Why Do The Insured Use More Health Care? *

Yingying Dong
Department of Economics
Boston College

Original: July 2007, Revised: August 2009

Abstract

Why do the insured use more health care? One obvious reason is that health care is cheaper for the insured. But additionally, having insurance can encourage unhealthy behavior via moral hazard. The effect of health insurance on medical utilization has been extensively studied; however, previous work has mostly ignored the effect of insurance on behavior and how that in turn affects medical utilization. This paper investigates these distinct effects. The increased medical utilization due to reduced prices may help the insured maintain good health, while that due to increased unhealthy behavior does not, so distinguishing these two effects has important policy implications. A two-period dynamic forward-looking model is used to derive the structural causal relationships among the decision to buy insurance, health behaviors (drinking, smoking, and exercise), and medical utilization. The model shows how exogenous changes in insurance prices and past behaviors can identify the direct and indirect effects of insurance on medical utilization. An empirical analysis also distinguishes between intensive and extensive margins (e.g., changes in the number of drinkers vs. the amount of alcohol consumed) of the insurance effect. Health insurance is found to encourage less healthy behavior, particularly heavy drinking, but this does not yield an immediate perceptible increase in doctor or hospital visits. The effects of health insurance are primarily found at the intensive margin, e.g., health insurance may not cause a non-drinker to take up drinking, while it encourages a heavy drinker to drink even more.

JEL Codes: C34, D12, I12

Keywords: Health insurance, Health care utilization, Health behavior, Structural model, Identification, Selection effect, Moral Hazard

*I would like to thank Arthur Lewbel, Peter Gottschalk, Shannon Seitz, Zhijie Xiao, and Donald Cox for their very helpful comments and suggestions. This paper has also benefited from comments from the participants at the 6th World Congress on Health Economics in Copenhagen, the Priester National Extension Health Conference, and seminars at Boston College, SUNY Albany Department of Public Administration and Policy, the Workers Compensation Research Institute (WCRI), IUPUI, CUNY Hunter College, UC Riverside, Boston FED, and Urban Institute. Any errors are my own.
1 Introduction

There is a large literature examining the effect of health insurance on health care. This literature generally shows that health insurance is associated with an increased use of health care, though the direction of causality and the ultimate effect on health are questionable.

Traditionally, researchers focus on two scenarios: First, individuals buying health insurance are likely to be those who anticipate greater need of health care. This is frequently referred to as adverse selection. Statistically, this means that insured individuals are a non-randomly selected sample of the population. They have observed and possibly unobserved characteristics that are correlated with a higher demand for medical care. Second, health insurance reduces the effective price of health care, so other things equal, the insured tend to use more health care. For example, individuals who are just indifferent between using and not using a certain medical service at uninsured rates will tend to use it if they have insurance. This is a direct price effect.

However, there is a third dimension to this issue: due to moral hazard, individuals once insured may become less cautious about their unhealthy or risky behaviors, which lead to more health problems, requiring more health care.\(^1\) That is, individuals may respond to their insurance status and change behaviors, and thereby consume more health care.

This paper takes a structural point of view and investigates the direct and indirect health insurance effects on care utilization, controlling for the selection effect. The adverse selection effect is a causal relationship running from potentially higher health care utilization to health insurance; whereas both the indirect effect that works through health-related behaviors and the direct price effect are causalities running from health insurance to higher care utilization. Even if the adverse selection is accounted for by econometric techniques or experimental designs, the latter two effects can not be separated without looking at the simultaneous relationships among health insurance, behaviors, and health care utilization.

Intuitively, health insurance may encourage individuals to engage in unhealthy behaviors, such as heavy drinking, because it lowers the offsetting cost of the negative health consequences associated with these unhealthy behaviors and may even make otherwise unaffordable medical care accessible. In addition, individuals with chronic

\(^1\)Commonly studied unhealthy behaviors are: drinking, smoking, inadequate physical activity, and unhealthy diet. This paper focuses on the first three types of unhealthy behavior, because there is no diet information available in the data set used.
diseases are more likely to rely on medication instead of behavioral improvement once the medication becomes cheaper. This is especially true when considering the emotional costs that behavioral improvement requires. Further, more accessible health care or treatments may distort the perceived risk of unhealthy activities, which can lead to less cautious behavior.

Some may argue that moral hazard may not be important in the health insurance context, because it only reduces the financial cost of illness, while health is irreplaceable. However, as Dave and Kaestner (2006) point out, this rationale fails to explain findings in other insurance contexts involving adverse health consequences. For example, they note an increase in car accidents when car insurance is more generous, and an increase in workplace injuries associated with increases in workers’ injury compensation.

The disincentive effect of health insurance on individuals’ healthy behaviors is defined as “behavioral moral hazard” in this paper. In contrast, “direct price effect” refers to health insurance lowering the price of medical care and hence inducing individuals to use more care ceteris paribus. Although both may lead to increased health care utilization, their ultimate effects on health are different. In the former case, the increased use of health care does not necessarily translate into better health because individuals actually need more health care; whereas in the latter case, increased health care is more likely to improve health.

An immediate policy implication is that mandating insurance coverage to improve a targeted population’s health status may not be fully efficient. The efficiency in part depends on how much individuals substitute medication for behavioral improvement. For example, Klick and Stratmann (2006) examine the effect of mandates in some states that required health insurance providers to cover diabetes treatment without increasing premiums. Their study shows that these mandates generate strong disincentives for individuals’ behavioral prevention and systematically increase the diabetics’ BMI in the affected states, which is taken as a result of engaging in worse diet and exercise practice. Therefore, distinguishing the two effects may further shed light on policy

---

2 These two terms are related to but not quite the same as the existing ex ante and ex post moral hazard concepts in health economics. Ex ante moral hazard refers to the moral hazard that takes place before a sickness episode, which is further classified into self-insurance (demand for preventive care) and self-protection (exercising, abstention from smoking etc). Ex post moral hazard refers to moral hazard that takes place after a sickness episode, i.e., using more health care when one gets sick. Here the focus is on distinguishing two different ways that health insurance may cause increased demand of medical care and exploring their relative roles. I consider the direct price effect as a rational economic behavior, i.e., individuals’ natural response to decreased prices of health care. It can be predicted ex ante by examining the price elasticity of care.
relevant parameters.

Distinguishing the two effects may also help explain a puzzle in the literature: while a lot of research shows that health insurance results in increased use of health care, there is little evidence that having heath insurance leads to improved health (Haas et al., 1993a, b, Perry and Rosen, 2001). Findings from the RAND Health Insurance Experiment (Newhouse, 1993) show that those for whom health care was free used about 40% more health services than those who had some cost sharing, but this resulted in “little or no measurable effect on health status for the average adult.”

This paper sets up a two-period dynamic forward-looking model, which incorporates rational addiction (Becker and Murphy, 1988), uncertainty, and marginal utility of consumption depending on health status. From the theoretical model, this paper derives the structural causal relationships among health behaviors (drinking, smoking, and exercise), medical utilization, and the decision to buy health insurance. The structural equations show that exogenous changes in the price of health insurance and past behaviors can help identify the direct and indirect effects of health insurance on medical utilization. Semi-reduced form equations are used in the empirical analysis. It is also shown that the structural parameter of interest, the direct and indirect effects of health insurance, can be recovered from the reduced-from parameters.

The data used in this study are from the US Health and Retirement Study (HRS). The HRS is a panel data set, so one can control for past health behaviors, which is important because of state dependence or the addictiveness of health behaviors. The empirical analysis adopts the generalized Tobit or sample selection specification with transformations on the dependent and lagged dependent variables. Such a specification deals with typical features of data on medical utilization and unhealthy behaviors, i.e., a large proportion of zero observations and a highly skewed distribution of positive observations. It also separately accounts for the extensive margin (changes in the percentage of individuals who participate in unhealthy behaviors) and the intensive margin (changes in the quantity of unhealthy behaviors by participants) of the insurance effect, which turns out to be empirically important.

Health insurance is found to encourage less healthy behavior, particularly heavy drinking, but this does not yield a short term perceptible increase in doctor or hospital visits. The effects of health insurance are primarily found at the intensive margin, e.g., health insurance may not cause a non-drinker to take up drinking, while it encourages a heavy drinker to drink even more. These results suggest that to counteract the behavioral moral hazard, health insurance should be coupled with incentives that target individuals
who currently engage in unhealthy behaviors, such as heavy drinkers.

The rest of the paper proceeds as follows. Section 2 reviews the literature. Section 3 sets up the theoretical model. Section 4 describes the features of data on medical utilization and unhealthy behaviors, as well as discusses the empirical specification. Section 5 describes the sample and addresses identification of the effects of health insurance. Empirical results are reported in Section 6, and concluding remarks are provided in Section 7.

2 Literature Review

There are a large number of studies examining the insurance effect on health care utilization (See Zweifel and Manning, 2000 and Buchmueller et al, 2005 for surveys on this literature). They do not consider the potential effect of health insurance on individuals’ health behaviors and how that further affects medical utilization. Below is an illustration of (1) the focus of this literature versus (2) the focus here.

A small strand of literature examines *ex ante* moral hazard, whereas the majority examines the effect of health insurance coverage on the receipt of preventative care, such as mammography, and prostate or cholesterol screening. These studies include Roddy et al. (1986), Lillard et al. (1986), Keeler and Rolph (1988), Cherkin et al. (1990), McWilliams et al. (2003), and Decker (2005). In addition, Kenkel (2000) examines the effect of health insurance on the use of preventive care as well as on health behaviors. Based on logit model estimation, Kenkel’s study suggests that people with private health insurance are more likely to engage in health promoting behaviors than those without insurance. However, the author points out that these results may be biased if insurance status is endogenous to health practices. Courbage and Coulon (2004) examine whether purchasing additional private health insurance modifies the probability of exercising,
smoking and undergoing regular check-ups in the UK.\textsuperscript{3} Using Probit and instrumental variable estimation, they find that having additional private insurance may in fact lead to healthier choices.

Card et al. (2004) separately examine the effect of Medicare eligibility on a large range of mostly discrete outcomes: usage of medical procedures, smoking, exercise, as well as self-reported health, obesity, and mortality rates. Their study exploits the exogenous increase in health insurance coverage at age 65, which is the Medicare eligible age. They show that eligibility for Medicare has a significant impact on health care utilization and a discernable effect on self-reported health, though reaching age 65 has no systematic effect on mortality rate and on probabilities of smoking, exercising and being obese.

Khwaja (2002, 2006) looks at individuals' health insurance decisions, health habits, and care utilization simultaneously, but does not consider the magnitude of the direct price effect versus the indirect behavioral moral hazard effect within the total effect of health insurance. Using a dynamic programming approach, Khwaja (2002) estimates dynamic discrete choice models and conducts simulations based estimated parameters. His study concludes that insurance coverage causes insignificant moral hazard in probabilities of having habits like smoking and drinking, and significant moral hazard in the probability of seeking medical treatment. Using the same approach, Khwaja (2006) finds that Medicare generates a low level of moral hazard in the probabilities of alcohol consumption, smoking and exercise among the elderly.

To summarize, most existing studies consider the insurance effect either on medical utilization or on health behaviors, but do not look at the simultaneous structural relationships among health insurance, behaviors and health care utilization. A few studies that do consider all three do not separate the direct and indirect insurance effects, which have important policy implications as discussed earlier. These studies also focus on discrete outcomes, i.e., they examine the insurance effect on the probability of having an unhealthy habit rather than the quantity. These studies find that health insurance does not have a significant effect on the probability of participating in unhealthy behavior (extensive margin); however, it is reasonable to believe that, given participation, health insurance may have a nonnegligible effect on the quantity of unhealthy behavior (intensive margin). For example, health insurance may not induce a non-smoker to become a smoker, while it is likely to affect how much a smoker smokes. Looking at discrete outcomes fails to identify the empirically important intensive margin of the insurance effect,

\textsuperscript{3}Public health insurance is provided to all residents in the UK.
which may lead to misleading conclusions about the behavioral moral hazard effect.

In contrast to the previous literature, this paper presents a dynamic continuous choice model and derives the structural causal relationships among health-related behaviors like drinking, health care utilization, and the decision to buy health insurance. This allows identification of both the direct price and indirect moral hazard effects of health insurance on medical utilization. Using quantitative instead of just qualitative data, this paper determines both the extensive and intensive margins of insurance effects. Distinguishing the price and behavioral moral hazard effects as well as the extensive and intensive margins provide more policy relevant implications than existing studies. For example, if health insurance makes drinkers drink more, but does not cause people to take up drinking, then policies should target current drinkers.

3 Theoretical Model

This section sets up a theoretical model and derives the structural causal relationships for the insurance decision, health behaviors, and medical utilization. Semi-reduced form equations are then obtained for health behaviors and medical utilization. Based on the relationship between the structural equations and the semi-reduced form equations, the structural parameters of interest, the direct and indirect insurance effects, are recovered from the reduced-form parameters. To keep the discussion short, derivation details are provided in the Appendix.

3.1 The Basic Theoretical Model

As Grossman (1972) and many others have noted, consumers value health. Better health may improve the efficiency of other goods consumption, whereas health care is merely a means to producing health or slowing its decline. Therefore, this paper assumes that a typical individual draws utility from his health, composite goods consumption and unhealthy behaviors. The unhealthy behaviors under consideration include drinking, smoking and insufficient physical activity. Unhealthy behaviors are addictive. Besides utility, they also generate disutility through their harmful effects on health.

Here health is viewed as a durable consumption good with value that depreciates with age. An individual may invest in health using medical care and health behaviors, and the individual is exposed to stochastic health shocks. That is, the individual’s health
evolves according to
\[ H_{t+1} = \delta H_t + \bar{H}(M_t, \mathbf{B}_t, s_t), \]  
(1)

where \( H_t \) = initial health status in period \( t \); \( \bar{H} \) = health production function; \( \delta = 1 \) - health depreciation rate; \( M_t \) = medical care; \( \mathbf{B}_t = (B^1_t, B^2_t, B^3_t)' \), a vector of health behaviors with three elements representing alcohol drinking, smoking, and exercise, respectively; \( s_t \) = health shock.

Individual heterogeneity (e.g., different genes) is not included explicitly in health production, because it can not be distinguished from individual permanent taste or fixed effects in the utility function. Instead of having fixed effects in the health production function \( \bar{H} \), this study later includes fixed effects in the utility function.

The health shock \( s_t \) could be any acute diseases, such as a heart attack, or injury. \( s_t \) may depend on health behaviors \( \mathbf{B}_t \) and initial health status \( H_t \), which allows the incidence (probability of \( s_t \) greater than 0) and severity (expected value of \( s_t \)) of the health shock to be a function of \( \mathbf{B}_t \) and \( H_t \). For example, if the individual is in bad health or engages in unhealthy behaviors excessively, he will be more vulnerable to health stocks. By this construction, \( \mathbf{B}_t \) may have a short-term effect on health by inducing a health shock immediately and a long-term effect by reducing the next period health stock \( H_{t+1} \), regardless of the health shock. Health care \( M_t \) may mitigate the negative health effects of behaviors \( \mathbf{B}_t \) and those of the shock \( s_t \).

Given the potential addictiveness of unhealthy behaviors such as drinking and smoking, it is assumed that the marginal utility of current unhealthy behaviors \( \mathbf{B}_t \) depends on past unhealthy behaviors \( \mathbf{B}_{t-1} \) (Becker and Murphy, 1988; Becker, Grossman and Murphy, 1994). Therefore, both \( \mathbf{B}_t \) and \( \mathbf{B}_{t-1} \) enter the current period \( t \) utility function.

The individual is assumed to be forward-looking; i.e., he cares about the addictiveness and negative health effects of his current unhealthy behaviors. For simplicity, assume that the individual, characterized by a fixed taste parameter \( \tau \),\(^5\) lives for two periods, \( t \) and \( t + 1 \), representing the present and future, respectively. The individual invests in health in period \( t \), and bears the addiction and adverse health consequences in period \( t + 1 \). There is no health investment in the second period, as the individual is no longer living afterwards.

In addition to health shocks, there are taste shifters. Taste shifters are revealed to the

---

\(^4\)For convenience of notation, an individual subscript is dropped from all the equations. Throughout the paper, boldfaced letters are used to denote vectors or matrices, and regular letters to denote scalars.

\(^5\)Later \( \tau \) is parameterized as a function of some fundamental individual characteristics, such as race, gender, education etc. It can also be taken directly as a vector of individual characteristics.
individual at the beginning of each period, so the individual knows his exact preference when he makes decisions. Without loss of generality, it is assumed that health shocks come in the middle of each period.\footnote{Additional health shocks could also come at the beginning of a period, i.e., before the individual makes behavior decisions, and hence would affect these decisions. In my model these beginning of the period of health shocks would be indistinguishable from the taste shifter, so one could just take these health shocks to be one dimension of the taste shifter.} In particular, it is assumed that when the individual makes his health behavior decisions, health shocks are unknown, while when he makes his medical utilization decision, the health shock in that period has resolved.

The individual’s decision-making is a three-stage procedure in the two periods. At the beginning of period $t$, the individual decides whether to buy insurance. Assume that he knows his current preference $\nu_t$, health status $H_t$, and habits $B_{t-1}$. Uncertainty arises because the current period health shock $s_t$ and the future shocks, taste shifter $\nu_{t+1}$ and health shock $s_{t+1}$, are unknown. The individual makes his insurance decision based on expected utility maximization. Then with his insurance status determined, the individual chooses his health behaviors $B_t$, medical care $M_t$, and composite goods consumption $C_t$. At the beginning of period $t + 1$, the individual’s taste shifter $\nu_{t+1}$ is revealed to him, and he enjoys a health status $H_{t+1}$ determined by his initial health status $H_t$, first period health inputs $M_t$ and $B_t$, and health shock $s_t$. The individual decides health behaviors $B_{t+1}$ and composite goods consumption $C_{t+1}$.

For computational tractability and simplicity, assume that utility is additively separable over time and that in each period it has a quasi-linear representation. Further, assume that the time preference rate equals the interest rate $r$. The individual’s utility function can then be written as,

$$C_t + U(B_{t-1}, B_t, H_t, \tau, \nu_t, s_t) + \beta[C_{t+1} + U(B_t, B_{t+1}, H_{t+1}, \tau, \nu_{t+1}, s_{t+1})],$$

subject to the budget constraint

$$(C_t + P_B I_t + P_B' B_t + (1 - d I_t) M_t) + \beta(C_{t+1} + P_B' B_{t+1}) = W_t,$$

where $B_t, H_t,$ and $s_t$ are defined as in equation (1); $\beta = 1/(1 + r)$ is the discount factor; $C_t =$ composite goods consumption, which is a numeraire; $\tau =$ fixed taste parameter; $\nu_t =$ taste shifter; $W_t =$ present value of total wealth; $I_t =$ insurance dummy; $P_B =$ insurance premium; $P_B = (P_B^1, P_B^2, P_B^3)'$ is the price vector for health behaviors; $d =$ insurer co-payment rate, with $d \in (0, 1]$.

The quasi-linear utility assumption greatly simplifies the model solution by omitting wealth effects on the individual’s choices. Since wealth effects drop out, how wealth is...
determined has no impact here. Therefore, the total wealth $W_t$ is not explicitly modeled.\footnote{For a general utility function, having $W_t$ be a function of current and expected future behaviors and health, and hence a function of medical care $I_t$, would not change the ultimate functional forms of $B_t$, $M_t$ and $I_t$. However, past behavior could affect $W_t$, and thereby affect all the individual’s choices here. Since it could only do so through past incomes, or more precisely savings, in this case one could have savings in the eventual behavior and medical utilization equations, whereas all the other results that follow from this quasi-linear utility would still hold.} However, income will later be included as one dimension of individual characteristics, so in the empirical specification, choices will still depend on income.

The modeling here focuses on individuals’ choices, i.e., the demand side of the insurance market, while insurance premiums are determined by both demand and supply; therefore, I defer the modeling of insurance premiums to later. In particular, it will be assumed that, as a reduced form, insurance prices are a function of individuals’ age, health condition, habits, and other characteristics as well as unobserved supply side factors. Since health shocks can generate immediate disutility, $s_t$ is directly included in the utility function. For example, the individual suffers utility loss from health shocks, even though medical care may be used to prevent health loss from those shocks. This discourages the individual from drinking or smoking without a limit in the last period of life.

In a dynamic forward-looking model, an individual is assumed to take into account the effect of his current choices on his future course of actions and consistently maximizes his expected utility over time. The model can therefore be solved by backward induction. In this paper’s context, the individual’s optimal choices in the second period are derived first, conditioning on his first period choices, and then the results are used to derive his optimal choices in the first period, conditional on his insurance decision at the beginning of this period. Finally, the individual’s insurance decision is derived based on his expected optimal choices at the following stages.

Substituting the budget constraint into the utility function yields

$$U(B_{t-1}, B_t, H_t, \tau, \nu_t, s_t) + \beta U(B_t, B_{t+1}, H_{t+1}, \tau, \nu_{t+1}, s_{t+1})$$
$$+ W_t - [P_t I_t + P'B_t + (1 - dI_t) M_t + \beta P'B_{t+1}] = 0.$$  \(4\)

In the second period, the individual chooses $B_{t+1}$ to maximize his expected utility given by equation (4), conditioning on his then information set $\mathcal{F}_{t+1} = \{B_t, H_{t+1}, \tau, \nu_{t+1}\}$. Denote the solution to this maximization problem as $B^*_{t+1} \equiv B^*_{t+1} (B_t, H_{t+1}, \tau, \nu_{t+1})$, where $H_{t+1} = \delta H_t + \tilde{H}(M_t, B_t, s_t)$. By backward induction, in the first period, the individual chooses health behaviors $B_t$ and medical care utilization $M_t$ to maximize
his expected utility now defined by equation (4) after replacing \( B_{t+1} \) with \( B_{t+1}^* \), and conditioning on his information set at this stage, \( \mathcal{F}_t = \{ B_{t-1}, H_t, I_t, \tau, \nu_t \} \). The first order conditions for this maximization problem yield structural equations for \( M_t \) and \( B_t \).

At the beginning of period \( t \), the individual decides whether to buy insurance, depending on which choice gives the higher expected utility. Therefore, for both \( I_t = 1 \) and \( I_t = 0 \), substitute into the utility function, equation (4), the corresponding optimal choices at the following stages in a backward manner. This yields the stochastic utility functions in the two cases. Conditional on his information set \( \mathcal{F}_0 = \{ B_t, H_t, \tau, \nu_t \} \), the individual compares the expected utility in the two cases, and buys insurance if and only if the expected value of buying health insurance exceeds that of not buying.

These steps produce structural equations for the insurance decision, health behaviors, and health care utilization. The resulting system of structural equations can be expressed as

\[
M_t = f_M(I_t, B_t, H_t, \tau, \nu_t, e_t),
\]

\[
B_l^t = f_{B_l}(I_t, B_{t-1}, M_t, H_t, \tau, \nu_t) \text{ for } l = 1, 2, 3,
\]

\[
I_t = 1 \left[ f_I(B_{t-1}, H_t, \tau, \nu_t) - P_{lt} > 0 \right],
\]

where \( 1[\cdot] \) is the indicator function that equals one if the bracketed term is true, and zero otherwise; \( e_t \) is the random component of the health shock, i.e., \( e_t = s_t - E(s_t | H_t, B_t) \).

Let \( P_{lt} = f_I(B_{t-1}, H_t, \tau, \nu_t) \), which is the price that makes the individual just indifferent between buying and not buying insurance, namely, the individual’s willingness-to-pay for insurance. Equation (7) implies that the individual buys health insurance if and only if his willingness-to-pay for insurance is higher than the actual price he needs to pay.

The above system of equations shows that \( M_t \) depends on \( B_t \) and \( I_t \), and \( B_t \) further depends on \( I_t \), so health insurance \( I_t \) can have both direct and indirect effects on medical utilization \( M_t \). More importantly, these structural equations show exclusion restrictions for the endogenous variables in the structural \( M_t \) equation. First, \( B_{t-1} \) appears in the structural \( B_t \) equations, but not in the structural \( M_t \) equation. The intuition is that past unhealthy behaviors \( B_{t-1} \) affect \( M_t \) only through their effects on the current health \( H_t \) and current unhealthy behaviors \( B_t \). After conditioning on these two variables, \( B_{t-1} \) has no direct effect on medical utilization \( M_t \). Alternatively, if two individuals behave differently in the past, then this difference in \( B_{t-1} \) will be reflected either in \( H_t \) or in
$B_t$; i.e., these two measures summarize past behaviors. Second, $P_{It}$ shows up only in the insurance equation, so it provides a source of identification for the effects of health insurance. These exclusion restrictions together imply that the structural $M_t$ equation is identified. It follows that the structural parameters of interest, the direct and indirect effects of health insurance, are identified.

Equation (7) is the individual’s demand function of insurance, where the insurance price is endogenously determined. For simplicity, a reduced-form function is used for the insurance price. Given that insurance companies in general charge new customers or adjust their charges for existing customers based on their age, pre-existing conditions, and health habits, insurance premiums may vary with these factors. Also there is a big price change at age 65, because almost all individuals in the US can enroll in Medicare either free or at a low cost when they turn age 65 (a further discussion is given in section 5.2). That is, the insurance price can be expressed as

$$P_{It} = P_{It}(D_{65}^t, \text{Ex}_t, B_{t-1}, H_t, \tau) + \tilde{\nu}_t,$$

where $D_{65}^t$ is an age dummy indicating age 65 or above; $\text{Ex}_t$ represents other variables except for $B_{t-1}$, $H_t$, and individual taste $\tau$; $\tilde{\nu}_t$ is the reduced-form error that may capture, for example, unobserved supply side factors. By substituting equation (8) for the insurance price in (7), equation (7) can be rewritten as

$$I_t = 1 \left[ f_I(B_{t-1}, H_t, \tau, \nu_t) - P_{It}(D_{65}^t, \text{Ex}_t, B_{t-1}, H_t, \tau) - \tilde{\nu}_t > 0 \right].$$

In addition, solving the structural equations (5) and (6) for $B_t$ and $M_t$ yields semi-reduced form equations for these choice variables. These equations are semi-reduced forms because they depend on the endogenously determined insurance decision $I_t$ except for other exogenous and predetermined variables.

These equations provide a simplified conceptual framework that can be used to guide empirical analysis. For example, in theory, equation (5) would require a perfect measure for the true health status, including both the observed component and the unobserved component, which might not be feasible in practice. However, one may redefine the time dimension so that this assumption can plausibly hold. In particular, one could let $B_t$ be measures of current health behaviors, which are not necessary those in one year, and $B_{t-1}$ be summary measures of behavioral history that have a sufficiently long lag so that their effects are already reflected in $H_t$. For example, $B_t$ could be the average amount of unhealthy behaviors in the most recent three years, and $B_{t-1}$ be the averages for the three or more years further back in time.

The coefficients in equation (6), the structural $B_t$ equation, can not be identified. However, identifying the parameters of interest regarding insurance effects does not require identifying the structural $B_t$ equation.
3.2 An Illustration with the Quadratic Utility Function

To illustrate, consider a linear health production function, $\tilde{H}(M_t, B_t, s_t) = h_1 M_t + h_2 B_t + h_3 s_t$, and a quadratic form of the nonlinear term in the utility function, $U(B_{t-1}, B_t, H_t, \tau, \nu_t, s_t) = Z_t' \Pi Z_t$, where $Z_t = (B_{t-1}', B_t', H_t, H_t, \tau, \nu_t, s_t)'$, $\Pi = \{u_{j,k}\}$, $j, k = 1, 2, \ldots, 6$, is the coefficient (Hessian) matrix. The elements in the coefficient matrix are conformably defined; i.e., $u_{j,k}$ is a $p \times q$ matrix, where $p$ ($q$) is 3 when $j$ ($k$) = 1 or 2, and 1 otherwise.

For simplicity, further assume that $s_t$ depends on $H_t$ and $B_t$ linearly; i.e., the initial health status $H_t$ and health behaviors $B_t$ can shift the distribution of health shock $s_t$. It can be shown that the semi-reduced form equations for $B_t$ and $M_t$ then have linear representations; i.e.,

$$B_t = b_0 + b_1 I_t + b_2 B_{t-1} + b_3 H_t + b_4 \tau + b_5 \nu_t + b_6 \epsilon_t, \quad (10)$$

$$M_t = m_0 + m_1 I_t + m_2 B_{t-1} + m_3 H_t + m_4 \tau + m_5 \nu_t + m_6 \epsilon_t, \quad (11)$$

where the coefficient $b_2$ is a $3 \times 3$ matrix; all the other coefficients in equation (10) and $m_2$ are $3 \times 1$ vectors, and the rest are scalars. $b_1$ and $m_1$ in equations (10) and (11) represent the average effects of health insurance on health behaviors and medical utilization.

Given the functional form assumptions and each possible realization of the shocks, the continuous choice variables in the two periods can all be expressed as linear functions of $I_t$, $B_{t-1}$, $H_t$, $\tau$, $\nu_t$ and shocks. It follows that the difference between the expected utility with insurance and without insurance is a linear function of $B_{t-1}$, $H_t$, $\tau$, $\nu_t$ and $P_{It}$. That is, the insurance model can be written as

$$I_t = 1 \left[ (\alpha_1 + \alpha_2 B_{t-1}' + \alpha_3 H_t + \alpha_4 \tau + \alpha_5 \nu_t) - P_{It} > 0 \right], \quad (12)$$

where $1[\cdot]$ is an indicator function as defined before. $\alpha_2$ is a $3 \times 1$ vector, and all the other coefficients are scalar. As in equation (9), $P_{It}$ is a function of $B_{t-1}$, $H_t$, $\tau$, the age dummy $D_{it}^{65}$, and other possible exclusion restrictions $Ex_t$, while the willingness-to-pay for insurance $P_{It}^*$ is now given by the bracketed term in equation (12).

Derivation of equations (10), (11), and (12) is provided in the Appendix. The empirical analysis later is based on these semi-reduced form equations.
3.3 The Direct and Indirect Effects of Health Insurance

The semi-reduced form equations (10), (11), and (12) will be numerically much simpler to estimate than the underlying structural model. The structural parameters of primary interest, the direct and indirect insurance effects, can be recovered from the semi-reduced form equations, because of identification of the structural $M_t$ equation and the exclusion restrictions. The following shows the details.

By equation (5), we have the structural $M_t$ equation,

$$M_t = f_M(I_t, B_t, H_t, \tau, \nu_t, e_t).$$  \hspace{1cm} (13)

Further, the semi-reduced form equations for $B_t$ and $M_t$ can be written as

$$B_t^l = g_{B^l}(B_{t-1}, H_t, I_t, \tau, \nu_t, e_t), \text{ for } l = 1, 2, 3,$$  \hspace{1cm} (14)

$$M_t = g_M(B_{t-1}, H_t, I_t, \tau, \nu_t, e_t).$$  \hspace{1cm} (15)

which are obtained by solving equations (5) and (6) for $B_t$ and $M_t$. For quadratic utility these would be equations (10) and (11) in the previous section.

Equations (13), (14), and (15) together imply the following decomposition

$$\frac{dM_t}{dI_t} = \frac{\partial M_t}{\partial I_t} + \frac{\partial B_t'}{\partial I_t} \frac{\partial M_t}{\partial B_t}. $$ \hspace{1cm} (16)

Equation (16) means that the total effect of health insurance on medical care utilization, \( \frac{dM_t}{dI_t} \), can be decomposed into a direct effect, \( \frac{\partial M_t}{\partial I_t} \), and an indirect effect through its effect on health behaviors, \( \frac{\partial B_t'}{\partial I_t} \frac{\partial M_t}{\partial B_t} \). \( \frac{dM_t}{dI_t} \) is obtained from (15), which corresponds to the coefficient of $I_t$ in equation (11) when utility is quadratic. The direct price effect \( \frac{\partial M_t}{\partial I_t} \) is not estimated directly, but can be recovered from equation (16). \( \frac{\partial B_t'}{\partial I_t} \frac{\partial M_t}{\partial B_t} \) represents the effect of health insurance on health behaviors. It is a $1 \times 3$ vector of elements given by (14). When utility is quadratic, these elements are the coefficients of $I_t$ in equation (10). \( \frac{\partial M_t}{\partial B_t} \) is a $3 \times 1$ vector describing the effects of each health-related behavior on the use of medical care. It involves structural parameters to be calculated.

Further, equations (13) and (14) indicate that

$$\frac{\partial M_t}{\partial B_{t-1}} = \frac{\partial B_t'}{\partial B_{t-1}} \frac{\partial M_t}{\partial B_t}, $$ \hspace{1cm} (17)

where \( \frac{\partial M_t}{\partial B_{t-1}} \) is a $3 \times 1$ vector of derivatives of the semi-reduced form $M_t$ equation, and

\[10\text{ For convenience of notation, expressions in this section ignore the fact that } I_t \text{ is discrete. For example, } \frac{\partial M_t}{\partial I_t} \equiv (M_t|I_t = 1) - (M_t|I_t = 0).\]
\[ \frac{\partial B'_t}{\partial B_{t-1}} \] is a 3 \times 3 matrix of derivatives of the semi-reduced form \( B_t \) equations. With quadratic utility, these are the coefficients of \( B_{t-1} \) in semi-reduced form equations (11) and (10), respectively. Notice (17) represents a system of linear equations in three unknowns, \( \frac{\partial M_t}{\partial B'_t}, \frac{\partial M_t}{\partial B_t}, \) and \( \frac{\partial M_t}{\partial B_{t-1}} \). Solving this system yields

\[
\frac{\partial M_t}{\partial B_t} = \left( \frac{\partial B'_t}{\partial B_{t-1}} \right)^{-1} \frac{\partial M_t}{\partial B_{t-1}}. \tag{18}
\]

Equation (18) provides a way to compute \( \frac{\partial M_t}{\partial B_t} \) based on the parameters in the semi-reduced form \( B_t \) and \( M_t \) equations. Since \( \frac{dM_t}{dt}, \frac{\partial B'_t}{\partial B_t}, \) and \( \frac{\partial M_t}{\partial B_t} \) can all be obtained either directly or indirectly from the semi-reduced form \( M_t \) and \( B_t \) equations, the direct price effect \( \frac{\partial M_t}{\partial I_t} \) can be computed based on equation (16).

4 Data Features and Empirical Specification

Before empirically examining the effect of health insurance on health-related behaviors and medical utilization, one needs to consider the complications of using data on medical utilization and unhealthy behaviors, such as cigarette smoking and alcohol drinking. These data are typically characterized by a nontrivial fraction of zero outcomes and a skewed distribution of the positive outcomes (Manning and Mullahy, 2001). The following addresses these issues separately.

4.1 A Large Proportion of Zero Observations

Zero outcomes may indicate non-participation. Participation decisions are generally considered different from consumption decisions and hence modeled separately in the literature. For example, Yen and Jones (1996) model the quitting decision of smokers based on the fixed costs and the expected benefits of quitting, where the benefits depend on how many cigarettes an individual would have smoked had he not quit (the consumption decision).

The previous section considers consumption decisions. Those equations could be estimated using participants’ data; however, one needs to deal with the sample selection issue that results from choosing a nonrandom sample of the population, such as drinkers and smokers. Therefore, this paper adopts a binary choice equation determining the probability of participation (non-zero consumption) and a regression equation determining the level of consumption, given participation. The participation equation could be
established in theory by following Yen and Jones (1996). Both the participation and consumption equations include health insurance as a dummy endogenous regressor. That is, this paper adopts sample selection models or generalized Tobit models. The generalized Tobit model has been used in Kenkel (1990) to estimate the effect of consumer health information on health care use. Having separate participation and consumption equations allows us to distinguish between the extensive and intensive margins of the health insurance effect.

Besides generalized Tobit, a variety of other models have been applied in the literature to separate participation and consumption. These include two-part models, standard Tobit, and variants of generalized Tobit models. For example, Fry and Pashardes (1994) use Logit instead of Probit for participation and then a linear regression for positive expenditures to estimate UK household tobacco demand. Examples of two-part models include Lewit et al. (1981), Wasserman et al. (1991), and Blaylock and Blisard (1992). Two-part models assume independence between errors in the participation and consumption equations, which is not intuitively appealing, because the utility of participation should depend on the utility derived from consumption.

Standard Tobit models have been used in William (2002) for alcohol consumption and in Kenkel (1991) for health behaviors including drinking, smoking and exercise. They rely on the restrictive assumption that zero outcomes and non-zero outcomes are generated by the same underlying process. Moreover, in the case of multiple non-negative choice variables, utility maximization does not yield standard Tobit models. Therefore, the standard Tobit is not used here.

4.2 A Skewed Distribution of Positive Observations

Implementing Tobit type models requires a normality assumption; however, as mentioned, data on cigarette and alcohol consumption as well as medical utilization have skewed distributions. Therefore, this paper applies transformations on the dependent and lagged dependent variables. One transformation is the Box-Cox transformation, which has been used by Yen and Jones (1996) to model cigarette consumption in the UK.

Another transformation is the inverse hyperbolic sine (IHS) transformation (John-

\[11\]

It can be shown that such maximization yields a system of simultaneous Tobit models, to which the solutions for nonnegative choice variable do not have a Tobit specification. Intuitively, when there are multiple nonnegative constraints, the optimal value of any choice variable falling on the corner will affect the optimal value of the other choice variables.
The IHS transformation has been proposed to accommodate non-normal errors in both regression models and limited dependent variable models (Burbidge et al. 1988; MacKinnon and Magee 1990; Horowitz and Neumann 1989). Unlike the Box-Cox transformation, which can not be applied to random variables that can take on zero values, the IHS transformation can be performed on random variables that can take on any value. Yen and Jensen (1995) use this transformation to model household alcohol expenditures in the US. Their results support using the IHS transformation in modeling alcohol consumption.

The IHS transformation of a random variable $y$ is

$$T(y, \theta) = \frac{\sinh^{-1}(\theta y)}{\theta} = \frac{\log(\theta y + (\theta^2 y^2 + 1)^{1/2})}{\theta},$$

(19)

where $\theta$ is a nonnegative scaling parameter that can be estimated along with the other parameters in a model. $T(y, \theta)$ is linear around the origin and approximates the logarithm in its right tail. The parameter $\theta$ governs the proportion of the function’s domain that approximates linear and the proportion that is close to logarithm. In particular, $T(y, \theta)$ goes to $y$ when $\theta$ approaches to 0.

Also applied is the logarithmic transformation, which is a limiting case of the Box-Cox transformation. Results from all three transformations are compared to check their robustness to different specifications. Since zero outcomes may convey different information from positive outcomes, a dummy variable indicating last period zero consumption is included as a covariate. Transformations are then applied to the positive values of lagged dependent variables. This allows for a more flexible elasticity profile of lagged health behaviors.

With the appropriate transformation, a two-step approach is used to estimate the generalized Tobit model. As insurance status is endogenous, the participation probability equation is jointly estimated along with the insurance equation in the first step; i.e., a Bivariate Probit is estimated. Then in the second step, a bias correction term ($\lambda$) that mimics the inverse Mill’s ratio is added to the consumption regression equation.\footnote{The same transformation parameter is involved in both the participation equation and the consumption equation, so this paper uses an iterative approach to obtain a consistent estimate of the transformation parameter in the two equations.} The same set of covariates are used in the probability and level equations; therefore, this paper estimates sample selection models, instead of self-selection models, as distinguished by Maddala (1985).

The full specification of the generalized Tobit or sample selection model with trans-
formations can be expressed as

\[ I_t = 1 [X'_t\alpha - P_t + \omega_t > 0], \]

\[ 1 [Y_t > 0] = 1 [X'_t\gamma_0 + I_t\gamma_1 + \varsigma_t > 0], \]

\[ T(Y_t, \theta_Y) = 1 [Y_t > 0] \cdot (X'_tR_0 + I_tR_1 + \pi_t), \]

where \( X_t = (1, T(B_{t-1}^T, \theta_B), H_t, \tau)' \); \( Y = M \) or \( B \); \( T(\cdot) \) represents one of the three transformations; \( \theta_B \) and \( \theta_M \) are the transformation parameters for alcohol consumption and medical utilization, respectively; \( \omega_t, \varsigma_t, \) and \( \pi_t \) are the error terms.

### 4.3 Other Data Issues

Given the fact that the dependent variables are the number of drinks and number of visits, count variable models, such as poisson or negative binomial models could be used. However, most count models have restrictive properties. For example, the Poisson model requires that the conditional variance equals the conditional mean, while real-life data frequently display overdispersion, i.e., the variance exceeds the mean. Moreover, both the Poisson model and the negative binomial regression model can not deal with excess zeros. Recently developed zero-inflated count models are essentially count data versions of selection or Tobit models. In particular, they also assume there are two separate data generation processes, with one generating only zero counts, and the other generating counts from either a Poisson or a negative binomial model.

This study adopts relatively flexible specifications and estimates appropriate transformations along with the other parameters in the model. Ophem (2000) shows that count data distributions and continuous data distributions can be related through transformation. In particular, the simple log transformation, a special case of the more general Box-Cox transformation, implies an exponential specification, which is related to the Poisson regression in a count data setting. The transformed Tobit model adopted here facilitates dealing with the endogeneity issues and the two margins of the effects of health insurance.
5 Sample Description and Model Identification

5.1 Sample Description

This study uses the 3rd, 4th and 5th waves of HRS data. The HRS is a national panel survey of individuals over age 50 and their spouses in the US. It has extensive information on demographics, income, family structure, health, health care utilization, and insurance etc. Data have been collected every two years since 1992. Seven waves of data have been made available so far. The first two waves have only yes or no information on individuals’ drinking behavior, while individuals in the last two waves are on average over 70, and are not suitable for this paper’s empirical analysis.

Individuals older than 70 are removed from the sample. With no uninsured counterparts, these individuals’ extremely high usage of medical care or low unhealthy behaviors may exaggerate the estimated insurance effects. For example, end-stage patients may use enormous amounts of medical care, which may not depend on their insurance status. Including these individuals would lead to upward bias in the estimated effects of insurance on medical use. For the same reason, those who receive social security disability insurance and those deceased within two years of being surveyed are also excluded. This yields a sample of size 14,289, including 13,016 insured individuals and 1,273 uninsured individuals.

Drinking behavior is measured by the number of alcoholic drinks consumed per week and medical utilization is the total number of doctor and hospital visits in a year. Limited information is available for smoking and exercise, so two dummies are used to indicate whether or not an individual smokes or participates in vigorous physical activity three times or more per week. The full generalized Tobit models with transformation are then estimated for drinking and medical utilization, while Probit models are estimated for smoking and exercise.

This paper uses two sets of measures for individuals’ health conditions: self-reported health status and physician diagnosed diseases, which include hypertension, diabetes, heart disease, cancer, stroke, lung disease, arthritis, and psychiatric disease. Self-reported health status was collected in five categories, but is summarized here by three

---

13 This refers to the RAND HRS, which is a contribution of the RAND Center for the Study of Aging. It contains cleaned and processed variables that are comparable across survey waves. Details can be found in the RAND HRS data documentation, version F.

14 The HRS measure of physical activity captures both going to the gym and physically demanding work. That is, it is a measure of overall physical activity level, so the estimated insurance effect on exercise is essentially that on a non-sedentary life style.
dummy variables, defined by excellent health, very good or good health, and fair or poor health (the default). Note that the health status in the model refers to that at the beginning of a period of interest, so the previous wave’s health measures are used.

Unless long enough panel data are available, it is impossible to distinguish between individual heterogeneity and state dependence, i.e., $B_t$ depends on $B_{t-1}$ (Heckman, 1981); therefore, individual heterogeneity $\tau$ is parameterized as a function of individual characteristics, including age, gender, race, education, and income. $^{15}$ Education is defined as one of three categories: college or above, high school or GED, and less than high school education (the default). The average of the respondent’s and the spouse’s current income is used as the income measure. $^{16}$ In addition, the number of children in the household is included to adjust for the actual financial status of the household. Summary statistics of the sample by insurance status are reported in Table 1.

Table 1 to be placed here.

As shown by Table 1, the insured individuals are more likely to visit a doctor or hospital, and they also have more visits on average than the uninsured. Specifically, 94.1% of insured individuals visited a doctor or hospital at least once in the past two years, compared with 81.1% of the uninsured. $^{17}$ The average number of visits per year among the insured is about 4.4, which is about 1.3 more than that of the uninsured. In addition, the insured in general have healthier behaviors than the uninsured: 18.4% of insured individuals smoke currently, in contrast to a much higher rate of 27.4% among the uninsured. 35.3% of insured individuals currently drink alcohol, which is about 6% higher than that of the uninsured; however, the uninsured drinkers on average drink 3 more alcoholic beverages per week than the insured drinkers. Therefore, the insured group consists of a larger fraction of drinkers, whereas they are mainly light drinkers. Finally, the proportion of individuals who exercise regularly is slightly higher among the insured than among the uninsured.

It seems that health insurance is associated with healthier behaviors. However, these simple correlations may reflect the selection rather than the causal effect of health

---

$^{15}$Since both the permanent taste $\tau$ and taste shifter $\nu_t$ are included in the utility function, any time varying characteristics such as age, income can be interpreted as observable parts of $\nu_t$.

$^{16}$To the extent that current income may be endogenous due to its dependence on current health, one may just replace current income with lagged ones. However, current income is more relevant to current purchase decisions, and the impact of this endogeneity if any should be small because health affect income mostly with lags and the model is conditioning on initial health status.

$^{17}$In HRS, individuals are asked to report their health care utilization for the past two years. This paper uses the average number of physician and hospital visits in one year by dividing the reported number by two.
insurance. First, similar differences exist in past behaviors between the insured and uninsured; i.e., the insured individuals may have a healthier lifestyle \textit{ex ante}. Second, the insured are on average about 2 years older than the uninsured. Age difference may partly explain their difference in drinking and smoking, because these behaviors tend to decline with age among the elderly (Card et al. 2004). Moreover, although the insured tend to report better health for the last period than the uninsured, overall, they are more likely to have diagnosed diseases associated with aging, such as hypertension, cancer, heart disease, lung disease, and arthritis. Therefore, the insured individuals’ physical condition may limit their drinking and smoking. Finally, the insured on average have higher education and higher household income, which may indicate better health knowledge and higher health demand, and so may also account for their healthier behaviors.

5.2 Exclusion Restrictions and Identification

Insurance status is endogenous in the behavior and medical utilization equations, so exogenous variation in insurance status is required to avoid identifying these equations purely through functional forms and distributional assumptions. The theoretical model in section 4 shows that the variables that determine insurance prices but not health behaviors and medical utilization can serve as exclusion restrictions, and that one such variable is the age dummy indicating age 65 or above. Here I introduce another potential exclusion restriction, self-employment status. Both are discussed in detail below.

The exogenous change in insurance price at age 65 generates a significant jump in health insurance coverage.\textsuperscript{18} Card et al. (2004) use this exogenous variation in a regression discontinuity estimation. The assumption is that medical care utilization and health behaviors would evolve smoothly with age in the absence of the discrete change in insurance coverage at age 65. Card et al. (2004) rules out discontinuities in potential confounding factors such as employment status, income, and family structure.

\textsuperscript{18}Medicare is available to individuals who are at least 65 years old and who worked at least 40 quarters in covered employment (or have a spouse who qualifies for coverage). For those who meet the eligibility criteria, Medicare hospital insurance (Part A) is available free of charge and Medicare Part B (Medicare insurance) is available for a modest monthly premium. Normally, individuals who are approaching their 65th birthday receive notice of impending eligibility and are informed that they have to enroll in the program and choose whether or not to accept Medicare Part B coverage. Medicare is also available to people under 65 who are receiving Disability Insurance, and those with kidney disease. Individuals who do not qualify on the basis of their own or their spouse’s work history may still enroll in Medicare at age 65 by paying monthly premiums for both Part A and Part B coverage. This option is limited to US citizens and legal aliens with at least five years of residency in the U.S (Card et al., 2004).
by examining the empirical age profile of these variables.\footnote{A detailed discussion about the validity of this assumption can be found in Card et al. (2004).}

In this paper’s sample, the insurance coverage rate for individuals aged 65 or above is 99.1%, in contrast to 88.1% for individuals under 65, so the age cutoff for Medicare eligibility has a big impact on the insurance holding probability. The insurance equation therefore includes the age dummy as a covariate, which is excluded from the other equations. Age, age squared, and an interaction term between age and the age dummy are included in all equations to capture the age profile of the dependent variables. The interaction term allows the first derivative of the dependent variables with respect to age to change at 65, but still maintains that there is no jump at this age point. Tentatively including the age dummy in the behavior and care utilization equations did not yield significant coefficients. Therefore, similar to Card et al. (2004), the age dummy is plausibly excluded from these quantity equations.

In addition, self-employed individuals have to pay the full insurance cost, so other things equal they are much less likely to buy insurance. In our sample, 78.7% of self-employed individuals have insurance, in contrast to 92.6% of those who are not self-employed. Since self-employment status affects the probability of having insurance through its impact on the insurance cost, it also serves as a potential exclusion restriction. Self-employment status is used as an instrumental variable in estimating the insurance effect on the use of medical care by Meer and Rosen (2003). The validity of this instrument is checked by analyzing panel data on the transition from wage-earning into self-employment and showing that individuals who become self-employed do not differ systematically from those who remain wage-earners in the use of health care. Deb and Trivedi (2006) use a set of employment characteristics, including self-employment status, as instruments to identify the effect of the insurance plan on medical care utilization.

To check the validity of self-employment as an instrument in the current context, this paper employs samples of individuals whose self-employment status changes between two waves of data and examines if the transition into or out of self-employment substantially changes individuals’ health behaviors. Considering there is a two years’ difference between two waves of data, the validity check is done using the difference-in-difference approach, i.e., comparing the probabilities and levels of unhealthy behaviors between those who transit into self-employment and those who do not. Similarly, those probabilities and levels are also compared between individuals who transit out of self-employment and those who keep being self-employed. None of the differences are statistically signif-

\footnote{A detailed discussion about the validity of this assumption can be found in Card et al. (2004).}
significant. These test results are reported in Table A1 in the Appendix. Further, several over-identification tests show that the exogeneity of the instrumental variables in the quantity equations of drinking and health care utilization can not be rejected. They provide evidence that these instrumental variables may be valid in the current setting.

6 Empirical Results

This section first reports the basic estimation results, and then based on these results conducts a further investigation on heavy drinking. Finally, it discusses increases in health care utilization caused by the moral hazard effect associated with drinking.

6.1 Basic Estimation Results

Table 2 presents the estimation results for the health insurance decision. Not surprisingly, these estimates show a significant increase in the probability of holding health insurance at the age of 65. Self-employment status is negatively correlated with having insurance. Smoking last period and higher drinking in the last period are associated with smaller probabilities of holding insurance, which reflect the selection effect of health insurance. That is, health insurance picks out a sample of individuals who have healthier behaviors per se.

Having a diagnosed disease in the previous period generally increases the probability of buying insurance, except for having stroke or psychiatric disease. These increases are significant for heart disease and lung disease. Individuals who self report having good or very good health are more likely to have insurance ceteris paribus than those reporting either excellent health or poor or fair health. These results may be due to two counteracting factors: the demand side’s adverse selection and the supply side’s screening. On the one hand, individuals who are in excellent health are less willing to buy insurance; on the other hand, individuals who are in poor health may have difficulty obtaining or affording health insurance. As expected, education and household income

\footnote{The validity of the exclusion restrictions is examined using Sargan’s and Basmann’s over-identification tests in the linear instrumental variable regression setting and Hansen’s $J$ test in a two-step efficient general method of moments (GMM) setting. For the alcohol consumption, the P-values of the three tests are 0.596, 0.597 and 0.622 respectively. For the medical care utilization, the P-values are 0.192, 0.193, and 0.182, respectively.}

\footnote{Note that the effect of age 65 or over in the insurance decision is given by the coefficient of the age dummy and that of the interacted term between age and the age dummy.}
have positive effects on the probabilities of having insurance, while number of children has a negative effect.

Table 2 to be placed here.

Tables 3a and 3b summarize the estimated insurance effects at the extensive margin and the intensive margin, respectively. Since reduced-form equations are estimated, these are the total effects of health insurance. In the probability equations, transformations are simply applied to the lagged dependent variables, which are less crucial, so Table 3a only reports the results with the log transformation. Table 3b reports the insurance effects on the quantities of alcohol drinking and health care utilization with all transformations. Full estimation results for the quantity equations are included in the Appendix (Tables A2 and A3). Also reported in Table 3b are results from two-part models. As shown by the two tables, health insurance increases both the probability of visiting a doctor or hospital and the number of visits conditional on visiting. The increase in the probability is only 0.8% and is not significant. In contrast, the increase in the yearly number of visits among those who do visit doctors or hospitals is 36.7%. After correcting for sample selection, this is reduced to 34.7%, which is equivalent to about 1.6 more visits per year.

In addition, health insurance decreases the probability of exercising regularly by 4.6% and increases the probability of smoking by 1.7%. Interestingly, it decreases the probability of drinking by 6.5%. However, the IHS regression shows that insurance induces those drinkers to drink around 12.2% more alcoholic beverages per week. Correcting for sample selection, the increase is reduced to 4.3%, which corresponds to about 0.3 more alcoholic drinks per week.

Tables 3a and 3b to be placed here.

Estimation of the Box-Cox and log models reveals similar effects of health insurance. In fact, the estimates of the transformation parameters indicate consistency of the three specifications: the IHS transformation parameter is far above zero, while the Box-Cox transformation parameter is close to zero;22 both suggest that the log model may be a good approximation. These results need to be interpreted with caution though, since none of the insurance effects on behaviors in these equation are statistically significant.

22 The IHS transformation parameters for the alcohol consumption and doctor/hospital visits are 4.24 and 9.01, respectively. In contrast, the corresponding Box-Cox transformation parameters are 0.025 and 0.006.
Further, as can be seen from Table A2 in the Appendix, last period health, including both self-reported health condition and diagnosed diseases, are the major determinants of doctor or hospital visits, which is not surprising. The coefficients of the bias correction term $\lambda$ is negative. This is unusual, but could be a small sample issue. In this paper’s sample, only about 7% of individuals have never seen a doctor or hospital for the past two years. The binary choice participation equation and hence $\lambda$ may be imprecisely estimated. However, because the sample of individuals with zero visits is small, ignoring this sample selection issue does not change the estimation results much. For example, the effect of health insurance on the number of visits goes down from 36.7% to 34.7%, as shown in Table 3b.

Table A3 in the Appendix shows that last period drinking, smoking, and chronic disease like diabetes, as well as gender and race are the significant predictors for current alcohol consumption. Moreover, it is well documented that income and alcohol consumption is positively correlated (Auld, 2005), which is consistent with the findings here. The estimated coefficient for the last period not drinking differs in the IHS alcohol consumption model from the Box-Cox and log models. This is related to the difference of the transformations in the neighborhood of origin. As can be seen, the difference in this estimated coefficient does not affect the estimates for all the other coefficients in the model. Finally, for both tables, $\rho$ is the correlation of the reduced form errors in the insurance and the outcome equations. The estimates are negative. The theoretical model does not generate any predictions regarding the sign or magnitude of $\rho$.

6.2 Further Investigation on Drinking

The previous section shows that insurance increases the probability of participating in unhealthy behaviors by a positive but insignificant amount. This is consistent with findings elsewhere in the literature. But some other findings here, while less statistically significant, are more unusual. First, the two margins of the insurance effect on drinking go in opposite directions; i.e., insurance appears to decrease the probability of drinking, but it also induces the drinkers to drink more. Second, insurance affects the probability of drinking and that of smoking in opposite directions.

A possible explanation for these results is that, unlike smoking, low levels of drinking are considered healthy (Dufour, 1996). The difference between non-smoking and smoking is similar to that between light drinking (including non-drinking) and heavy drinking, rather than between not drinking at all and drinking any amount. This is consistent
with what is observed in the data: compared with the uninsured, the insured have a smaller proportion of smokers, but a larger proportion of drinkers. However, the insured drinkers tend to be light drinkers. Therefore, cleanly identifying the disincentive effect of health insurance on healthy behaviors requires distinguishing between light drinking and heavy drinking. Mixing light drinkers with heavy drinkers does not identify the moral hazard associated with drinking.

The following analysis distinguishes between heavy drinking (unhealthy drinking) and non-heavy drinking. Heavy drinking is defined as weekly alcohol consumption being above some percentile of the alcohol consumption distribution. Different percentiles are tried as cut-offs. The results are reported in Table 4. As shown in the table, focusing on a sample of heavier drinkers increases the estimated insurance effects. From the sample of drinkers above the 50th percentile or median to the sample of drinkers above the 90th percentile, the effect of health insurance on the weekly number of alcoholic drinks goes from about 1.6 to more than 6. All these effects are statistically significant.

Table 4 to be placed here.

6.3 The Magnitude of the Indirect Insurance Effect

In this section I apply the results in section 4.3 to evaluate the indirect effects of health insurance on health care utilization. This entire indirect effect cannot be estimated because of data limitation. In particular, I cannot estimate quantity equations for smoking and exercise. However, as reported in the previous section, health insurance appears to make people less cautious regarding heavy drinking, so this section investigates how much insurance-induced drinking increases doctor or hospital visits. The following analysis focuses on above-the-median heavy drinkers.

Rewrite equation (16) in section 3.3 as

\[
\frac{dM_t}{dI_t} = \frac{\partial M_t}{\partial I_t} + \frac{\partial M_t}{\partial B^1_t} \frac{\partial B^1_t}{\partial I_t} + \frac{\partial M_t}{\partial B^2_t} \frac{\partial B^2_t}{\partial I_t} + \frac{\partial M_t}{\partial B^3_t} \frac{\partial B^3_t}{\partial I_t}. \tag{23}
\]

The increased medical utilization caused by insurance-induced drinking is given by \(\frac{\partial M_t}{\partial B^1_t} \frac{\partial B^1_t}{\partial I_t}\) in equation (23), as \(B^1_t\) represents drinking. In addition, the first equation

\(^{23}\)For Tobit type models, the marginal effect of a covariate needs to take into account its effects in both the probability equation and the quantity equation. Specifically, it is the weighted sum of the two effects. The weights are the conditional mean and the probability of participation, respectively. To get population mean effects, all the relevant terms in the formula needs to be calculated in this manner.
in (17) is given by
\[
\frac{\partial M_t}{\partial B_{t-1}^1} = \frac{\partial M_t}{\partial B_t^1} \frac{\partial B_t^1}{\partial B_{t-1}^1} + \frac{\partial M_t}{\partial B_t^2} \frac{\partial B_t^2}{\partial B_{t-1}^1} + \frac{\partial M_t}{\partial B_t^3} \frac{\partial B_t^3}{\partial B_{t-1}^1}.
\] (24)

As indicated by the estimation results, the effect of the last period drinking \(B_{t-1}^1\) on the current period drinking \(B_t^1\) is much larger than the effect on current smoking \(B_t^2\) or exercise \(B_t^3\); therefore, the last two terms on the right hand side of equation (24) are relatively small and \(\frac{\partial M_t}{\partial B_{t-1}^1}\) is mostly determined by \(\frac{\partial M_t}{\partial B_t^1} \frac{\partial B_t^1}{\partial B_{t-1}^1}\). Further, the last two terms are non-negative, because research shows that unhealthy behaviors are complements (Cameron and Williams, 2001). It follows that \(\frac{\partial M_t}{\partial B_t^1} \frac{\partial B_t^1}{\partial B_{t-1}^1} \leq \frac{\partial M_t}{\partial B_{t-1}^1}\). Simple algebra further gives the result \(\frac{\partial M_t}{\partial B_t^1} \frac{\partial B_t^1}{\partial t} \leq \frac{\partial M_t}{\partial B_{t-1}^1} \left(\frac{\partial B_t^1}{\partial B_{t-1}^1}\right)^{-1} \frac{\partial B_t^1}{\partial t}\); i.e., the increase in medical utilization caused by the insurance-induced drinking is at most \(\frac{\partial M_t}{\partial B_{t-1}^1} \left(\frac{\partial B_t^1}{\partial B_{t-1}^1}\right)^{-1} \frac{\partial B_t^1}{\partial t}\).

Using the sample of above-the-median drinkers, it can be shown that health insurance on average increases doctor or hospital visits per year by 27.9%, and that the increased visits caused by insurance-induced drinking is at most 0.3%.\(^{24}\) Averaging this over the whole population will further decrease it below 0.3%. Moreover, if the number of visits caused by insurance-induced changes in smoking and exercise is of the same size, then the increased number of visits caused by insurance-induced unhealthy behaviors as summarized by drinking, smoking and insufficient exercise will be less than 1%. The total effect of health insurance on doctor or hospital visits is mainly the direct price effect.

7 Conclusions

This paper investigates the direct and indirect effects of health insurance that can lead to increased use of health care. A two-period dynamic forward-looking model is constructed to derive the structural causal relationships among the decision to buy insurance, health behaviors including drinking, smoking, and exercise, and health care utilization. Semi-reduced form equations are then obtained for health behaviors and care utilization as functions of the endogenous health insurance. The structural parameters of interest, the (direct) price and (indirect) behavioral moral hazard effects of insurance, are recovered from the reduced form parameters.

\(^{24}\)The numbers reported here are estimates from the IHS Tobit model. The results do not vary much in the Log and Box-Cox models.
Using data from the HRS, this study estimates the effects of health insurance on drinking, smoking, exercising, and doctor or hospital visits. An additional analysis then distinguishes between light and heavy drinking, focusing on the effect of insurance on heavy drinking, and then analyzing possible increases in health care utilization due to the behavioral moral hazard associated with heavy drinking.

The results show that having health insurance encourages individuals’ unhealthy behaviors, particularly heavy drinking, but the resulting immediate increase in individuals’ doctor or hospital visits appears to be negligible. Within the total effect of health insurance on care utilization, the direct price effect is dominant. The empirical results also suggest that insurance effects at the extensive margin tend to be smaller than that at the intensive margin. For example, the insurance effects on probabilities of heavy drinking, smoking, exercising, and visiting a doctor or hospital are small and insignificant; whereas those on quantities of heavy drinking and visiting a doctor or hospital are much larger and statistically significant. Therefore, access to insurance does not appear to induce people to take up drinking, but apparently encourages those who are already heavy drinkers to drink even more. It is worth emphasizing that these reported medical use increases due to behavioral changes are short run effects. The finding that these effects are small is therefore not surprising. The resulting long run effects accumulated over time would likely be much larger.

Since health insurance affects the amount of unhealthy behavior individuals engage in, it would be beneficial to couple health insurance with incentives to counteract this behavioral moral hazard. Specifically, the difference between the two margins suggests that to effectively do so, policies should target current unhealthy behavior participants, such as heavy drinkers. In July 2006, three West Virginia counties adopted a pilot Medicaid program. Members could sign an agreement valid for a year that required them to engage in positive health behaviors, such as maintaining healthy weight, exercising, quitting tobacco use. Those who signed and adhered to the agreement were then rewarded with additional benefits (Tworek and Horn, 2007). The results here suggest that such a program would be most beneficial if it targets people who currently actively engage in unhealthy behaviors, because these people’s behaviors are most affected by health insurance.

Several issues that are not covered here can be topics for future research. For example, the current evidence on the moral hazard effect is mainly on alcohol consumption. It would be interesting to see if similar effects exist for other behaviors that affect health. In addition, doctor or hospital visits here include both visits for curative purposes and
those for preventive purposes; it would be useful to investigate whether health insurance affects the two types of visits differently. Addressing these questions would require more detailed information than is available in the current data set. Another issue is that since Canada and some European countries have universal health care coverage, it would be useful to look at cross-country comparisons.
Table 1 Summary statistics of the sample by insurance status

<table>
<thead>
<tr>
<th></th>
<th>Insured (n=13,016)</th>
<th>Uninsured (n=1,273)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Visiting a doctor/hospital(^1)</td>
<td>.941</td>
<td>.235</td>
</tr>
<tr>
<td># of visits in a year</td>
<td>4.35</td>
<td>7.74</td>
</tr>
<tr>
<td>Current period behavior:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drinking</td>
<td>.330</td>
<td>.470</td>
</tr>
<tr>
<td>Smoking</td>
<td>.167</td>
<td>.373</td>
</tr>
<tr>
<td>Exercising(^2)</td>
<td>.504</td>
<td>.500</td>
</tr>
<tr>
<td># of alcoholic drinks(^3)</td>
<td>2.35</td>
<td>5.92</td>
</tr>
<tr>
<td>Last period behavior:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drinking</td>
<td>.353</td>
<td>.478</td>
</tr>
<tr>
<td>Smoking</td>
<td>.184</td>
<td>.387</td>
</tr>
<tr>
<td>Exercising</td>
<td>.523</td>
<td>.499</td>
</tr>
<tr>
<td># of alcoholic drinks</td>
<td>2.47</td>
<td>6.04</td>
</tr>
<tr>
<td>Last period health status:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair/Poor</td>
<td>.166</td>
<td>.372</td>
</tr>
<tr>
<td>Good/Very good</td>
<td>.652</td>
<td>.476</td>
</tr>
<tr>
<td>Excellent</td>
<td>.182</td>
<td>.386</td>
</tr>
<tr>
<td>Diagnosed disease:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>.367</td>
<td>.482</td>
</tr>
<tr>
<td>Diabetes</td>
<td>.096</td>
<td>.294</td>
</tr>
<tr>
<td>Cancer</td>
<td>.065</td>
<td>.246</td>
</tr>
<tr>
<td>Heart disease</td>
<td>.119</td>
<td>.324</td>
</tr>
<tr>
<td>Stroke</td>
<td>.022</td>
<td>.147</td>
</tr>
<tr>
<td>Lung disease</td>
<td>.050</td>
<td>.218</td>
</tr>
<tr>
<td>Psychiatric disease</td>
<td>.085</td>
<td>.278</td>
</tr>
<tr>
<td>Arthritis</td>
<td>.440</td>
<td>.496</td>
</tr>
<tr>
<td>Age (in year)</td>
<td>61.65</td>
<td>5.19</td>
</tr>
<tr>
<td>Male</td>
<td>.406</td>
<td>.491</td>
</tr>
<tr>
<td>Non-white(^4)</td>
<td>.158</td>
<td>.365</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.067</td>
<td>.251</td>
</tr>
<tr>
<td>Education: Less than high-school</td>
<td>.192</td>
<td>.394</td>
</tr>
<tr>
<td>High-school or GED</td>
<td>.386</td>
<td>.487</td>
</tr>
<tr>
<td>College or above</td>
<td>.422</td>
<td>.494</td>
</tr>
<tr>
<td>Income($1000)(^5)</td>
<td>62.80</td>
<td>86.19</td>
</tr>
<tr>
<td>Number of children</td>
<td>3.45</td>
<td>2.05</td>
</tr>
</tbody>
</table>

1. A dummy indicating whether or not an individual visits a doctor or hospital at least once since the last survey. 2. A dummy indicating whether or not the individual participates in vigorous physical activity three or more times per week. 3. The average number of alcoholic drinks consumed in a week for the last three months. 4. This includes black people and a group of individuals categorized as other (races). 5. In 1997 dollars.
<table>
<thead>
<tr>
<th>Last period behavior:</th>
<th>Health Insurance</th>
<th>Coeff. (SE)</th>
<th>Marginal effect (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log # of alcoholic drinks</td>
<td>-0.130 (.030)***</td>
<td>-0.012 (.003)***</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.197 (.041)***</td>
<td>-0.020 (.004)***</td>
<td></td>
</tr>
<tr>
<td>Exercising</td>
<td>-0.034 (.035)</td>
<td>-0.003 (.003)</td>
<td></td>
</tr>
<tr>
<td>Last period health status:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good/Very good</td>
<td>0.127 (.048)***</td>
<td>0.012 (.005)***</td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>0.094 (.064)</td>
<td>0.008 (.005)</td>
<td></td>
</tr>
<tr>
<td>Last period diagnosed disease:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.040 (.038)</td>
<td>0.004 (.003)</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.090 (.060)</td>
<td>0.008 (.005)</td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>0.053 (.078)</td>
<td>0.005 (.006)</td>
<td></td>
</tr>
<tr>
<td>Heart disease</td>
<td>0.199 (.063)***</td>
<td>0.016 (.004)***</td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>-0.048 (.114)</td>
<td>-0.004 (.011)</td>
<td></td>
</tr>
<tr>
<td>Lung disease</td>
<td>0.238 (.088)***</td>
<td>0.018 (.005)***</td>
<td></td>
</tr>
<tr>
<td>Psychiatric disease</td>
<td>-0.029 (.059)</td>
<td>-0.003 (.005)</td>
<td></td>
</tr>
<tr>
<td>Arthritis</td>
<td>0.027 (.037)</td>
<td>0.002 (.003)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.162 (.039)***</td>
<td>0.014 (.003)***</td>
<td></td>
</tr>
<tr>
<td>Non-white</td>
<td>-0.173 (.044)***</td>
<td>-0.017 (.005)***</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.524 (.053)***</td>
<td>-0.068 (.009)***</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.135 (.046)***</td>
<td>0.012 (.004)***</td>
<td></td>
</tr>
<tr>
<td>Age^2.10^{-2}</td>
<td>-0.123 (.042)***</td>
<td>-0.011 (.004)***</td>
<td></td>
</tr>
<tr>
<td>Age*(Age≥65)</td>
<td>0.116 (.059)**</td>
<td>0.010 (.005)**</td>
<td></td>
</tr>
<tr>
<td>Education:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or GED</td>
<td>0.297 (.045)***</td>
<td>0.025 (.004)***</td>
<td></td>
</tr>
<tr>
<td>College or above</td>
<td>0.372 (.049)***</td>
<td>0.032 (.004)***</td>
<td></td>
</tr>
<tr>
<td>Log income($1000)</td>
<td>0.282 (.018)***</td>
<td>0.025 (.002)***</td>
<td></td>
</tr>
<tr>
<td>Number of children</td>
<td>-0.024 (.008)***</td>
<td>-0.002 (.001)***</td>
<td></td>
</tr>
<tr>
<td>Not drinking last period</td>
<td>-0.219 (.061)***</td>
<td>-0.018 (.005)***</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-5.07 (1.28)***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.
### Table 3a The insurance effects on the probabilities of visiting a doctor/hospital and participating in health behaviors

<table>
<thead>
<tr>
<th></th>
<th>Coeff. of Insurance(SE)</th>
<th>Marginal Effect(SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visiting a doctor/hospital</td>
<td>.037(.146)</td>
<td>.008(.006)</td>
</tr>
<tr>
<td>Exercising</td>
<td>-.115(.102)</td>
<td>-.046(.041)</td>
</tr>
<tr>
<td>Smoking</td>
<td>.166(.192)</td>
<td>.017(.014)</td>
</tr>
<tr>
<td>Drinking</td>
<td>-.187(.145)</td>
<td>-.065(.053)</td>
</tr>
</tbody>
</table>

### Table 3b The insurance effects on the number of doctor/hospital visits per year and the number of alcoholic drinks consumed per week

<table>
<thead>
<tr>
<th></th>
<th>IHS Percent change (SE)</th>
<th>Change in #</th>
<th>Box-Cox Percent change (SE)</th>
<th>Change in #</th>
<th>Log Percent change (SE)</th>
<th>Change in #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of visits</td>
<td>.367(.069)***</td>
<td>1.67</td>
<td>.365(.069)***</td>
<td>1.66</td>
<td>.368(.056)***</td>
<td>1.68</td>
</tr>
<tr>
<td># of alcoholic drinks</td>
<td>.122(.100)</td>
<td>.900</td>
<td>.120(.098)</td>
<td>.886</td>
<td>.108(.095)</td>
<td>.794</td>
</tr>
<tr>
<td>Model (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of visits</td>
<td>.347(.069)***</td>
<td>1.58</td>
<td>.359(.068)***</td>
<td>1.63</td>
<td>.348(.057)***</td>
<td>1.58</td>
</tr>
<tr>
<td># of alcoholic drinks</td>
<td>.043(.097)</td>
<td>.318</td>
<td>.041(.095)</td>
<td>.302</td>
<td>.029(.093)</td>
<td>.214</td>
</tr>
</tbody>
</table>

1. Model (1): Two-part model, without sample selection bias correction; Model (2): Sample selection model, with sample selection bias correction;
2. * Significant at the 10% level; ** Significant at the 5% level; ***Significant at the 1% level.
Table 4 Insurance effects on heavy drinking

<table>
<thead>
<tr>
<th>Cutoff percentile</th>
<th>Heavy drinking # of drinks</th>
<th>Sample mean</th>
<th>Probability Level (IHS)</th>
<th>Marginal Coeff. of Insurance</th>
<th>Percent increase in #</th>
<th>Level (Box-Cox) Percent increase in #</th>
<th>Level (Log) Percent increase in #</th>
</tr>
</thead>
<tbody>
<tr>
<td>50(^{th})</td>
<td>&gt; 4</td>
<td>12.8</td>
<td>.093 (.175)</td>
<td>.011</td>
<td>.131 (.066)**</td>
<td>.128 (.066)**</td>
<td>.130 (.048)**</td>
</tr>
<tr>
<td>60(^{th})</td>
<td>&gt; 6</td>
<td>14.9</td>
<td>.190 (.165)</td>
<td>.013</td>
<td>.155 (.071)**</td>
<td>.173 (.073)**</td>
<td>.181 (.070)**</td>
</tr>
<tr>
<td>70(^{th})</td>
<td>&gt; 7</td>
<td>16.8</td>
<td>.127 (.168)</td>
<td>.007</td>
<td>.120 (.070)**</td>
<td>.107 (.072)</td>
<td>.126 (.068)**</td>
</tr>
<tr>
<td>80(^{th})</td>
<td>&gt; 12</td>
<td>21.4</td>
<td>.229 (.188)</td>
<td>.005</td>
<td>.177 (.074)**</td>
<td>.141 (.076)**</td>
<td>.180 (.017)**</td>
</tr>
<tr>
<td>85(^{th})</td>
<td>&gt; 14</td>
<td>27.2</td>
<td>.307 (.211)</td>
<td>.003</td>
<td>.232 (.100)**</td>
<td>.176 (.104)**</td>
<td>.239 (.107)**</td>
</tr>
<tr>
<td>90(^{th})</td>
<td>&gt; 15</td>
<td>28.5</td>
<td>.330 (.278)</td>
<td>.002</td>
<td>.245 (.099)**</td>
<td>.218 (.099)**</td>
<td>.246 (.116)**</td>
</tr>
</tbody>
</table>

1. Standard errors are in the parentheses; * Significant at the 10% level; ** Significant at the 5% level; ***Significant at the 1% level;
2. The mean weekly # of alcoholic drinks for all the drinkers is 7.37. Using the mean as the cutoff yields the same results as using the 70\(^{th}\) percentile.
Appendix

I Derivation of the Structural Equations for \( M_t, B_t \) and \( I_t \) Given a General Quasi-linear Utility Function: Eqs. (5) - (7)

Based on backward induction, solve for the last stage choices first: conditioning on the information set \( F_{t+1} = \{ B_t, H_{t+1}, \tau, \nu_{t+1} \} \), the individual chooses \( B_{t+1} \) to maximize his expected utility; i.e.,

\[
\max_{\{ B_{t+1}|F_{t+1} \}} E_{s_{t+1}} \left[ U(B_{t-1}, B_t, H_t, \tau, \nu_t, s_t) + \beta U(B_t, B_{t+1}, H_{t+1}, \tau, \nu_{t+1}, s_{t+1}) \right] + W_t - \{ P_t I_t + P'_B B_t + (1 - d I_t) M_t + \beta P'_B B_{t+1} \}.
\]

(25)

Let \( U_t = U(B_{t-1} , B_t , H_t, \tau, \nu_t, s_t) \) and \( U_{t+1} = U(B_t, B_{t+1}, H_{t+1}, \tau, \nu_{t+1}, s_{t+1}) \). Under certain regularity conditions so that derivative and integral are exchangeable, the first order condition (FOC) for \( B_{t+1} \) is given by

\[
\int_{s_{t+1}} \left( \frac{\partial U_{t+1}}{\partial B_{t+1}} + \frac{\partial s_{t+1}}{\partial B_{t+1}} \frac{\partial U_{t+1}}{\partial s_{t+1}} \right) dF(s_{t+1}|H_{t+1}, B_{t+1}) = P_B,
\]

(26)

where \( F(\cdot|\cdot) \) denotes the conditional distribution function. Solving the above FOC for \( B_{t+1} \) yields the optimal choice \( B^*_t = B^*_t(B_t, H_t, \tau, \nu_t) \), where \( H_{t+1} = \delta H_t + \tilde{H}(M_t, B_t, s_t) \).

In the first period, given his information set \( F_t = \{ B_{t-1}, H_t, I_t, \tau, \nu_t \} \), the individual chooses \( B_t \) and \( M_t \) to maximize his expected utility; i.e.,

\[
\max_{\{ B_t|F_t \}, \{ M_t|F_t, s_t \}} E_{s_{t+1}, \nu_{t+1}, s_{t+1}} \left[ U(B_{t-1}, B_t, H_t, \tau, \nu_t, s_t) + \beta U(B_t, B^*_t, H_{t+1}, \tau, \nu_{t+1}, s_{t+1}) \right] + W_t - \{ P_t I_t + P'_B B_t + (1 - d I_t) M_t + \beta P'_B B^*_t \}.
\]

(27)

Under the same regularity conditions as mentioned above, the FOC for \( M_t \) is given by

\[
\int_{\nu_{t+1}} \int_{s_{t+1}} \beta \frac{\partial H_{t+1}}{\partial M_t} \left\{ \frac{\partial B^*_t}{\partial H_{t+1}} \frac{\partial U_{t+1}}{\partial B^*_t} + \frac{\partial U_{t+1}}{\partial H_{t+1}} \frac{\partial B^*_t}{\partial H_{t+1}} - \frac{\partial B^*_t}{\partial H_{t+1}} P_B + \frac{\partial s_{t+1}}{\partial H_{t+1}} \frac{\partial U_{t+1}}{\partial s_{t+1}} \right\} dF(s_{t+1}|H_{t+1}, B^*_t) dF(\nu_{t+1}) = \beta - d I_t,
\]

(28)

where \( s_t = E(s_t|H_t, B_t) + e_t \) and \( e_t \) is the random component of the health shock \( s_t \), as \( s_t \) depends on \( H_t \) and \( B_t \), and the FOC’s for \( B_t \) are given by

\[
\int_{s_t} \int_{\nu_{t+1}} \int_{s_{t+1}} \left\{ \frac{\partial U_t}{\partial B_t} + \frac{\partial s_t}{\partial B_t} \frac{\partial U_t}{\partial s_t} + \beta \frac{\partial U_{t+1}}{\partial B_t} + \beta \frac{\partial B^*_t}{\partial B_t} \left( \frac{\partial U_{t+1}}{\partial B^*_t} - P_B + \frac{\partial s_{t+1}}{\partial B^*_t} \frac{\partial U_{t+1}}{\partial s_{t+1}} \right) \right\}
\]

34
Given the Quadratic Utility in Section 3.2: Eqs. (10) and (11) and structural equations (28), (30), and (31) correspond to (5) - (7) in the text.

II Derivation of the Structural and Semi-reduced Form

Note that in (31), behaviors to maximize his expected utility, so they are the structural equations for (28) and (30) define how the individual chooses health care utilization and unhealthy behaviors. The negative health effects can be offset by medical care. Assume that the individual takes this into account, then the second parenthesized term can be substituted by (28). Equation (29) can then be written as

\[
\int_{s_t} \int_{\nu_{t+1}} \int_{s_{t+1}} \left\{ \frac{\partial U_{t+1}}{\partial B_t} + \frac{\partial s_{t+1}}{\partial B_t} \frac{\partial U_{t+1}}{\partial s_{t+1}} + \frac{\partial s_{t+1}}{\partial s_{t+1}} \frac{\partial U_{t+1}}{\partial s_{t+1}} + \beta \frac{\partial B_{t+1}}{\partial B_t} \right\} \cdot dF(s_{t+1}|H_{t+1}, B_{t+1})dF(\nu_{t+1})dF(s_t|H_t, B_t) = P_B, \quad (29)
\]

where \( \frac{\partial U_{t+1}}{\partial B_{t+1}} \) represents \( \frac{\partial U_{t+1}}{\partial B_{t+1}} \big|_{B_{t+1}=B_{t+1}^*} \) and \( \frac{\partial s_{t+1}}{\partial B_{t+1}} \) represents \( \frac{\partial s_{t+1}}{\partial B_{t+1}} \big|_{B_{t+1}=B_{t+1}^*} \). The above FOC’s show that the individual chooses \( M_t \) and \( B_t \) so that the marginal utilities equal the marginal costs. Note that the term in the second parentheses in (29) represents the utility loss due to the health effects of unhealthy behaviors. The negative health effects can be offset by medical care. Assume that the individual takes this into account, then the second parenthesized term can be substituted by (28). Equation (29) can then be written as

\[
\int_{s_t} \int_{\nu_{t+1}} \int_{s_{t+1}} \left\{ \frac{\partial U_{t+1}}{\partial B_t} + \frac{\partial s_{t+1}}{\partial B_t} \frac{\partial U_{t+1}}{\partial s_{t+1}} + \beta \frac{\partial B_{t+1}}{\partial B_t} \right\} \cdot dF(s_{t+1}|H_{t+1}, B_{t+1}^*)dF(\nu_{t+1})dF(s_t|H_t, B_t) = P_B - \frac{\tilde{H}_2(M_t, B_t, s_t)}{\tilde{H}_1(M_t, B_t, s_t)}(1 - dI_t), (30)
\]

(28) and (30) define how the individual chooses health care utilization and unhealthy behaviors to maximize his expected utility, so they are the structural equations for \( M_t \) and \( B_t \).

For both \( I_t = 1 \) and \( I_t = 0 \), substitute into the utility function the optimal choices in the following two stages in a backward manner. Denote these functions as \( V_1(\cdot) = V_1(W_t - P_t, B_{t-1}, H_t, \tau, \nu_t, s_t, \nu_{t+1}, s_{t+1}) \) and \( V_0(\cdot) = V_0(W_t, B_{t-1}, H_t, \tau, \nu_t, s_t, \nu_{t+1}, s_{t+1}) \), respectively. The individual buys insurance if and only if he expects \( V_1 \) is greater than \( V_0 \), i.e.,

\[
I_t = \begin{cases} 
1 & \text{iff } E_{s_t,\nu_{t+1},s_{t+1}}[V_1(\cdot)] - E_{s_t,\nu_{t+1},s_{t+1}}[V_0(\cdot)] > 0; \\
0 & \text{otherwise.} 
\end{cases} \quad (31)
\]

Note that in (31) \( W_t \) will drop off because the utility function is quasi-linear. The structural equations (28), (30), and (31) correspond to (5) - (7) in the text.

II Derivation of the Structural and Semi-reduced Form \( B_t \) and \( M_t \) Equations

Given the Quadratic Utility in Section 3.2: Eqs. (10) and (11)

For simplicity, let \( s_t = \zeta_1 H_t + \zeta_2 B_t + e_t \). Given the quadratic utility and linear health production function assumptions in section 4.2, equation (26), the FOC for \( B_{t+1} \), can
be expressed as

\[
\begin{align*}
\textbf{u}_{2,1} \textbf{B}_t + \textbf{u}_{2,2} \textbf{B}^*_t + \textbf{u}_{2,3} \textbf{H}_{t+1} + \textbf{u}_{2,4} \tau + \textbf{u}_{2,5} \nu_{t+1} + \textbf{u}_{2,6} (\zeta_1 \textbf{H}_{t+1} + \zeta'_2 \textbf{B}^*_t)+ \\
\zeta_2 \left[ \textbf{u}_{6,1} \textbf{B}_t + \textbf{u}_{6,2} \textbf{B}^*_t + \textbf{u}_{6,3} \textbf{H}_{t+1} + \textbf{u}_{6,4} \tau + \textbf{u}_{6,5} \nu_{t+1} + \textbf{u}_{6,6} (\zeta_1 \textbf{H}_{t+1} + \zeta'_2 \textbf{B}^*_t) \right] &= \textbf{P}_B. 
\end{align*}
\]

(32)

Solving for \( \textbf{B}^*_t \) yields

\[
\begin{align*}
\textbf{B}^*_t &= (\textbf{u}_{2,2} + \textbf{u}_{2,6} \zeta'_2 + \zeta_2 \textbf{u}_{6,2} + \zeta_2 \textbf{u}_{6,6} \zeta'_2)^{-1} (\textbf{P}_B - (\textbf{u}_{2,1} + \zeta_2 \textbf{u}_{6,1}) \textbf{B}_t) \\
&= -(\textbf{u}_{2,3} + \zeta_2 \textbf{u}_{6,3} + \zeta_2 \textbf{u}_{6,6} \zeta'_1) \textbf{H}_{t+1} - (\textbf{u}_{2,4} + \zeta_2 \textbf{u}_{6,4} \tau - (\textbf{u}_{2,5} + \zeta_2 \textbf{u}_{6,5}) \nu_{t+1}).
\end{align*}
\]

(33)

Equation (28), the FOC for \( \textbf{M}_t \), can then be written as

\[
h_1 c_2 M_t + c_3 \textbf{B}_t + c_4 \tau + h_3 c_2 e_t = (\beta h_1)^{-1}(1-dI_t) + c_1 \textbf{P}_B
\]

(34)

where

\[
\begin{align*}
c_1 &= -(\textbf{u}_{3,2} + \textbf{u}_{6,6} \zeta'_2 + \zeta_1 \textbf{u}_{6,2} + \zeta_1 \textbf{u}_{6,6} \zeta'_2)(\textbf{u}_{2,2} + \textbf{u}_{2,6} \zeta'_2 + \zeta_2 \textbf{u}_{6,2} + \zeta_2 \textbf{u}_{6,6} \zeta'_2)^{-1} \\
c_2 &= c_1[\textbf{u}_{2,3} + \textbf{u}_{6,3} \zeta_2 + \zeta_1 \textbf{u}_{6,2} + \zeta_1 \textbf{u}_{6,6} \zeta'_2] + u_{3,3} + u_{3,6} \zeta_1 + \zeta_1(u_{6,3} + u_{6,6} \zeta_1), \\
c_3 &= (c_1 \textbf{u}_{2,1} + u_{3,1}) + (\zeta_1 + c_1 \zeta_2) \textbf{u}_{6,1} + c_2(h_2 + h_3 \zeta'_2), \\
c_4 &= (c_1 \textbf{u}_{2,4} + u_{3,4}) + (\zeta_1 + c_1 \zeta_2) u_{6,4}.
\end{align*}
\]

Equation (30), the FOC for \( \textbf{B}_t \), can be written as

\[
(\textbf{u}_{2,1} + \zeta_2 \textbf{u}_{6,1}) \textbf{B}_{t-1} + c_5 \textbf{B}_t + c_7 \textbf{M}_t + [c_6 + c_7(\delta + h_3 \zeta_1)] \textbf{H}_t + c_9 \tau
\]

(35)

where

\[
\begin{align*}
c_5 &= -(\textbf{u}_{1,2} + \textbf{u}_{1,6} \zeta'_2)(\textbf{u}_{2,2} + \textbf{u}_{2,6} \zeta'_2 + \zeta_2 \textbf{u}_{6,2} + \zeta_2 \textbf{u}_{6,6} \zeta'_2)^{-1}, \\
c_6 &= u_{2,3} + \zeta_2 \textbf{u}_{6,3} + u_{2,6} \zeta_1 + \zeta_2 \textbf{u}_{6,6} \zeta_1, \\
c_7 &= \beta(u_{1,3} + u_{1,6} \zeta_1 + c_5 c_6), \\
c_8 &= c_7(h'_2 + h_3 \zeta'_2) - c_5^{-1}(\textbf{u}_{1,2} + \textbf{u}_{1,6} \zeta'_2) + \beta \textbf{u}_{1,1} + \beta c_5 (\textbf{u}_{2,1} + \zeta_2 \textbf{u}_{6,1}), \\
c_9 &= u_{2,4} + \zeta_2 \textbf{u}_{6,4} + \beta u_{1,4} + \beta c_5 (u_{2,4} + \zeta_2 \textbf{u}_{6,4}).
\end{align*}
\]

(34) and (35) yield the following structural equations for \( \textbf{M}_t \) and \( \textbf{B}_t \):

\[
\begin{align*}
\textbf{M}_t &= (h_1 c_2)^{-1} \left[ h_1 c_1 \textbf{P}_B + (\beta h_1)^{-1}(1-dI_t) - c_3 \textbf{B}_t - c_2 (\delta + h_3 \zeta_1) \textbf{H}_t - c_4 \tau - h_3 c_2 e_t \right],
\end{align*}
\]

(36)

and

\[
\begin{align*}
\textbf{B}_t &= c_8^{-1} \left[ (1 + \beta c_5) \textbf{P}_B - h_1^{-1}h_2(1-dI_t) - (\textbf{u}_{2,1} + \zeta_2 \textbf{u}_{6,1}) \textbf{B}_{t-1} - c_7 h_1 \textbf{M}_t \\
&= -(c_6 + c_7 \delta + c_7 h_3 \zeta_1) \textbf{H}_t - c_9 \tau - (u_{2,5} + \zeta_2 \textbf{u}_{6,5}) \nu_{t} \right].
\end{align*}
\]

(37)

where \( c_1, \ c_2, \ldots, \ c_9 \) are defined as above. Solving the simultaneous equation system
(36) and (37) for $M_t$ and $B_t$ yields the following semi-reduced form equations:

$$B_t = b_0 + b_1 I_t + b_2 B_{t-1} + b_3 H_t + b_4 \tau + b_5 \nu_t + b_6 e_t,$$

$$M_t = m_0 + m_1 I_t + m_2 B_{t-1} + m_3 H_t + m_4 \tau + m_5 \nu_t + m_6 e_t,$$

where

$$b_0 = (c_8 - c_7 c_2^{-1} c_3)^{-1}[(1 + \beta c_5 - c_7 c_2^{-1} h_1 c_1) P_B - h_1^{-1} h_2 - c_7 (\beta c_2 h_1)^{-1}],$$

$$b_1 = (c_8 - c_7 c_2^{-1} c_3)^{-1} d h_1^{-1} [h_2 + c_7 (\beta c_2)^{-1}],$$

$$b_2 = -(c_8 - c_7 c_2^{-1} c_3)^{-1} (u_{2,1} + \zeta_2 u_{6,1}),$$

$$b_3 = -(c_8 - c_7 c_2^{-1} c_3)^{-1} e_6,$$

$$b_4 = -(c_8 - c_7 c_2^{-1} c_3)^{-1} (c_9 - c_7 c_2^{-1} c_4),$$

$$b_5 = -(c_8 - c_7 c_2^{-1} c_3)^{-1} (u_{2,5} + \zeta_2 u_{6,5}),$$

$$b_6 = c_8^{-1} c_7 h_3,$$

$$m_0 = h_1^{-1} (c_2 - c_3 c_8^{-1} c_7)^{-1} [h_1^{-1} (\beta^{-1} + c_3 c_8^{-1} h_2) + c_1 P_B - c_3 c_8^{-1} (1 + \beta c_5) P_B],$$

$$m_1 = -h_1^{-2} (c_2 - c_3 c_8^{-1} c_7)^{-1} d [\beta^{-1} + c_3 c_8^{-1} h_2],$$

$$m_2 = h_1^{-1} (c_2 - c_3 c_8^{-1} c_7)^{-1} c_3 c_8^{-1} (u_{2,1} + \zeta_2 u_{6,1}),$$

$$m_3 = h_1^{-1} (c_2 - c_3 c_8^{-1} c_7)^{-1} c_3 c_8^{-1} e_6 - h_1^{-1} (\delta + h_3 \zeta_1),$$

$$m_4 = h_1^{-1} (c_2 - c_3 c_8^{-1} c_7)^{-1} (c_3 c_8^{-1} c_9 - c_4),$$

$$m_5 = h_1^{-1} (c_2 - c_3 c_8^{-1} c_7)^{-1} c_3 c_8^{-1} (u_{2,5} + \zeta_2 u_{6,5}),$$

$$m_5 = h_1^{-1} h_3.$$

### III Derivation of the $I_t$ Equation Given the Quadratic Utility: Eq. (12)

The previous section shows that given the functional form assumptions $B_t, M_t, B_{t+1}$, and hence $H_{t+1}$ are all linear functions of $B_{t-1}, H_t, \tau, \nu_t$, as well as $s_t$ and $\nu_{t+1}$. Denote the functions of those variables in the insured state ($I_t = 1$) and the uninsured state ($I_t = 0$) as $B_t^I, M_t^I, B_{t+1}^I$, and $H_{t+1}^I$, for $I = 0, 1$. Similarly, denote health shocks in the two states as $s_t^I$ and $s_{t+1}^I$, as health shocks depend on health behaviors. It is easy to show that $B_t^I - B_t^0, B_{t+1}^I - B_{t+1}^0, H_{t+1}^I - H_{t+1}^0, M_t^I - M_t^0$, and $E[s_t^I | H_t, B_t^I] - E[s_t^0 | H_t, B_t^0], E[s_{t+1}^I | H_{t+1}, B_{t+1}^I] - E[s_{t+1}^0 | H_{t+1}, B_{t+1}^0]$ are all constants.

The difference between the individual’s expected utility when $I_t = 1$ and when $I_t = 0$,

$$E_{s_{t+1}^I, s_{t+1}^0} [V_1(\cdot)] - E_{s_{t+1}^I, s_{t+1}^0} [V_0(\cdot)],$$

can then be written as

$$E_{s_{t+1}^I, s_{t+1}^0} \begin{bmatrix} U(B_{t-1}, B_t^I, H_t, \tau, \nu_t, s_t^I) - U(B_{t-1}, B_t^0, H_t, \tau, \nu_t, s_t^0) \\
+ \beta \left[ U(B_{t+1}^I, B_{t+1}^I, H_{t+1}^I, \tau, \nu_{t+1}, s_{t+1}^I) - U(B_t^0, B_{t+1}^0, H_{t+1}^0, \tau, \nu_{t+1}, s_{t+1}^0) \right] \\
- P_B (B_t^I - B_t^0) - \beta P_B (B_{t+1}^I - B_{t+1}^0) - (M_t^I - M_t^0) + dM_t^I - P_{it} \end{bmatrix}. \quad (40)$$
Since $U(\cdot)$ is quadratic, it is easy to show that the first three terms in the above expression are linear functions of $B_{t-1}, H_t, \tau,$ and $\nu_t$. Further, $B_{t+1}^1 - B_{t+1}^0, B_{t+1}^1 - B_{t+1}^0, M_t^1 - M_t^0$ are constants, and $M_t^1$ is linear in $B_{t-1}, H_t, \tau,$ and $\nu_t$. It follows that (40) is a linear function of $B_{t-1}, H_t, \tau, \nu_t$ and $P_{It}$. Therefore, $I_t$ is given by eq. (12) under the quadratic utility and the linear health production assumptions.
Table AI-1 The impact of transition out of self-employment on health behaviors

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1</th>
<th>Control 1</th>
<th>Diff.-In-Diff.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in drinking (level)</td>
<td>.271 (.300)</td>
<td>-.087 (.197)</td>
<td>.358 (.359)</td>
<td>.319</td>
</tr>
<tr>
<td>Change in drinking (prob.)</td>
<td>-.028 (.021)</td>
<td>-.029 (.014)</td>
<td>.001 (.025)</td>
<td>.965</td>
</tr>
<tr>
<td>Change in smoking (prob.)</td>
<td>-.009 (.008)</td>
<td>-.014 (.007)</td>
<td>.005 (.010)</td>
<td>.637</td>
</tr>
<tr>
<td>Change in exercising (prob.)</td>
<td>-.050 (.033)</td>
<td>-.027 (.018)</td>
<td>-.023 (.038)</td>
<td>.536</td>
</tr>
</tbody>
</table>

1. All the changes refer to those from wave 4 to wave 5; standard errors are in the parentheses;
2. Treatment 1: individuals who transit out of self-employment (n=321); Control 1: individuals who are self-employed in both waves (n=755); Treatment 2: individuals who transit into self-employment (n=175); Control 2: individuals who are not self-employed in both waves (n=10,595)

Table AI-2 The impact of transition into self-employment on health behaviors

<table>
<thead>
<tr>
<th></th>
<th>Treatment 2</th>
<th>Control 2</th>
<th>Diff.-In-Diff.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in drinking (level)</td>
<td>-.509 (.419)</td>
<td>-.207 (.041)</td>
<td>-.301 (.421)</td>
<td>.476</td>
</tr>
<tr>
<td>Change in drinking (prob.)</td>
<td>-.022 (.028)</td>
<td>-.030 (.004)</td>
<td>-.008 (.028)</td>
<td>.787</td>
</tr>
<tr>
<td>Change in smoking (prob.)</td>
<td>-.029 (.015)</td>
<td>-.012 (.002)</td>
<td>-.017 (.015)</td>
<td>.267</td>
</tr>
<tr>
<td>Change in exercising (prob.)</td>
<td>.029 (.041)</td>
<td>-.014 (.005)</td>
<td>.043 (.041)</td>
<td>.300</td>
</tr>
</tbody>
</table>

1. All the changes refer to those from wave 4 to wave 5; standard errors are in the parentheses;
2. Treatment 1: individuals who transit out of self-employment (n=321); Control 1: individuals who are self-employed in both waves (n=755); Treatment 2: individuals who transit into self-employment (n=175); Control 2: individuals who are not self-employed in both waves (n=10,595)
### Table A2 The number of doctor/hospital visits per year (Sample selection model)

<table>
<thead>
<tr>
<th>Transformation on dependent &amp; lagged dependent variables</th>
<th>log</th>
<th>IHS</th>
<th>Box-Cox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health insurance</td>
<td>.348 (.057)**</td>
<td>.347 (.069)**</td>
<td>.359 (.068)**</td>
</tr>
<tr>
<td>Last period behavior:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of alcoholic drinks</td>
<td>.012 (.013)</td>
<td>.011 (.013)</td>
<td>.015 (.017)</td>
</tr>
<tr>
<td>Smoking</td>
<td>-.001 (.025)</td>
<td>.000 (.024)</td>
<td>-.009 (.024)</td>
</tr>
<tr>
<td>Exercising</td>
<td>-.065 (.015)**</td>
<td>-.065 (.015)**</td>
<td>-.065 (.015)**</td>
</tr>
<tr>
<td>Last period health status:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good/Very good</td>
<td>-.275 (.025)**</td>
<td>-.274 (.024)**</td>
<td>-.280 (.024)**</td>
</tr>
<tr>
<td>Excellent</td>
<td>-.457 (.034)**</td>
<td>-.456 (.033)**</td>
<td>-.467 (.033)**</td>
</tr>
<tr>
<td>Last period diagnosed disease:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>.168 (.023)**</td>
<td>.167 (.023)**</td>
<td>.179 (.023)**</td>
</tr>
<tr>
<td>Diabetes</td>
<td>.331 (.029)**</td>
<td>.330 (.029)**</td>
<td>.341 (.029)**</td>
</tr>
<tr>
<td>Cancer</td>
<td>.168 (.031)**</td>
<td>.167 (.032)**</td>
<td>.175 (.032)**</td>
</tr>
<tr>
<td>Heart disease</td>
<td>.197 (.026)**</td>
<td>.196 (.026)**</td>
<td>.204 (.026)**</td>
</tr>
<tr>
<td>Stroke</td>
<td>.212 (.047)**</td>
<td>.211 (.050)**</td>
<td>.217 (.050)**</td>
</tr>
<tr>
<td>Lung disease</td>
<td>.251 (.036)**</td>
<td>.250 (.036)**</td>
<td>.255 (.036)**</td>
</tr>
<tr>
<td>Psychiatric disease</td>
<td>.250 (.030)**</td>
<td>.249 (.029)**</td>
<td>.259 (.029)**</td>
</tr>
<tr>
<td>Arthritis</td>
<td>.163 (.016)**</td>
<td>.163 (.016)**</td>
<td>.166 (.016)**</td>
</tr>
<tr>
<td>Male</td>
<td>-.071 (.020)**</td>
<td>-.070 (.021)**</td>
<td>-.081 (.020)**</td>
</tr>
<tr>
<td>Non-white</td>
<td>.026 (.022)</td>
<td>.026 (.021)</td>
<td>.029 (.021)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-.010 (.033)</td>
<td>-.009 (.032)</td>
<td>-.012 (.032)</td>
</tr>
<tr>
<td>Age</td>
<td>.001 (.024)</td>
<td>.001 (.024)</td>
<td>.016 (.023)</td>
</tr>
<tr>
<td>Age(10^{-2})</td>
<td>-.002 (.022)</td>
<td>-.002 (.021)</td>
<td>-.014 (.021)</td>
</tr>
<tr>
<td>Age-(Age≥65)</td>
<td>.000 (.000)</td>
<td>.000 (.000)</td>
<td>.000 (.000)</td>
</tr>
<tr>
<td>Education:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or GED</td>
<td>.008 (.026)</td>
<td>.007 (.025)</td>
<td>.015 (.025)</td>
</tr>
<tr>
<td>College or above</td>
<td>.085 (.031)**</td>
<td>.084 (.030)**</td>
<td>.096 (.030)**</td>
</tr>
<tr>
<td>Log income($1000)</td>
<td>-.007 (.011)</td>
<td>-.007 (.011)</td>
<td>.001 (.011)</td>
</tr>
<tr>
<td>Number of children</td>
<td>-.005 (.004)</td>
<td>-.005 (.004)</td>
<td>-.004 (.004)</td>
</tr>
<tr>
<td>Not drinking last period</td>
<td>.032 (.025)</td>
<td>.057 (.050)</td>
<td>.033 (.025)</td>
</tr>
<tr>
<td>Constant</td>
<td>.916 (.717)</td>
<td>3.87 (.695)</td>
<td>.350 (.686)</td>
</tr>
<tr>
<td>Lambda</td>
<td>-.814 (.193)**</td>
<td>-.818 (.187)**</td>
<td>-.679 (.187)**</td>
</tr>
<tr>
<td>(\rho)</td>
<td>-.058 (.029)**</td>
<td>-.058 (.059)</td>
<td>-.063 (.059)</td>
</tr>
</tbody>
</table>

1. Standard errors are computed using the Delta method; * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level;
2. \(\rho\) represents the correlation of the unobservables in the health insurance and the outcome equations.
<table>
<thead>
<tr>
<th></th>
<th>Transformation on dependent &amp; lagged dependent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log</td>
</tr>
<tr>
<td>Health insurance</td>
<td>.029 (.093)</td>
</tr>
<tr>
<td>Last period behavior:</td>
<td></td>
</tr>
<tr>
<td># of alcoholic drinks</td>
<td>.882 (.039)***</td>
</tr>
<tr>
<td>Smoking</td>
<td>.201 (.028)***</td>
</tr>
<tr>
<td>Exercising</td>
<td>.015 (.021)</td>
</tr>
<tr>
<td>Last period health status:</td>
<td></td>
</tr>
<tr>
<td>Good/Very good</td>
<td>.008 (.042)</td>
</tr>
<tr>
<td>Excellent</td>
<td>.055 (.050)</td>
</tr>
<tr>
<td>Last period diagnosed disease:</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>-.007 (.024)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>-.147 (.055)***</td>
</tr>
<tr>
<td>Cancer</td>
<td>.027 (.043)</td>
</tr>
<tr>
<td>Heart disease</td>
<td>-.011 (.038)</td>
</tr>
<tr>
<td>Stroke</td>
<td>-.086 (.085)</td>
</tr>
<tr>
<td>Lung disease</td>
<td>.010 (.059)</td>
</tr>
<tr>
<td>Psychiatric disease</td>
<td>-.035 (.046)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>-.015 (.022)</td>
</tr>
<tr>
<td>Male</td>
<td>.252 (.028)***</td>
</tr>
<tr>
<td>Non-white</td>
<td>-.113 (.044)***</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.001 (.060)</td>
</tr>
<tr>
<td>Age</td>
<td>-.068 (.032)***</td>
</tr>
<tr>
<td>Age2·10⁻²</td>
<td>.061 (.028)***</td>
</tr>
<tr>
<td>Age·(Age≥65)</td>
<td>-.002 (.001)***</td>
</tr>
<tr>
<td>Education:</td>
<td></td>
</tr>
<tr>
<td>High school or GED</td>
<td>-.031 (.043)</td>
</tr>
<tr>
<td>College or above</td>
<td>.078 (.048)</td>
</tr>
<tr>
<td>log Income($1000)</td>
<td>.037 (.016)***</td>
</tr>
<tr>
<td>Number of children</td>
<td>-.002 (.006)</td>
</tr>
<tr>
<td>Not drinking last period</td>
<td>-.917 (.197)***</td>
</tr>
<tr>
<td>Constant</td>
<td>1.02 (.917)</td>
</tr>
<tr>
<td>Lambda</td>
<td>1.14 (.190)***</td>
</tr>
<tr>
<td>ρ</td>
<td>-.110 (.061)*</td>
</tr>
<tr>
<td>Transformation parameter:</td>
<td></td>
</tr>
<tr>
<td># of alcoholic drinks (θₐ)</td>
<td>4.24 (5.85)</td>
</tr>
</tbody>
</table>

1. Standard errors are computed using the Delta method; * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.
2. ρ represents the correlation of the unobservables in the health insurance and the outcome equations.
References


Dufour M.C., 1996. Risks and benefits of alcohol use over the life span. Alcohol
Health & Research World 20(3), 145-151.


Maddala, G.S., 1985. A survey of the literature on selectivity bias as it pertains


