# The impact of private insurance on healthcare expenditure

**June 2023**

**Tutor: Peter Foldvari**

**Tutorial (WG03)**

**Assignment Group 18:**

Alwar Rahul Govindarajan (13564811)

Bowen Ma (12960780)

Ionel Popescu (13899430)

Duc Minh Nguyen (13569120)

Faculty Economics and Business
University of Amsterdam

**Abstract**

This report analyzes the impact of holding private insurance on healthcare expenditure. Initially, OLS regression was used, but endogeneity issues with the private insurance variable led to the adoption of IV estimation. Additionally, Heckman's Two-Stage selection estimation was applied to investigate the effects of excluding observations with zero healthcare expenditure.

# 1 Introduction

This paper addresses the current and contentious issue of how health insurance impacts healthcare expenditure in the United States, as noted by Shen (2013). Our main objective is to analyze the relationship between insurance coverage and health expenditures. Recognizing that private insurance coverage is a personal decision, we consider it to be endogenous and employ Instrumental Variables (IV) to address the endogeneity concern. Furthermore, we also investigate the potential selection bias resulting from excluding observations with zero healthcare expenditure by employing Heckman's Two-Stage selection estimation method.

# 2 Theoretical background and method

The factors that are most crucial in determining healthcare expenditure are those that are directly related to the physical well-being of an individual (Bajari, Dalton, Hong, and Khwaja, 2014). Similar to Shen (2013), we take into consideration a number of explanatory variables which cover the demographic, socioeconomic, and health-related characteristics of the sample. As Shen (2013) suggests, we also take into consideration here that insurance and expenditure are related to each other. There is considerable literature which suggests that if a person is medically insured, he/she is more likely to spend more on healthcare (Bajari, Dalton, Hong, and Khwaja, 2014). There is also the concept of adverse selection, which refers to the scenario in which higher-risk or sick individuals, who have greater coverage needs, purchase health insurance, while healthy people delay or decide to abstain (Spence and Zeckhauser, 1971). This suggests that the choice of insurance depends on the anticipated healthcare expenditure; hence we treat the choice of insurance as an endogenous variable.

# 3 Data

The dataset utilized in this report is sourced from the Medical Expenditure Panel Survey (MEPS) (Agency for Healthcare Research and Quality, 2011). It contains information on an individual's medical expenditure, demographic and socio-economic characteristics. In this report, we only consider individuals aged between 31 and 40, due to limited accessibility. We start with 4635 data, after excluding unusable and negative values there are 4323 observations left. We introduce dummy variables for mental health status, marital status, ethnicity, and type of job. In contrast with Shen (2013), where no individuals' physical conditions are included, we consider the suggestion of Ward (2021) that concluded that Body Mass Index (BMI) does affect the healthcare costs of an individual. We classified BMI in accordance with the standards published by the U.S. Department of Health & Human Services.

By analyzing the mean spending on healthcare among different groups from Table 1 we found that the mean spending on health among privately insured groups is approximately 4 times more than the group without insurance. That is where the inspiration for this report from.

Table 1: Mean values of health expenditures for different groups($)

| | mean |
|---|---|
| private_insurances | 3087.2 |
| public_insurances | 4202.5 |
| no_insurances | 784.9 |

# 4 Results

The starting model of our analysis is based on the model suggested by Shen (2013) that describes a log-linear relation between health expenditure ($y_i$) and the choice of insurance. The most significant difference in our model is that we choose to include publicly insured individuals present in our data as an exogenous explanatory variable (Shen, 2013):

$$log(y_i) = X_i^{'}\delta + \beta \,\text{private insurance}_i + \epsilon_i \tag{1}$$

Given the potential endogeneity of the dummy variable $private\ insurance$ (Shen, 2013), we expect that OLS would return an inconsistent estimator of $\beta$. Thus, based on previous research, we suggest IV estimation. So, we assume a reduced form model (Heij et al., 2004) for the potentially endogenous variable (stage 1 regression):

$$\text{Private insurance}_i = Z_i^{'}\gamma + u_i \tag{2}$$

The instruments in $Z$ matrix used for IV estimations are $family\ income$ and $use\ of\ seat\ belt$, which is assumed to be exogenous. The argument for this is that the healthcare expenditure itself does not directly influence the desire to wear seat belts, and neither directly determining the family income. The Sargan-Hansen test in Table 1 does not reject the null hypothesis that the instruments are exogenous (p-value=0.585), which confirms our assumptions. The F-test for weak instruments is rejected with a large F-statistic of 19.22 (see Table 1), so $family\ income$ and $use\ of\ seat\ belt$ are relevant.

The IV estimation summarized in Table 2 suggests that the coefficient for individuals with private insurance (3.123) is approximately 50% more than the value of the corresponding OLS estimator (2.173). This is indicative of the endogeneity of $private\ insurance$ with respect to healthcare expenses, which points towards a downward bias of the OLS. Furthermore, by applying equation (2), we obtain an R-squared value of 0.421 (Table 2), and our analysis returned low values for the variance inflation factor, suggesting there is no multicollinearity in our instruments matrix.

We did not exclude the observations with 0 healthcare expenditure, unlike Shen (2018), since this can lead to selection bias (Heij et al., 2004). A way to see this problem is to apply Heckman's Two-Stage selection estimation (Harmon, C. & Walker, I. 1995). The structural equation of this model is the same as the one used for OLS. The selection equation has a dependent variable of a dummy for positive expenditure and includes an extra variable $white\ collar$ in the regressors. The results in Table 3 show that the Heckman corrected private_insurance coefficient (1.322) is significantly less than that of OLS (2.173), indicating an overestimation of OLS due to selectivity. Moreover, the Mill's ratio is significant and positive, indicating an upward selection bias.

**Table 2: OLS and IV estimation**

| VARIABLES | OLS<br>log expenditure | OLS<br>private insurance | IV<br>log expenditure |
|---|---|---|---|
| family income /1000 | | 0.00103*** | |
| | | (0.000167) | |
| use of seat belt | | -0.0127 | |
| | | (0.0201) | |
| private insurance | 2.173*** | | 3.123** |
| | (0.124) | | (1.318) |
| public insurance | 2.247*** | -0.520*** | 2.747*** |
| | (0.163) | (0.0121) | (0.713) |
| underwheight | 0.229 | 0.0417 | 0.194 |
| | (0.434) | (0.0596) | (0.431) |
| overweight | -0.180 | 0.0187 | -0.192* |
| | (0.112) | (0.0145) | (0.114) |
| obesity | 0.0805 | 0.0378** | 0.0503 |
| | (0.115) | (0.0150) | (0.121) |
| marital status | 0.147 | 0.0796*** | 0.0456 |
| | (0.0955) | (0.0139) | (0.176) |
| health_m | -0.828*** | 0.0133 | -0.842*** |
| | (0.0961) | (0.0127) | (0.0986) |
| smoker | -0.00144 | -0.00662 | 0.00706 |
| | (0.122) | (0.0167) | (0.124) |
| no. of comorbidities | 0.878*** | 0.0268*** | 0.853*** |
| Constant | 2.855 | -0.904 | 3.760 |
| | (7.619) | (0.982) | (7.777) |
| Observations | 4,323 | 4,323 | 4,323 |
| R-squared | 0.284 | 0.422 | 0.273 |
| Adj R squared | 0.281 | 0.419 | |
| Log Likelihood | -10701 | -1892 | |
| F-test first-stage | | | 19.22 |
| F test df | | | 2 |
| Sargan-Hansen test p-value | | | 0.585 |
| Sargan-Hansen test df | | | 1 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 3: OLS and Heckman two stage method**

| VARIABLES | OLS<br>log expenditure | Heckman correction<br>log expenditure | Selection Equation Probit<br>positive expenditure |
|---|---|---|---|
| white collar | | | 0.154** |
| | | | (0.0611) |
| private insurance | 2.173*** | 1.322*** | 0.770*** |
| | (0.124) | (0.180) | (0.0556) |
| public insurance | 2.247*** | 1.369*** | 0.813*** |
| | (0.163) | (0.193) | (0.0758) |
| age | -0.140 | -0.211 | -0.00798 |
| | (0.431) | (0.300) | (0.218) |
| age squared | 0.00158 | 0.00281 | -8.87e-05 |
| | (0.00606) | (0.00422) | (0.00306) |
| underwheight | 0.229 | 0.129 | 0.0640 |
| | (0.434) | (0.304) | (0.228) |
| overweight | -0.180 | -0.212*** | -0.0244 |
| | (0.112) | (0.0788) | (0.0557) |
| obesity | 0.0805 | -0.0988 | 0.110* |
| | (0.115) | (0.0814) | (0.0585) |
| health_m | -0.828*** | -0.594*** | -0.285*** |
| | (0.0961) | (0.0802) | (0.0493) |
| smoker | -0.00144 | -0.0647 | 0.0397 |
| | (0.122) | (0.0838) | (0.0595) |
| no. of comorbidities | 0.878*** | 0.542*** | 0.377*** |
| | (0.147) | (0.100) | (0.0717) |
| lambda | | 1.131*** | |
| | | (0.393) | |
| Constant | 2.855 | 7.419 | -0.978 |
| | (7.619) | (5.323) | (3.842) |
| Observations | 4,323 | 4,323 | 4,323 |
| R-squared | 0.284 | | |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

# 5 Conclusions

In this paper, we estimate the effect of the choice of having private insurance on healthcare expenditures, for US individuals between 31-40 years old. The two different approaches we use to analyze this problem indicate significant biases in the OLS estimation of this effect. What we find interesting is that the endogeneity of $private\ insurance$ induces a negative bias: IV estimation returning a more than 50% larger estimate. Whereas the sample selectivity bias induces a positive bias: Heckman's method suggests an approximately twice lower estimate. The question of which bias direction is stronger would present an interesting topic for further research.

**References:**

Agency for Healthcare Research and Quality (2011) Medical Expenditure Panel Survey Data Overview. Available at: https://meps.ahrq.gov/mepsweb/data_stats/data_overview.jsp (Accessed: 18 June 2023).

Bajari, P., & Dalton, C. (2023). Moral hazard, adverse selection, and health expenditures: A semiparametric analysis. https://www.jstor.org/stable/43186480

Colm Harmon, & Walker, I. (2023). Estimates of the Economic Return to Schooling for the United Kingdom.https://www.jstor.org/stable/2950988

Shen, C. (2013). Determinants of Health Care Decisions: Insurance, Utilization, and Expenditures. Review of Economics and Statistics, 95(1), 142–153. https://doi.org/10.1162/REST_a_00232

Spence, M., & Zeckhauser, R. (1978). Insurance, Information, and Individual Action. In Uncertainty in Economics (pp. 333–343). Elsevier. https://doi.org/10.1016/B978-0-12-214850-7.50027-9

Ward, Z. J., Bleich, S. N., Long, M. W., & Gortmaker, S. L. (2021). Association of body mass index with health care expenditures in the United States by age and sex. PLOS ONE, 16(3), e0247307.https://doi.org/10.1371/journal.pone.0247307