Undergraduate Economic Review

Volume 7 | Issue 1 Article 14

2011

Effects of Age and Income on Individual Health Insurance Premiums

Jessie Li

Massachusetts Institute of Technology, jessieli@alum.mit.edu

Effects of Age and Income on Individual Health Insurance Premiums

Abstract
This paper examines the effect of an individual's age and income on the premiums that she pays in the individual (non-employer sponsored) health insurance market. After controlling for medical conditions, insurance plan type, and demographic characteristics, it was found using both OLS and 2SLS that older and wealthier people pay higher premiums than younger and less wealthy individuals, which raises the possibility of adverse selection in the individual insurance market. These results may have important policy implications for regulation in the individual health insurance market.

Keywords
health insurance, premiums, adverse selection

Cover Page Footnote
I would like to thank Professor Jerry Hausman and Alp Simsek for helping me with this paper.

1 Introduction

Over the past decade, health insurance premiums have been on the rise, and the percentage of uninsured individuals has increased [Davis2007]. Furthermore, with the aging of the baby-boomer generation, the demand for health services has increased, which brings up the question of whether the higher average age of the US population would increase average health premiums charged by insurance companies [DAY2008]. Considering the recent debate over health care reform, it would also be interesting to examine whether there is any correlation between a person's income and the premiums that she pays. In an attempt to answer some of these questions, this paper examines the effects of age and income on out-of-pocket premiums in the individual (non-employer sponsored) health insurance market, while controlling for medical conditions, demographic characteristics, and the insurance plan type. The hypothesis is that older and wealthier people pay higher premiums, keeping other factors constant.

2 Background

Health care in America can be divided into roughly two categories. One category is called traditional care (or fee-for-service), in which prices for health care rendered are determined solely between the physician and the patient, and the role of the insurance company is simply to reimburse the physician for treatments of acute illnesses or injuries. On the contrary, in managed care, the individual pays a monthly premium to group of health providers in a managed care organization (MCO). In return, the MCO will negotiate deals with a network of physicians to provide care at a lower price to the members of the organization, but with certain restrictions on the number of times that the members can see the physicians. The MCO provides coverage for primary and preventive care in addition to acute illnesses and injuries [Knight1998].

The managed care model provides more cost control than the traditional model by limiting the number of unnecessary medical procedures prescribed by the physician. In the traditional care model, physicians have a financial incentive to prescribe unnecessary medical procedures, leading to rising health care costs for the patient [Knight1998]. In the managed care model, physicians need to maintain a good reputation in order to remain in the network, and they are held accountable for their actions. When they enter into an agreement with the MCO, they promise to reduce the per unit cost of services for members in the MCO and to limit the number of unnecessary medical procedures. In return, the physicians usually receive compensation in the form of capitation, in which the MCO prepays the physician for the treatment of the patient. This compensation reduces the physician's incentive to prescribe unnecessary procedures.

Another way that managed care provides cost control is by limiting the types of physicians that patients can visit at the cheaper price. In the traditional care model, patients can visit any physician they want, and as long as their illness qualifies as an "acute illness," the insurance company would have to provide coverage. On the contrary, the managed care model restricts the physicians that patients can visit to those physicians in the network, if the patients would like the cheaper price. Doing so allows the MCO to keep costs down since the MCO had already negotiated a lower price with the physicians in the network. Furthermore, in some managed care plans, members must select a primary care provider (PCP) and always see the PCP first before seeing any specialist. The rationale for selecting a PCP is that the PCP would first determine whether it is necessary for the patient to see the more expensive specialist. If the PCP can handle the case himself, the MCO would save on medical costs.

2.1 Types of Managed Care Organizations

There are three main types of managed care organizations: Health Management Organization (HMO), Preferred Provider Organization (PPO), and Point of Service (POS). HMO is a type of MCO which offers health care services to its members for a fixed monthly fee. HMOs offer two types of managed care plans, closed-panel HMO and open-access HMO. Under the closed-panel HMO plan, patients are required to select a primary care physician who must authorize the patient before he can receive any specialist treatment. Patients pay a minimal copayment for office visits, and the HMO covers hospitalizations in full [Knight1998]. In the open-access HMO plan, patients have the option of seeing a specialist directly for a higher copayment. In both closed-panel and open-access HMO plans, the HMO does not provide coverage for visits to physicians outside the network.

PPOs are networks of doctors and health professionals that provide health services to individuals for a negotiated fee [Knight1998]. In contrast to HMOs, PPOs do not restrict patients to seeing their primary care provider first; patients can visit specialists without having to pay a higher copayment. Furthermore, PPOs provide coverage to individuals seeking care from physicians not enrolled in the network, although patients have to pay a high deductible and a high coinsurance.

A POS plan combines features of the HMO and the PPO plans. As in the HMO plan, the patient is required to select a PCP who must authorize any visits to specialists. Like the PPO plan, the POS plan provides out-of-network coverage at a high deductible and coinsurance rate.

Based on a study conducted by the Henry J. Kaiser Family Foundation and the Health Research and

Educational Trust (HRET) in 2006, premiums for PPO plans are typically higher than premiums for POS and HMO plans. One reason for the higher premiums is that PPO plans have less cost control since patients can directly see specialists who usually charge more than primary care physicians [Claxton et al.2006]. POS plans usually have higher premiums than HMO plans because POS plans partially cover out-of-network visits, which are usually more expensive than in-network visits.

Compared to traditional care plans, managed care plans tend to have higher premiums because managed care plans provide better coverage. Based on a 2005 survey by the Henry J. Kaiser Family Foundation and the Health Research and Educational Trust (HRET), the average monthly premium for family coverage was 832 dollars for a traditional care plan, 871 for a HMO plan, 924 for a PPO plan, and 900 for a POS plan [Gabel et al.2005]. PPO has the highest premium because it offers the flexibility of seeing a specialist before seeing a primary care physician.

Studies have shown that managed care may have the problem of adverse selection, in which the managed care organizations only allow healthy individuals to become members of the organization [Frank et al.2000]. When deciding what premiums to charge plan participants, managed care organizations have wide latitude in determining individuals' premiums. Factors that MCO's consider include age, income, and preexisting conditions. Although there are many different factors that contribute to a person's health, one possible factor is the person's age. Insurance companies may think that older people are more likely to have illnesses and therefore charge them higher premiums. To see if in fact older people do pay higher premiums, I examined data on out-of-pocket premium costs for people in the individual (non-employer sponsored) insurance market.

3 Data

The data was collected from the National Health Interview Survey, an annual survey conducted by the US Census Bureau on individual persons regarding their medical expenditures, personal health, and demographic characteristics. Cross-sectional individual data from the years 2004, 2005, 2006, 2007, and 2008 were used for this paper. Health insurance variables of interest include the annual out-of-pocket premium cost for the private insurance plan, the plan type such as HMO, PPO, POS, or fee-for-service, and whether the plan was paid for individually or by the employer. Demographic characteristics include the individual's annual earnings, age, gender, race, education, and citizenship. Health status variables include whether the individual has trouble walking or remembering and whether he is limited in any other way.

Because the individual's income was reported as a number between 1 and 10, each corresponding to a

specific income bracket, the values needed to be replaced with the national mean income in each bracket. Data from the Consumer Expenditure Survey was used to determine the national mean income values.

Since over twenty variables for various medical conditions existed, adding all of the variables into the regression would over-parameterize the model. Therefore, a principal component was created to provide a weighted average of some medical conditions. Because many medical condition variables had a non-response rate over 80%, using a principal component created out of those variables in the regressions would reduce the degrees of freedom significantly. Therefore, only those medical conditions with response rates greater than 50% were included in generating the principal component. The principal component used in this paper combines the dummy variables for the following questions: does the person have trouble walking, does the person have trouble remembering, and is the person limited in any other way?

Table 1 summarizes the variables used in analysis.

Table 1: Explanatory Variables

PREMIUM	Annual out-of-pocket premium cost
AGE	Age in number of years
Income	Annual earnings
EDUCATION	Number of years of schooling
USCITIZEN	Dummy for whether the person is a US citizen
MALE	Dummy for whether the person is male
HISPANIC[1]	Dummy for whether the person is hispanic
MEDICALCONDITION	principal component for medical conditions
HMO	Dummy for HMO plan
PPO	Dummy for PPO plan
POS	Dummy for POS plan
FEE	Dummy for fee-for-service plan
OTHER	Dummy for other plan

3.1 Empirical Specifications

First, ordinary least squares was used to regress log of out-of-pocket premium costs on log age, log income, a variety of demographic characteristics, a principal component for medical conditions, and dummy variables for plan type.

The OLS specification is the following:

$$\begin{aligned}\ln(PREMIUM) &= \beta_1 \ln(AGE) + \beta_2 \ln(INCOME) + \beta_3 EDUCATION + \beta_4 USCITIZEN + \\ &\quad \beta_5 MALE + \beta_6 HISPANIC + \beta_7 MEDICALCONDITION + \beta_8 HMO + \\ &\quad \beta_9 PPO + \beta_{10} POS + \beta_{11} FEE + \beta_{12} OTHER + \varepsilon_i\end{aligned}$$

The idea of this specification is to determine if older or wealthier people tend to pay higher premiums, keeping medical conditions, plan type, and demographic characteristics constant. If they do, then either the insurance companies could be charging them higher premiums, or the older or wealthier people could be demanding good coverage plans with higher premiums. If it turns out that older people do not demand plans with typically higher premiums, then it would be reasonable to assume that insurance companies are charging older people higher premiums, which could signify an adverse selection problem.

After OLS was performed, 2SLS was performed on separate years of data and pooled data. The instruments for income were a variable for whether the individual qualified for food stamps, a variable for whether the individual received income from pensions, and a variable for whether the individual received social security or Rairoad Rewards. An overidentification test was performed to verify the exclusion restriction, and a Hausman test was performed to test for errors in variables [Hausman1978].

Both OLS and 2SLS found that older and wealthier individuals pay higher premiums, keeping other variables constant. In order to test if older or wealthier people demand plans with higher premiums, a nested logit regression was performed on the five different insurance plans: HMO, PPO, POS, fee-for-service, and other. Figure 1 shows the nested logit tree structure.

HMO, PPO, and POS both fall under the category of managed care, while fee-for-service and other fall under the category of traditional care. The case specific variables are income and age, and unfortunately, there are no non-case specific variables due to lack of data on the individual plans. Nested logit regressions using income and age as the top level variables were performed separately on each year.

4 Results

4.1 Separate Year OLS

Table 2 shows the OLS regression results for each year. As we can see, the coefficients on log income and log age are positive and highly significant in all years, which suggests that wealthier and older people

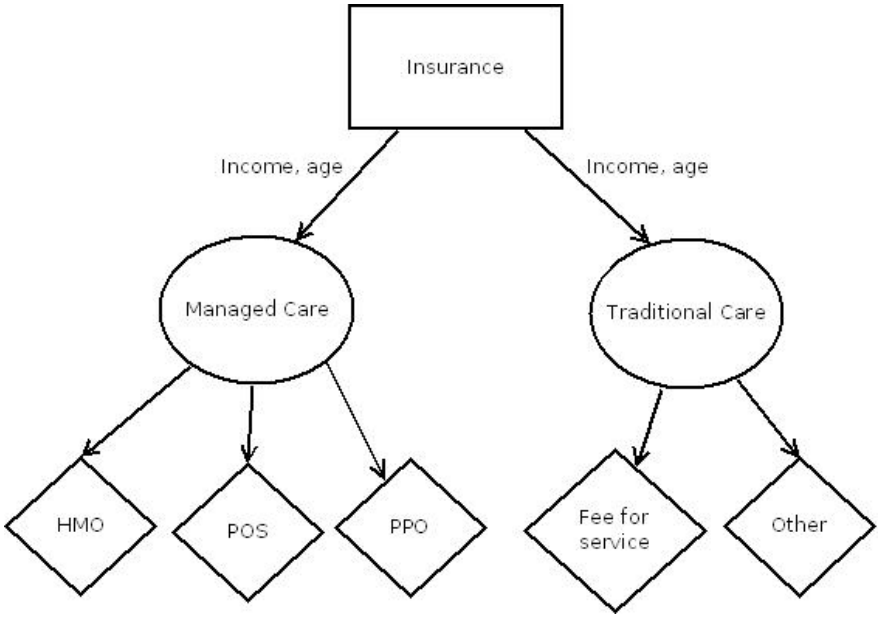

Figure 1: Nested Logit Tree for Choice of Different Insurance Plan Types

Table 2: Separate Year OLS Results, Dependent Variable: Log of Out-of-Pocket Premiums

Variable	2004	2005	2006	2007	2008
$\ln(AGE)$	.4204174	.4151197	.4538275	.3900424	.3581655
	(.0581931)	(.0622168)	(.0752566)	(.0664133)	(.06177)
$\ln(INCOME)$	.0415217	.0557738	.0305913	.0387983	.0535124
	(.0176526)	(.0166717)	(.0166773)	(.0171019)	(.0187026)
$EDUCATION$	.0304392	.0461535	.0466376	.0311253	.0235074
	(.0066975)	(.007141)	(.0086489)	(.0088529)	(.0065983)
$USCITIZEN$	.1221257	.2999292	.1211093	.2544095	.3497774
	(.0972408)	(.0954521)	(.0926383)	(.114078)	(.0834037)
$MALE$	.0771887	.0375758	.1011693	.0386174	.071505
	(.0388674)	(.0403688)	(.0453056)	(.0469243)	(.042062)
$HISPANIC$	−.2304746	−.117611	−.2036018	−.1141885	.0494393
	(.0631677)	(.0590579)	(.0824605)	(.0658442)	(.0599691)
$MEDICALCONDITION$	−.0123592	−.0046269	.0448208	.0251137	.0196089
	(.0275418)	(.0141291)	(.0206096)	(.0292753)	(.0248476)
$dummy\_for\_hmo$	.0879942	.0748514	.1457179	.3420263	.1525497
	(.0718251)	(.0694337)	(.0740579)	(.0914668)	(.071746)
$dummy\_for\_ppo$	.3129684	.1489021	.2489223	.3720535	.203056
	(.0659825)	(.063657)	(.0675239)	(.0837974)	(.0662049)
$dummy\_for\_pos$	.1790635	.234418	.139051	.2061048	−.0768062
	(.1591573)	(.1132297)	(.1766579)	(.168013)	(.1348936)
$dummy\_for\_fee\_for\_service$	−.0075153	−.0225754	−.3068197	−.1012478	−.1120505
	(.1151747)	(.1306332)	(.224515)	(.1565182)	(.1833302)
$dummy\_for\_other$	−.0245825	−.0242369	−.1191394	−.1173949	.1675383
	(.0868418)	(.0996339)	(.0973304)	(.1147397)	(.0993519)
$Numberof\ Observations$	3068	2804	2184	2420	2658
R^2	0.0662	0.0611	0.0740	0.0550	0.0425

do in fact pay higher premiums. The large increase in premiums with age could be due to the fact that elderly people are more prone to illnesses so the insurance company charges them higher premiums in order to compensate for the risk that it is bearing.

Looking at the variable for education, we see that the coefficients for all years are positive and highly significant, which suggests that better educated people pay higher premiums. This result is plausible because better educated people may desire insurance of better quality so they choose the expensive health insurance with extensive coverage.

US citizens also tend to pay higher premiums than non-US citizens, conditional on income, age, and other characteristics being constant. A possible explanation for this finding is that US citizens have faster wage growth than non-US citizens, which means that US citizens can afford better health insurance [Bratsberg et al.2002].

The coefficient on the variable for male is positive but not significant in the years 2005 and 2007, which suggests that the effect of gender on premiums is unclear. Although it is true that men on average earn more than women, it is not necessarily the case that men prefer the insurance plans with the higher premiums.

The coefficient for the Hispanic dummy is negative and significant in all years except 2008. The negative sign can perhaps be explained by the fact that Hispanic people on average earn less than non-Hispanics, which suggests that Hispanics are less likely to pay for expensive health insurance [Reimers1983].

For the medical condition principal component, the coefficient is positive in the years 2006, 2007, and 2008, but negative in the years 2004 and 2005. However, the t-statistics for the coefficients are less than one in all years except 2006, which suggests that health status does not seem to have a significant effect on insurance premiums.

Examining the coefficients on the dummies for the plan types yields some interesting trends. The dummy for the PPO plan has a positive coefficient that is significant at the 5% level for all years, which suggests that PPO plan subscribers pay higher premiums than non-PPO subscribers, keeping other variables constant. Also, it is interesting to note that the magnitudes of the PPO dummy coefficients are larger than the magnitudes of the coefficients for the other managed care plans in all years. This suggests that PPO plans charge higher premiums than HMO or POS plans to individuals with identical characteristics. This finding is consistent with the fact that PPO plans incur the greatest costs for the insurance company since patients

have the flexibility to see a specialist directly before consulting with a primary care physician.

When we look at the coefficients for the POS plan, we see that the coefficients are not significant, which means we cannot conclude anything about whether POS plan subscribers pay higher premiums than non-POS plan subscribers. Similarly, the coefficients of the fee-for-service dummy are not significant and neither are the coefficients for the "other" plan type dummy. Since "other" encompasses different varieties of health insurance, it is not surprising that there is no clear trend in premiums for that group.

The separate year regression results have shown us that wealthier and older people pay higher premiums, conditional on other factors staying constant. Also, US citizens and non-Hispanics pay more for health insurance than non-US citizens and Hispanics. Moreover, PPO plans typically charge the highest premiums out of the three types of managed care plans, which is probably because the insurance company has the hardest time with cost control in the PPO plan.

Since the coefficients on income and age do not differ by much over the years, it might be reasonable to run a pooled regression and see if the same effects for income and age emerge. In order to determine if pooled regression would produce the same results on age and income as the separate year regressions, we need to determine if the coefficients on income and age stay the same across all years.

For this purpose, year dummies were created, and the data was pooled. During the pooling process, income and premiums were deflated to 2008 dollars using the CPI for all urban consumers using all items. Then, log income and log age were interacted against the year dummies for 2007, 2006, 2005, and 2004. The following OLS regression was performed:

$$\begin{aligned}\ln(PREMIUM) \quad = \quad & \beta_1 \ln(AGE) + \beta_2 \ln(INCOME) + \beta_3 EDUCATION + \beta_4 USCITIZEN + \\ & \beta_5 MALE + \beta_6 HISPANIC + \beta_7 MEDICALCONDITION + \beta_8 HMO + \\ & \beta_9 PPO + \beta_{10} POS + \beta_{11} FEE + \beta_{12} OTHER + \\ & interaction\_terms\_between\_age\_and\_year + \\ & interaction\_terms\_between\_income\_and\_year + \varepsilon_i\end{aligned}$$

Then, a Chow test was performed on the null hypothesis that the coefficients for the interaction terms are equal to zero. Jointly testing that the interaction terms are equal to zero is equivalent to testing that the coefficients on income and age are the same across all years. The Chow-statistic with 8 degrees of freedom in the numerator and 13109 degrees of freedom in the denominator from that test was equal to 0.28, which corresponds to a p-value of 0.9714. Therefore, we cannot reject our null hypothesis that the coefficients for

income and age are the same across all years. Pooling the data would thus give the same OLS results on income and age as the separate year OLS regressions.

4.2 Pooled OLS Results

Table 3 shows the pooled OLS results.

As we can see, the pooled OLS coefficients for age and income are similar to the coefficients obtained in the separate year OLS regressions. The coefficient for age is 0.40, which lies within the 0.35 and 0.45 range for the separate year coefficients. Similarly, the coefficient for income (0.045) lies within the 0.03 to 0.06 range for the separate year coefficients. The coefficients on the other variables remain significant and retain the same sign as in a majority of the separate year regressions.

4.3 2SLS Results

The OLS results have given support to the fact that older and wealthier people pay more for health insurance, keeping other factors constant. However, the OLS specification may produce inconsistent estimates in the presence of errors in variables. Since we only observe an individual's income bracket instead of the actual income, we suspect measurement error in the income variable, which will cause OLS income estimates to be biased downward. In order to test for errors in variables, two stage least squares was performed on each year's of data using the following three instruments for income: FOODSTAMP which measures whether the individual qualified for foodstamp programs, SSRR which measures whether the individual qualified for social security or Railroad Rewards, and PENSION which measures whether the individual qualified for pension programs. An overidentification test was performed to test a joint exclusion restriction that all of the instruments are uncorrelated with the error term. Table 4 shows the 2SLS results for each year.

All the overidentification statistics are less than the chi-squared statistic with 2 degrees of freedom at the 30% level, so therefore the exclusion restriction holds. Also, in the first stage regressions, the coefficients on the instruments are significant, which indicates that the relevance condition is satisfied. Intuitively, it makes sense for FOODSTAMP, PENSION, and SSRR to be correlated with income since whether or not one receives foodstamps is correlated with one's income. Similarly, whether or not people receive pensions or social security depends on whether or not one has retired, which is correlated with one's age and income, but not the premiums that one pays.

If we compare Table 4 to Table 2, we see that the coefficients on income are greater in the 2SLS specification than in the OLS specification, which indicates the presence of measurement error in the income variable. Interestingly, the age coefficient is smaller in the 2SLS specification than in the OLS specification,

Table 3: Pooled OLS Results, Dependent variable: log of out-of-pocket premiums

Variable (Standard Error)	Pooled OLS
$\ln(AGE)$	.405053 (.0287427)
$\ln(INCOME)$	.0446586 (.0078165)
$EDUCATION$	.0349317 (.003324)
$USCITIZEN$	.2353483 (.0433879)
$MALE$	.0608728 (.018999)
$HISPANIC$	−.1184009 (.0297783)
$MEDICALCONDITION$	.0125261 (.0107097)
$dummy\_for\_hmo$	.153905 (.0339851)
$dummy\_for\_ppo$	.2552552 (.0312121)
$dummy\_for\_pos$	.138467 (.0676477)
$dummy\_for\_fee\_for\_service$	−.0919594 (.071846)
$dummy\_for\_other$	−.0247504 (.0446242)
$year2004$	4.914516 (.1404129)
year2005	4.998325 (.1415235)
year2006	4.98287 (.1419267)
year2007	4.993465 (.1420729)
year2008	4.922989 (.141527)
$Numberof\ Observations$	13134
R^2	0.9828

Income and Premiums were deflated to 2008 dollars using CPI for all items.

Table 4: 2SLS Results for each year

Variable	2004	2005	2006	2007	2008
ln(AGE)	.2270316	.2570671	.3073388	.2292546	.2269913
	(.0672485)	(.0690353)	(.0796716)	(.0788999)	(.087212)
ln($INCOME$)	.4715432	.4334326	.4840748	.3339356	.2858897
	(.064938)	(.0609526)	(.0825295)	(.0698639)	(.1091581)
$EDUCATION$	−.0093745	.0115651	.003576	.003419	.002331
	(.0093035)	(.0097041)	(.011818)	(.0108912)	(.0123945)
$USCITIZEN$	.1115718	.2781028	.0972838	.1728524	.291594
	(.0900009)	(.0949033)	(.1037401)	(.0983807)	(.0928799)
$MALE$	−.176253	−.2056726	−.2053188	−.1571799	−.0743381
	(.0560732)	(.0578194)	(.0745904)	(.0657968)	(.080502)
$HISPANIC$	−.3015365	−.1999004	−.2694939	−.1990006	.0039647
	(.0677656)	(.0731439)	(.0795579)	(.0769117)	(.0697193)
$MEDICALCONDITION$	−.1031864	−.0480112	.0038743	−.0330214	−.0330578
	(.031579)	(.0222164)	(.0292037)	(.0329017)	(.036995)
Dummy HMO	−.0721288	−.0533092	−.0881288	.2397646	.0587676
	(.0783737)	(.0768556)	(.0970516)	(.0874482)	(.0892901)
Dummy PPO	.1737884	.0673431	.0135002	.2764946	.1265857
	(.0744768)	(.0702699)	(.0917101)	(.0802111)	(.0802169)
Dummy POS	−.0024859	.1127285	−.0994332	.2258198	−.1047753
	(.1385513)	(.1384202)	(.1678121)	(.1574213)	(.1392627)
Dummy FEE	.023998	.1168527	−.3527816	−.069674	−.0660702
	(.1317011)	(.1248004)	(.1506026)	(.1563709)	(.1599011)
Dummy OTHER	−.002216	.0274523	−.0831459	−.053187	.2611311
	(.0974525)	(.1057953)	(.1240865)	(.121285)	(.1240023)
$Numberof\ Observations$	3068	2804	2184	2420	2658
Overid Statistic	.466826	.411061	.467722	1.38813	.573292
p-value	0.7918	0.8142	0.7915	0.4995	0.7508

which suggests that some of the variation in the age coefficient was absorbed by the instruments. This makes sense because PENSION and SSRR are positively correlated with age.

The coefficients on education lose their significance when we use 2SLS, which suggests that much of the variation in education could be explained by the instruments. Once we include the instruments in the model, the effect of education on premiums goes away.

The coefficients on citizenship remain positive and significant as in the OLS model, which confirms the fact that keeping other factors constant, US citizens pay more for health insurance than non-US citizens.

Interestingly, the coefficient on male switches sign when we use IV, which suggests that the effect of gender on premiums is indeterminate.

The coefficient on Hispanic remains negative and significant in all years except 2008, which confirms the OLS result that Hispanics pay lower premiums, keeping other variables fixed.

The medical condition coefficient becomes negative and significant in years 2004, 2005, and 2007, and it is insignificant in other years. A possible explanation for this effect is that the principal component variables of whether the person has difficulty walking or remembering are not indicative of one's health status. However, due to the high non-response rates for other medical conditions, a tradeoff was made in favor of preserving the number of degrees of freedom.

In order to verify that the OLS estimates are inconsistent, a Hausman test was performed on each year's OLS and 2SLS results [Hausman1978]. The Hausman statistics with 1 degree of freedom are shown in Table 5.

Table 5: Hausman Statistics for Test of Errors in Variables

	2004	2005	2006	2007	2008
Hausman statistic with 1 degree of freedom	46.74	41.32	31.71	19.17	4.65

As we can see, all of the Hausman statistics are greater than the chi-squared statistics with one degree of freedom, which means we reject the null hypothesis that OLS is consistent.
Next, 2SLS was performed on the pooled data using the same instruments. Table 6 compares the pooled OLS and 2SLS results.

Since the overidentification statistic of 1.00 is less than the chi-squared statistic with two degrees of freedom at the 20% significance level, we cannot reject the null hypothesis that the joint exclusion restriction

Table 6: Pooled OLS and 2SLS: Dependent Variable: log premiums

Variable (Standard Error)	Pooled OLS	Pooled 2SLS
$\ln(AGE)$	.405053 (.0287427)	.2372954 (.033888)
$\ln(INCOME)$	.0446586 (.0078165)	.407034 (.0367217)
$EDUCATION$	.0349317 (.003324)	.0014867 (.0048946)
$USCITIZEN$	.2353483 (.0433879)	.1884017 (.0423754)
$MALE$	.0608728 (.018999)	−.1654679 (.0306198)
$HISPANIC$	−.1184009 (.0297783)	−.1922166 (.0326707)
$MEDICALCONDITION$	.0125261 (.0107097)	−.045 (.0131676)
$dummy\_for\_hmo$	.153905 (.0339851)	.0114374 (.0381438)
$dummy\_for\_ppo$	.2552552 (.0312121)	.132894 (.0352123)
$dummy\_for\_pos$	.138467 (.0676477)	.0409361 (.0656601)
$dummy\_for\_fee\_for\_service$	−.0919594 (.071846)	−.039643 (.0633929)
$dummy\_for\_other$	−.0247504 (.0446242)	.0330499 (.0498907)
$year2004$	4.914516 (.1404129)	−.077244 (.0316351)
year2005	4.998325 (.1415235)	.0064345 (.0322337)
year2006	4.98287 (.1419267)	5.145605 (.1411761)
year2007	4.993465 (.1420729)	.0222087 (.0333382)
year2008	4.922989 (.141527)	−.063389 (.0326214)
$Numberof\ Observations$	13134	13134
R^2	0.9828	0.9826
Overid Statistic		.999881
Hausman Statistic		6349.3504

Instruments: FOODSTAMP, PENSION, and SSRR

holds, giving support to the validity of the instruments. The Hausman statistic of 6349 confirms the presence of an errors in variables problem.

The OLS and 2SLS coefficients on US Citizen differ by about 20% but retain their sign and significance, confirming the fact that US citizens pay more for health insurance than non-US citizens, keeping other factors constant. The coefficient on the PPO dummy remains positive and significant between the two specifications, confirming that PPO plans charge higher premiums than other plans.

By comparing the OLS and 2SLS results on year-by-year and pooled data, we have seen greater income coefficients and smaller age coefficients in the 2SLS specification than in the OLS specification. Despite the large differences in the magnitudes of the coefficients, they nevertheless remain positive no matter what specification is used. It would be good to provide some explanation for why after keeping demographics and indicators of potential health problems constant, older and wealthier people are likely to pay more for health insurance than younger and poorer people. Two possible explanations exist. First, it is possible that older and wealthier people are more risk averse and prefer the plans with the higher premiums, which usually have better coverage than the plans with the lower premiums. Second, it is possible that adverse selection is present, in which the insurance companies charge older people higher premiums because they believe that older people are more prone to medical illnesses. In order to disentangle these two effects, a nested logit regression was performed to determine if older people prefer managed care plans over traditional care plans. If it turned out that older people prefer the more expensive managed care plans over traditional care plans, then it is plausible that older people are choosing the plans with the higher premiums (perhaps because they would like better coverage). But if it turned out that older people actually favored traditional care over managed care but are still paying higher premiums, then it could be that adverse selection based on age is present.

4.4 Nested Logit Results

In order to determine the plans favored by the elderly and wealthy individuals, nested logit regression was performed. Table 7 shows the nonnormalized nested logit results where income and age are case-specific variables at the top level of the nested logit tree.

As we can see, the coefficient on the interaction term between income and traditional care is negative and significant at the 1% level for all years. This strongly suggests that wealthier people prefer managed care over traditional care. A possible explanation is that wealthier people desire plans that provide primary and preventive care coverage in addition to coverage for severe illnesses since wealthier people may routinely visit a physician for a regular checkup. Or it could be that wealthier people place more trust in the physicians

Table 7: Nested Logit Results with Income and Age as Top level Variables

Variable (standard error)	2004	2005	2006	2007	2008
dummy for hmo	1.383061 (20.97903)	1.467541 (.3298407)	1.414274 (.3806032)	1.61868 (11.87173)	1.71384 (.1187688)
dummy for ppo	1.786377 (20.97902)	2.055911 (.3289393)	2.020572 (.3795774)	2.315035 (11.8717)	2.434116 (.115598)
dummy for pos	−.9544614 (20.97927)	−.7203107 (.3454905)	−.7950022 (.3984572)	−.6262243 (11.8723)	−.4548043 (N/A)
dummy for fee for service	−.9080833 (.1196628)	−.4347192 (.1212794)	−.5632885 (.1418441)	−.7149679 (.1479859)	−.8244304 (.1588591)
traditional care * ln(age)	1.282378 (.1807629)	2.153326 (.2245949)	2.277842 (.2697607)	2.075714 (.2526602)	1.807026 (.2479094)
traditional care * ln(income)	−.2573889 (.0429034)	−.2501298 (.0454457)	−.2516521 (.0509986)	−.2276727 (.0512258)	−.3247668 (.0511939)
Numberof Observations	3068	2804	2184	2420	2658

that are part of the network of a managed care organization and therefore enroll in the managed care plans.

The coefficients for the interaction term between age and traditional care are positive and highly significant across all years. This suggests that older people prefer traditional care over managed care. A possible explanation is that older people may be accustomed to the traditional care plans that were popular in the 1980s and consequently decided to stay in the same plans that they had in their youth. Or, it could mean that older people would like the traditional care plans with the better deals even if they do not provide as good of coverage as the managed care plans.

Based on the nested logit results, we cannot say that adverse selection based on income exists because wealthier people are choosing the more expensive managed care plans. However, it is possible that adverse selection based on age exists because the older people prefer the less expensive traditional care plans, and yet they are paying higher premiums than younger people, keeping medical conditions, income, and other demographics constant.

4.5 Answers to possible objections

4.5.1 Omitted Variable: Regional differences in health care premiums

A possible argument is that regional differences in health care plans may contribute to the different premiums that individuals face and that leaving out the regional effects might introduce bias if one of the explanatory variables was correlated with the region variable. To test this argument, dummies for the four regions (Northeast, Midwest, South, and West) were added into the pooled least squares regression. After

adding in the dummies, coefficients on all variables retained the same sign, and their magnitudes did not change by more than 15%. Therefore, omitting regional dummies from the specification should not be a problem.

4.5.2 Variables for the principal component

Some may argue that the variables used to determine the principal component are not indicative of serious medical conditions. It is true that the existence of any limitations in one's ability to walk may not necessarily indicate any serious medical problems. However, due to the high non-response rate for questions on medical conditions, there was a tradeoff between degrees of freedom and quality of indicators of medical conditions. In order to preserve a reasonably high number of degrees of freedom (2000 as opposed to 150), only the variables with the highest response rates were used.

4.5.3 Heteroskedasticity

Initially, there were concerns over whether heteroskedasticity would be a problem because the White Test performed on the least squares results rejected null hypothesis of homoskedasticity. However, when robust standard errors were used, the p-values did not change by more than 15%, which suggests that heteroskedasticity should not affect the results.

4.5.4 Conditional Logit

Some may wonder why the conditional logit was not used in determining the preferences of wealthy and elderly people for managed care versus traditional care. The reason is that conditional logit regression relies on the assumption of the Independence of Irrelevant Alternatives, which means that individuals' preferences for a particular insurance plan should not depend on the other insurance plans. In order to test whether this assumption was true, a Hausman-McFadden test was performed on an unrestricted and restricted conditional logit model. In the restricted logit model, the HMO plan was dropped from the choice set. Table 8 shows the Hausman statistics from that test.

Table 8: Hausman statistics for test of IIA

	2004	2005	2006	2007	2008
Hausman statistic with 6 degrees of freedom	-964.80138	-102.52895	-238.15736	332.04984	586.91815

As we can see, the Hausman statistics are either larger than their threshold chi-squared values, or they are negative, suggesting that the asymptotic assumptions of the Hausman-McFadden test are not met. Since

none of the statistics are positive and less than their threshold chi-squared values, we cannot say that the Independence of Irrelevant Alternatives assumption is satisfied. Therefore, it is better to use the nested logit model which does not rely on the IIA assumption.

5 Conclusion

This paper has examined the effects of the purchaser's age and income on her health insurance premiums in the individual (non-employer sponsored) market. After OLS and 2SLS regressions were performed, highly significant positive effects were found for both age and income when controlling for medical conditions, plan type, and other demographic characteristics. These results confirm the original hypothesis that older and wealthier people tend to pay higher premiums, after controlling for medical condition and demographics.

In an attempt to determine if the higher premiums are a result of adverse selection by the insurance company or of demand for more expensive insurance plans by wealthier and older buyers, a nested logit regression was performed to determine which plans were favored by wealthier and older individuals. It was found that wealthier people favor the more expensive managed care plans, which provide better coverage; this finding suggests that greater demand for managed care plans by wealthy individuals may be driving up the premiums for these plans. It was also found that older people actually prefer the less expensive traditional care plans over managed care plans, which suggests that adverse selection based on age may exist because insurance companies are charging older people higher premiums even when they prefer the less expensive plans. A possible explanation for this adverse selection effect is that insurance companies believe older people are more prone to serious illnesses and therefore need to insure against the greater expected losses by charging higher premiums.

The result of this study on the effects of age and income on out-of-pocket health insurance premiums in the individual market has important implications in light of the aging baby-boomer generation. According to the results of this study, as the average age of the US population increases, average premiums in the individual market would also increase. Furthermore, the results showing the preferences of older individuals for less expensive traditional care indicate the possibility of adverse selection based on age. The possible existence of adverse selection should have important policy implications for deciding whether it is necessary to have more governmental regulation in the individual health insurance industry.

References

[Bratsberg et al.2002] Bratsberg, B. et al. (2002). The effect of naturalization on wage growth: A panel study of young male immigrants. *Journal of Labor Economics*, 20(3):568–597.

[Claxton et al.2006] Claxton, G. et al. (2006). Health benefits in 2006: Premium increases moderate, enrollment in consumer-directed health plans remains modest. *Health Affairs*, 25:w476–w485.

[Davis2007] Davis, K. (2007). Uninsured in america:problems and possible solutions. *British Medical Journal*, 334:346–348.

[DAY2008] DAY, J. C. (2008). Population profile of the united states. `http://www.census.gov/population/www/pop-profile/natproj.html`.

[Frank et al.2000] Frank, R. et al. (2000). Measuring adverse selection in managed health care. *Journal of Health Economics*, 19(6):829–854.

[Gabel et al.2005] Gabel, J. et al. (2005). Health benefits in 2005: Premium increases slow down, coverage continues to erode. *Health Affairs*, 24(5):1273–1280.

[Hausman1978] Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

[Knight1998] Knight, W. (1998). *Managed Care: What it is and how it works*. Aspen Publishers, Inc.

[Reimers1983] Reimers, C. (1983). Labor market discrimination against hispanic and black men. *The Review of Economics and Statistics*, 65(4):570–579.