on the patient’s insurance status but varies with agent’s income. The idea behind the specification of the cost factor $\beta_{ij}$ is the fact that in addition to actual medical expenses which may or may not be covered by the insurance, in order to see a doctor, the patient needs to invest some time, which is costly. For an employed person, this cost could be directly related to the lost income from absenteeism, for a retired person it is the value of time in the next best alternative (e.g., baby-sitting a grandson). Notice that without $\epsilon_i$, for patients with supplemental insurance, $\beta_{ij}$ would be zero and the model would predict the consumption of infinite amount of health care.
The first order conditions for the maximization of utility (1.1) subject to budget constraint (1.2) are as follows:

\[
\frac{\partial L}{\partial c_i} = (1 - \theta_i)c_i^{\gamma_1} - \lambda = 0, \\
\frac{\partial L}{\partial m_i} = \theta_i m_i^{\gamma_2} - \lambda \beta_{ij} = 0, \\
\frac{\partial L}{\partial \lambda} = y_i - P_j - c_i - m_i \beta_{ij} = 0.
\]

The standard optimization result shows that the marginal rate of substitution between aggregate consumption and health care consumption equals to the ratio of prices. However, the price of health care that a consumer faces is only the co-payment portion \(\alpha_{ij}\) of the actual price plus the opportunity cost of time for the hospital visit \(\epsilon_i\). As a result, the relative price of medical services to aggregate consumption is equal to \(\beta_{ij}\) and the MRS becomes:

\[
MRS = \frac{\theta_i (1 - \theta_i)c_i^{\gamma_1}}{(1 - \theta_i)c_i^{\gamma_1}m_i^{\gamma_2}} = \alpha_{ij} + \epsilon_i = \beta_{ij}
\] (1.4)

Since a patient only needs to pay \(\alpha_{ij} < 1\) portion of the actual health care cost, the price ratio between health care and aggregate consumption is biased towards favoring health care and against general consumption. This “excess” of health care consumption is a standard measure of moral hazard associated with insurance. However, the presence of \(\epsilon_i\) in Equation (1.4) changes the calculation because it makes the health care consumption less attractive, thereby potentially mitigating the moral hazard effect. For some non-serious illness it could actually make a difference between seeing and not seeing a doctor.

Since the utility function is strictly increasing, the budget constraint binds at the optimal bundle. Therefore \(c_i\) is determined by the equation \(c_i = y_i - p_j - m_i \beta_{ij} = y_i - p_j - m_i (\alpha_{ij} + \epsilon_i)\), which substituted into Equation (1.4) gives:

\[
\frac{\theta_i}{(1 - \theta_i)} \frac{(y_i - p_j - m_i (\alpha_{ij} + \epsilon_i))^{\gamma_1}}{m_i^{\gamma_2}} = \alpha_{ij} + \epsilon_i.
\] (1.5)

From Equation (1.5), we could derive an expression for health shock parameter \(\theta_i\), which is then used to estimate the risk parameters \(\gamma_1\), \(\gamma_2\) and the idiosyncratic cost \(\epsilon_i\) from the observable variables:

\[
\theta_i = \frac{(\alpha_{ij} + \epsilon_i)m_i^{\gamma_2}}{(y_i - p_j - m_i (\alpha_{ij} + \epsilon_i))^{\gamma_1} + (\alpha_{ij} + \epsilon_i)m_i^{\gamma_2}}.
\] (1.6)

Here, it is important to realize that even without the supplemental insurance, the patient will almost never pay the full cost of the medical service, because some portion of it is always paid by the universal (compulsory) insurance.
Equation (1.6) is used to estimate the distribution of $\theta$ for groups of patients with various insurance coverages. One could expect that the population who bought the supplemental insurance have different $\theta$ distribution than population with no insurance. If the test statistics support that two $\theta$ distribution are drawn from different populations, this will constitute the empirical evidence of selection.

1.4 Estimation

The estimation of Equation (1.6) requires individuals’ income which is not available from the invoices data and hence needs to be forecasted. With the imputed income variable, insurance premia and other observed variables, we could estimate the risk parameters $\gamma_1$, $\gamma_2$ and the opportunity cost of time as a percentage of medical care cost $\epsilon_i$ using a generalized method of moments (GMM) estimator. With these estimated model parameters, we can back out the distribution of $\theta$ from Equation (1.6), which will then subsequently be used to construct the moral hazard counterfactuals and test for the presence of adverse selection.

1.4.1 Income Prediction

The income variable for all individuals in the invoices data set is forecasted using the CHBS data.\(^4\) The overlapping variables from the two data sets are age, gender, eligibility category for compulsory insurance, and county. All patients in the invoices data are assumed to be from the same county. In the first step we estimate the income equation using the CHBS data based on the following specification:\(^5\)

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 \text{age}_i + \beta_3 \text{age}^2_i + \beta_4 \text{male}_i + \beta_5 k_{2i} + \beta_6 k_{3i} + \beta_7 k_{4i} + \beta_8 k_{5i} + \beta_9 k_{6i} + \upsilon_i,$$

where $D_i$ is the county dummy for the region where the hospital is located, male is the gender indicator for men, $k_{1i}$ through $k_{6i}$ are indicators of the eligibility categories for compulsory insurance previously defined and $\upsilon_i$ is the error term. The regression results are presented in Table 1.3. All coefficients have the expected signs and all but male dummy are statistically significant. One can see that the average income in the county where our hospital patients reside is lower than the national average by HRK 3,534.32.

In the second step, we predict income variable in the invoices data set, using the estimated coefficients from Table 1.3. The summary statistics of the predicted income variable for patients

\(^4\)The similar approach has been used by Fang, Keane, Silverman (2008) who combine the Medicare Current Beneficiary Survey (MCBS) with the Health and Retirement Study (HRS) to predict medical expenditure in the HRS sample.

\(^5\)We also tried two other model specifications: first without the county dummy, and second with the county dummy and all its interaction terms with other variables. The obtained income prediction results are very similar.
in the invoices data set are presented in Table 1.4. Comparing the forecasts with the income variable in the CHBS data from Table 1.2, we see that the prediction model works reasonably well, at least as far as pinning down the mean income goes. Subtracting the estimate of the dummy variable coefficient (-3,534.32) from the mean incomes from Table 1.2, one obtains results which are reasonably close to the forecasted values in Table 1.4 both in active and retired groups. The amount of monthly premia for active and retired groups based on the predicted income are displayed in the fifth column of Table 1.4. The result shows that in both active and retired groups, the predicted income is always below the threshold (5,108 HRK/month) required for the insureds to pay higher premia. Hence, for all patients in the invoices data, their income is low enough to make them eligible for the lower premium. This result is somewhat unexpected, although certainly possible, in light of the fact that this county’s income is below the national average.

### 1.4.2 Model Identification

The structural model parameters $\gamma_1$, $\gamma_2$ and $\epsilon_i$ are estimated using GMM. We model $\epsilon_i$ as the individual’s of the health care cost:

$$\epsilon_i = \frac{t v_i y_i}{T m_i}, \quad (1.7)$$

where parameter $t$ captures the required amount of time per visit, $T = 2000$ measures the total number of hours per year which amount to a full-time equivalent employment, $v_i$ is the number of visits per patient, $y_i$ is income and $m_i$ is the total cost of those visits.\(^6\) Therefore, the opportunity cost of time associated with hospital visits $\frac{t v_i y_i}{T}$ is measured in terms of lost income.

In order to identify the unknown parameters $\gamma_1$, $\gamma_2$, $t$, we need at least four moment conditions. We randomly divide the data set into two groups, i.e, each observation is randomly assigned to one of the two groups. Therefore, the $\theta$ distribution for the two groups should be the same. By setting the first four moments of health status distribution for two groups to be equal, GMM optimal estimators $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{t}$ minimize the sum of all moment conditions.\(^7\) Since we nonparametrically match the shape of the distributions, we drop observations whose medical cost are beyond 95% in the right tail of the expenditure distribution. The remaining part of the sample has 21,758 patients.

---

\(^6\)We assume the required amount of time per visit is the same for all patients. This assumption is reasonable because the region from which the hospital draws its patients is rather small so the travel time to reach the hospital should not vary a lot across patients. Secondly, unlike for big hospitals in the large cities, the facilities utilization of regional hospitals in Croatia is extremely low, so the waiting time is likely to be equally short regardless of the required procedure.

\(^7\)Bajari et al. (2014) used the same approach. However they have three years worth of data and they assume that $\theta$ distribution does not change from one year to the next.
The mean of the health status distribution for each of the two groups can be written as:

$$\mu_{\theta_k}(\gamma, t) = \sum_{i=1}^{N_k} \frac{1}{N_k} \theta_i(\gamma, t), \quad k = 1, 2;$$

(1.8)

where $N_k$ is the number of patients in group $k$; $\gamma$ is a vector $[\gamma_1 \ \gamma_2]$. The variance of health status distribution for each of the two groups can be written as:

$$\operatorname{Var}_{\theta_k}(\gamma, t) = \sum_{i=1}^{N_k} \frac{1}{N_k} (\theta_i(\gamma, t) - \mu_{\theta_k}(\gamma, t))^2, \quad k = 1, 2.$$  

(1.9)

The skewness of health status distribution for each of the two groups can be written as:

$$\operatorname{Sk}_{\theta_k}(\gamma, t) = \sum_{i=1}^{N_k} \frac{1}{N_k} \frac{(\theta_i(\gamma, t) - \mu_{\theta_k}(\gamma, t))^3}{\operatorname{Var}_{\theta_k}(\gamma, t)^{3/2}}, \quad k = 1, 2.$$  

(1.10)

and, finally, the kurtosis of health status distribution for each of the two groups can be written as:

$$\operatorname{Ku}_{\theta_k}(\gamma, t) = \sum_{i=1}^{N_k} \frac{1}{N_k} \frac{(\theta_i(\gamma, t) - \mu_{\theta_k}(\gamma, t))^4}{\operatorname{Var}_{\theta_k}(\gamma, t)^2} - 3, \quad k = 1, 2.$$  

(1.11)

To simplify the notation, we introduce vector $w = (m, y, p_j, \alpha_j)$, and define moment conditions as $h(w, \gamma, t)$. In case of the mean we have one sample moment conditions:

$$h_1(w, \gamma, t) = \mu_{\theta_1}(\gamma, t) - \mu_{\theta_2}(\gamma, t)$$  

(1.12)

The sample moment conditions for variance, skewness, kurtosis are of the similar form. Since two groups are drawn from same population, it follows that $E[h(w, \gamma, t)] = 0$. The GMM estimators, $\hat{\gamma}_1, \hat{\gamma}_2, \hat{t}$ minimize the sum of all 4 squared sample moment conditions, which can be written compactly as the following objective function:

$$[\hat{\gamma}_1, \hat{\gamma}_2, \hat{t}] = \arg\min_{\gamma_1, \gamma_2, t} \sum_{k=1}^{4} h_k^2.$$  

(1.13)

We found the minimum of the objective function and hence the estimates $\hat{\gamma}_1, \hat{\gamma}_2, \hat{t}$ by a grid search over the parameter values $\hat{\gamma}_1 \in [0, 10]$, $\hat{\gamma}_2 \in [0, 10]$ and $\hat{t} \in [0.5, 8]$. The grid increment for $\hat{\gamma}_1, \hat{\gamma}_2$ was set at 0.01 and the grid increment for $t$ was set at 0.1. We restrict $t$ to be in the interval between half an hour and 8 hours because patients came to the hospital for outpatient services only.
1.4.3 Moral Hazard

To estimate the magnitude of moral hazard, we run two counterfactual experiments. First, we take away the supplemental insurance from people in $Bought$ group and calculate their counterfactual health expenses in the absence of supplemental insurance. In the first experiment, a patient has to pay the part of cost that was originally covered by supplemental insurance out of pocket. This means that their out of pocket expenses as a percentage of the health care cost which was only $\epsilon_i$ (the price ratio is $\epsilon_i$), in the absence of supplemental insurance, increases to $(\alpha_{ij} + \epsilon_i)$. The consumer is also provided with a lump sum income transfer equal to the amount originally covered by supplemental insurance to ensure the original consumption bundle $(c_1, m_1)$ is affordable. This removes the income effect from the price change in health care relative to other goods and allows us to focus on the substitution effect.\footnote{The patients in the $Free$ group are not used in this experiment because of the data coding problem. In many instances the entire cost of the visit has been charged to the compulsory insurance and zero to the supplemental insurance. Hence, when losing the insurance the patient in the $Free$ group is now required to pay certain percentage of the cost as a co-payment, but any percent of zero is still zero. Notice, however, that this data coding problem does not affect the estimation for $\hat{\gamma}_1, \hat{\gamma}_2, \hat{t}$ because for estimation purposes we only need total cost per patient rather than a part covered by compulsory and a part covered by the supplemental insurance.}

Next, we need to explain how to calculate $m_2$ for each individual in $Bought$ group. Denote the lump sum income transfer as $T_i = \alpha_{ij}m_1i$, where $\alpha_{ij}$ is the portion covered by supplemental insurance (observed in the data). In the counterfactual scenario, patient $i$ pays the amount $\alpha_{ij}m_2i$ out of pocket and the budget constraint becomes:

$$c_{2i} + m_{2i}(\alpha_{ij} + \epsilon_i) \leq y_i - p_j + T_i$$

Next, we need to solve the optimization problem. With the binding budget constraint, $c_2$ can be expressed as $c_{2i} = y_i - p_j - m_{2i}(\alpha_{ij} + \epsilon_i) + T_i$ and the agent’s utility function that needs to be maximized becomes:

$$U(m_{2i}, p_j, y_i; \theta_i, \epsilon_i, \gamma) = (1 - \theta_i) \frac{(y_i - p_j - m_{2i}(\alpha_{ij} + \epsilon_i) + T_i)^{1-\gamma_1}}{1 - \gamma_1} + \theta_i \frac{m_{2i}^{1-\gamma_2}}{1 - \gamma_2}.$$  \hspace{1cm} (1.15)

The first order conditions set the marginal rate of substitution between aggregate consumption and health care equal to the price ratio. In the absence of supplemental insurance, a consumer pays his share $(\alpha_{ij} + \epsilon_i)$ out of pocket. The price ratio is then equal to $(\alpha_{ij} + \epsilon_i)$ and the marginal rate of substitution becomes:

$$MRS = \frac{\theta_i}{(1 - \theta_i)} \frac{(y_i - p_j - m_{2i}(\alpha_{ij} + \epsilon_i) + T_i)^{1-\gamma_1}}{m_{2i}^{1-\gamma_2}} = \alpha_{ij} + \epsilon_i.$$  \hspace{1cm} (1.16)

Since the health status parameter $\theta_i$, risk parameters $\gamma_1, \gamma_2$, and the idiosyncratic cost $\epsilon_i$ are
all unchanged, we can use their estimates obtained in Section 4.2 and solve for the only unknown variable, \( m_{2i} \), using Equation (1.16). The difference between the observed health care consumption \( m_{1i} \) and the counterfactual health care consumption \( m_{2i} \) is caused by the change in price ratio and represents a measure of moral hazard. The standard errors of estimates for moral hazard are obtained through bootstrapping. We calculate estimate of moral hazard with the updated \( \gamma_1, \gamma_2, \hat{t} \) in each iteration of the bootstrapped invoice samples. Standard errors are calculated among these estimates.

In the second counterfactual scenario we give supplemental insurance to the patients in the No group and calculate their counterfactual health expenses assuming they have supplemental insurance. In this scenario the patients only need to pay the share \( \epsilon_i m_i \) out of pocket. There is also an income transfer, \( T_i = -\alpha_{ij} m_{1i} \), which is negative because it takes away part of the income which was originally spent on health care consumption. Then, the budget constraint becomes:

\[
c_{2i} + m_{2i} \epsilon_i \leq y_i - p_j + T_i.
\] (1.17)

Substituting the binding budget constraint for \( c_{2i} \) in the utility function yields:

\[
U(m_{2i}, p_j, y_i; \theta_i, \epsilon_i, \gamma) = (1 - \theta_i) \frac{(y_i - p_j - m_{2i} \epsilon_i + T_i)^{1-\gamma_1}}{1 - \gamma_1} + \theta_i \frac{m_{2i}^{1-\gamma_2}}{1 - \gamma_2}.
\] (1.18)

Finding maximum would reveal the optimal allocation bundle achieved at the point where the marginal rate of substitution between aggregate consumption and health care equals the price ratio \( \epsilon_i \) and the marginal rate of substitution will once again become:

\[
MRS = \frac{\theta_i}{(1 - \theta_i)} \frac{(y_i - p_j - m_{2i} \epsilon_i + T_i)^{\gamma_1}}{m_{2i}^{\gamma_2}} = \epsilon_i.
\] (1.19)

Same as before, for each observation, the only unknown variable is \( m_{2i} \) which could be solved for each patient using Equation (1.19). We expect the counterfactual health care \( m_{2i} \) to be greater than the observed \( m_{1i} \). Thus, the difference between the counterfactual \( m_{2i} \) and the observed \( m_{1i} \) is the amount a patient would over-consume as a result of price distortion due to the supplemental insurance and it represents an alternative measure of moral hazard.

### 1.4.4 Adverse Selection

In the presence of adverse selection, the patients with no supplemental insurance (No) and the patients who bought the supplemental insurance (Bought) should have different health status distributions. We first estimate the relationship between patients’ characteristics and insurance status using logistic regression. The outcome variable is the insurance: for the Bought group \( Y_i = 1 \) and for the No group \( Y_i = 0 \). The results give the marginal effect of the regressors on
the probability of purchasing the insurance. The first specification includes only demographic covariates: age, gender and income. We expect the marginal effects of age and income to be positive and gender (male) to be negative, meaning that older, richer and female patients are more likely to buy the supplemental insurance relative to their counterparts. The second logit model includes demographic covariates and estimated health status $\hat{\theta}$. A larger value of $\theta$ denotes relatively worse underlying health status. We expect that the estimated health status $\hat{\theta}$ has a positive marginal effect because patients who bought the supplemental insurance did so in anticipation of frequently using it.

In addition to logistic regression, the presence of selection can be also tested by comparing estimated $\theta$ distributions between insurance plans. In the presence of adverse selection, we would expect distribution of the latent health status be significantly different between plans, with higher $\theta$ patients choosing more insurance and lower $\theta$ patients choosing less insurance. The absence of adverse selection would be indicated by the health status distributions being identical between insurance plans. We perform the Kolmogorov-Smirnov test for the equality of $\theta$ distributions between insurance plans. The null hypothesis is that $\theta$ distributions for two plans come from the same population. We also use Kolmogorov-Smirnov test to determine whether the cdf of $\theta$ distribution for one plan is stochastically dominating that for the other. The plan with a dominating cdf has more consumers with lower value of $\theta$ (healthier). One would expect that in the presence of adverse selection, the patients in the $\text{No}$ group should be comparatively healthier (lower value of $\theta$) and the cdf of $\theta$ distribution in the $\text{No}$ group should stochastically dominate the cdf of the $\theta$ distribution for the $\text{Bought}$ group.

Finally, one cannot rule out the presence of favorable selection either. This is because people can buy more or less insurance for reasons which are not only related to their risk-aversion and health status but could be additionally impacted by other sources of unobserved heterogeneity.$^9$ In the presence of favorable selection, the cdf of $\theta$ distribution of the $\text{Bought}$ group should stochastically dominate the cdf of the $\text{No}$ group.

1.5 Empirical Results

The standard errors of the estimates are obtained through bootstrapping. In each iteration of the bootstrapping exercise, we randomly draw observations with replacement from CHBS data and use the bootstrapped sample to estimate the income prediction coefficients. Then we randomly draw observations with replacement from invoice data and apply the new income prediction coefficients to the bootstrapped invoice sample. We run the grid search with the

---

$^9$Fang, Keane and Silverman (2008) found evidence of advantageous selection in the Medigap insurance market and suggested that the sources of this advantageous selection include the insureds' income, education, longevity expectations, financial planning horizons and especially the cognitive ability.
new sample and predicted income, a new set of estimates $\hat{\gamma}_1, \hat{\gamma}_2, \hat{t}$ would be found. We repeat this procedure for 100 times and obtain the standard errors of the estimates among the 100 iterations.

The estimates of the model parameters and standard errors are displayed in Table 1.5. The results show that people are more risk averse when it comes to their health compared to risk aversion associated with aggregate consumption. The estimate of $t$ was found to be $\hat{t} = 2.70$, indicating that, on average, each visit to the hospital will cost a patient two hours and forty-two minutes worth of his income. Using $\hat{t}$, we can calculate the opportunity cost of time as a percentage of health care cost using Equation (1.7). On average, it amounts to 42% of the health care cost (with the standard error of 0.35) and is larger than the average co-payment rate for patients in the No group (30%). Therefore, the opportunity cost of time is a critical component of the total cost of hospital visits and must not be ignored.

Next, we use the estimates to quantify the asymmetric information effects. As mentioned before, there are two ways to measure the effects of moral hazard. One way is to take away supplemental insurance from people in Bought group and calculate the counterfactual health care consumption in the absence of supplemental insurance. For each patient, the moral hazard is measured by the difference between the counterfactual and the original health care expenditure, $\Delta m_i = m_{2i} - m_{1i}$ or as a percent change in the original health care expenditure, $\%\Delta m_i = \frac{m_{2i} - m_{1i}}{m_{1i}}$. The averages across all patients in different groups are reported in Table 1.6. The results show that people in Bought group would spend 31.86 HRK or 8.3% less on health care if we take away their supplemental insurance. Because our data is limited only to the observations of people who came to the hospital and sought medical help, the estimates of moral hazard would normally be impacted by the “users only” data problem. However, in this particular counterfactual experiment, the estimate of moral hazard is unbiased because taking away insurance from non-users do not change their behavior. People who did not visit the doctor when they had insurance, surely would not visit the doctor when they don’t have insurance.

Using this counterfactual experiment it is also meaningful to sum up all individual measures of moral hazard to come up with the aggregate measure of moral hazard for the county/hospital as a whole. As seen from Table 1.6, the aggregate 4-months hospital level reduction in health care consumption due to elimination of moral hazard amounts to 375 thousand HRK (about 72 thousand dollars) or 9.09%.

An alternative measure of moral hazard is obtained by giving the supplemental insurance to people in the No group and see how much would they spent on health care if they had the supplemental insurance. The difference between the counterfactual and the actual health

---

$^{10}$The results are comparable to the results obtained by Bajari et al. (2014) in the sense that the risk aversion parameter for health care consumption is larger than that of general consumption. Their estimates of $\gamma_1$ are in the range of [1.88,1.98] and the estimates of $\gamma_2$ are in the range of [3.12, 3.27].
care consumption measures the effect of moral hazard. As seen from Table 1.6, people in the
No group would have spent 27.25 HRK more on health care once they were provided with
the supplemental insurance. This measure of moral hazard suffers from the "users only" data
problem because in this counterfactual experiment we cannot give the insurance to the people
in the No group who never showed in the hospital (non-users) because we don’t know who
they are and how many of them there are. Therefore this measure surely understates the true
magnitude of moral hazard in the general population.

Let’s turn to the adverse selection tests. The results for logistic regression are displayed in
Table 1.7. It reports the marginal effects of consumer characteristics and health status on the
insurance plan choice. The first model includes only demographic characteristics, whereas the
second model includes demographic characteristics and the estimate of the latent health status
variable $\hat{\theta}$. Most results in Model 1 are as expected. The marginal effect of age is positive and
the marginal effect of gender (male) is negative. It means that older and female patients are
more likely to purchase the supplemental insurance than their counterparts. An unexpected
result is that income has negative marginal effect which means that richer individuals are less
likely to buy the supplemental insurance than the poor ones. This result can be interpreted by
realizing that rich people can self-insure themselves against adverse effects of illness. This is
especially true in Croatia where the compulsory insurance provides a very generous safety net
for all citizens whereas the supplemental insurance covers expenses which are unlikely to cause
a serious financial hardship for more affluent citizens. The results in Model 2 which includes
the estimated latent health status do not reverse any of the results from Model 1. As far as
the health status is considered, it has positive effect on Bought group. With larger value of $\theta$
indicating worse health, this result suggests that people without insurance tend to be relatively
healthier and people who decided to buy the insurance tend to be in worse health.

Next, we test the adverse selection effect by comparing $\theta$ distributions. The two-sided and
one-sided test results are reported in Table 1.8. The two-sided test results show that patients in
two groups have different health status distributions and the results are statistically significant
at 1%. Finally, we test whether the cdf of the $\theta$ distribution for No group is dominating that of
Bought group. The group with the dominating cdf of the $\theta$ distribution has more people on the
lower end of the distribution, which means it has larger percentage of healthier people. The null
hypotheses that the distributions are the same are tested against the alternative that the cdf of
the $\theta$ distribution for No is greater than the cdf of the $\theta$ distribution for Bought. The one-sided
test results show that the cdf of the $\theta$ distribution for the No group is significantly dominating
that of Bought group. It means healthier people self-selected themselves into No group, while
people with comparatively worse health self-select themselves into the Bought group. These
results provide strong, statistically significant, evidence of adverse selection.

The enlarged versions of cdf distributions of $\theta$ for two groups are plotted in Figure 1.1. The
graph clearly shows that the cdf of $\theta$ distribution for Bought group is stochastically dominated by the cdf of the $\theta$ distribution for No group, indicating that people in No group are generally healthier than people in the Bought group.

### 1.6 Conclusion

This paper uses a structural approach to estimate clean moral hazard effect in health insurance, net of possible influences of adverse selection. We found positive and economically meaningful effect of moral hazard and also a statistically significant effect of adverse selection in the population of hospital patients.

An important methodological contribution of our paper is to take into account individuals’ opportunity cost of time associated with medical care consumption. To the best of our knowledge this is the first attempt in the literature to incorporate such cost into a structural estimation of health care demand. We argue that this cost does not depend on the providers’ price nor the insurance type but depends on agent’s income. The idea behind this specification is the fact that in addition to actual medical expenses which may or may not be covered by the insurance, different people could have different transaction costs associated with visiting a hospital or a doctor, which for certain minor illnesses could deter the health care consumption thus mitigating the moral hazard problem.

The generalization of the obtained results for the purposes of country level and international comparisons is somewhat hampered by the fact that hospital invoices data is comprised of users only, i.e., we don’t have data for insured and uninsured citizens with zero medical consumption in the given period. As it turns out, the obtained structural model parameter estimates ($\hat{\gamma}_1, \hat{\gamma}_2, \hat{t}$) are still consistent because the structural model is assumed for users and non-users alike and equation (1.6) used to estimate $\theta_i$ for individuals with different insurance coverage is also valid for both users and nonusers. However, two important consequences of the users only data still remain.

First, the two counterfactual moral hazard scenarios are markedly different. The scenario where we take away the supplemental insurance from the insured users and simulate their counterfactual health expenses in the absence of this insurance gives the correct estimate of the moral hazard effect at the aggregate (hospital) level. This is true in light of the fact that the insured nonusers are not going to be impacted by taking away the insurance from them because if they did not go to the hospital when they had insurance, they surely will not go when they don’t have the insurance. The second scenario where we give the insurance to uninsured users suffers from the fact that uninsured users group does not encompass people who did not seek medical attention precisely because they lacked the insurance. Therefore, giving the insurance to uninsured citizens would impact not only users but potentially also nonusers who do not
appear in our data. Hence the second scenario is surely going to under-estimate the magnitude of the moral hazard. The difference between two scenarios is attributable to the users only data problem.

Second, the tests for adverse selection are also impacted by the users only data problem because these tests should be based on the distribution of the unobserved health status $\theta$ for the entire population whereas ours are based on users only. Our estimates of $\theta_i$ overstate the true health status in the overall population, making it worse than it really is. There are two reasons for this phenomenon, both pooling in the same direction. First, among those insured, it is reasonable to assume that most of those that need medical help will actually demand it and hence will show up as users. Only those, with relatively minor illnesses and high opportunity costs of visits will not seek medical attention when they need it and will not show up as users. The rest of the insured population simply did not need medical attention and never showed up in the hospital. Hence, the actual health status of the entire Bought group is likely to be better (lower $\theta$) than the actual patients (users only) based estimates. Next, the users only based estimates of $\theta_i$ for the No group is also likely to overstate the true health status of this entire group because it is possible that some people who do not have the insurance did not show up in the hospital to seek medical attention precisely because they did not have the insurance. The unobserved health status of those individuals (nonusers) is likely to be better than those uninsured individuals who actually sought medical attention since their condition was too serious to forgo the treatment. Whether performing the adverse selection tests based on the distributions of unobserved health status reflective of the entire population would change the previously obtained results is impossible to predict.
Table 1.1: Summary Statistics by Insurance Category

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Patients(#)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost per patient (HRK)</td>
<td>257.11</td>
<td>5.28</td>
<td>2,451</td>
</tr>
<tr>
<td>Co-payment per patient (HRK)</td>
<td>77.42</td>
<td>69.81</td>
<td>2,451</td>
</tr>
<tr>
<td>Visits per patient</td>
<td>1.94</td>
<td>0.04</td>
<td>2,451</td>
</tr>
<tr>
<td>Age</td>
<td>40.51</td>
<td>0.29</td>
<td>2,451</td>
</tr>
<tr>
<td>Male (%)</td>
<td>0.52</td>
<td>0.01</td>
<td>2,451</td>
</tr>
<tr>
<td><strong>Free</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost per patient (HRK)</td>
<td>338.11</td>
<td>3.78</td>
<td>7,512</td>
</tr>
<tr>
<td>Co-payment per patient (HRK)</td>
<td>0</td>
<td>0</td>
<td>7,512</td>
</tr>
<tr>
<td>Visits per patient</td>
<td>2.68</td>
<td>0.03</td>
<td>7,512</td>
</tr>
<tr>
<td>Age</td>
<td>54.74</td>
<td>0.20</td>
<td>7,512</td>
</tr>
<tr>
<td>Male (%)</td>
<td>0.41</td>
<td>0.01</td>
<td>7,512</td>
</tr>
<tr>
<td><strong>Bought</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost per patient (HRK)</td>
<td>350.69</td>
<td>3.04</td>
<td>11,795</td>
</tr>
<tr>
<td>Co-payment per patient (HRK)</td>
<td>0</td>
<td>0</td>
<td>11,795</td>
</tr>
<tr>
<td>Visits per patient</td>
<td>2.86</td>
<td>0.02</td>
<td>11,795</td>
</tr>
<tr>
<td>Age</td>
<td>53.63</td>
<td>0.15</td>
<td>11,795</td>
</tr>
<tr>
<td>Male (%)</td>
<td>0.41</td>
<td>0.01</td>
<td>11,795</td>
</tr>
</tbody>
</table>

Table 1.2: Summary Statistics: CHBS Data

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Income</td>
<td>33,569.91</td>
<td>30,345.11</td>
<td>0</td>
<td>162,650</td>
<td>4,623</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>39.55</td>
<td>41.00</td>
<td>18.00</td>
<td>67.00</td>
<td>4,623</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.51</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>4,623</td>
</tr>
<tr>
<td>Retired</td>
<td>Income</td>
<td>30,976.62</td>
<td>28,396.80</td>
<td>0</td>
<td>161,317</td>
<td>2,216</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>70.35</td>
<td>70.00</td>
<td>55.00</td>
<td>98.00</td>
<td>2,216</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.38</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>2,216</td>
</tr>
</tbody>
</table>
Table 1.3: Regression Results for Income in the Survey Data

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>28618.61</td>
<td>1469.80</td>
<td>19.47</td>
<td>0.00</td>
</tr>
<tr>
<td>$D$</td>
<td>-3534.32</td>
<td>1010.92</td>
<td>-3.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Age</td>
<td>294.96</td>
<td>62.16</td>
<td>4.74</td>
<td>0.00</td>
</tr>
<tr>
<td>$Age^2$</td>
<td>-1.86</td>
<td>0.63</td>
<td>-2.94</td>
<td>0.00</td>
</tr>
<tr>
<td>Male</td>
<td>390.88</td>
<td>419.18</td>
<td>0.93</td>
<td>0.35</td>
</tr>
<tr>
<td>$k_2$</td>
<td>-12759.49</td>
<td>976.95</td>
<td>-13.06</td>
<td>0.00</td>
</tr>
<tr>
<td>$k_3$</td>
<td>-8609.764</td>
<td>632.06</td>
<td>-13.62</td>
<td>0.00</td>
</tr>
<tr>
<td>$k_4$</td>
<td>-13386.89</td>
<td>799.04</td>
<td>-16.75</td>
<td>0.00</td>
</tr>
<tr>
<td>$k_5$</td>
<td>-24444.04</td>
<td>2017.79</td>
<td>-12.11</td>
<td>0.00</td>
</tr>
<tr>
<td>$k_6$</td>
<td>-2425.734</td>
<td>986.27</td>
<td>-2.46</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Note: Adjusted $R^2$=0.083.

Table 1.4: Summary Statistics: Predicted County Income and Premia

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Premium/month</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>29,191.13</td>
<td>32,084.15</td>
<td>5,347.36</td>
<td>37,171.75</td>
<td>80</td>
<td>13,869</td>
</tr>
<tr>
<td>Retired</td>
<td>27,736.97</td>
<td>28,093.84</td>
<td>12,102.77</td>
<td>34,750.91</td>
<td>50</td>
<td>7,889</td>
</tr>
</tbody>
</table>

Table 1.5: Estimates of Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>5.98</td>
<td>0.30</td>
<td>19.93***</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>8.20</td>
<td>0.38</td>
<td>21.58***</td>
</tr>
<tr>
<td>$t$</td>
<td>2.70</td>
<td>0.66</td>
<td>4.09***</td>
</tr>
</tbody>
</table>

Note: Standard errors are calculated based on 100 bootstrap iterations. *** indicates 1 % significance level; ** 5 % significance level and * 10 % significance level.
### Table 1.6: Counterfactual Effects of Moral Hazard

<table>
<thead>
<tr>
<th>Group</th>
<th>Insured (Bought)</th>
<th>Uninsured (No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ($\Delta m_i$)</td>
<td>-31.86 (7.15)</td>
<td>27.25 (6.89)</td>
</tr>
<tr>
<td>Mean (%$\Delta_i$)</td>
<td>-8.30 (1.94)</td>
<td>8.81 (2.35)</td>
</tr>
<tr>
<td>Total Reduction (HRK)</td>
<td>-375,838 (84,691)</td>
<td>–</td>
</tr>
<tr>
<td>Percent Reduction (%)</td>
<td>9.09 (2.04)</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: $\Delta m = m_2 - m_1$; $\% \Delta m = \frac{m_2 - m_1}{m_1}$; standard errors are in the parentheses, obtained with 100 bootstraps. Total reduction is the sum of $\Delta m_i$. Percent reduction is total reduction divided by total cost.

### Table 1.7: Testing for Adverse Selection: Logistic Regression of Insurance Choices

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender (male=1)</td>
<td>-0.0525*** (0.0059)</td>
<td>-0.0524*** (0.0059)</td>
</tr>
<tr>
<td>Age</td>
<td>0.0065*** (0.0002)</td>
<td>0.0065*** (0.0002)</td>
</tr>
<tr>
<td>Log(income)</td>
<td>-0.1723*** (0.0147)</td>
<td>-0.1667*** (0.0148)</td>
</tr>
<tr>
<td>$\hat{\theta}$</td>
<td>0.7970** (0.3417)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Marginal effects are presented. Standard errors are in parentheses.

### Table 1.8: Tests of Adverse Selection: K-S Statistics for equality of $\theta$ distribution for No and Bought

<table>
<thead>
<tr>
<th>Group (No and Bought)</th>
<th>Statistics</th>
<th>P-value</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-sided</td>
<td>0.159</td>
<td>0.000</td>
<td>Reject</td>
</tr>
<tr>
<td>One-sided</td>
<td>0.159</td>
<td>0.000</td>
<td>Reject</td>
</tr>
</tbody>
</table>

20
Empirical CDF of Health Status $\theta$ for Bought and No

Figure 1.1: Empirical CDF of $\theta$ for two groups (Enlarged)
Chapter 2

Risk Aversion, Moral Hazard and Gender Differences in Health Care Utilization

2.1 Introduction

As the development of medical science progresses, human longevity prolongs with time. Due to certain biological, behavioral and economical reasons, women have longer life expectancy than men. Women are genetically less likely to have cardiovascular, infectious, and cancerous diseases. Behaviorally, men are more likely to smoke and drink, which are related to lung and liver diseases. Economically, men are more engaged in various business activities which could lead to more pressure and stress. Distinction in life expectancy is an important factor in explaining the difference in overall health care consumption between genders. There are many studies in the medical literature that have found gender difference in health care consumption. The comparison are usually based on medical records such as discharge rates, hospitalization, etc., without controlling for socioeconomic characteristics of patients. This paper investigates gender difference in health care consumption and seeks to explain it by the gender differences in moral hazard due to widespread use of health insurance. Moral hazard has been blamed for the rapid increase of health care expenditure as more people become insured and as insurance coverages become more comprehensive.

The literature on gender difference in health care consumption can be grouped into two clusters: medical sciences and social sciences (economics). Different sexes are vulnerable to different kinds of diseases or require different kinds of medical treatments in different age cohorts. As a result of genetic differences, men have higher infant mortality (Naeye et al., 1971). Also, men
are more susceptible to infections and diseases generally in their early age due to the genetic immunological disadvantages (Washburn, Medearis and Childs, 1966; Michaels and Rogers, 1971). During the prime of adult life, the gender differences in health care utilization are more pronounced because of child-bearing. Nathanson (1977) investigated hospital discharge rates by sex and conditions in Canada and United States. Hospital discharge rate, as a measure of hospital utilization, looks at the number of patients who leave a hospital after receiving care. The study found the total discharge rate for women is about one third larger than that of men in both countries. However, after excluding discharge rates related to obstetrical conditions, discharge rate of women is only one tenth larger than that of men in Canada and one twentieth larger in US. In addition, men and women have different hormonal balance. Some researchers believe hormonal factors are responsible for lower rates of coronary heart disease among women prior to the menopause (Moriyama, Krueger and Stamler, 1971) and it may also explain women’s low susceptibility to certain cancers following the menopause (Lilienfeld, Levin and Kessler, 1972). To reduce the confounding variability in background characteristics, Gold et al. (2002) did a study on older different-sex twins. They found men have more life-threatening and cardiovascular conditions than women.

From the economics perspective, at least four major explanatory frameworks can be constructed to explain gender differences in health care utilization: (1) preferences towards health, (2) life style, (3) opportunity cost of time and (4) risk aversion. Preferences and attitudes towards health are arguably different in women and men. Verbrugge (1985) found that women are more sensitive to their health needs. They pay more attention to coming symptoms and are more willing to address them. Bertakis et al. (2000) found that after controlling for health status and socio-demographic characteristics, women still have higher medical expenses for all categories of charges except hospitalizations. Using 2008 Medical Expenditure Panel Survey, Vaidya et al. (2012) found that men are significantly less likely to utilize preventive care services (e.g., flu shots, blood pressure and cholesterol check-ups, dental exams) than women. Men are more likely to have a physical exam if it is required for work or insurance purposes rather than voluntarily (Andersen and Anderson 1967).

The second explanation relies on the gender differences in life styles. For example, in many cultures it is more socially acceptable for men to smoke and drink. Also, traditional male occupations require more physical endurance and are more likely to result in injuries. The health statistics in U.S. show the average annual injury-related visits to hospital emergency departments for men are larger than for women. In the period of 2007-2008, men’s injury-related visits amounted to 16,640,000, whereas those of women amounted to 14,688,000. In the 18-44 age cohort, the disparity is the most pronounced. Men of this age cohort had 7,173,000 injury-related visits, while women had 5,823,000 (Health, United States, 2011).

The third possible explanation for the differences in health care utilization could be explained
by the difference in the opportunity cost of time. Higher opportunity cost of time could make
a visit to a doctor relatively more costly and could ultimately deter a sick person to seek
medical attention in cases that are perceived as less serious. Gilleskie (2010) analyzed the
gender differences in work absenteeism and medical care decisions during an episode of acute
illness. Based on the 1987 National Medical Expenditure Survey data, she found that men
are more responsive to variations in sick leave and health insurance coverage than women. A
change from no sick leave coverage to covered sick leave induced a 45% increase in illness-related
absences in men, but only an 11% increase among women. A change from no insurance coverage
for physician visits to free physician visits increases medical care consumption by 20% among
men and 15% among women. Reducing these benefits also has smaller effect on the absenteeism
and medical care consumption of women then men.

Finally, in this paper we argue that the gender differences in health care consumption could
be systematically related to the gender differences in risk aversion. Given that most people
carry some level of health insurance, a straightforward theoretical link between health care
utilization and risk-aversion is established through moral hazard. One of the fundamental result
from the standard asymmetric information model (principal-agent) establishes an increasing
relationship between risk-aversion and moral hazard, where the latter is defined as the welfare
difference between the first-best and the next-best contracting outcome where the first-best is
not attainable because of the unobservable agent’s action. However, if the agent is risk-neutral,
then the asymmetric information problem is inert and the first-best outcome generally obtains.
Interestingly enough, this microeconomic theory definition of moral hazard is not used by the
health economists. Instead, when they talk about moral hazard, they mean the difference in
health care consumption caused by health insurance via the reduction in prices of health services
enabled by the insurance coverage. Therefore, the concept of moral hazard, as standardly used
in health economics, has nothing to do with risk-aversion. Instead, central to the definition
of moral hazard is the price elasticity of demand for health care services: the more elastic
the demand curve, the higher the moral hazard effect. The main challenge of this paper is to
establish a link between the two concepts by exploring the relationship between price elasticity
of demand for health care services and risk-aversion.

In conducting this research, another challenge we face comes from the data. We use the
dataset which consists of invoices for all outpatient services from a regional hospital in Croatia
during a four month period in 2009. Because the data contains only users of medical services,
we do not have any information about people who didn’t show up at the hospital in the period
covered by the data. To address this problem, we rely on a truncated count model and simulated
maximum likelihood estimator. We hypothesize that the gender difference in moral hazard, if
there is any, is due to the difference in risk aversion. The commonly inferred opinion that women
are more risk averse then men leads to a theoretical prediction that women’s health consumption
suffers from a smaller moral hazard effect than men’s. After adjusting for the sample selection in
the estimation, we found a significant evidence of moral hazard but no evidence of statistically
significant gender difference in moral hazard due to health insurance.

2.2 Theoretical Framework

The conventional definition of moral hazard used in health economics literature is tied to the
case of price elasticity of demand for health care services. Since the demand function is typi-
cally downward sloping, the more elastic the demand (the smaller the elasticity coefficient), the
higher the moral hazard effect because the reduction in price caused by health insurance cover-
age will push the quantity consumed further to the right along the quantity axis. Two extreme
cases are determined by an infinitely elastic (horizontal) demand curve causing infinitely large
moral hazard and a perfectly inelastic (vertical) demand function causing zero moral hazard.

To establish a sought after relationship between the above definition of moral hazard and
risk-aversion, we start with a model of a possibly risk-averse economic agent who maximizes
his or her utility function by choosing health care \( m \) and a composite commodity \( c \), subject to
a budget constraint. The utility function is additive in aggregate consumption and health care,
each in the constant relative risk aversion (CRRA) functional form:

\[ U(c, m, \theta, \gamma) = (1 - \theta) \ln c + \theta \frac{m^{1-\gamma}}{1-\gamma} \]  

(2.1)

where \( \gamma > 0 \) is the risk-aversion coefficient for health care consumption.\(^1\) The larger the coefficient, the more risk averse the person with respect to the variation in health care consumption. The utility also depends on latent health status parameter \( \theta \in [0, 1] \). It can be interpreted as measuring the importance an agent places on health care and aggregate consumption.

A consumer’s budget constraint requires that his expenditure on aggregate consumption
and health care must not be greater than his income minus the insurance premium:

\[ c + \alpha m \leq y - p, \]  

(2.2)

where prices are normalized to unity, \( y \) is income, \( p \) is the insurance premium and \( \alpha \) is the
co-payment rate, i.e., the percentage of the cost that patients have to pay out of pocket which
depends on the insurance coverage. The more generous the insurance coverage, the smaller is \( \alpha \) which a consumer needs to pay out of pocket.

The first order conditions for the maximization of utility (2.1) subject to budget constraint

\(^1\)This specification of the utility function is a simplified version of the utility function from Bajari et al. (2014)
where the risk-aversion coefficient for aggregate consumption is assumed to be 1 for simplicity.
(2.2) are as follows:

\[
\frac{\partial L}{\partial c} = (1 - \theta) \frac{1}{c} - \lambda = 0, \\
\frac{\partial L}{\partial m} = \theta m^{-\gamma} - \lambda \alpha = 0, \\
\frac{\partial L}{\partial \lambda} = y - p - c - \alpha m = 0.
\]

The standard optimization result shows that the marginal rate of substitution between aggregate consumption and health care consumption equals to the ratio of prices. However, the price of health care that a consumer faces is only the co-payment portion \( \alpha \) of the actual price. As the result, the relative price of medical services to aggregate consumption is equal to \( \alpha \) and the MRS becomes:

\[
MRS = \frac{\theta}{(1 - \theta)} \frac{c}{m^{\gamma}} = \alpha. \tag{2.3}
\]

Since a patient only needs to pay \( \alpha < 1 \) portion of the actual health care cost, the price ratio between health care and aggregate consumption is biased towards favoring health care and against general consumption. This “excess” of health care consumption is a standard measure of moral hazard associated with insurance.

Since the utility function is strictly increasing, the budget constraint binds at the optimal bundle. Therefore \( c \) is determined by the equation \( c = y - p - \alpha m \), which substituted into Equation (2.3) gives:

\[
\frac{\theta}{(1 - \theta)} \frac{y - p - \alpha m}{m^{\gamma}} = \alpha. \tag{2.4}
\]

The expression of the price elasticity of demand for health care \( \eta \) can be derived from Equation (2.4) using implicit differentiation:

\[
\eta = \frac{\partial m}{\partial \alpha} \frac{\alpha}{m} = -\frac{1 + \frac{\alpha m}{y - p - \alpha m}}{\gamma + \frac{\alpha m}{y - p - \alpha m}} < 0. \tag{2.5}
\]

The nonzero elasticity of demand for health care gives rise to the existence of moral hazard. The more elastic the demand for health care (smaller \( \eta \)), the larger the moral hazard effect. From Equation (2.5) we can see that the measure of elasticity, hence the magnitude of moral hazard, depends on the risk aversion coefficient \( \gamma \). To determine the relationship between price elasticity of demand and risk aversion we compute its derivative with respect to \( \gamma \):

\[
\frac{\partial \eta}{\partial \gamma} = \frac{1 + \frac{\alpha m}{y - p - \alpha m}}{\left( \gamma + \frac{\alpha m}{y - p - \alpha m} \right)^2} > 0. \tag{2.6}
\]

As seen from expression (2.6), higher risk aversion leads to less elastic demand for health
services and hence smaller moral hazard effect. A highly risk averse patient will seek medical attention regardless of the price, so her demand for health care is inelastic leading to smaller moral hazard. On the other hand, less risk-averse individual would tend to ignore minor health problems when faced high cost of medical care, hence her demand function would be elastic giving rise to larger moral hazard. In other to generate testable predictions about gender differences in moral hazard associated with with health care we first need predictions about the gender differences in risk aversion. For this we turn to the existing literature.

2.3 Gender Difference in Risk Aversion

The existing literature on the subject of gender differences in risk aversion seem to favor the conclusion that women are more risk averse, especially when it comes to financial decision making. A systematic summary of this literature is found in Table 2.1. As seen from the table, no studies have found that men are more risk averse then women. However, many studies have found that women are more risk averse than men in various risky environments, such as for example, risk associated with financial decisions (Arano, Parker and Terry (2010), Charness and Gneezy (2011), Eckel and Grossman (2002), Bajtelsmin and Van Derhei (1997), Jianakoplos and Bernasek (1998)); risk associated with alcohol and drug use (Spigner, Hawkins, and Loren, 1993) and risk associated with insurance for loss (Powell and Ansic, 1997).

Finally, there are also studies that found no significant gender difference in risk aversion. For example, the study by Gneezy et al. (2009) found no statistically significant gender difference in risk attitude in less developed societies (Maasai society in Africa and Khasi society in South Asia). Flynn et al. (1994) found that white men perceived risk as much more acceptable than did white women but the perception of risk between nonwhite men and women were quite similar. Schubert et al. (1999) compared the gender difference in the abstract gambling experiment and in the contextual investment or insurance experiment. They conducted their experiments with undergraduates from different fields. They found different risk propensities between genders in abstract gambles, but no gender differences in contextual experiments. Since financial decisions are always contextual in practice, the abstract gambling experiments might be inadequate for the analysis of gender difference in risk attitudes toward financial decisions.

The most comprehensive comparative study of the literature in this area is Nelson (2014). She found that the statement about women being more risk averse is not supported by the actual empirical evidence. Quantitative measures of substantive difference (Cohen’s d) and substantive overlap (Index of Similarity) were applied to the data on men’s and women’s risk distributions used in 35 studies in economics, finance and decision making literature. Cohen’s d is a measure of the substantive magnitude of a difference, it expresses the difference between means in standard deviation units. She also proposed the Index of Similarity to measure the
overlap of any discrete distribution. This index can be intuitively interpreted as the proportion of the women and men that are identical, which means their characteristics or behaviors exactly match up with someone in the opposite sex group. The results show fewer statistically significant differences but more overlapping than has been commonly claimed in the literature.

2.4 Institutional Framework and Data Description

The health care system in Croatia is dominated by a single public health insurance fund: the Croatian Institute for Health Insurance (HZZO). The HZZO offers two types of insurance: the compulsory insurance and the supplemental insurance. The compulsory insurance, which is funded by a 15% payroll tax, covers two kinds of medical care services: one with full coverage, the other with a system of co-payments. Full coverage medical care services are provided to children up to 18 years of age and pregnant women and for everybody else for life threatening types of illnesses such as infectious diseases, psychiatric care, surgeries, cancers, mandatory vaccinations, etc. All other health services (including but not limited to primary care, hospitals stays and prescription drugs) are subject to a system of co-payments. The patients are required to pay 20% of the full price of medical care, with the largest out-of-pocket cost share amount set at 3,000.00 HRK per invoice.\footnote{The figures are reflective of the year 2009 which is the year covered by the dataset. The exchange rate for the local currency, Croatian Kuna (HRK), as of June 20, 2009 was 1USD=5.19 HRK. A more detailed description of the health insurance system in Croatia can be found in Liu, Nestic and Vukina (2012).}

The supplemental insurance is a voluntary insurance that can be acquired by a person 18 years or older, having compulsory insurance, by signing a contract with the HZZO. Certain categories of citizens are entitled to the supplemental insurance free of charge, i.e., their premiums are covered from the state budget. The list of people entitled to free supplemental insurance includes, among others, the full time secondary school and college students. For those not entitled to free supplemental coverage, the premia range from 50 to 130 HRK per month depending on whether the insured is active or retired and income. A person having the supplemental insurance is entitled to full waiver of all medical expense co-payments mentioned before.

The original data set consists of all invoices for all outpatient services from a regional hospital in Croatia during the period from March 1 to June 30, 2009. The data set consists of 105,646 observations. Each observation reflects the invoice for one hospital visit. The data contains the following set of variables: a numeric code for the type of hospital service provided, compulsory health insurance number, supplemental insurance number (if the patient has one), period covered by the supplemental insurance, numeric code for categories entitled to supplemental insurance free of charge, eligibility category for compulsory insurance ($k_1$–employed, $k_2$–farmers, $k_3$–pensioners, $k_4$–unemployed, $k_5$–living on social welfare, $k_6$–self-employed and other), cost
of hospital service, part of the cost covered by compulsory insurance, part of cost covered by supplemental insurance, part of cost covered by participation (co-payment), date of birth and sex of the patient.

In this paper we focus only on the supplemental insurance and its impact on health care utilization. For this reason, patients who visited the hospital only because of the illnesses that are fully covered by the compulsory insurance are excluded. We also delete patients that are entitled to supplemental insurance for free. The rest of the sample has two groups of patients: those who bought the supplemental insurance (Bought) and those who did not buy it (No). The working data set consist of 14,991 patients.

The summary of the data by insurance type and gender is given in Table 2.2. As seen, women have significantly more hospital visits than men in the Bought group (0.15). The summary of the data by insurance type and employment status is given in Table 2.3. We denote employed, farmers and self-employed as the Active group and pensioners, unemployed and living on social welfare as the Inactive group. Active group has significantly more hospital visits than Inactive group in both Bought (0.27) and No (0.57) samples.

Finally, we looked at the reasons why people come to the hospital. Top ten diagnoses that are subject to co-payments are summarized by gender in Table 2.4. As it turns out, the list of diagnoses for which people seek medical attention is quite similar for both sexes. Both men and women record outpatient physical therapy and medical biochemistry as top two reasons for coming to the hospital whereas other components of the top ten diagnoses list overlap with somewhat different rank. The only differences between men and women are traumatic injuries (probably job related) and urology (probably prostate screening) on the men side and clinical cytology (probably breast cancer tissue tests) and ultrasound (probably pregnancies related) on the women side.

2.5 Estimation

We are interested in how health care utilization and insurance status are associated with people’s socioeconomic characteristics $x_i$. Let $y_i$ denote the number of hospital visits and $I_i$ is the insurance indicator. In order to estimate $E(y_i, I_i|x_i) = g(x_i\beta)$, we need to specify a full conditional distribution of $(y_i, I_i)$ given $x_i$. However, because of the data limitation, we only observe $(y_i, I_i, x_i)$ when $y_i > 0$. We can use the density of $(y_i, I_i)$ conditional on $x_i$ and $y_i > 0$.

\footnote{The estimation is a modification of Chapter 17.3 in Wooldridge 2002.}
Therefore, we need to derive the density (likelihood function) for \( f(y_i, I_i|y_i > 0) \):

\[
\begin{align*}
f(y_i, I_i|y_i > 0) &= \frac{f(y_i, I_i, y_i > 0)}{f(y_i > 0)} \cdot \frac{f(y_i, I_i > 0|I_i) f(I_i)}{Pr(y_i > 0)} \\
&= \frac{f(y_i|I_i) f(I_i)}{Pr(I_i = 1) Pr(y_i > 0|I_i = 1) + Pr(I_i = 0) Pr(y_i > 0|I_i = 0)} \quad (2.7)
\end{align*}
\]

Here, \( f(y_i, y_i > 0|I_i) = f(y_i|I_i) \) because for the observed \( y_i \), \( y_i \) belongs to the set \( y_i > 0 \).

In order to obtain the probability terms in Equation (2.7), we need to specify models for \( y_i \) and \( I_i \). Assuming \( y_i \) follows a Poisson distribution with the mean:

\[
E(y_i|I_i, h_i) = \exp(\beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 male_i + \beta_4 active_i + \beta_5 I_i + \beta_6 h_i \\
+ \beta_7 age_i \ast I_i + \beta_8 age_i^2 \ast I_i + \beta_9 male_i \ast I_i + \beta_{10} active_i \ast I_i + \beta_{11} h_i \ast I_i)
\]

which depends on age, gender, and active status. Active is defined as equal to 1 if the individual falls into the employed, self-employed or farmers category, and equal to 0 if the individual falls into the unemployed, pensioners, or living on social welfare category. \( h \) denotes the latent (general or expected) health status. We assume that \( h \) follows standard normal distribution. If \( y_i > 0 \), then we observe \( y_i \), otherwise, we do not. Therefore,

\[
f(y_i|I_i) = \int f(y_i|I_i, h_i) \phi(h_i)dh_i \\
= \int \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \phi(h_i)dh_i \quad (2.8)
\]

since for Poisson distribution with mean \( \lambda \), the density is given by \( f(y_i) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \).

In order to calculate the integral, for each \( i \), we draw \( h \) from the standard normal distribution 100 times, calculate the integrand then take the average, i.e.,

\[
f(y_i|I_i) = \frac{1}{100} \sum_{j=1}^{100} \frac{\lambda_{ij}^{y_i} \exp(-\lambda_{ij})}{y_i!} \quad (2.9)
\]

where

\[
\lambda_{ij} = \exp(\beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 male_i + \beta_4 active_i + \beta_5 I_i + \beta_6 h_i^2 \\
+ \beta_7 age_i \ast I_i + \beta_8 age_i^2 \ast I_i + \beta_9 male_i \ast I_i + \beta_{10} active_i \ast I_i + \beta_{11} h_i \ast I_i)
\]

\( j = 1, 2, ..., 100 \).

Now, the probability of observing the patient with insurance visiting the hospital can be
computed as:

\[
\Pr(y_i > 0|I_i = 1) = \int \Pr(y_i > 0|I_i = 1, h_i) \phi(h_i) dh_i
= \int \{1 - \exp(-\lambda_i)\} \phi(h_i) dh_i
= 1 - \int \exp(-\lambda_i) \phi(h_i) dh_i
= 1 - \frac{1}{100} \sum_{j=1}^{100} \exp(-\lambda_{ij})
\] (2.10)

since the probability \( \Pr(y_i > 0) = 1 - \Pr(y_i = 0) = 1 - e^{-\lambda} \), and we randomly draw \( h \) from standard normal distribution. Here,

\[
\lambda_{ij} = \exp(\beta_0 + \beta_1 \text{age}_i + \beta_2 \text{age}_i^2 + \beta_3 \text{male}_i + \beta_4 \text{active}_i + \beta_5 \times 1 + \beta_6 h_j^i
+ \beta_7 \text{age}_i \times 1 + \beta_8 \text{age}_i^2 \times 1 + \beta_9 \text{male}_i \times 1 + \beta_{10} \text{active}_i \times 1 + \beta_{11} h_j^i \times 1)
\]
\[j = 1, 2, \ldots, 100.\]

Similarly, observing the patient without the insurance visiting the hospital can be computed as:

\[
\Pr(y_i > 0|I_i = 0) = \int \Pr(y_i > 0|I_i = 0, h_i) \phi(h_i) dh_i
= 1 - \int \exp(-\lambda_i) \phi(h_i) dh_i
= 1 - \frac{1}{100} \sum_{j=1}^{100} \exp(-\lambda_{ij})
\] (2.11)

where:

\[
\lambda_{ij} = \exp(\beta_0 + \beta_1 \text{age}_i + \beta_2 \text{age}_i^2 + \beta_3 \text{male}_i + \beta_4 \text{active}_i + \beta_5 \times 1 + \beta_6 h_j^i
\]
\[j = 1, 2, \ldots, 100.\]

Next, we need to model the decision whether to acquire insurance or not. Let \( I_i^* \) denotes the tendency to purchase insurance,

\[
I_i^* = \gamma_0 + \gamma_1 \text{age}_i + \gamma_2 \text{male}_i + \gamma_3 \text{active}_i + h_i,
\]

\[
I_i = 1 \quad \text{if} \quad I_i^* > 0
= 0 \quad \text{otherwise.}
\]
Same as in the previous model, \( h_i \) is the latent health status, assumed to follow a standard normal distribution. The purchasing decision depends on individual’s age, gender and employment status. The estimates of \( \gamma \)’s are obtained by probit regression. Based on the model, gender also plays a role in the selection of insurance plan. Since we have controlled for the health status, other potential sources of selection in the gender category could be risk aversion or perhaps some other unobserved systematic differences between sexes.

Based on the distribution of \( I_i \),

\[
Pr(I_i = 1) = Pr(I_i^* > 0) = Pr(h_i > -\gamma_0 - \gamma_1 \text{age}_i - \gamma_2 \text{male}_i - \gamma_3 \text{active}_i) = \Phi(\gamma_0 + \gamma_1 \text{age}_i + \gamma_2 \text{male}_i + \gamma_3 \text{active}_i) \tag{2.12}
\]

and

\[
Pr(I_i = 0) = 1 - \Phi(\gamma_0 + \gamma_1 \text{age}_i + \gamma_2 \text{male}_i + \gamma_3 \text{active}_i) \tag{2.13}
\]

where \( \Phi \) denotes the cdf of standard normal distribution.

Finally,

\[
f(I_i) = Pr(I_i = 1)^{I_i} Pr(I_i = 0)^{1-I_i} \tag{2.14}
\]

and the log-likelihood function for any \((y_i, I_i)\) in the sample can be obtained by plugging all the probability terms, Equation (2.9)–(2.14) into Equation (2.7) and taking the logarithm. \( \beta \)'s are found by maximizing the sum of all log likelihood functions.

### 2.6 Empirical Results

The coefficients and their marginal effects from the probit regression is displayed in Table 2.5. Each additional year of age increases the probability of purchasing the insurance by 0.004. Males have 0.05 less probability of purchasing than females. Active people have 0.13 less probability of purchasing than inactive people because the inactive group is dominated by pensioners (80%).

The estimates and standard errors for the maximum likelihood estimation are displayed in Table 2.6. Standard errors and test statistics are obtained through bootstrapping. The marginal effects represent the change in the expected outcome due to a one-unit change in the regressor. For example, other things being equal, each additional year of age increases the expected number of visits by 0.265 for insured group \((I_i = 1)\), while 0.201 for uninsured group \((I_i = 0)\).\(^4\) For people in different employment status, insured active people have 0.307 more hospital visits than

---

\(^4\)Keep in mind that uninsured here only means no supplemental insurance because all citizens of Croatia are always insured by the compulsory insurance.
insured inactive people, whereas uninsured active people have 3.429 more visits than uninsured inactive people. Therefore, active people have lower moral hazard effect, which makes sense due to the fact that active people have higher opportunity cost of time. Insured people have 0.705 more visits than uninsured people, which is the evidence for moral hazard effect.

Regarding the gender difference in the hospital visits, the estimates show that insured males have 0.056 fewer visits than insured females, whereas uninsured males have 0.152 more visits than uninsured females. But these results are insignificant. Therefore, after correcting for the sample selection issue, our empirical results provides no evidence of gender difference in moral hazard effect.

The coefficient for health status $h$ shows that each additional unit increase in $h$ increases the expected number of hospital visits by 4.144 for insured people, 4.380 for the uninsured (higher $h$ indicates worse health). Since $h$ is also the error term in the insurance equation, higher $h$ means higher likelihood of purchasing the insurance. Therefore, there is also an evidence of adverse selection.

### 2.7 Conclusion

The lack of the consistent explanation for the empirically observed gender differences in health care consumption in itself serves as strong motivation for an innovative empirical research in this area. More importantly, to the best of our knowledge, the economics literature appears to be completely silent when it comes to explaining the gender differences in health care consumption by a possible gender differences in the magnitude of asymmetric information effects. Since asymmetric information problems, be they of moral hazard or adverse selection nature, have been recognized as one of the important drivers of health care costs, if men and women are different with respect to how they respond to incentives, then knowing what those differences are, could be important for correct pricing of insurance policies. However, differential pricing of insurance policies with respect to gender may not be allowed by law. For example, the Patient Protection and Affordable Care Act (Obama-care) prohibits “gender rating”, starting in 2014 (Arons 2012). Women cannot be discriminated against in the health insurance market, with higher premiums, lost maternity coverage and denials of coverage for gender-related pre-existing conditions. The results of this research could provide important insights about possible cost of such regulatory requirements.

Another challenge is that we only have users data such that the sample selection issue needs to be dealt with. We use truncated count model and simulated maximum likelihood approach to estimate gender differences in moral hazard in health care utilization relying on the hospital invoices data for non life-threatening diagnoses. Lack of information of non-users who did not show up at the hospital, we derive conditional likelihood density to adjust this
sample selection issue in the estimation. The results show no evidence of statistically significant gender difference in hospital visits and moral hazard effects in the population we studied. But we did find significant moral hazard effect. People who bought the insurance have 35% more hospital visits than people who do not have the insurance.\textsuperscript{5}

\textsuperscript{5}Insured people have about 0.7 more visits, the average visits for uninsured group were about 2.
Table 2.1: Summary of Literature about Gender Difference in Risk Aversion

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Field of Study</th>
<th>More Risk Averse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arano et al.(2010)</td>
<td>Retirement asset allocation</td>
<td>Women</td>
</tr>
<tr>
<td>Charness and Gneezy(2012)</td>
<td>Financial investment</td>
<td>Women</td>
</tr>
<tr>
<td>Bajtelsmin and Van Derhei(1997)</td>
<td>Pension investment</td>
<td>Women</td>
</tr>
<tr>
<td>Powell and Ansic(1997)</td>
<td>Insurance for loss</td>
<td>Women</td>
</tr>
<tr>
<td>Gneezy et al.(2009)</td>
<td>Risk associated with selecting into competitive environments</td>
<td>Same</td>
</tr>
<tr>
<td>Flynn et al.(1994)</td>
<td>Environmental and health risk</td>
<td>Women in white</td>
</tr>
<tr>
<td>Schubert et al.(1999)</td>
<td>Financial decisions</td>
<td>Same when facing contextual decisions.</td>
</tr>
<tr>
<td>Byrnes et al.(1999)</td>
<td>General risk taking</td>
<td>Women</td>
</tr>
<tr>
<td>Harris et al.(2006)</td>
<td>General risk taking</td>
<td>Women</td>
</tr>
<tr>
<td>Fehr-Duda et al.(2006)</td>
<td>Lottery experiments</td>
<td>Women</td>
</tr>
<tr>
<td>Barsky et al.(1997)</td>
<td>General risk tolerance</td>
<td>Women</td>
</tr>
<tr>
<td>Bernasek and Shwiff(2001)</td>
<td>Asset Investment</td>
<td>Women</td>
</tr>
<tr>
<td>Lindquist and Save-Soderbergh(2011)</td>
<td>Game show wagers</td>
<td>Women</td>
</tr>
<tr>
<td>Booth and Nolen(2012)</td>
<td>Gambling experiment</td>
<td>Women in coed schools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Same in single-sex schools</td>
</tr>
</tbody>
</table>

35
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Field of Study</th>
<th>More Risk Averse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ronay and Kim (2006)</td>
<td>General risk taking</td>
<td>Same on individual level</td>
</tr>
<tr>
<td>Beckmann and Menkhoff (2008)</td>
<td>Investment decisions</td>
<td>Women</td>
</tr>
<tr>
<td>Dohmen et al. (2011)</td>
<td>General risk taking</td>
<td>Women</td>
</tr>
<tr>
<td>Olsen and Cox (2001)</td>
<td>Investment attitudes</td>
<td>Women</td>
</tr>
<tr>
<td>Ertac and Gurdal (2012)</td>
<td>Financial decisions</td>
<td>Women</td>
</tr>
<tr>
<td>Meier-Pesti and Penz (2008)</td>
<td>Investment decisions</td>
<td>Women</td>
</tr>
<tr>
<td>Charness and Gneezy (2010)</td>
<td>Asset investment</td>
<td>Women</td>
</tr>
<tr>
<td>Sunden and Surette (1998)</td>
<td>Retirement asset allocations</td>
<td>Women</td>
</tr>
<tr>
<td>Barber and Odean (2001)</td>
<td>Stock Investment</td>
<td>Women</td>
</tr>
<tr>
<td>Finucane et al. (2000)</td>
<td>Hazards risk</td>
<td>Women</td>
</tr>
<tr>
<td>Kahan et al. (2007)</td>
<td>Environmental risk</td>
<td>Women</td>
</tr>
<tr>
<td>Eriksson and Simpson (2010)</td>
<td>Lottery experiments</td>
<td>Women</td>
</tr>
<tr>
<td>Hartog et al. (2002)</td>
<td>Financial Lottery</td>
<td>Women</td>
</tr>
<tr>
<td>Rivers et al. (2010)</td>
<td>Health and environmental risk</td>
<td>Women</td>
</tr>
<tr>
<td>Charness and Genicot (2009)</td>
<td>Risk sharing experiment</td>
<td>Women</td>
</tr>
<tr>
<td>Borghans et al. (2009)</td>
<td>Betting experiment</td>
<td>Women</td>
</tr>
<tr>
<td>Dreber et al. (2010)</td>
<td>Financial investment</td>
<td>Women</td>
</tr>
</tbody>
</table>
Table 2.2: Summary Statistics by Insurance Category and Gender

<table>
<thead>
<tr>
<th>Sample</th>
<th>Male</th>
<th>Female</th>
<th>∆ in gender (M-F)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Visit per patient</td>
<td>N</td>
<td>Visit per patient</td>
</tr>
<tr>
<td>Bought</td>
<td>3.231</td>
<td>5,202</td>
<td>3.380</td>
</tr>
<tr>
<td></td>
<td>(3.689)</td>
<td></td>
<td>(4.075)</td>
</tr>
<tr>
<td>No</td>
<td>2.007</td>
<td>1,316</td>
<td>1.954</td>
</tr>
<tr>
<td></td>
<td>(2.033)</td>
<td></td>
<td>(2.185)</td>
</tr>
</tbody>
</table>

Note: ∆ in gender denotes the difference of means between genders, calculated by mean of male minus mean of female. Standard errors are in the parentheses. *** indicates 1 % significance level; ** 5 % significance level and * 10 % significance level.

Table 2.3: Summary Statistics by Insurance Category and Employment Status

<table>
<thead>
<tr>
<th>Sample</th>
<th>Active</th>
<th>Inactive</th>
<th>∆ in Employment (A-I)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Visit per patient</td>
<td>N</td>
<td>Visit per patient</td>
</tr>
<tr>
<td>Bought</td>
<td>3.458</td>
<td>6,124</td>
<td>3.184</td>
</tr>
<tr>
<td></td>
<td>(4.203)</td>
<td></td>
<td>(3.622)</td>
</tr>
<tr>
<td>No</td>
<td>2.066</td>
<td>2,121</td>
<td>1.490</td>
</tr>
<tr>
<td></td>
<td>(2.219)</td>
<td></td>
<td>(1.126)</td>
</tr>
</tbody>
</table>

Note: ∆ in employment denotes the difference of means between employment groups, calculated by mean of active group minus mean of inactive group. Standard errors are in the parentheses. *** indicates 1 % significance level; ** 5 % significance level and * 10 % significance level.
<table>
<thead>
<tr>
<th>Diagnoses subject to co-payment (top 10)</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Outpatient physical therapy</td>
<td>7.84%</td>
</tr>
<tr>
<td></td>
<td>Medical biochemistry</td>
<td>3.20%</td>
</tr>
<tr>
<td></td>
<td>Radiology</td>
<td>2.62%</td>
</tr>
<tr>
<td></td>
<td>Traumatology</td>
<td>2.32%</td>
</tr>
<tr>
<td></td>
<td>Ophthalmology</td>
<td>2.31%</td>
</tr>
<tr>
<td></td>
<td>Otorhinolaryngology</td>
<td>1.92%</td>
</tr>
<tr>
<td></td>
<td>Urology</td>
<td>1.81%</td>
</tr>
<tr>
<td></td>
<td>Dermatology and venereology</td>
<td>1.78%</td>
</tr>
<tr>
<td></td>
<td>Orthopedy</td>
<td>1.34%</td>
</tr>
<tr>
<td></td>
<td>Physical medicine and rehabilitation</td>
<td>1.32%</td>
</tr>
<tr>
<td></td>
<td>Other diagnoses</td>
<td>15.34%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>41.80%</td>
</tr>
</tbody>
</table>
Table 2.5: Probit Regression: Marginal Effects of Independent Variables to the Probability of Purchasing Insurance

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Coefficient</th>
<th>$\frac{dy}{dx}$</th>
<th>t-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_i$</td>
<td>$Age_i$</td>
<td>0.021 (0.001)</td>
<td>0.004 (0.000)</td>
<td>22.28</td>
</tr>
<tr>
<td></td>
<td>$Male_i$</td>
<td>-0.236 (0.026)</td>
<td>-0.051 (0.006)</td>
<td>-9.04</td>
</tr>
<tr>
<td></td>
<td>$Active_i$</td>
<td>-0.605 (0.033)</td>
<td>-0.131 (0.007)</td>
<td>-18.37</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>0.502 (0.061)</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: Marginal effects are obtained through probit estimation; standard errors are in the parentheses.

Table 2.6: Maximum Likelihood Estimation: Hospital Visits

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Coefficients ($\beta$)</th>
<th>Standard Error</th>
<th>t-statistics</th>
<th>Marginal Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_i$</td>
<td>$Age_i$</td>
<td>0.066</td>
<td>0.014</td>
<td>4.667</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>$Age_i^2$</td>
<td>-0.001</td>
<td>0.000</td>
<td>-3.372</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>$Male_i$</td>
<td>0.049</td>
<td>0.113</td>
<td>0.433</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>$Active_i$</td>
<td>1.106</td>
<td>0.183</td>
<td>6.036</td>
<td>3.429</td>
</tr>
<tr>
<td></td>
<td>$I_i$</td>
<td>1.357</td>
<td>0.343</td>
<td>3.961</td>
<td>4.207</td>
</tr>
<tr>
<td></td>
<td>$h_i$</td>
<td>1.413</td>
<td>0.050</td>
<td>28.225</td>
<td>4.380</td>
</tr>
<tr>
<td></td>
<td>$Age_i * I_i$</td>
<td>0.021</td>
<td>0.016</td>
<td>1.287</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>$Age_i^2 * I_i$</td>
<td>-0.000</td>
<td>0.000</td>
<td>-0.892</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>$Male_i * I_i$</td>
<td>-0.067</td>
<td>0.130</td>
<td>-0.515</td>
<td>-0.208</td>
</tr>
<tr>
<td></td>
<td>$Active_i * I_i$</td>
<td>-1.007</td>
<td>0.194</td>
<td>-5.186</td>
<td>-3.122</td>
</tr>
<tr>
<td></td>
<td>$h_i * I_i$</td>
<td>-0.076</td>
<td>0.061</td>
<td>-1.251</td>
<td>-0.236</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>-3.876</td>
<td>0.266</td>
<td>-14.562</td>
<td>-12.016</td>
</tr>
</tbody>
</table>

Note: Coefficients are obtained through maximum likelihood estimation; standard errors are calculated based on 200 bootstrap iterations.
Chapter 3

Using Age-Based Insurance Eligibility Criterion to Estimate Moral Hazard in Medical Care Consumption

3.1 Introduction

A large body of health economics literature documents a strong association between health insurance status and patterns of health care utilization. People with more generous insurance coverage tend to consume more health care, a phenomenon known as moral hazard. These results suggest that when individuals lose health insurance, they change their consumption of health care services, i.e. they seek medical attention less frequently. But would the uninsured consume more health care if they had health insurance? Because, compared to those insured, uninsured individuals are likely to have different health conditions, attitudes towards risk, wealth, etc., the casual inference becomes rather difficult. To overcome this problem, we exploit quasi-experimental variation in insurance status brought about by the rules the insurance companies use to establish the coverage eligibility of dependents. One of the largest segments of the population that lacks the health insurance are young adults (age 19 to 29). For example, in the United States, 29% of the uninsured are young adults (Schwartz and Schwartz, 2008). Because many health insurance contracts cover depends until the age of 18 and only cover older dependents if they are full time students, significant number of teenagers become uninsured after they reach the threshold birthday. The fact that young adults are relatively healthier than older population, it is reasonable to assume that they would consume relatively...
low levels of health care regardless of whether they are insured or uninsured and that the social cost of extending the coverage to the uninsured young adults should be relatively modest.

In this paper, we rely on the fuzzy regression discontinuity (RD) design to estimate the effect of losing the insurance on health care consumption in young adults using invoices data for outpatient hospital services from a regional hospital in Croatia. Croatian case in interesting because of the unique feature of the state-run health insurance system which is dominated by a single public health insurance fund: the Croatian Institute for Health Insurance (HZZO). The HZZO offers two types of insurance: the compulsory and the supplemental insurance. Compulsory insurance covers two kinds of medical care services: one with full coverage, the other with a system of co-payments. Full coverage of medical care services are provided to children younger than 18, whereas the rest of the population is fully covered only for relatively serious types of illnesses. All other health services are subject to a system of co-payments. In order to avoid paying those co-payments a person needs to have the supplemental insurance. The compulsory insurance coverage is universal whereas the supplemental insurance can be either bought or is extended automatically free of charge to some categories of citizens such as, for example, full time students. In this context, the 18th birthday represents a threshold for the supplemental insurance coverage where young adults crossing the threshold will loose the full coverage and face three options: stay in school and continue being fully covered for free, buy the supplemental insurance and continue enjoying full coverage or refuse to buy the supplemental coverage, stay uninsured and pay co-payments as required. Since crossing the 18th birthday threshold is not a unique determinant of the assignment into the treatment (losing the coverage), the problem fits into the fuzzy RD design. RD design can be used to determine the treatment effects in quasi-experimental settings where treatment is determined by a forcing variable exceeding the threshold. The probability of being treated at the threshold jumps from zero to one (see: Lee and Lemieux, 2010). In the fuzzy RD design, the forcing variable does not exclusively determine the treatment assignment, hence the discontinuity in the probability of being treated at the threshold is less than one.

Regression discontinuity design has been used extensively to determine the causal effect from an intervention. For example, Card, Dobkin, and Maestas (2009) compared the health care consumption among people just before and just after the age of 65, the threshold for Medicare eligibility, in a sharp RD design. They found that the Medicare Eligibility causes a discontinuous increase in health care utilization. Van Der Klaauw (2002) identified the causal effect of financial aid on college enrollment decisions using fuzzy RD approach. The enrollment rate increased by about 0.2 at the financial aid eligibility threshold. De Paola and Scoppa (2014) applied the fuzzy RD design to estimate the effects of remedial courses on student achievement. Students who have received a test score below the threshold are assigned to the treatment, taking the remedial courses. They found that the treated students gain more
credits and have a lower probability of dropping out than students who were just above the threshold. Vardardottir (2013) studied ability peer effects among students who were just above the treatment-determining grade cutoff and assigned to high-ability classes and students who were just below the cutoff and assigned to normal classes in a fuzzy RD design. Being assigned to a high-ability class increases students’ year grade and spring exam results. Stancanelli and Van Soest (2012) estimate the effect of retirement at age 60 on housework hours in a fuzzy RD design. They found a substantial increase in housework hours both for men and women upon their own retirement. Chen and van der Klaauw (2008) evaluated the effect of disability insurance program on labor force participation among disability insurance beneficiaries in a fuzzy RD design. They found that the labor force participation rate among those beneficiaries would have been 20% higher if none had received the benefits. Anderson, Dobkin, Gross (2012) used the age 19 as an instrument to identify the causal effect of losing health insurance at age 19 on health care consumption. They found that not having insurance leads to a decreasing level of health care consumption, including a 40 percent reduction in emergency department visits and a 61 percent reduction in inpatient hospital admissions.

A distinct feature of our data set which considerably complicates the estimation is that it consists of users only. People that did not use the medical services of that hospital during the time period covered by the data do not show up in the data. The problem of estimating the moral hazard effect with users only data is caused by the fact that some of the sick people would not seek medical attention at all or would seek medical attention less often precisely because they do not have the insurance. Therefore, using users only data would clearly underestimate the moral hazard effect. To deal with this attenuation bias we use instrumental variables (IV) approach from Anderson, Dobkin and Gross (2012) that relies on the assumption that the net change in the observed hospital visits after turning 18 is driven only by individuals who lost the insurance coverage, an assumption implied by the standard IV exclusion restriction. We found a statistically significant reduction in the number of hospital visits for young adults who lost the supplemental insurance after turning 18, confirming the moral hazard hypothesis. The hospital visits dropped either 89% or 95% depending on whether the effect was estimated with a one-year or a two-year data window surrounding the 18th birthday.

3.2 Institutional Framework and Data Description

Croatian health insurance system, a relic of the socialist past time, is an extremely generous and, of course, an expensive system to maintain. The compulsory insurance, which is funded by a 15% payroll tax, covers two kinds of medical care services: one with full coverage, the other with a system of co-payments. Full coverage medical care services are provided to children up to 18 years of age and pregnant women and for everybody else for life threatening types of
illnesses such as infectious diseases, psychiatric care, surgeries, cancers, mandatory vaccinations, etc. All other health services (including but not limited to primary care, hospitals stays and prescription drugs) are subject to a system of co-payments. The patients are required to pay 20\% of the full price of medical care, with the largest out-of-pocket cost share amount set at 3,000.00 HRK per invoice.\(^1\) The supplemental insurance is a voluntary insurance that can be acquired by a person 18 years or older, having compulsory insurance, by signing a contract with the HZZO. Certain categories of citizens are entitled to the supplemental insurance free of charge, i.e., their premiums are covered from the state budget. The list of people entitled to free supplemental insurance includes, among others, the full time secondary school and college students. For those not entitled to free supplemental coverage, the premia range from 50 to 130 HRK per month depending on whether the insured is active or retired and income. A person having the supplemental insurance is entitled to full waiver of all medical expense co-payments mentioned before.

The original data set consists of all invoices for all outpatient services from a regional hospital in Croatia during the period from March 1 to June 30, 2009.\(^2\) The data set consists of 105,646 observations. Each observation reflects one hospital visit (invoice). The data contains the following set of variables: a numeric code for the type of hospital service provided, compulsory health insurance number, supplemental insurance number (if the patient has one), period covered by the supplemental insurance, numeric code for categories entitled to supplemental insurance free of charge, eligibility category for compulsory insurance, cost of hospital service, part of the cost covered by compulsory insurance, part of cost covered by supplemental insurance, part of cost covered by participation (co-payment), date of birth and sex of the patient. The invoices do not record the exact date when the patient visited the hospital, but are chronologically ordered. To determine the date of the visit, we first divided all invoices into 122 days (March 1 to June 30) and designated the first 866 invoices as March 1st invoices, the next 866 invoices as March 2nd invoices, etc. This is a fairly innocuous simplification because the actual number of daily appointments is determined by the hospital’s outpatient capacity and it is, therefore, reasonable to assume that the average number of patients treated each day is approximately the same. Since we know the date of birth of each patient and knowing the day of the visit, we can calculate the number of days each patient was away from the 18th birthday when visited the hospital. If the patient is younger than 18, the days away from the 18th birthday are recorded as a negative number and if the patient is older than 18 then those days are recorded as a positive number. Finally, we create the variable $\text{week}$ by converting the number of days away from the 18th birthday into the number of weeks away from the 18th birthday.

\(^1\)The figures are reflective of the year 2009 which is the year covered by the dataset. The exchange rate for the local currency, Croatian Kuna (HRK), as of June 20, 2009 was 1USD=5.19 HRK. A more detailed description of the health insurance system in Croatia can be found in Liu, Nestic and Vukina (2012).

\(^2\)The name and the location of the hospital are suppressed for confidentiality reasons.
For the purposes of this study all invoices for the patients outside the 17 to 19 years of age were dropped, leaving the sample of 1,586 invoices in total. The summary statistics for this one-year window surrounding the patients’ 18th birthday are presented in Table 3.1. We first look at the invoice level. If the invoice shows that the patient visited the hospital with both compulsory and supplemental insurance, we call it an insured visit, otherwise it is an uninsured visit. Notice, that because of the existence of compulsory insurance, even the uninsured visits are in fact insured by the compulsory insurance, these patients only lack the supplemental coverage. So, in everything that we do in this paper, an insured visit identifies a visit to the hospital where the patient did not have the supplemental insurance. Therefore, the younger than 18 group \((Y - 18)\) has only insured visits whereas the older than 18 group \((O - 18)\) has both insured and uninsured visits. The number of uninsured visits in the \(O - 18\) group is only 2.40% of the total visits. In the right-hand-side panel of Table 3.1 we collapsed the sample to the patient level to compare the health care utilization between two groups. The \(O - 18\) group has less people visiting the hospital, less visits per patient compared to the \(Y - 18\) group. We also expand to a two-year window surrounding the patients’ 18th birthday and present the summary statistics in Table 3.5.

3.3 Regression Discontinuity Design

In the basic setting for the regression discontinuity design researchers are interested in the causal effect of a binary intervention or treatment. Individuals are either exposed or not exposed to a treatment. Let \(W_i \in \{0, 1\}\) denote the treatment received, with \(W_i = 1\) if unit \(i\) was exposed to the treatment, and \(W_i = 0\) otherwise. Let \(Y_i(w_i)\) denote the outcome; \(Y_i(0)\) is the outcome without exposure to the treatment and \(Y_i(1)\) is the outcome with exposure to the treatment. We are interested in the difference \(Y_i(1) - Y_i(0)\). Since one can never observe \(Y_i(0)\) and \(Y_i(1)\) at the same time, one needs to use the average effects of the treatment over population or sample to estimate the average treatment effect.

It’s critical to distinguish two RD settings, the sharp and the fuzzy regression discontinuity designs. Based on Imbens and Lemieux (2008), in a sharp regression discontinuity (SRD) design, the assignment \(W_i\) is a deterministic function of one of the covariates, i.e., the forcing (or treatment-determining) variable \(X\); \(w_i = 1\) if \(X_i \geq c\), \(w_i = 0\) if \(X_i < c\). Therefore, all units with the value of the forcing variable \(X_i\) at least \(c\) are assigned to the treatment and participation is mandatory; all units with the value of forcing variable \(X_i\) less than \(c\) are assigned to the control group. In the SRD design, the interest is to use the discontinuity in the conditional expectation of the outcome given the forcing variable in order to estimate the average treatment effect:

\[
\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x],
\] (3.1)
which is interpreted as the average causal effect of the treatment at the discontinuity point

$$\tau_{SRD} = E[Y_i(1) - Y_i(0)|X_i = c]. \quad (3.2)$$

In the fuzzy regression discontinuity (FRD) design, however, the probability of being treated does not need to jump from zero to one at the threshold. The design allows other factors to influence the assignment to treatment, besides the forcing variable. Therefore, only a jump in the probability of assignment to the treatment at the threshold is required:

$$\lim_{x \downarrow c} \Pr(W_i = 1|X_i = x) \neq \lim_{x \uparrow c} \Pr(W_i = 1|X_i = x). \quad (3.3)$$

The average causal effect of the treatment can be identified by the ratio of the jump in the outcome at the threshold to the jump in the treatment indicator at the threshold:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y|X = x] - \lim_{x \uparrow c} E[Y|X = x]}{\lim_{x \downarrow c} E[W|X = x] - \lim_{x \uparrow c} E[W|X = x]}. \quad (3.4)$$

### 3.3.1 Instrumental Variable Approach

To estimate the ratio in Equation (3.4) we follow the approach of Anderson, Dobkin and Gross (2012). To explain the procedure, consider the following reduced form model of the effects of health insurance coverage on health care utilization:

$$Y_i = \gamma_0 + \gamma_1 D_i + \epsilon_i, \quad (3.5)$$

where $Y_i$ indicates health care utilization of individual $i$ and $D_i$ is an indicator variable equal to one if individual $i$ has health insurance. All other determinants of health care utilization are summarized in the error term $\epsilon_i$. Therefore, the coefficient $\gamma_1$ denotes the causal effect of health insurance on health care utilization, also known as the moral hazard effect.

However, this moral hazard effect is contaminated by other factors because the insurance coverage variable $D_i$ is correlated with other determinants of health care utilization in $\epsilon_i$, for example, the unobserved health status of that individual. An individual chooses to acquire health insurance based on idiosyncracies that simultaneously affect the decision to be insured and the consumption of health care. Therefore the estimate for $\gamma_1$ is inconsistent due to endogeneity. To solve the problem one typically relies on the IV approach. The objective is to estimate the causal effect of loosing health insurance coverage on the number of hospital visits. Using crossing the 18th birthday as an instrument, we estimate the first stage – the share of young adults who loose the insurance coverage while crossing the age-18 threshold and the reduced form – the change in the number of hospital visits associated with crossing the 18th birthday. We identify
the effect of the moral hazard (coefficient $\gamma_1$) by dividing the reduced form estimate (the effect of turning 18 on the number of visits) by the first stage estimate (the effect of turning 18 on health insurance coverage).\footnote{This strategy is analogous to using the age-18 discontinuity as an instrument to identify the causal effect of health insurance, see Hahn, Todd and Van der Klaauw (2001).}

Notice however, that having hospital invoices data introduces sample selection bias in the first stage estimation because we only observe the insurance status for individuals who show up in the hospital and consume some services for which they or their insurers are charged. Regression estimate of the change in the proportion of uninsured after crossing the age-18 threshold understates the true size of this change. Because losing insurance reduces the likelihood of a hospital visit and therefore affects the probability of appearing in the sample, this selection mechanism leads to attenuation bias when estimating the change in the insurance coverage because newly uninsured individuals are more likely to leave the sample.

To correct for the bias in the first-stage estimates we assume that the net change in the observed hospital visits after crossing the age-18 threshold is driven only by individuals who lost insurance coverage. This assumption is implied by the standard IV exclusion restriction and is quite reasonable. Because all our patients in the $Y - 18$ group are fully insured, those types that did not go to the hospital before 18 (when they had insurance) and hence did not show up in the data as users, have no reason to visit hospital immediately after turning 18 (even if they have insurance) and will surely not go if they don’t have insurance.

Let $D_i$ denote the insurance coverage indicator and $A_i$ denote age of an individual, then the effect of crossing age-18 threshold on insurance coverage at the population level can be expressed as:

$$\lim_{a \downarrow 18} E[D_i|A_i = a] - \lim_{a \uparrow 18} E[D_i|A_i = a].$$

(3.6)

Next, let $D_i(1) = 1$ indicate an individual older than 18 with the insurance coverage, $D_i(0) = 0$ an individual older than 18 without the insurance, $D_i(0) = 1$ and $D_i(0) = 0$ are defined similarly for individuals younger than 18. Let $Y_i(1) = 1$ denote an individual older than 18 who visited the hospital, $Y_i(0) = 0$ an individual older than 18 who did not visit the hospital, $Y_i(0) = 1$ and $Y_i(0) = 0$ are defined similarly for individuals younger than 18. Since we could only observe individuals who visited the hospital (users), the effect of turning 18 on insurance coverage among users is estimated as:

$$E[D_i(1)|Y_i(1) = 1] - E[D_i(0)|Y_i(0) = 1].$$

(3.7)

However, we would like to estimate the effect of turning 18 on insurance coverage at the popu-
The desired effect is estimated as follows. First we denote the number of visits made before 18 as $y_0$, the number of insured visits made before 18 as $d_0$, the number of visits made after 18 as $y_1$ and the number of insured visits made after 18 as $d_1$. The ratios $\frac{d_0}{y_0}$ and $\frac{d_1}{y_1}$ represent the fractions of insured visits before and after 18 respectively. The corresponding fractions of uninsured visits before and after 18 are denoted as $(1 - \frac{d_0}{y_0})$, $(1 - \frac{d_1}{y_1})$ respectively. It can be shown that the bias-corrected estimate for the effect of turning 18 on the insurance coverage is obtained as:

$$\frac{d_1 - d_0}{y_0} \cdot \frac{p}{1-p} \cdot E[D_i(1) - D_i(0)|Y_i(0) = 1],$$

which equals the population level estimate we want in Equation (3.8).

Using the assumption that number of visits per patient is constant, Equation (3.9) can be estimated as:

$$\frac{d_0 - d_1}{y_0} \cdot \frac{1}{y_1} = \lim_{a \downarrow 18} E[D_i|A_i = a] \cdot \lim_{a \downarrow 18} E[Y_i|A_i = a] - \lim_{a \uparrow 18} E[D_i|A_i = a].$$

Since we are interested in estimating how an increase in the proportion of the uninsured affects the change in hospital visits as people turn 18, we need to estimate the following equation:

$$\frac{d_0 - d_1}{y_0} \cdot \frac{1}{y_0} = \lim_{a \downarrow 18} E[U_i|A_i = a] \cdot \lim_{a \downarrow 18} E[Y_i|A_i = a] - \lim_{a \uparrow 18} E[U_i|A_i = a].$$

where $U_i$ is an indicator equal to one if individual $i$ is uninsured, i.e., $U_i = 1 - D_i$.

In the second step, we estimate the reduced form equation, i.e. the percentage decline in visits due to crossing the age-18 threshold, which can be written as:

$$\frac{y_1 - y_0}{y_0},$$

and obtained by the following estimator:

$$\lim_{a \downarrow 18} E[Y_i|A_i = a] - \lim_{a \uparrow 18} E[Y_i|A_i = a].$$

The proof of this result is contained in the web support to Anderson, Dobkin, Gross (2012).
It can be shown that (3.12) converges to:

$$\frac{y_1 - y_0}{y_0} \xrightarrow{p} E[Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = -1, Y_i(0) = 1] * E[D_i(1) - D_i(0)|Y_i(0) = 1]$$  (3.14)

where the first term $E[Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = -1, Y_i(0) = 1]$ denotes the clean moral hazard for individuals who visited the hospital before turning 18 and lost the insurance coverage after turning 18 and the second term $E[D_i(1) - D_i(0)|Y_i(0) = 1]$ denotes the first stage estimate from Equation (3.9), i.e., the change in insurance coverage due to turning 18.\footnote{The proof of this result is also contained in the web support to Anderson, Dobkin, Gross (2012).} Therefore, the clean moral hazard is the ratio between the reduced form estimate and the first stage estimate.

### 3.4 Estimation and Results

We start with graphical presentation of the data which represents a helpful and informative tool in determining whether the quasi natural experiment that generated the data fits the RD design. We collapse the data based on the variable *week* and count the number of total visits and uninsured visits in each week surrounding the patients’ 18th birthday. The scatter plot and fitted values of the percentages of uninsured visits in each week away from age-18 are depicted in Figure 3.1. The fitted lines are estimated separately on each side of the cutoff point. There is a clear jump in the percentage of uninsured visits at the cutoff point at age-18. The proportion of uninsured visits is zero for people younger than 18 because all are fully covered by both compulsory and supplemental insurance. The scatter plot and fitted values of the log of the total visits in each week surrounding the 18th birthday are displayed in Figure 3.2. There is a sharp drop in hospital visits at the cutoff point at age-18. The graphs provide a powerful visual instrument to the credibility of the regression discontinuity design.

The primary concern in the RD design is that factors other than insurance coverage such as high school graduation, starting college, obtaining driving licence, entering legal drinking age, starting employment, etc., could change discontinuously at the age of 18 and dramatically alter the need to access health care. Because we measure age at the weekly level, only factors that change sharply within few weeks around the age-18 threshold will bias our estimates. As it turns out, most of those obvious confounders should not bias our estimates. First, high school graduations occur in June and universities start classes in September or October but 18th birthdays are randomly distributed throughout the year. Second, unlike in the United States, Croatia has no meaningfully defined and enforced drinking age. Finally, a potentially interesting factor is the legal driving age. At age-18 young adults in Croatia become eligible to drive. However, the fact that drivers’ education process is quite lengthy and expensive, most of the young adults will not obtain their licence exactly on their 18th birthday but sometimes...
during that year or even later.

Based on the framework developed before, in the first step we estimate the limit expectations in Equation (3.11) by Seemingly Unrelated Regression (SUR) using the following two equations:

\[(1 - \frac{d_j}{y_j}) = \alpha_0 + \alpha_1 I_{over18, j} + \alpha_2 aweek_{j} + \alpha_3 aweek_{j} \ast I_{over18, j} + \epsilon_j \tag{3.15}\]
\[\ln y_j = \beta_0 + \beta_1 I_{over18, j} + \beta_2 aweek_{j} + \beta_3 aweek_{j} \ast I_{over18, j} + \mu_j; \quad j = 1, 2, \ldots, 105. \tag{3.16}\]

where \(d_j\) and \(y_j\) are weekly counts of insured visits and total visits respectively, \(aweek_{j}\) is age in week equals to \(\frac{1}{52}\) if the visits are made in the first week after turning 18, equals to \(\frac{2}{52}\) if the visits are made in the second week after turning 18, etc., \(aweek_{j}\) equals to 0 if the visits are made on the eighteenth birthday, \(aweek_{j}\) equals to \(-\frac{1}{52}\) if the visits are made in the first week before turning 18, etc., \(I_{over18, j}\) is an indicator variable equal to one if observation \(j\) is older than 18 and \(\epsilon_j, \mu_j\) are random errors with \(E(\epsilon_j) = 0, E(\mu_j) = 0\) and \(\text{Cov}(\epsilon_j, \mu_j) \neq 0\). The dependent variables are the weekly percent of uninsured visits \((1 - \frac{d_j}{y_j})\) and the natural log of weekly number of visits \(\ln y_j\). We first perform the estimation with the data within a one-year window around the 18th birthday.

Using estimated coefficients from Equation (3.15), the limit expectations in Equation (3.11) can be recovered as follows:

\[\lim_{a \uparrow 18} E[U_i | A_i = a] = \hat{\alpha}_0 + \hat{\alpha}_1 \]
\[\lim_{a \downarrow 18} E[Y_i | A_i = a] = \exp(\hat{\beta}_0 + \hat{\beta}_1) \]
\[\lim_{a \uparrow 18} E[Y_i | A_i = a] = \exp(\hat{\beta}_0) \]
\[\lim_{a \downarrow 18} E[U_i | A_i = a] = \hat{\alpha}_0 \]

where \(\hat{\alpha}_2, \hat{\alpha}_3, \hat{\beta}_2,\) and \(\hat{\beta}_3\) drop out because \(aweek = 0\) at the limit of age-18. Therefore, the first stage estimator for the increment in the proportion of uninsured at age 18 becomes:

\[\frac{(\hat{\alpha}_0 + \hat{\alpha}_1)\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} + 1 - \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} - \hat{\alpha}_0. \tag{3.17}\]

In the second step we estimate the limit expectations in Equation (3.13) using the following equation:

\[\ln y_j = \beta_0 + \beta_1 I_{over18, j} + \beta_2 aweek_{j} + \beta_3 aweek_{j} \ast I_{over18, j} + \mu_j \tag{3.18}\]

which turns out to be identical to the second equation from Equation (3.15). Therefore, the reduced form estimator for the percentage change in hospital visits for individuals turning 18
is:

\[
\frac{\exp(\hat{\beta}_0 + \hat{\beta}_1) - \exp(\hat{\beta}_0)}{\exp(\hat{\beta}_0)} = \exp(\hat{\beta}_1) - 1
\]  

(3.19)

where \( \hat{\beta}_2, \hat{\beta}_3 \) drop out in the limit.

Based on Equation (3.14), it is straightforward to see that the moral hazard effect \( \gamma_1 \) is identified by dividing the effect of turning 18 on hospital visits, i.e. the reduced form estimator from Equation (3.19), by the effect of turning 18 on insurance coverage, i.e. the first stage estimator from Equation (3.17). Thus, the IV estimator can be written as:

\[
\tau_{IV} = \frac{\exp(\hat{\beta}_1) - 1}{(\hat{\alpha}_0 + \hat{\alpha}_1)\exp(\hat{\beta}_0 + \hat{\beta}_1)} + 1 - \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} - \hat{\alpha}_0
\]

(3.20)

All results for one-year window data are displayed in Tables 3.2 to 3.4. The SUR results for Equation (3.15) are presented in Table 3.2. The top panel which displays the results for the percentage of uninsured visits equation indicate that the percentage of uninsured visits is increasing after people turn 18 by about 8%. Splitting the sample by gender reveals exactly the same pattern of behavior for young men and women. We would also expect the coefficient of \( a_{week_j} \times I_{over18_j} \) to be negative to indicate that with the passage of time after the 18th birthday, people who lost the supplemental insurance would gradually become more inclined to purchase the coverage. The coefficient has the right sign, but it is not significantly different from zero, indicating that, at least during the first year after the 18th birthday, the percentage of young adults with the supplemental insurance will not increase. The results seem to be indicating that all those that have the supplemental insurance beyond the 18th birthday are those that have it by default, i.e. by maintaining their full time student status rather then actually purchasing the policy.

The bottom panel of Table 3.2 displays the results for the log of the hospital visits equation. The results indicate that the number of hospital visits decrease by about 1.65 visits (ln(0.5)) as the young adults cross the age-18 birthday threshold. Again there is hardly any difference in the behavior between sexes. The OLS regression results of the reduced form, Equation (3.18), are displayed in Table 3.3. The coefficients are the same as those obtained using SUR which are displayed in the bottom panel of Table 3.2 but the SUR results are more efficient.\(^6\)

Recall that the above results presented in in Tables 3.2 and 3.3 were obtained with hospital users only data and as such are not generally valid, whereas we are interested in the results that would be valid for the population as a whole. Table 3.4 summarizes the results of our analysis. The asymptotically correct (bias corrected) estimates of the increment in the percentage of

---

\(^6\)The SUR standard errors are smaller which can be only seen if one prints the results with 4 or more digits after the decimal point.
uninsured at age 18 based on Equation (3.17) are displayed in the left-hand-side column. We estimated a 44% increase in the percentage of uninsured in the whole population. There is slightly higher increase in the female population (44%) than in the male population (43%). The percentage change in hospital visits at age 18 based on Equation (3.19) are displayed in the middle column of Table 3.4. All three segments of the population are experiencing a 39% decrease in the number of visits. The right-hand-side column of Table 3.4 presents the estimates of the clean moral hazard effect, i.e., the reduction in hospital visits due to losing insurance for individuals who visit the hospital before age 18 and lost the insurance after crossing age 18. There is a 89% decrease in visits among this group of people. This reduction accounts for both the increase in probability of becoming uninsured and the drop in the numbers of hospital visits.

So, how do we interpret the obtained results? Relying on the summary statistics from Table 3.1, we see that the number of uninsured visits in the $O–18$ group is 38 which is the consequence of the reduction in the number of visits due to losing insurance by 89%. So without the moral hazard, the number of visits would have been $38/(1-0.89) = 345$, so 307 visits never happened because people who lost the insurance either never came to the hospital or came less frequently. Another way to look at the obtained result is by taking the total number of visits in the $Y–18$ group of 915 and reduce it by 39% which is the reduced form estimate of the percent reduction in the total number of visits to obtain 357 visits. Then take the number of insured visits in the $Y–18$ which is also 915 because all visits are insured visits and multiply it by 44% which is the first stage estimate of the percent increase in uninsured visits from 0 to 402. Finally dividing $357/402 = 0.89$ gives us the population wide estimate of the reduction in the number of visits due to the loss of insurance.

Since the data used in the estimation covers only a 4-month period, in order to obtain an annual estimate of moral hazard at the hospital level, we need to multiply the number of visits that never happened by 3 to obtain an estimate of 921 visits per year. Finally, the reduction in the number of visits can be converted into a money metric measure of moral hazard by using the average cost per visit of HRK 162.5 to obtain approximately HRK 150,000 or about 28.8 thousand U.S. dollars worth of savings at the hospital level per year.\footnote{Of course, a more precise money metric estimate of the moral hazard effect is possible by looking at the change in the cost structure of visits before and after crossing the age-18 threshold. However, since our estimation procedure relies on the visits data and not the cost data, this calculation would be less reliable.}

Finally, as a robustness check, we repeated the estimation with the two-year window data. The results are presented in Tables 3.5 to 3.8. This time the clean moral hazard effect amounts to 95% decrease in visits among people who visit the hospital before the age 18 and lost the insurance after turning 18. From Table 3.5 we see that the number of uninsured visits in the $O–18$ group is 122, so without the moral hazard the number of visits would have been
\[ \frac{122}{1 - 0.95} = 2,440, \] and the estimated moral hazard amounts to \( 2,440 - 122 = 2,318 \) visits for the 4 month period or 6,954 visits or slightly over 1 million Kunas \( (6,954 \times 147.92 = 1,028,636) \) per year at the hospital level.

The two sets of results reveal substantially different estimates of moral hazard. However, these results are not directly comparable because they are based on the different number of people. The one-year window estimates are based on patients that 18 years old and the two-year window are based on people that are 18 and 19 years old, so clearly there is more people in the two-year window data and hence the the estimate of moral hazard should be larger. Assuming that the natural convergence region for the regional hospital in question coincides with the population of the county (županija) where the hospital is located, we can approximate the potential hospital’s patients base. Relying on the 2011 Census, we see that the population of 18 years old in the county under consideration is 1,401 and the population of 19 year old is 1,497. Using the one-year window results, the estimated moral hazard per person is HRK 107 \( (921 \times 162.5/1401) \) and using the two-year window results, the estimated moral hazard per person amounts to HRK 355 \( (1,028,636/(1,401 + 1,497)) \). So, which of the two estimates are more reliable and would it be advisable to conduct the estimation with a three-year window surrounding the threshold? There are no unique answers to these questions because adding an additional year around the threshold involves a trade-off. One one hand, it allows the sought after causal effect to better materialize (loss of insurance on the health care utilization); on the other hand it decreases the likelihood that people on both sides of the threshold are going to be exactly the same when it comes to their unobserved health status, a critical assumption that underlies the RD design.

### 3.5 Conclusion

In this paper we implemented a fuzzy regression discontinuity design to estimate the moral hazard effect in health care consumption of young adults relying on the invoices data from one small regional hospital in Croatia. The invoices data suffers from the fact that we only observe people who used medical care services within the time period covered by the data. To deal with this sample selection bias, we estimate the causal effect of insurance on medical care consumption using the 18th birthday as an instrumental variable. The 18th birthday represents a threshold where the young adults will, by default, loose the full coverage unless they remain full time students or decide to buy the coverage on their own. The estimation uses weekly counts of hospital visits. Results from a one-year window estimation show that there is an 89% decrease in hospital visits due to losing insurance among individuals who visit the hospital before age 18 and lost the insurance after crossing age 18. Based on the two-year window the reduction in hospital visits amounts to 95%.
So, how economically significant is this effect at the national level? Based on the one-year window estimates and the 2011 Census figures for the 18 year old population of Croatia of 47,960, the total moral hazard effect would be HRK 5.1 million. Therefore, the total cost of extending the supplemental insurance for all young adults in the country from the current 18th birthday to the 19th birthday would consist of two components: (a) the direct loss in collected premiums for supplemental insurance and the indirect cost of moral hazard. Assuming the highest premium for supplemental insurance of HRK 130 per month gives us a total cost of HRK 6.7 million per year in lost premiums plus the indirect cost of moral hazard of HRK 5.1 million per year, for the total of HRK 11.8 million. Alternatively, extending the free supplemental insurance to all young adults until they turn 20 would cost HRK 12.8 million in lost premiums and HRK 35 million in indirect cost of moral hazard, for the total of HRK 47.8 million.\(^8\)

Now let’s try to relate the above numbers to an alternative program recently promulgated by the Croatian government. Given a severe youth unemployment problem\(^9\), starting January 1, 2015, the government has launched a program whereby the employers are excused from paying the payroll taxes of 17.2\% (health insurance earmarked tax rate of 15\% plus some other taxes of 2.2\%) for the period of up to 5 years for every new employee below the age of 30. For an employee with an average net salary of HRK 8,000 per month the total savings for the employer amount to HRK 16,512 per year. Of course, the amount of money saved by an employer represents a loss to HZZO or indirectly the state budget. Relying on the data published by the Croatian Unemployment Bureau (Hrvatski zavod za zaposljavanje – HZZ), the total unemployment in the 15-29 age cohort amounts to 102,483 persons. Assuming an optimistic scenario whereby the newly introduced measure will reduce the youth unemployment by about 10\% or 10,250 people per year, the total cost to the HZZO or the state budget can be estimated at HRK 169 million per year.

The comparison of the two programs reveal several interesting points: first, the cost of moving the threshold for free supplemental insurance from 18 to 20 years of age would cost about a third of the employment stimulus package currently in place. Second, the significant part of that cost is the cost of moral hazard, which gets frequently ignored in discussions about the cost of health care, and not the cost due to lost premiums. Finally, the national health effect of such a program could be substantial. Recall that loosing the supplemental insurance at the age of 18 produced the reduction in hospital visits, some of those lost visits could have profound impacts on early disease detection and prevention and could actually save a lot of money to the health insurance system down the road.

\(^8\)This number is obtained using the larger estimate of moral hazard obtained with the two-year window data of HRK 355 per young adult and the 2011 Census figures for the 18 year old population of 47,960 and the 19 year old population of 50,790.

\(^9\)According to Eurostat the youth unemployment in the fourth quarter of 2013 was a staggering 48.6\%
### Table 3.1: Summary Statistics: One-year Window Around the Age-18

<table>
<thead>
<tr>
<th>Sample</th>
<th>Patients</th>
<th>Sample</th>
<th>Invoices</th>
<th>Insured</th>
<th>Uninsured</th>
<th>Cost per visit</th>
<th>Number</th>
<th>Visits per patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y-18</td>
<td>915</td>
<td>915</td>
<td>0</td>
<td>134.36 (174.01)</td>
<td>332</td>
<td>2.76 (3.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O-18</td>
<td>671</td>
<td>633</td>
<td>38</td>
<td>200.76 (664.68)</td>
<td>287</td>
<td>2.02 (1.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1,586</td>
<td>1,548</td>
<td>38</td>
<td>162.45 (453.10)</td>
<td>619</td>
<td>2.42 (2.77)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The numbers in parenthesis are standard deviations.

### Table 3.2: Seemingly Unrelated Regression Results: One-Year Window

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Total</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1 - \frac{d_{ij}}{y_{ij}})$</td>
<td>$I_{over18j}$</td>
<td>0.08 (0.02)***</td>
<td>0.08 (0.02)***</td>
<td>0.08 (0.02)***</td>
</tr>
<tr>
<td></td>
<td>$aweek_{ij}$</td>
<td>0.00 (0.03)</td>
<td>0.00 (0.03)</td>
<td>0.00 (0.03)</td>
</tr>
<tr>
<td></td>
<td>$aweek_{ij} * I_{over18j}$</td>
<td>-0.05 (0.04)</td>
<td>-0.05 (0.04)</td>
<td>-0.05 (0.04)</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>0.00 (0.02)</td>
<td>0.00 (0.02)</td>
<td>0.00 (0.02)</td>
</tr>
<tr>
<td>$\ln y_{ij}$</td>
<td>$I_{over18j}$</td>
<td>-0.50 (0.17)***</td>
<td>-0.49 (0.17)***</td>
<td>-0.50 (0.17)***</td>
</tr>
<tr>
<td></td>
<td>$aweek_{ij}$</td>
<td>0.54 (0.20)***</td>
<td>0.50 (0.20)**</td>
<td>0.54 (0.20)***</td>
</tr>
<tr>
<td></td>
<td>$aweek_{ij} * I_{over18j}$</td>
<td>-0.58 (0.29)**</td>
<td>-0.54 (0.29)*</td>
<td>-0.58 (0.29)**</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>3.01 (0.12)***</td>
<td>3.00 (0.12)***</td>
<td>3.01 (0.12)***</td>
</tr>
</tbody>
</table>

Note: The dependent variables in the regressions are proportion of uninsured visits and log of visits at each age in weeks. Standard errors are in parentheses. *** - 1 % significance level; ** - 5 % significance level; * - 10 % significance level.
Table 3.3: OLS Regression Results: One-Year Window

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Total</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>$lny_{ij}$</td>
<td>$I_{over18_{ij}}$</td>
<td>-0.50 (0.17)***</td>
<td>-0.49 (0.17)***</td>
<td>-0.50 (0.17)***</td>
</tr>
<tr>
<td></td>
<td>$aweek_{ij}$</td>
<td>0.54 (0.20)***</td>
<td>0.50 (0.21)**</td>
<td>0.54 (0.20)***</td>
</tr>
<tr>
<td></td>
<td>$aweek_{ij} * I_{over18_{ij}}$</td>
<td>-0.58 (0.29)**</td>
<td>-0.54 (0.29)*</td>
<td>-0.58 (0.29)**</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>3.01 (0.12)***</td>
<td>3.00 (0.12)***</td>
<td>3.01 (0.12)***</td>
</tr>
</tbody>
</table>

Note: The dependent variable in the regression is log of visits at each age in weeks. Standard errors are in parentheses.

Table 3.4: Percent Change in Uninsured (First Stage) and Total Visits (Reduced Form) and the Moral Hazard (IV): One-Year Window

<table>
<thead>
<tr>
<th>Sample</th>
<th>First Stage</th>
<th>Reduced Form</th>
<th>IV Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.44 (0.09)***</td>
<td>-0.39 (0.10)***</td>
<td>-0.89 (0.06)***</td>
</tr>
<tr>
<td>Male</td>
<td>0.43 (0.09)***</td>
<td>-0.39 (0.10)***</td>
<td>-0.89 (0.06)***</td>
</tr>
<tr>
<td>Female</td>
<td>0.44 (0.09)***</td>
<td>-0.39 (0.10)***</td>
<td>-0.89 (0.06)***</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.

Table 3.5: Summary Statistics: Two-year Window Around the Age-18

<table>
<thead>
<tr>
<th>Sample</th>
<th>Invoices</th>
<th>Insured</th>
<th>Uninsured</th>
<th>Cost per visit</th>
<th>Number</th>
<th>Visits per patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y-18</td>
<td>1,646</td>
<td>1,646</td>
<td>0</td>
<td>135.03 (179.72)</td>
<td>620</td>
<td>2.66 (2.91)</td>
</tr>
<tr>
<td>O-18</td>
<td>1,370</td>
<td>1,248</td>
<td>122</td>
<td>163.42 (476.56)</td>
<td>580</td>
<td>2.04 (2.29)</td>
</tr>
<tr>
<td>Total</td>
<td>3,016</td>
<td>2,894</td>
<td>122</td>
<td>147.92 (347.77)</td>
<td>1,200</td>
<td>2.36 (2.64)</td>
</tr>
</tbody>
</table>

Note: The numbers in parenthesis are standard deviations.
Table 3.6: Seemingly Unrelated Regression Results: Two-Year Window

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Total</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>((1 - \frac{q_j}{y_j}))</td>
<td>(I_{\text{over18},j})</td>
<td>0.03 (0.02)</td>
<td>0.03 (0.02)</td>
<td>0.03 (0.02)</td>
</tr>
<tr>
<td></td>
<td>(a\text{week}_j)</td>
<td>0.00 (0.01)</td>
<td>0.00 (0.01)</td>
<td>0.00 (0.01)</td>
</tr>
<tr>
<td></td>
<td>(a\text{week}<em>j \times I</em>{\text{over18},j})</td>
<td>0.05 (0.02)**</td>
<td>0.05 (0.02)**</td>
<td>0.05 (0.02)**</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>0.00 (0.01)</td>
<td>0.00 (0.01)</td>
<td>0.00 (0.01)</td>
</tr>
</tbody>
</table>

Note: The dependent variables in the regressions are proportion of uninsured visits and log of visits. Standard errors are in parentheses.

Table 3.7: OLS Regression Results: Two-Year Window

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Total</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\ln y_j)</td>
<td>(I_{\text{over18},j})</td>
<td>-0.44 (0.12)**</td>
<td>-0.45 (0.12)**</td>
<td>-0.44 (0.12)**</td>
</tr>
<tr>
<td></td>
<td>(a\text{week}_j)</td>
<td>0.25 (0.07)**</td>
<td>0.26 (0.07)**</td>
<td>0.25 (0.07)**</td>
</tr>
<tr>
<td></td>
<td>(a\text{week}<em>j \times I</em>{\text{over18},j})</td>
<td>-0.20 (0.10)**</td>
<td>-0.21 (0.10)**</td>
<td>-0.20 (0.10)**</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>2.89 (0.08)**</td>
<td>2.90 (0.08)**</td>
<td>2.89 (0.08)**</td>
</tr>
</tbody>
</table>

Note: The dependent variable in the regression is log of visits. Standard errors are in parentheses.
Table 3.8: Percent Change in Uninsured (First Stage) and Total Visits (Reduced Form) and the Moral Hazard (IV): Two-Year Window

<table>
<thead>
<tr>
<th>Sample</th>
<th>First Stage Estimator</th>
<th>Reduced Form Estimator</th>
<th>IV Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.37 (0.07)***</td>
<td>-0.35 (0.08)***</td>
<td>-0.95 (0.05)***</td>
</tr>
<tr>
<td>Male</td>
<td>0.38 (0.07)***</td>
<td>-0.36 (0.08)***</td>
<td>-0.95 (0.05)***</td>
</tr>
<tr>
<td>Female</td>
<td>0.37 (0.07)***</td>
<td>-0.35 (0.08)***</td>
<td>-0.95 (0.05)***</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.

Figure 3.1: Uninsured visits surrounding patients’ 18th birthday (two-year window)
Figure 3.2: Hospital visits surrounding patients’ 18th birthday (two-year window)
Chapter 4

General Conclusion

This dissertation proposed three approaches to estimate asymmetric information effects in health insurance and health care markets using a case study in Croatia. The first chapter uses a structural approach to estimate clean moral hazard effect in health insurance, net of possible influences of adverse selection. We found positive and economically meaningful effect of moral hazard and also a statistically significant effect of adverse selection in the population of hospital patients. An important methodological contribution of this chapter is to take into account individuals’ opportunity cost of time associated with medical care consumption. To the best of our knowledge this is the first attempt in the literature to incorporate such cost into a structural estimation of health care demand. The generalization of the obtained results for the purposes of country level and international comparisons is somewhat hampered by the fact that hospital invoices data is comprised of users only, i.e., we don’t have data for insured and uninsured citizens with zero medical consumption in the given period. As it turns out, the obtained structural model parameter estimates are still consistent because the structural model is assumed for users and non-users alike and the equation used to estimate $\theta_i$ for individuals with different insurance coverage is also valid for both users and nonusers. However, two important consequences of the users only data still remain. First, the two counterfactual moral hazard scenarios are markedly different. Second, the tests for adverse selection are also impacted by the users only data problem because these tests should be based on the distribution of the unobserved health status $\theta$ for the entire population whereas ours are based on users only.

Thus, we proposed two empirical approaches to correct for the sample selection issue brought by the data in the following two chapters. The second chapter is motivated by the lack of a consistent explanation for the empirically observed gender differences in health care consumption, which calls for an innovative empirical research in this area. More importantly, to the best of our knowledge, the economics literature appears to be completely silent when it comes to explaining the gender differences in health care consumption by a possible gender differences.
in the magnitude of asymmetric information effects. Since asymmetric information problems, because of moral hazard or adverse selection nature, have been recognized as one of the important drivers of health care costs, if men and women are different with respect to how they respond to incentives, then knowing what those differences are, could be important for correct pricing of insurance policies. However, differential pricing of insurance policies with respect to gender may not be allowed by law. For example, the Patient Protection and Affordable Care Act (Obama-care) prohibits ”gender rating”, starting in 2014 (Arons 2012). Women cannot be discriminated against in the health insurance market, with higher premiums, lost maternity coverage and denials of coverage for gender-related pre-existing conditions. The results of this research could provide important insights about possible cost of such regulatory requirements.

Another challenge is that we only have users data such that the sample selection issue needs to be dealt with. We use truncated count model and simulated maximum likelihood approach to estimate gender differences in moral hazard in health care utilization relying on the hospital invoices data for non life-threatening diagnoses. Lack of information of non-users who did not show up at the hospital, we derive conditional likelihood density to adjust this sample selection issue in the estimation. The results show no evidence of statistically significant gender difference in hospital visits and moral hazard effects in the population we studied. But we did find significant moral hazard effect. People who bought the insurance have 35% more hospital visits than people who do not have the insurance.

In the third chapter we implemented a fuzzy regression discontinuity design to estimate the moral hazard effect in health care consumption of young adults. To deal with this sample selection bias, we estimate the causal effect of insurance on medical care consumption using the 18th birthday as an instrumental variable. The 18th birthday represents a threshold where the young adults will, by default, loose the full coverage unless they remain full time students or decide to buy the coverage on their own. The estimation uses weekly counts of hospital visits. Results from a one-year window estimation show that there is an 89% decrease in hospital visits due to losing insurance among individuals who visit the hospital before age 18 and lost the insurance after crossing age 18. Based on the two-year window the reduction in hospital visits amounts to 95%.
REFERENCES


