On the Anatomy of Adverse Selection in Health Insurance Market: Evidence from MEPS

Cinzia Di Novi*

August 5, 2007

Abstract

We use the 2003/2004 Medical Expenditure Panel Survey in conjunctions with the 2002 National Health Interview Survey to test for adverse selection in the USA health insurance market. The test is conducted by estimating the correlation between the completeness of insurance an individual buys and his ex-post risk experience, conditional on the observable characteristics which are used in pricing insurance policies. Completeness of health insurance plan is measured by health insurance reimbursement. Since reimbursement is only defined for those who participate in insurance and have positive health care expenditure, the model may suffer from sample selection bias. To obtain consistent parameter estimates, model is estimated using Wooldridge's (1995) two step estimation procedure extended to the case in which selectivity is due to two sources rather than one.

JEL classification: I11, I18, D82.

Keywords: adverse selection, health insurance, risk profile.

*Department of Public Policy and Public Choice, University of Eastern Piedmont, via Cavour, 84, Alessandria, Italy. Collegio Carlo Alberto, via Real Collegio, 30, Moncalieri, Torino. E-mail: dinovi@econ.unito.it

1 Introduction

Since the seminal paper of Rothschild and Stiglitz (1976) greater attention has been devoted to the problem of asymmetric information among agents. An important form of asymmetric information between consumers and insurers is adverse selection. In health insurance market adverse selection may occur when consumers' true health-cost risk is private information: insurance company may know that consumers vary in the level of risk, but, on principle, is not able to discern who are high and who are low risk profile individuals within a group of potential insured. (Akerlof, 1970; Rothschild and Stiglitz, 1976). Identifying risks accurately is not an easy task and requires that insurance company incurs some costs. Insured parties are heterogeneous in terms of expected costs and have more information about their risks. Naturally, high-risk individuals are not encouraged to "reveal" their risk to the insurance company; this asymmetry is a serious problem since may lead insurance company to face large differences in expected health costs due to heterogeneity in demographics and the incidence of illness.

As the insurers has imperfect information on the individuals' health status, the cover and the premium will be set somewhere between what is required by the low and the high risk profile users. However, low risk users may feel they are paying too much with the respect to their needs. Low risk individuals tend to drop out of the risk pool, then, the average risk in the pool rises causing premium to rise and yet more people to drop out and so on. This may leave to the case in which only high risk profile individuals buy insurance and pay "average" rate.

To counteract to this problem, insurance companies may offer separate contracts with different coverage and prices, making claimant pay part of the claim (with coinsurance rate, deductible etc.) so that individuals should reveal their risks. Hence, risky individuals who expect high health care costs would tend to purchase insurance with higher premium but lower excesses since they are more likely to be claiming on a regular basis. On the other hand lower risk users, who expect low costs, would prefer a less complete insurance, with a lower premium and a higher excess in the unlikely event that they have to claim.[1] The phenomenon described above is known as *ex ante adverse selection*". (Fang et.al, 2006)[2]. The "*positive correlation property"* between the individual riskiness and the completeness of a health insurance plan, which characterize this phenomena, forms the basis for our empirical test for adverse selection. This test is conducted by using data from the 2003/2004 Medical Expenditure Panel Survey – Household Component (MEPS-HC) in conjunction with the previous year's National Health Interview Survey (NHIS). Many empirical work use information on coinsurance rate, health insurance benefits, stop-loss and deductible to measure generosity of health insurance plan (see, for instance, Browne and Doerpinghaus,1993). Unfortunately our data do not contain information about the insurance contract ; hence, we measure health insurance plan completeness by using health insurance reimbursement that is the vertical difference between total health expenditure and out-of-pocket expenditure on health care paid by consumers.

Health insurance reimbursement, however, is only defined for a subset of individuals from the overall population since we observe it only for those who participate in insurance and have positive health care expenditure. Thus, the model may suffer from sample selection bias and straightforward regression analysis may lead to inconsistent parameters estimate. Another problem that arises from the estimation is the presence of unobserved heterogeneity in the equations of interest. Wooldridge (1995) has proposed an estimator which deals with both sources of estimation bias. We extend this estimation method to the case in which selectivity is due to two sources rather than one

[1]This form of allocation has been proved superior (in terms of economic efficiency) to that in which a mean price is paid by all individuals. The main work in this area is attributed to Rothschild and Stiglitz (1976).

[2]This is also known as adverse selection effect à la Rothschild-Stiglitz: high risk agents, knowing they are more likely to have an accident, self-select by choosing contracts entailing a more comprehensive coverage.

(participation in insurance and participation in health care expenditure). The nature of the test is similar to the one in Browne and Doerpinghaus (1993).

We find no systematic relation between illness of individuals and insurance choice. We think that a possible explanation can be found in the so called "cream skimming" practise: health plans may have an incentive to alter their policy to attract the healthy and repeal the sick (Newhouse, 1996; Ellis, 1997). Then, individuals enrolled are relatively healthy people and this lead to the failure of the correlation test.

The remainder of this paper is organized as follows. Section 2 briefly surveys the empirical related literature. Section 3 describes the data and variables. In Section 4 we perform the empirical analysis, explain the test in detail and present our main results. Section 5 concludes the paper with a discussion. The definition of the variables, descriptive statistics and tables with estimation coefficients are in Appendix .

2 Related Literature

There is substantial empirical literature examining adverse selection in health insurance markets. However, there is conflicting evidence on the presence of adverse selection: the results are mixed. We briefly summarize these studies here.

Cameron and Trivedi (1991), for instance, use Australian data to estimate a logit model for the choice between a standard package and a more generous insurance plan. They find no significant effect of health condition variables on insurance choice. Marquis (1992), in a study of data from the RAND Health Insurance Experiment, finds that individuals who select more generous health insurance plans are more likely to have large health expenditures: this result is consistent with presence of adverse selection. Also Browne and Doerpinghaus (1993) find evidence for adverse selection: their results show that low and high risks purchase a pooling insurance policy and low risks subsidize the insurance purchase of high risk insured individuals. This supports the prediction by Miyazaki's (1977) theory of adverse selection. Cardon and Hendel (2001) test for a correlation between health care spending and insurance coverage using a two-stage model of the demand for health insurance. In their setup, individuals first receive a private signal that is correlated with their future health. Basing on this signal individuals make their choice about how much insurance to purchase. In the second stage, individuals consume health care. Their empirical analyses revealed that the joint insurance/health care consumption decision is largely explained by observed characteristics (such as income, education etc.) rather than unobserved health status. Then, they conclude that apparently there is no private information that insureds can use against the insurers and hence no adverse selection.

Bajari et al. (2006) use the Health Retirement Study to estimate a structural model of the demand for health insurance and medical care. They find evidence of moral hazard but not of adverse selection. Goldman et al. (2006) estimate independent effects of medical and drug benefits on plan selection. They find that while generosity of the medical benefit played an important role in choosing a plan, choices did not vary significantly by health status. In contrast, their data support a significant correlation between health status and plans with generous drug benefits: sicker individuals tend to enroll in plans with generous drug benefits, while healthier choose less generous plan. Basing on their founding, they assert that drug coverage may be more susceptible to adverse selection than medical coverage .

In insurance markets other than health, evidence for adverse selection is considerably contradictory too. Puelz and Snow (1994) presented empirical evidence of adverse selection in the market for automobile collision insurance. Using data from a private insurer, they find strong evidence of adverse selection in the insurer's portfolio. Chiappori and Salanié (2000) use data on contracts and accidents to examine the extent of asymmetric information in the French market for automobile insurance. They examine a relatively homogenous group of drivers with less than four years' driving experience. Their test do not reveal evidence of risk-related adverse selection. They find that when choosing their automobile insurance contracts, individuals behave as though they had no better knowledge of their risk than insurance companies, as adverse selection hypothesis would require.

Cawley and Philipson (1999) test for adverse selection in the market for life insurance; they first show that the death rate among those who purchase life insurance is lower than those who do not, moreover they find that who expect to die soon do not buy more complete life insurance plan. This is clearly in contrast with the basic adverse selection theory.

Finally, Makki and Somwaru (2001) analyze farmers' choices of crop insurance contracts. Their analysis offers empirical evidence of adverse selection by showing that high-risk farmers are more likely to select revenue insurance contracts and higher coverage levels with the respect to low–risk farmers.

Most of the studies we have come across have been using discrete choice models to model for health insurance purchase decision. They have used logit or probit specifications to analyze this decision problems in which the dependent variable has often two outcomes: buying or not buying health insurance. Few studies have gone to the next level and tried to explain which factors affect the extent of insurance purchase. Moreover, in most of the studies which test for adverse selection two important estimation issues such unobserved heterogeneity and selection bias, are traditionally treated separately[3]. The aim of this paper is to find factors which affect the extent of insurance purchase with particular attention to individuals' risk profile. In

[3]It is often mathematically complex to combine these two issues together, a large burden of computer programming and a set of strong distributional assumptions are need for the combination. The model presented in this paper, however, is estimated with the common statistical software STATA 9. Also the statistical assumptions needed for Wooldridge's model in this paper is relatively weaker than the other methods.

our model, we control for selection bias and at the same time for unobserved heterogeneity issue.

3 Data and Variables

We use data from the 2003/2004 Medical Expenditure Panel Survey – Household Component (MEPS-HC) and 2002 National Health Interview Survey (NHIS). MEPS is an on-going survey sponsored by the Agency of Health Care Policy Research (AHCPR). MEPS provides a nationally representative sample of US civilian non-institutionalized population. MEPS is self-reported and contains detailed information on health care consumption and demographics including age, sex, marital status, income, work status and geographic location. In addition data contain information on the respondents' health status, health conditions, health charges and payments, access to care, health conditions, health insurance coverage.

Each year's sample for MEPS is drawn from respondents to the previous year's NHIS that is conducted annually by the National Centers for Health Statistics (NCHS), Center for Disease Control and Prevention (CDC). NHIS provides rather detailed information about health status, diseases, life-style, education and other individual characteristics. We use the 2002 NHIS in conjunction with 2003/2004 MEPS with MEPS as our primary database because it contains information on health insurance reimbursement that is the dependent variable of interest in this paper, as well as the detailed information on health insurance.

After correcting for the missing values, the sample was reduced to 890 individuals resulting in 1780 observations. Observations containing veterans and individuals who are covered by Champus/ ChampVa insurance are removed from the data set since their medical services demand and access to medical services distinctly differs from the general population[4].

[4]The health care system in US is characterized by: private insurance, Medicare and

Table 2 presents summery statistics for demographics and health insurance information. The sample of 1780 individuals is divided into insured and uninsured. Only 8% of the sample is uninsured. As showed in Table 2, uninsured are younger, and poorer. Health care expenditure is important relative to total income, around 11% for insured and 13% for uninsured. The average expenditure for full sample is 4,300 $. The distribution of the expenditures is highly skewed, as expected. Insured spent 50% more in health care than uninsured (4314$ versus 2001 $).

Table 2 shows that 89% of insured report that their health is good versus the 82% of uninsured. Insured behave in a healthier way: the percentage of smoker and the percentage of heavy alcohol consumers are lower; on average they present a lower BMI, and they practice physical activity more often than uninsured.

At the first sight, it seems that there are no symptoms consistent with adverse selection: a substantial fraction of the sample is insured and among insured about 90% of individuals enjoy good health. [5]

Medicaid and Military health insurance.

Private medical insurance is the largest component of the health care system: insured pay a fee-for-service reimbursement basis; they pay directly the medical treatments and be reimbursed at a later date by the insurer. Medicare is a program funded by the government through social security payments. It was created mainly for people 65 years of age and older, some disabled people under 65 years of age, and people with end-stage renal disease. This scheme is extremely basic with very few services offered with much of the cost still having to be met by the patient. Since Medicare has a number of gaps in coverage, most enrolls own supplemental insurance coverage.

Medicaid is funded jointly by the federal and state authorities and is available for individuals of all ages and families with low incomes and resources who cannot afford proper medical care.

Champus (now known as Tricare) is a health care benefits program for active duty and retired members of the military.

ChampVa is a health care benefits program for permanently disabled veterans and their dependents.

[5] A possible explanation of the higher percentage of healthy individuals among insured can be found in the insurance plan characteristics. Plans may have incentives to distort their offering to attract the healthy and repel the sick.

Seeking favourable risk is often referred as cream skimming. These strategic behavior can take a variety of forms including designing insurance benefits packages in such a way

3.1 Risk Profile Variables

To perform the correlation test, first we classify individuals as being high and low risk profile individuals. Individuals are classified as being low-risk if their health status is good. As a measure of health status we use two indicators: a subjective and an objective one. In particular, following the Berger and Leigh(1988), we choose blood pressure as indicator of overall health, since it is the most important predictor of cardiovascular disease which is the greatest killer in the U.S. We create a binary variable (hypertension) that takes value one if respondents suffer from high blood pressure and zero otherwise. We classify individuals as high-risk profile individuals if they report that they suffer from hypertension. Moreover, we use as a measure of overall health SAH (self-assessed health)[6] that is a five category variable rating from poor to excellent. We construct a binary variable (health) with the value one if individuals report that their health status is excellent, very good, good and zero otherwise (fair or poor). Then, we classify as high-risk individuals those whose self-reported health is fair or poor.

In addition, individuals are classified as being characterized by a high-risk profile if they follow an unhealthy life-style. Life-style variables measure the effort that individuals use to prevent an illness and at the same time they are good predictor of future illness. The behavioral variables employed are indicator of smoking, alcohol consumption, physical activity practice and BMI[7] . Individuals are classified as being characterized by a high risk profile if

as to be more attractive to healthy persons than unhealthy by, for instance, excluding particular prescription drugs or offering health club memberships which appeal to the low risks. The result is that individuals enrolled in health insurance are relatively healthy people.

[6]Self- reported health status is a very good indicator of overall health. It has been showed to be an important predictor of subsequent mortality and medical services use, and is widely used as a measure for the stock of health in pervious studies that analyze empirical determinants of health. (Contoyannis and Jones 2004, Contoyannis et al. 2004).

[7]BMI (Body Max Index) is used as measure of obesity. Obesity is considered a risk

they smoke, usually consume heavy drinks, practice vigorous physical activity less than once per week and if their reported BMI is higher than 25.0000.

3.2 Other Characteristics

In addition to the health and life-style indicators, the independent variables, used to control for differences in policy, can be grouped in the following categories: demographic variables (age, sex, race), socioeconomic variables (education, marital status, employment status, income) preferences (risk aversion). Moreover, we control for total annual expenditure, out-of-pocket annual premium and whether individuals suffer from any form of disabilities that limit their activities (such as working, studying etc.)

Because older individuals tend to use more medical services and may have higher medical expenditure, we expect a positive relationship between age and the amount of reimbursement. Since men tend to use less medical services than female we expect a negative coefficient for male. A positive relationship between the variables black and other race and the completeness of coverage is expected because of the higher need of medical services among non whites caused by a higher morbidity rate.

According to the "marriage protection hypothesis" (which states that the actual process of living with a spouse confers health benefits to both partners) we expect that married people tend to use less medical services. Thus, we expect a negative correlation between the variable "married" and the dependent variable that measure the generosity of health plan.

The variables which are indicators of education, employment status and income are included in the analysis to account differences, other than risk type, which may affect the amount of insurance purchase by the insured. We expect a negative relationship between degree of education and the amount of

factor for several diseases. It is often associated with aspects of an individual's life-style such as an insufficient physical activity and inappropriate nutrition. Those who are a BMI $>$ or equal than 25.0000 are overweight and at risk of obesity and are expected to have poorer health.

insurance purchased: individuals with a higher level of schooling are observed to be healthier than the others[8]. Hence, we expect that individuals with a higher degree of education use less medical services and purchase a less complete insurance plan. Similarly, the coefficients for income and employed are expected to be negative.

We include also a measure of risk aversion. Higher risk aversion translates into a willingness to pay more to eliminate financial risk. For a given premium, we expect a positive coefficient for the variable that measures risk aversion since more risk-averse insured will be willing to tolerate higher deductible, coinsurance rate, stop-loss than someone who is less risk-averse[9].

The variable that we use as indicator of limited activity controls for the portion of risk observable to the insurer. The activity limitations indicator is expected to be positively related to the generosity of the health insurance place, because be limited increases the likelihood of need for medical care.

[8]One explanation of this empirical regularity is that education increases the productivity of producing health i.e. more health can be produced for the same inputs (Gerdtham et al., 1999, Berger and Leigh, 1989). Schooling helps people choose healthier life-styles by improving their knowledge of the relationship between health behaviors and health outcomes. (Kenkel, 1991). A more educated person may have more knowledge about the harmful effects of cigarette smoking, alcohol consumption or about what constitutes an appropriate, healthy diet. Furthermore, schooling increases information about the importance of having regular exams or screening tests to prevent an illness or at least to minimize disease.

[9]Chiappori and Salaniè in their recent work "Testing for Asymmetric Information in Insurance Markets" stressed the importance of including risk aversion among explanatory variables:

[... more risk averse drivers tend to both buy more insurance and to drive cautiously; this would even suggest a negative correlation between insurance coverage and accident frequency...]. Then, if do not control for individuals risk aversion, we may obtain spurious correlation between individuals' risk profile and completeness of coverage.

4 Estimation Strategies and Empirical Results

4.1 Wooldridge Two-Step Estimation

To test for differences in insurance purchases by high and low risk profile individual we use as a measure of completeness of coverage the natural logarithm of health care reimbursement as dependent variable which is constructed by taking the natural logarithm of the total health care expenditure paid by private insurance, Medicare and Medicaid. The assumption of lognormality better fits the expenditure reimbursement and has precedents (see, for example, Keeler et al., 1977, Browne and Doerpinghaus, 1993).

Health insurance reimbursement is only defined for a subset of individuals from the overall population since we observe it only for those who participate in insurance and face positive health care expenditure. Hence the model suffers from sample selection bias and straightforward regression analysis may lead to inconsistent parameters estimate.

Another problem that arises from the estimation is the presence of unobserved heterogeneity in the equations of interest. Wooldridge (1995) has proposed an estimator which deals with both sources of estimation bias; this estimator requires panel data and produces consistent parameter estimates under a set of assumptions. It does not impose distributional assumptions about the error terms but requires specifying the functional form of the conditional mean of the individual effects in the structural equation. We extend this method to the case in which selectivity is due to two sources rather than one.

We start by sketching in the following Wooldridge (1995) sample selection model with one selection criterion, then we present a specification of this model in which the selection process is based on two selection criteria rather than one.

Following M.Rochina-Barrachina (1999), we considere the following problem:

$$\begin{array}{c} d_{it}^{*} = z_{it}\gamma + \mu_i + u_{it} \\ d_{it} = 0 \quad if \quad d_{it}^{*} \leq 0 \\ d_{it} = 1 \quad if \quad d_{it}^{*} > 0 \end{array} \tag{1}$$

$$\begin{array}{c} y_{it}^{*} = x_{it}\beta + \alpha_i + \varepsilon_{it} \\ y_{it} = y_{it}^{*} \quad if \quad d_{it} = 1 \\ y_{it} \ \textit{not observed otherwise} \end{array} \tag{2}$$

where equation (1) defines the selection rule while equation (2) is the primary equation. i $(i = 1, ...n)$ denotes the individuals while t $(t = 1, ..., t)$ denotes the panel. x_{it} and z_{it} are vector of exogenous variables. The dependent variable in the primary equation, y_{it}, is observed only for the observations satisfying the selection rule. Wooldridge suggests employing Chamberlain (1980) characterization, by assuming the conditional mean of the individual effects in the selection equation as a linear projections on the leads and lags of observable variables:

$$\mu_i = z_{i1}\delta_1 + ... + z_{it}\delta_t + c_i \tag{3}$$

where c_i is a random component. By substituting Chamberlain characterization into the selection equation yields:

$$d_{it}^{*} = z_{it}\gamma + z_{i1}\delta_1 + ... + z_{it}\delta_t + v_{it} \tag{4}$$

where $v_{it} = c_i + u_{it}$. v_{it} is independently distributed of z_{it} and is normally distributed with zero mean and σ^2 variance. The regression function of α_i on z_{it} and v_{it} is linear, accordingly:

$$E\left[\alpha_i \left| z_{it}, v_{it} \right.\right] = x_{i1}\psi_1 + ... + x_{it}\psi_t + \phi_t v_{it} \tag{5}$$

We do not observe v_{it}, but only the binary indicator d_{it}. Then, we replace $E\left[\alpha_i \left| z_{it}, v_{it} \right.\right]$ with:

$$E\left[\alpha_i \left| z_{it}, d_{it} = 1 \right.\right] = x_{i1}\psi_1 + ... + x_{it}\psi_t + \phi_t E\left[v_{it} \left| z_{it}, d_{it} = 1 \right.\right] \quad (6)$$

Wooldridge assumes that ε_{it} is mean independent of z_{it} conditional on v_{it} and its conditional mean is linear on v_{it}:

$$E\left[\varepsilon_{it} \left| z_{it}, v_{it} \right.\right] = E\left[\varepsilon_{it} \left| v_{it} \right.\right] = \rho_t v_{it} \quad (7)$$

By the Law of Iterated Expectation:

$$E\left[\varepsilon_{it} \left| z_{it}, d_{it} = 1 \right.\right] = \rho_t E\left[v_{it} \left| z_{it}, d_{it} = 1 \right.\right] \quad (8)$$

From the above assumption, Wooldridge derives an explicit expression for

$$\begin{gathered} E\left[\alpha_i + \varepsilon_{it} \left| z_{it}, d_{it} = 1 \right.\right] = E\left[\alpha_i \left| z_{it}, d_{it} = 1 \right.\right] + E\left[\varepsilon_{it} \left| z_{it}, d_{it} = 1 \right.\right] = \\ = x_{i1}\psi_1 + ... + x_{it}\psi_t + \left(\phi_t + \rho_t\right) E\left[v_{it} \left| z_{it}, d_{it} = 1 \right.\right] \end{gathered} \quad (9)$$

where

$$E\left[v_{it} \left| z_{it}, d_{it} = 1 \right.\right] = \lambda\left(z_{i1}\gamma_1 + ... + z_{it}\gamma_t\right) \quad (10)$$

So, for each period, Wooldridge suggests to estimate a cross-sectional probit model for participation and compute the Inverse Mills Ratio (IMR), then, estimate the structural equation:

$$y_{it} = x_{i1}\psi_1 + ... + x_{it}\psi_t + x_{it}\beta + \left(\phi_t + \rho_t\right)\lambda\left(z_{i1}\gamma_1 + ... + z_{it}\gamma_t\right) \quad (11)$$

by using fixed effect OLS or pooled OLS for the sample for which $d_{it} = 1$ (Vella, 1998).

Concerning the health insurance reimbursement model, we consider the following characterization of Wooldridge's sample selection model where selectivity bias is a function of two indices:

$$\begin{aligned} d^*_{it_1} &= z_{it_1}\gamma_1 + \mu_{i_1} + u_{it_1} \\ d_{it_1} &= 0 \quad if \quad d^*_{it_1} \leq 0 \\ d_{it_1} &= 1 \quad if \quad d^*_{it_1} > 0 \end{aligned} \tag{12}$$

$$\begin{aligned} d^*_{it_2} &= z_{it_2}\gamma_2 + \mu_{i_2} + u_{it_2} \\ d_{it_2} &= 0 \quad if \quad d^*_{it_2} \leq 0 \\ d_{it_2} &= 1 \quad if \quad d^*_{it_2} > 0 \end{aligned} \tag{13}$$

$$\begin{aligned} y^*_{it} &= x_{it}\beta + \alpha_i + \varepsilon_{it} \\ y_{it} &= y^*_{it} \quad if \quad d_{it} = 1 \\ y_{it} &\ not\ observed\ otherwise \end{aligned} \tag{14}$$

Let d_{it_1} be an unobserved variable denoting insurance participation decision and d_{it_2} an unobserved variable denoting health care expenditure participation decision. z_{it_1} , z_{it_2} and x_{it} are vector of exogenous variables. y_{it} denotes the natural logarithm of health insurance reimbursement. y_{it} is observed only for the sample for which $d_{it_1} = 1$ and $d_{it_2} = 1$.

Sample selection is now based on two criteria. The method of estimation relies crucially on the relationship between v_{it_1} and v_{it_2}[10], in particular, the estimation depends on whether the two error terms are independent or correlated, that is whether or not $Cov\left(v_{it_1}, v_{it_2}\right) = 0$. The simplest case is when the disturbances are uncorrelated (Maddala,1983, Vella, 1998). If $Cov\left(v_{it_1}, v_{it_2}\right) = 0$ we can easily extend Wooldridge's two-step estimation method to this model. The correction term to include as regressor in the primary equation is:

[10]From Chamberlain trasformation of the individual effects: $v_{it_1} = c_{i_1} + u_{it_1}$ and $v_{it_2} = c_{i_2} + u_{it_2}$

$$E\left[\varepsilon_{it}\left|z_{it}, d_{it_1}=1, d_{it_2}=1\right.\right] = \rho_{t_1}\lambda_1\left(z_{i1_1}\gamma_{1_1}+...+z_{it_1}\gamma_{t_1}\right)+ \\ +\rho_{t_2}\lambda_2\left(z_{i1_2}\gamma_{1_2}+...+z_{it_2}\gamma_{t_2}\right) \tag{15}$$

Then, we estimate the following model:

$$y_{it} = x_{i1}\psi_1+...+x_{it}\psi_t+x_{it}\beta+\left(\phi_{t_1}+\rho_{t_1}\right)\lambda_1\left(z_{i1_1}\gamma_{1_1}+...+z_{it_1}\gamma_{t_1}\right)+ \\ +\left(\phi_{t_2}+\rho_{t_2}\right)\lambda_2\left(z_{i1_2}\gamma_{1_2}+...+z_{it_2}\gamma_{t_2}\right) \tag{16}$$

The procedure consists in first estimating, for each period, by two single a cross-sectional probit model, the selection equation one and the selection equation two. Than, the two corresponding Inverse Mills Ratio can be imputed and included as correction terms in the primary equation. Thus, by fixed effect or pooled OLS[11], estimate of the resulting primary equation corrected for selection bias can be done for the sample for which $d_{it_1}=1$ and $d_{it_2}=1$.

In the case v_{it_1} and v_{it_2} are correlated, so that $Cov\left(v_{it_1}, v_{it_2}\right) = \sigma^2$, [... the expression get very messy...] (Maddala, 1983) and we have to use for each period cross-sectional bivariate probit methods to estimate γ_{it_1} and γ_{it_2}. Further,

$$E\left[\varepsilon_{it}\left|z_{it_1}, z_{it_2}d_{it_1}=1, d_{it_2}=1\right.\right] = \rho_{t_1}M_{12}+\rho_{t_2}M_{21} \tag{17}$$

where $M_{ij}=\left(1-\sigma_{12}\right)^{-1}\left(P_i-\sigma_{12}P_j\right)$ and

[11]In this analysis fixed effect however presents a significant limitation with the respect to pooled OLS : we can not assess the effect of variables that do not vary very much within group: i.e. degree of education, race, region, etc. that can impact significantly the health insurance reimbursement. Also, explanatory variables whose change across time is constant – e.g. age – can not be included.

$$P_j = \frac{\int_{-\infty}^{z_{it_1}\gamma_{t_1}} \int_{-\infty}^{z_{it_2}\gamma_{t_2}} v_{it_1} v_{it_2}\, f\big(v_{it_1}, v_{it_2}\big) dv_{it_1} dv_{it_2}}{F\big(z_{it_1}\gamma_{t_1}, z_{it_2}\gamma_{t_2}\big)}.^{12}$$

4.1.1 Bivariate Probit Model for Care Expenditure and Insurance

To test whether v_{it_1} and v_{it_2} are correlated we run for each year a "preliminary" bivariate probit between insurance and health care expenditure participation. In our model the dependent variable employed to predict the probability of facing positive health care expenditure is a binary variable that takes value one if individuals incur in positive health care expenditure during the year of interview, and zero otherwise.

The independent variables employed can been categorized into three dimensions: need for care (need to see a specialist or have treatments or tests and an indicator of health status[13]), predisposition to use health services (age, sex, marital status, race) and enabling factors (education, insurance, income, employment status, region and residential location[14]). Among enabling factor, we consider insurance participation. An insured individual, in fact, may consume more medical services and have a greater expenditure compared to an uninsured one (i.e.*moral hazard effect*). (Arrow, 1963; Pauly, 1968; Dowd et al.,1991). In this study, the situation is further complicated by the fact that insurance participation itself may be affected by the likelihood of having positive health expenditure. The choice of insurance coverage may be affected by planned medical expenditure and expectations about medical care utilization (i.e. *adverse selection effect*).

[12]There is only one cross-sectional example in the literature that is due to Fishe et. al.(1981). They estimated the selection equations by bivariate probit method and evaluated the above expression by numerical methods.

[13]We adopt as indicator of health status the objective measure of health that is "hypertension" since it seems to work better.

[14]The variables MSA (Metropolitan Statistical Area) and the indicators of regions control for medical cost differences between metropolitan and no-metropolitan statistical area, as well as by region of the country.

To test the potential endogeneity of health insurance and at the same time whether the covariance between health insurance choice and health expenditure participation is significantly different of zero, we run for each year a cross sectional recursive bivariate probit models (Maddala, 1999).

For each period, the recursive structure builds on a first reduced form equation for the potentially endogenous dummy measuring insurance participation and a second structural form equation determining the expenditure participation:

$$d^*_{it_1} = z_{i1_1}\gamma_{1_1} + ... + z_{it_1}\gamma_{t_1} + v_{it_1} \tag{18}$$

$$\begin{aligned} d^*_{it_2} &= z_{i1_2}\gamma_{1_2} + ... + z_{it_2}\gamma_{t_2} + v_{it_2} = \\ &= z_{i1_2}\gamma_{1_2} + ... + d_{it_1}\zeta + w_{it}\xi + v_{it_2} \end{aligned} \tag{19}$$

where $d^*_{it_1}$ and $d^*_{it_2}$ are latent variables, and d_{it_1} and d_{it_2} are dichotomous variables observed according to the rule:

$$\begin{cases} d_{it_j} = 0 \quad if \quad d^*_{it_j} \leq 0 \\ d_{it_j} = 1 \quad if \quad d^*_{it_j} > 0 \end{cases} ; \quad j = 1, 2 \tag{20}$$

z_{it_1} , the lags of z_{it_j} and w_{it} are vector of exogenous variables, γ and ξ are parameter vectors, ζ is a scalar parameter. The dependent variable d_{it_1} used to predict the probability of being insured is again a dummy variable that takes value one if respondents are insured and zero otherwise. The vector of explanatory variables z_{it_1} used to predict the probability of being insured includes both exogenous variables that are determinants of health expenditure and personal attributes that are only determinative of health insurance choice[15] (i.e. risk aversion) . We assume that, for each period, the error

[15]Estimation of a recursive bivariate probit model requires some considerations for the identification of the model parameters: at least one of the insurance equation exogenous variables has not to be included in the expenditure equation as explanatory variable (Maddala, 1983). Following Maddala's approach we include among explanatory variables in the

terms v_{it_1} and v_{it_2} are distributed as bivariate normal, with zero mean and variance covariance matrix Σ. Σ has values of 1 on the leading diagonal and correlations $\rho_{12} = \rho_{21}$ as off-diagonal elements:

$$\begin{pmatrix} v_{it_1} \\ v_{it_2} \end{pmatrix} \sim IIDN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix}\right) \tag{21}$$

In the above setting, the exogeneity condition is stated in terms of the correlation coefficient , which can be interpreted as the correlation between the unobservable explanatory variables of the two different equations. The two selection equations can be estimated separately as single probit models only in the case of independent error terms v_{it_1} and v_{it_2} i.e. the coefficient ρ_{jk} is not significantly different of zero ($k = 1, 2$). If the error terms v_{it_1} and v_{it_2} are independent we can deal with the above model as independent equations (Maddala, 1983) and apply the model in the equation (16)[16].

Table 3 shows the correlation coefficients and the p-value for each year sample: the null hypothesis of $Cov\left(v_{it_1}, v_{it_2}\right) = 0$ is not rejected; hence, we can deal with the model in the equation (16) and compute Inverse Mills Ratio by using the two selection equations as single probit models.

Tables 4 and 5 show coefficients for insurance choice and expenditure participation equation estimated using bivariate probit specification. Our findings do not support adverse selection in insurance choice: no unobservable that affect the health care expenditure significantly affect insurance participation, while being insured has a positive influence on the probability

insurance equation a measure of risk aversion assuming that risk aversion has direct effect on insurance choice while it has only an indirect effect on health care expenditure through insurance participation. In addition we esclude from insurance participation equation "need for care" variable to avoid causality problems with the dependent variable.

[16] The estimation of the model is carried out using STATA 9 software by which it is possible to run a bivariate probit with the command **biprobit**. STATA provides the statistics $z = \dfrac{\hat{\rho}}{S_{\hat{\rho}}}$ to test the hypothesis $H : \rho = 0$. If the error terms are independent the bivariate probit estimation is equivalent to the separate probit estimations.

of facing positive health care expenditure [17]. It is worth noting that while socioeconomic variables influence the probability of being insured they do not impact significantly the probability of positive expenditure.

4.2 Structural Equation Estimation

Tables 6, 7 and 8 show coefficients for the structural insurance reimbursement equation estimated using pooled OLS specification. The test for completeness of insurance coverage purchased by high risk profile individuals includes three pooled OLS models each of which contains a different measure of risk: a subjective measure of health (self-assessed health), an objective measure of health (hypertension) and independent variables that measure the individual life-style[18]. In each model the dependent variable is the natural logarithm of health insurance reimbursement.

We find a little evidence for adverse selection: table 6 shows that the coefficient estimate for the variable "health" is negative but is not statistically significant. Life-style variables do not influence the choice of health plan with exception of the variable "exercise" that, however, presents a positive coefficient. Table 7 shows that the variable that measures whether individuals suffer from high blood pressure is positively and significantly correlated with the health insurance reimbursement. The reason of this positive correlation may be found in the fact that more than half of all hypertensive Americans are covered by Medicaid or Medicare. Medicare and Medicaid are essentially universal health insurance programs for this segment of the population, however, these type of programs present a number of gaps in coverage: for instance, despite Medicare and Medicaid have a prescription

[17] We have tested for multicollinearity in both probit models (health care expenditure and insurance model) by using the Variance Inflation Factor (VIF) and Tolerance(1/VIF)(Wooldridge, 2000). We find that VIF for al the independent variables in boh the equations are quite low. Therefore, we can safely assume that there are no problems of multicollinerity.

[18] We have constructed three different sub-models since the three measures of risk are strongly correlated and may generate problems of multicollinearity.

drug benefit, often face restrictions in the number of covered medications. Since this restrictions, many persons will exceed the initial drug benefit cap and may remain at risk for inadequate blood pressure control. (Duru et al., 2007). Hence, many hypertensives are forced to buy own supplemental insurance coverage which offers hypertensive prescription drugs; normally, that plans are more complete then the others that do not provide or provide less generous prescription drug benefits.

As expected the variable that measures whether respondents suffer from disabilities which limit their activities presents a positive and significant coefficients. The variable that measures individual risk aversion presents a positive sign but the coefficient is not statistically significant.

Concerning the variables that measure total income, education degree and employment status the parameter estimates have the expected sign, but only the parameter for educational degree that is statistically significant. In the empirical literature we can observe that higher educational degree is often associated to a better health status; in particular it seems that education improves indirectly health status helping people choose healthier life-styles by improving their knowledge of the relationship between health behaviors and health outcomes (Kenkel, 1991). Then, people with more schooling tend to choose less complete insurance plans since they tend to enjoy good health.

Other than regular variables as independent variables two independent variables here are the IMR (Inverse Mills Ratio) which have been estimated from the first and second probit equation. When added to the outcome equation as additional regressors, they measure the sample selection effect due to lack of observations on the non-health insurance purchasers and non-health expenditure participants. These variables should be statistically significant to justify the use of Wooldridge two-step estimation. Since in our models they are statistically significant there may be sample selection problem in the data and we need to use Wooldridge method (Bath and Jain, 2006).

5 Summary and Conclusions

We have used the 2003/2004 Medical Expenditure Panel Survey in conjunctions with the 2002 National Health Interview Survey to asses whether US health insurance market is affected by adverse selection. We have conducted a positive correlation test which estimates the correlation between the amount of insurance an individual buys and his ex-post risk experience. We have employed three measures of risk: perceived health status, blood pressure and individuals' life-style. In addition, we have controlled for a number of enrollee characteristics including age, sex, race, education and family size which are used in pricing insurance policies. As indicator of generosity and completeness of health plan, we have employed health care expenditure reimbursement which measures the vertical difference between total health care expenditure and out-of-pocket expenditure on health care paid by consumers. Since health insurance reimbursement is only defined for those who participate in insurance and have positive health care expenditure the model is estimated using Wooldridge's (1995) two step estimation procedure. We have extended this method to the case in which selectivity is due to two sources rather than one.

The evidence for adverse selection seems to be lacking. Our findings do not support the existence of a systematic relation between illness of individuals and insurance choice. There is no separating equilibrium: high risk individuals do not purchase more complete insurance than low risk profile individuals.

The absence of correlation between individuals' risk- profile and completeness of health insurance can be explained by the fact that individuals may choose a health insurance plan based not only on their expected health status but also on their preferences such as the geographic location, whether they can continue to see doctors with whom they have already established relationships, whether friends recommended plans etc. If such preferences exert sufficient influence, risk-based selection is a minor consideration; as they become less important, adverse selection increases.

Arguably, another explanation for these results may be found in health plans risk selection practise. The distribution of health expenditure is highly skewed. Only a small fraction of individuals account for most of nations' health care spending. Because of this, insurers may have a strong incentives to distort their offering to avoid enrollment of high cost individuals. Then, insurers may practice a kind of "reverse adverse selection": they would try attempt to increase their profits by refusing to write policies for the worst risks in an insurance pool (see Siegelman, 2004). These strategic behavior can take a variety of forms including: designing insurance benefits packages in such a way as to be more attractive to healthy persons than unhealthy one for instance by excluding particular prescription drugs, offering numerous pediatrician (families with children are better risks) or by excluding cancer specialist visits. In such cases health plan may also refuse to sell an applicant insurance altogether. If health plans cream healthy individuals, those who are enrolled in health insurance are relatively healthy people and the correlation between risk- profile and the generosity of health insurance plans becomes insignificant.

Cream skimming may be very dangerous for the society as whole. The larger will be profits resulting from cream skimming and the greater will be the incentive for health plans to repel the sick. Plan may also reject innovation that improve quality of the health care if they attract the "wrong people" since cream the "wrong people" may be more profitable than improving efficiency (Cutler, Zeckhouser). On the one hand, if health plans do not repeal the sick and specialize in care for high risk, they have to ask a higher premium; however if regulation imposes a nation- wide maximum premium, health plans that attract sicker people may go bankrupt. On the other hand, health plans which cream tend to give poor services to the chronically ill and choose not to contract with physicians or hospital to specialize themselves in treatment in chronic illness. Thus, sick individuals may not receive treatments that respond to their need.

The problem of adverse selection is generally acknowledged to plague the market for health insurance also leading to an increase in competition to insure the lower than average risk consumers. Government may respond to this dangerous inefficiency caused by adverse selection: for instance, it can forbid risk selection by, for, example requiring open enrolment. However, other subtle forms of risk selection may exist and may be hard to eradicate them.

References

[1] G.Akerlof (1970), “The Market for ’Lemons’: Quality Uncertainty and the Market Mechanism”, Quarterly Journal of Economics, **84**: 488-500.

[2] K.J.Arrow (1985), ”The Economics of Agency,” in John W. Pratt and Richard J.Zeckhauser (Eds.), Principals and Agents: The Structure of Business, Boston, MA: Harvard Business School Press: 37-51.

[3] P. Bajari, H.Hong, Khwaja (2006), ”Moral Hazard, Adverse Selection and Health Expenditure: A Semiparametric Approach Analysis” National Bureau of Economics Research, Working Paper N.12445.

[4] M.Berger, J.P.Leigh, (1989), ”Schooling,Self-Selection, and Health”. *Journal of Human Resources*, **24** :433-455

[5] M.J. Browne, H.I.Doerpinghaus (1993), ”Information Asymmetries and Adverse Selection in the Market for Individual Medical Expense Insurance”, *The Journal of Risk and Insurance*, 60: 300-312.

[6] A.C. Cameron , P.K.Trivedi (1991), ”The Role of Income and Health Risk in the Choice of Health Insurance: Evidence from Australia”, *Journal of Public Economics*, **45**:1-28.

[7] J.H. Cardon; I.Hendel (2001), ”Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey”, *The RAND Journal of Economics*, **32**: 408-42.

[8] J.Cawley, T.Philiphoson, ”An Empirical Examination of Information Barriers to Trade Insurance” National Bureau of Economics Research,Working Paper N.5669.

[9] P.S.K. Chi, P.L. Hsin (1999), ”Medical Utilization and Health Expenditure of the Elderly in Taiwan”, *Journal of Family and Economic Issues*, **20**: 251-270.

[10] P.A Chiappori, B. Salanié (2000), "Testing Contract Theory: A Survey of Some Recent Work", invited lecture World Congress of the Econometric Society Seattle, August 2000.

[11] P.A Chiappori, B. Salanié (2000), "Testing for Asymmetric Information in Insurance Markets", *The Journal of Political Economy*, **108**: 56-78.

[12] P.Contoyannis, A. M. Jones, (2004), "Socio-Economic Status, Health and Lifestyle", *Journal of Health Economics*, **23**: 965-995.

[13] D.M.Cutler, R.J.Zeckhouser (1997), "Adverse Selection in Health Insurance", National Bureau of Economics Research, Working Paper N.6107.

[14] D.M.Cutler, R.J.Zeckhouser (2000)," The Anatomy of Health Insurance ", Handbook of Health Economics. A.J. Culyer and J. P. Newhouse. North Holland, Elsevier Science B.V. 1A: 563-643.

[15] D.M.Cutler, S.J.Reber (1998), "Paying for Health Insurance: the Trade-Off between Competition and Adverse Selection", *The Quartely Journal of Economics,* **113**: 433-466.

[16] G.Dionne, C.Gouriéroux, C. Vanasse (1998) "Evidence of Adverse Selection in Automobile Insurance Markets" École des Hautes Études Commerciales (HEC) Montréal, Working Paper 98-09.

[17] G.Dionne, C.Gouriéroux, C. Vanasse (2001), "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment", *The Journal of Political Economy*, **109**: 444-453.

[18] K.A.Desmond, T. Rice, P.D.Fox (2006), "Does Greater Medicare HMO Enrollment Cause Adverse Selection into Medigap?", *Health Economics, Policy and Law*, **1**: 3-21.

[19] O.KDuru, R.B.Vargas, D.Kermah, D.Pan, K.C.Norris (2007) "Health Insurance Status and Hypertension Monitoring and Control in the

United States, *American Journal of Hypertension*, **20**:348–353.

[20] R.P.Ellis (1998), "Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins", *Journal of Health Economics*, **17**: 537-555.

[21] A.Finkelstein, J.Poterba(2006) "Testing for Adverse Selection with "Unused Observables"", mimeo.

[22] D.Genesove(1993), "Adverse Selection in the Wholesale Used Car Market", *The Journal of Political Economy*, **101**: 644-665.

[23] A.M.Gill, R.J. Michels (1992), "Does Drug Use Lower Wage?", *Industrial and Labor Relation Review,* **45**: 419-434.

[24] M.Godfried , H. Oosterbeek , F.Van Tulder (2001), " Adverse Selection and the Demand for Supplementary Dental Insurance", *De Economist*, **149**: 177-190.

[25] Keeler, E. B., J. P. Newhouse, and C. E. Phelps (1977), "Deductibles and the Demand for Medical Care Services: The Theory of a Consumer Facing a Variable Price Schedule Under Uncertainty", *Econometrica*, **4**: 641-656.

[26] D.S. Kenkel (1994), "The Demand for Preventive Medical Care", *Applied Economics*, **26**: 313-325.

[27] C.Koç (2005), "Health-Specific Moral Hazard Effects", *Southern Economic Journal*, **72**: 98-118.

[28] G.S.Maddala (1983), "Limited Dependent and Qualitative Variables in Econometrics", Cambridge University Press.

[29] S.S.Makki.;A. Somwaru (2001), "Evidence of Adverse Selection in Crop Insurance Markets", *The Journal of Risk and Insurance*, **68**: 685-708.

[30] M.S.Marquis (1992), "Adverse Selection with Multiple Choice among Health Insurance Plans: a Simulation Analysis" *Journal of Health*

Economics, **11**:129-151.

[31] H.Miyazaki (1977), "The Rat Race and Internal Labor Markets," *Bell Journal of Economics* **8**: 394-418.

[32] J. P.Newhouse(1996), "Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection", *Journal of Economic Literature*, **34**:1236-1263.

[33] P. Olivella, M.V.Hernández, "Testing for Adverse Selection into Private Medical Insurance" , The Institute for Fiscal Studies, Working Paper N. 06/02.

[34] M. V.Pauly (1974). "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection." *Quarterly Journal of Economics* **88**: 44-62.

[35] M.V.Pauly, S.Nicholson (1999), "Adverse Consequences of Adverse Selection", *Journal of Health Politics, Policy and Law*, **24**: 921-930.

[36] M.V.Pauly (2007), The Truth about Moral Hazard and Adverse Selection in Health Insurance " Center for Policy Research (CPR) at the Maxwell School of Syracuse University, Working Paper N.36/2007.

[37] P. A.Puhani (2000), "The Heckman Correction for Sample Selection and Its Critique", *Journal of Economic Survey*, **14**: 53-68.

[38] M.Rochina-Barrachina (1999) "A New Estimator for Panel Data Sample Selection Models," Annales. d'Economie et de Statistique, 55/56.

[39] M.Rothschild, J.Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", *The Quarterly Journal of Economics*, **90**: 629-649.

[40] P.Siegelman (2004), "Adverse Selection in Insurance Markets: An Exaggerated Threat", *The Yale Law Journal*, **113**: 1223-1281.

[41] F.Vella (1998), "Estimating Models with Sample Selection Bias: A Survey", *The Journal of Human Resources*, **33**: 127-169.

[42] W.P.M.M. van de Ven , R.P. Ellis (1999), "Risk Adjustment in Competition Health Plan Markets", Handbook of Health Economics. A.J. Culyer and J. P. Newhouse. North Holland, Elsevier Science Chapter 17.

[43] J.Wooldridge (1995), "Selection Correction for Panel Data Models under Conditional Mean Independence Assumptions", *Journal of Econometrics*, **68**: 115-132.

6 Appendix

Table 1: Variables Name and Definition

Variables Name	*Variables Definition*
age	age in years
male	1 if male, 0 otherwise
white	1 if white, 0 otherwise
black	1 if black, 0 otherwise
other_race	1 if other race, 0 otherwise
northeast	1 if lives in Northeast region, 0 otherwise
midweast	1 if lives in Midweast region, 0 otherwise
west	1 if lives in West region, 0 otherwise
south	1 if lives in South region, 0 otherwise
msa	1 if lives in Metropolitan Statistical Area, 0 otherwise
income	total annual income
employed	1 if employed, 0 otherwise
education	1 if had high_school, master or PhD degree when entered in MEPS, 0 otherwise
expenditure	total annual health care expenditure
reimbursement	total annual health care expenditure paid by insurance
family size	family size
married	1 if married, 0 otherwise
health	1 if current health is excellent, very good, good, 0 otherwise
activity limitations	1 if has limited in any activities because health problems, 0 otherwise
hypertention	1 if suffers from high blood pressure, 0 otherwise
smoke	1 if is current smoker, 0 otherwise
alcohol	1 ifcorrent consumes heavy alcohol, 0 otherwise
execise	1 if participates in vigorous physical activity at least once at week, 0 otherwise
obese	body max index >= 25.0000
need care	1 if needs for care during the year of interview, 0 otherwise
insured	1 if insured, 0 otherwise
risk aversion	1 if is not likely to take risk, 0 otherwise
mills1	mills ratio insurance partecipation
mills2	mills ratiohealth care expenditure partecipation

Table 2: Summary Statistics

	All	Insured	Uninsured
Age	48.18	48.58	43,30
Male	0.317	0.316	0.331
Income	38,062.86	39,924.13	15,563.32
Total health care expenditure	4,120.202	4,295.44	2,001.882
Annual premium		1,736.688	
Northeast	0.168	0.177	0.066
South	0.351	0.341	0.471
West	0.203	0.196	0.279
Midwest	0.278	0.285	0.184
White	0.859	0.869	0.728
Black	0.092	0.085	0.184
Other Race	0.049	0.046	0.088
Metropolitan statistical area	0.806	0.818	0.669
Health status	0.892	0.897	0.824
Hypertension	0.262	0.265	0.221
Activity limitations	0.318	0.314	0.360
Smoke	0.167	0.159	0.272
Alcohol	0.056	0.047	0.169
Bmi	27.44	26.96	33.18
Exercise	0.479	0.493	0.309
Risk aversion	0.788	0.799	0.662
Number of observations	1780	1644	136

Table 3: Bivariate Probit Correlation Coefficients

Dependent Variables	*pho*	*p-value*
Positive Expenditure/ Be Insured 2003	-0.5299	0.260
Positive Expenditure/ Be Insured 2004	-0.9496	0.541

Table 4: Cross-Sectional Bivariate Probit Estimation Coefficients
(p-value in parentheses)

	Expenditure 2003	Be Insured 2003
intercept	0.9643 (0.282)	
age	0.0168 (0.183)	0.0151 (0.014)
male	-0.9162 (0.000)	-0.1509 (0.357)
black	-0.2153 (0.503)	-0.5149 (0.018)
other_race	-0.1419 (0.769)	-0.3913 (0.190)
family size	-0.2158 (0.011)	0.0514 (0.447)
msa	-0.1917 (0.529)	0.4337 (0.010)
northeast	0.2356 (0.482)	0.4700 (0.097)
midwest	0.7171 (0.066)	0.0817 (0.676)
west	1.0350 (0.025)	-0.3043 (0.123)
insured	2.0831 (0.010)	
income	2.48e-06(0.608)	0.0000 (0.000)
employed	-0.9426 (0.073)	0.2692 (0.169)
education	-0.1600 (0.685)	0.7078 (0.000)
married	0.0911 (0.749)	0.4618 (0.007)
need care	0.1413 (0.361)	
hypertension	-0.0309 (0.922)	0.1954 (0.312)
activity limit.	-0.0227 (0.945)	0.0006 (0.997)
risk aversion		0.2727 (0.125)

Note: sample size 890; statistically significant at the 0.05 level.

Table 5: Cross-Sectional Bivariate Probit Estimation Coefficients
(p-value in parentheses)

	Expenditure 2004	Be Insured 2004
intercept	1.5073 (0.090)	
age	0.0040 (0.716)	0.0177 (0.003)
male	-1.064 (0.000)	-0.0717 (0.655)
black	-0.4799 (0.235)	-0.4255 (0.068)
other_race	-0.0838 (0.868)	-0.3941 (0.176)
family size	-0.2521 (0.005)	0.0495 (0.530)
msa	-0.4831 (0.126)	0.3653 (0.027)
northeast	-0.1615 (0.706)	0.4407 (0.094)
midwest	-0.3098 (0.702)	0.2263 (0.247)
west	-0.3386 (0.288)	-0.2170 (0.260)
insured	2.0491 (0.005)	
income	8.62e-06 (0.045)	0.0000 (0.000)
employed	-0.6465 (0.100)	0.0262 (0.892)
education	-0.1996 (0.690)	0.5928 (0.002)
married	0.1268 (0.693)	0.3238 (0.074)
need care	0.2761 (0.059)	
hypertension	0.6086 (0.150)	0.0780 (0.683)
activity limit.	0.2282 (0.516)	0.0733 (0.665)
risk aversion		0.2129 (0.208)

Note: sample size 890; statistically significant at the 0.05 level.

Table 6: Pooled OLS Regression Results.

Risk Variable: Self-Assessed Health.

Preidictor Variables	*Coefficients*	*p-values*
intercept	6.710288	0.000
age	0.0021	0.479
male	-0.1328	0.111
married	-0.0544	0.452
black	-0.0849	0.505
other race	0.0123	0.932
education	-0.4279	0.002
income	-2.33e-06	0.078
employed	-0.0418	0.671
premium	-0.0000	0.004
expenditure	0.0001	0.000
activity limitations	0.2518	0.001
health	-0.1613	0.162
risk aversion	0.1370	0.113
mills1	-2.2891	0.000
mills2	-1.1936	0.001

Note: sample size 1613; $R^2 = 0.4239$; Adjusted $R^2 = 0.4185$; statistically significant at the 0.05 level.

Table 7: Pooled OLS Regression Results.

Risk Variable: Hypertension

Preidictor Variables	*Coefficients*	*p-values*
intercept	6.5331	0.000
age	0.004	0.887
male	-0.1535	0.066
married	-0.0442	0.541
black	-0.1352	0.292
other race	0.0114	0.944
education	-0.4367	0.001
income	-2.09e-06	0.115
employed	-0.0185	0.851
premium	-0.0000	0.008
expenditure	0.0001	0.000
activity limitations	0.2405	0.001
hypertension	0.2514	0.002
risk aversion	0.1379	0.109
mills1	-2.1702	0.000
mills2	-1.0837	0.004

Note: sample size 1613; R^2 =0.4265; Adjusted R^2 = 0.4212; statistically significant at the 0.05 level.

Table 8: Pooled OLS Regression Results.

Risk Variable: Life-Style Indicators

Preidictor Variables	*Coefficients*	*p-values*
intercept	6.3921	0.000
age	0.0034	0.260
male	-0.1537	0.069
married	-0.619	0.393
black	-0.0529	0.680
other race	0.0507	0.752
education	-0.4623	0.001
income	-2.52e-06	0.058
employed	-0.7782	0.436
premium	-0.0000	0.006
expenditure	0.0001	0.000
activity limitations	0.2672	0.001
smoke	0.1384	0.137
obese	0.0393	0.570
alcohol	-0.2403	0.118
exercise	0.1982	0.004
risk aversion	0.1458	0.091
mills1	-2.3287	0.000
mills2	-1.1903	0.001

Note: sample size 1613; $R^2 = 0.4277$; Adjusted $R^2 = 0.4212$; statistically significant at the 0.05 level.