Living Rationally Under the Volcano?
An Empirical Analysis of Heavy Drinking and Smoking*

Peter Arcidiacono
Duke University

Holger Sieg
Carnegie Mellon University and NBER

Frank Sloan
Duke University and NBER

May 27, 2005

*We would like to thank Eugene Choo, Ron Gallant, Wilbert van der Klaauw, Tom Mroz, John Pepper, Krista Perreira, V. Kerry Smith, Steve Stern, Don Taylor and seminar participants of seminars at Duke University, Harvard University, the University of Virginia, the Conference for Computational Economics at Yale, the SEA meeting in Washington D. C., the SITE workshop in Stanford, and the Triangle Econometrics Conference. We also thank Lynn van Scoyoc for research assistance. This research is supported in part by the National Institute of Alcohol Abuse and Alcoholism (R01 AA12162-01). Sieg also acknowledges support from the Alfred P. Sloan Foundation.

Abstract

This study investigates whether models of forward looking behavior explain the observed patterns of heavy drinking and smoking of men in late middle age in the Health and Retirement Study better than myopic models. We develop and estimate a sequence of nested models which differ by their degree of forward looking behavior. Our empirical findings suggest that forward looking models fit the data better than myopic models. These models also dominate other behavioral models based on out-of-sample predictions using data of men aged 70 and over. Myopic models predict rates of smoking for old individuals which are significantly larger than those found in the data on elderly men.

1 Introduction

Two competing theories have been proposed for explaining the consumption of potentially harmful and addictive goods. Early approaches typically attributed consumption of these goods to irrational or myopic behavior (Winston, 1980; Thaler and Sheffrin, 1981).[1] More recently, Stigler and Becker (1977) and Becker and Murphy (1988) have forcefully argued that addiction can be modeled as an outcome of rational behavior of forward looking individuals with stable preferences.[2] These two theories primarily differ in their assumptions regarding the length of the planning horizon which is attributed to individuals. The myopic model assumes that the planning horizon is short and consists – in the limiting case – of only one time period. Individuals care about today, but ignore tomorrow and hence do not internalize the negative effects of smoking and drinking on health in the future. In contrast, rational addiction theory relies on the notion that individuals are forward looking. Thus, individuals take into consideration the future risks associated with smoking or heavy drinking.

Most prior empirical studies of the rational addiction theory follow Becker and Murphy (1988) and analyze first order conditions that prices and quantities need to satisfy, given individuals' (quadratic) utility functions.[3] Chaloupka (1991) and Becker, Grossman, and Murphy (1991, 1994) apply this methodology and find that tobacco consumption typically responds to lagged, current and future price changes as predicted by rational addiction theory.[4] However, the empirical literature seems to suggest that price effects are likely to be less important for older individuals, that are the focus of this study, than for teenagers

[1] One of most fascinating early analyses of alcohol abuse is due to Lowry (1947), which was published in its original version with a 17 year delay in the Prairie Schooner, XXXVII, 4, Winter 1963/64.

[2] An alternative to both approaches are models based on recent work by Laibson (1997), Harris and Laibson (2001) using hyperbolic discounting. In these models, individuals are forward looking, but put less weight on future events than in standard forward looking models which can give rise to time inconsistent consumption paths. See also Gruber and Koszegi (2001).

[3] Exceptions are recent work by Gilleskie and Strumpf (2004) who use linear approximation of decision rules in a discrete choice model and Choo (2000) and Khwaja (2001) who estimate models using full-solution dynamic programming techniques.

[4] Chaloupka and Warner (2000) provide an overview of the existing empirical literature of the rational addiction model.

and young adults.[5] Our test of the rational addition model is, therefore, not based on the response of individuals to current and future prices.

This study differs from previous empirical studies in a number of important ways. Previous research has largely focused on young adults who are subject to experimentation, habit formation, and reinforcement (Orphanides and Zervos, 1995). However, an important characteristic of tobacco consumption, in particular consumption of cigarettes, is the long latency period between time of initiation and onset of adverse events.[6] Relatively few adverse health events occur in the first half of life. To illustrate, at age 35, the cumulative probability of survival is the same for males who have never smoked and smokers. At age 45 (65, 85), the corresponding ratio is 1.02 (1.18, 2.11) (Hodgson, 1992). Perfectly forward looking individuals may therefore engage in heavy consumption of harmful substances at young ages because events in the distant future are heavily discounted. It is therefore difficult to distinguish between the two theories using data of young adults.

We focus our analysis on a sample of men in late middle age from the Health and Retirement Study (HRS). Persons over the age of 50 start to experience negative health shocks which are, at least partially, due to smoking and heavy drinking in the past. Much of the uncertainty about future health and the link between smoking or drinking and health outcomes is resolved during later years in life. We can thus study whether individuals rationally update their consumption behavior as they experience negative health shocks and provide a new test of the rational addiction hypothesis.[7] The responses of individuals to negative health shocks, therefore, provides the main sources of variation in the data that we exploit in the empirical analysis.

Earlier studies have typically assumed that individuals are either myopic or forward looking. In contrast, we adopt a dynamic discrete choice framework that allows us to control

[5] See, for example, the discussion in Sloan, Smith, and Taylor (2003).

[6] The latency period for alcohol can be substantially less than for smoking, for example, due to accidents while being intoxicated.

[7] Previous tests of the rational addiction model have been based on Euler equations, which rest on the implicit assumption that individuals must consume strictly positive amounts of alcohol or cigarettes. This is problematic if a large number of individuals do not consume tobacco or alcohol in each time period.

for varying degrees of forward looking behavior of individuals.[8] The myopic model is a special case of forward looking models. Our framework thus nests the competing behavioral theories. Instead of *assuming* that individuals are rational, we compare a sequence of nested models which differ by their degree of forward looking behavior.[9]

The HRS is a sample of older individuals. Due to habit formation and the difficulty to quit, it is desirable to estimate econometric models which condition on previous smoking and drinking histories. Based on measured variables in the HRS, we can constructed variables that partially capture these difference in initial conditions. However, any observed measures of past smoking or drinking are unlikely to capture all heterogeneity within the panel. We therefore assume individuals differ by (partially) unobserved characteristics at the start of the panel using a semi-nonparametric approach which allows for a finite mixture of types, each comprising a fixed proportion of the population.[10] The type probabilities depend on time invariant state variables such as observed measures of past smoking and drinking histories. We assume that the remaining time varying state variables are exogenous conditional on type when analyzing the decisions observed in the HRS. We estimate models with up to four different (partially) unobserved types and report the results in the paper

Our empirical findings suggest that forward looking models with moderately high values of the annual discount factor fit the data the best. To gain additional insights in the fit of the different model specifications, we predict the behavior for a sample of elderly individuals from the Asset and Health Dynamics Among the Oldest Old (AHEAD) data. We find that the out-of-sample predictions of our preferred forward looking model clearly dominate those of the myopic model. We thus conclude that forward looking models provide better within

[8]Dynamic discrete choice estimation was first used by Wolpin (1984), Miller (1984), Pakes (1986) and Rust (1987). Recent applications include Hotz and Miller (1993), Keane and Wolpin (1994, 1997), Rust and Phelan (1997), Gilleskie (1998), Eckstein and Wolpin (1999), Brien, Lillard, and Stern (2000), and Aguirregabiria and Mira (2002).

[9]In that sense our paper is similar in spirit to the analysis by Fang and Silverman (2004) who analyze whether welfare recipients adopt time-inconsistent or type consistent forward looking plans.

[10]This approach was introduced into the econometric literature by Heckman and Singer (1984). Keane and Wolpin (1997) adapted these techniques in a dynamic discrete choice framework. Arcidiacono and Jones (2003) discuss how to implement these estimators using an EM algorithm.

sample fits and out-of-sample predictions than their myopic counterparts.

Comparing myopic models to forward looking models yields substantial differences in the consumption patterns of alcohol and tobacco over time. Larger discount factors imply larger declines in consumption with age. In forward looking models, individuals take into account that the marginal adverse health effects of heavy drinking and smoking are higher later in life. The main drawback of the myopic model is that it predicts rates of smoking and heavy drinking for old individuals which are significantly larger than those found in the AHEAD. Myopic and forward looking models also have very different behavioral interpretations. Estimates of the myopic model make alcohol and tobacco appear unattractive, particularly for the unhealthy. However, in forward looking models, alcohol and tobacco are more attractive, but are not consumed because of losses in future utility caused by adverse health effects. This is particularly true for the unhealthy as the marginal adverse health effects of heavy drinking and smoking increase as the health of the individual deteriorates.

The rest of the study is organized as follows. Section 2 discusses the data used in this study which is based on the Health and Retirement Study. In Section 3, we develop a sequence of forward looking models of decision making under uncertainty and discuss identification and estimation of these models. Section 4 reports the estimation results. In Section 5, we analyze the fit of the models using within sample and out-of-sample predictions. Section 6 discusses how individuals would react to technological changes in health care. Section 7 offers some concluding remarks.

2 Data

The data used in this study come from the Health and Retirement Study. The HRS is a national panel study of birth cohorts 1931 through 1941 and their spouses, if married. Participants in the HRS have been interviewed every two years since 1992. We use the first four waves of the survey that have been completed and released. Individuals in the first wave of the HRS range from 51 to 61 years of age with some spouses being younger or older than this. The average age in wave 1 of individuals in our sample is 58.2. We analyze smoking and heavy drinking behavior of males.

We exclude women for two reasons. First, women differ significantly from males in preferences for alcohol and tobacco, in their attitudes towards risk, and their willingness to engage in harmful consumption. Furthermore, women live longer and have different tolerance levels for alcohol and tobacco than males. Hence, transitions for health status and mortality differ significantly between males and females. Controlling for these differences using additive dummy variables is likely to be insufficient. Second, the computational complexity of the analysis would be a lot more challenging if we also included women since the relevant state space for solving the dynamic model would be much larger.[11]

We use the panel structure of the data and include only persons who were not lost due to attrition and had complete information for the most relevant variables of this study. Our initial sample consists of 5,735 males. 1,281 men are lost due to attrition.

Drinking behavior questions in the HRS allow us to categorize wave 1 respondents as current drinkers or non-drinkers, and for drinkers the average number of drinks per day. Subsequent waves allow for tracking drinking status across time. For purposes of this study, we categorize individuals as drinkers and non-drinkers, and, among drinkers, by drinks per day: less than 1, 1-2, 3-4, and more than 4. We define heavy drinking as consuming three or more drinks on average per day.[12]

The group of non-heavy drinkers thus largely consists of individuals who are best characterized as moderate drinkers, i.e. individuals who consumer two or less drinks of alcohol per week. Thus we are primarily analyzing behavior of moderate and heavy drinkers. A number of individuals in the HRS consume zero drinks in all four waves. These individuals abstain from alcohol consumption for a variety of reasons such as religious objections, medical necessity (i.e. adverse side effects of medications), or lack of taste. Since the HRS does not contain detailed information about the causes of abstinence, we exclude these in-

[11] More research is clearly needed to address the issue of how spouses influence each others smoking, drinking, and health maintenance decisions.

[12] See Graham (1985) for a discussion of alternative measures of alcohol consumption.

dividuals from the sample. This, and missing values, gives us 2,784 males on average in each wave, a total number of 8,352 male person-waves.

Table 1: Sample Means

		Wave 1	Wave 2	Wave 3	Wave 4
Number of drinks:	$= 0$	.157	.226	.260	.306
	< 1	.558	.505	.505	.469
	$1 - 2$	.178	.178	.166	.158
	$3 - 4$	.080	.065	.050	.046
	> 4	.027	.026	.019	.020
Smoking:	yes - no	.281	.256	.230	.203
Health status:	excellent	.238	.200	.187	.128
	very good	.311	.311	.339	.308
	good	.287	.292	.290	.326
	fair	.114	.141	. 135	.172
	poor	.050	.056	.048	.066
Total household income		65,593	58,917	56,232	52,997

Table 1 shows that there is a steady decline in alcohol consumption, smoking, health status, and total income[13] during the four waves of the panel. For example, the percentage of individuals who do not drink increases from 15.7 percent to 30.6 percent during the observation period. Heavy drinking declines from 10.7 percent in wave 1 to 6.6 percent in wave 4.[14] Smoking, meanwhile, declines from 28.1 percent to 20.3 percent. Some of the trends in Table 1 are explained by the aging of the individuals in our sample. Older individuals are more likely to be in worse health and have lower income levels than younger

[13]The HRS includes detailed information on both labor and non-labor income. In this paper, we define total household income as a sum of the respondent's wage income, the spouse's wage income, household capital income and other household income. Income is measured in constant 1999 dollars.

[14]There is an issue of whether alcoholics give valid self-reports regarding their alcohol consumption. While there is some evidence in the literature that alcohol consumption is measured with error, there do not seem to be any systematic trends of over- or underreporting. See, for example, Watson, Tilleskjor, Hoodecheck-Schow, Purcel, and Jacobs (1984).

individuals. As a consequence, these individuals are more likely to reduce their alcohol and cigarette consumption as they get older. The response of individuals in our sample to negative health shocks provides one of the main sources of variation in the data that we exploit in this analysis.

Table 2: Heavy Drinking and Smoking Transitions

		Lagged Consumption			
		d=0,s=0	d=0,s=1	d=1,s=0	d=1,s=1
	d=0,s=0	5535 94.9%	294 16.6%	204 47.6%	21 6.6%
	d=0,s=1	137 2.3%	1384 78.0%	6 1.4%	133 42.0%
Current Consumption	d=1,s=0	151 2.6%	8 0.5%	211 49.2%	17 5.4%
	d=1,s=1	9 0.2%	88 5.0%	8 1.9%	146 46.1%
	Total	5832	1774	429	317

Additional information about alcohol and tobacco consumption is given in Table 2 which reports the transitions between heavy drinking (no, $d = 0$ and yes, $d = 1$) and smoking (no, $s = 0$ and yes, $s = 1$) combinations. We find that the elements on the diagonal of the transition matrix are larger in magnitude than the elements on the off-diagonals, with the elements on the diagonal ranging from 54 and 91 percent. Thus, in spite of the general trend of reducing alcohol and tobacco consumption evident in the sample means, many men have relatively stable drinking and smoking patterns which is likely due to habit formation. However, there is still a fair amount of transitions between the different heavy drinking and smoking states, as reflected by the off-diagonal elements of the matrix.

The HRS also contains some qualitative information about past problem drinking and smoking behavior. This information is used in the empirical analysis to control for past habit formation. The first questionnaire of the HRS includes the CAGE instrument for clinical assessment of alcohol disorders. The acronym CAGE represents four questions that comprise the instrument: Have you ever felt you should *Cut down* on your drinking (32.7 percent)? Have people *Annoyed* you by criticizing your drinking (16.9 percent)? Have you ever felt bad or *Guilty* about drinking (21.1 percent)? Have you ever had a drink first thing in the morning (*Eye-opener*) to calm your nerves or to get over a hang-over (8.7 percent)? Item responses on the CAGE are scored 0 or 1, with a high score indicating the presence of an alcohol problem. We define "problem drinkers" as those persons with a score of 2 or more on the CAGE scale. This is a conservative measure since most of the difference is between 0 and 1. Edwards, Marshall, and Cook (1997) conclude that "these tests (describes another test–MAST as well as CAGE) generally tend to pick up the more extreme rather than the early cases ... they show remarkably good sensitivity and specificity (refers to type 1 and type 2 errors in the language of doctors) for 'excessive drinking' as well as 'alcoholism' and may be superior to laboratory tests when used as screening instruments" (p. 197). 23.6 percent of the males in our sample are classified as problem drinkers.[15]

With regard to past smoking behavior, the HRS asks when former smokers quit. To capture past smoking behavior, we construct a time invariant dummy variable which equals one if the individual quit less than 10 years ago or smoked at wave 1 and is zero otherwise. 34.5 percent of the individuals in the sample are classified as "problem smokers," people who entered the survey with smoking in their recent history.

High levels of alcohol use and smoking are highly correlated in our data and in other samples.[16] Thus, many of the same men who were at risk of poor health because of their alcohol consumption patterns were also at high risk because of their smoking. Although use of each substance carries its own risk when used independently of each other, some joint effects are appreciable.

---

[15]The CAGE testing procedure is also discussed in King (1986) and Bernadt, Taylor, Smith, and Murray (1982).

[16]See, for example, Rosengren, Wilhelmsen, and Wedel (1988), Schlecht, Franco, J. Pintos, and et. al. (1999) and Palfai, Ostafin, Monti, and Hutchinson (2000).

Self-reported health status is measured on a 5-point Likert scale ranging from excellent (1) to poor (5). Table 1 shows the sample means for each of the five categories as they evolve over time. Self-reported health status declines throughout most of the observational period. However, the sample means do not decline monotonically as one may have expected, especially for adjacent response categories, which may be due to measurement error in the variable. We therefore aggregate the information in self-reported health status to a simple bivariate health indicator which is equal to one if the individual is in good health (self-reported health status less than 4) and zero otherwise.[17]

We also control for a number of demographic characteristics. Educational attainment is measured by an indicator variable equal to one if the individual has a college education and is zero otherwise. 25.9 percent of men in our sample have a college degree. We also create a variable for the person's expected longevity. We experimented with different family background variables and chose an indicator based on the life span of the mother. 23.8 percent of the sample report that their mothers died by age 70.[18]

We supplement the HRS data with price information for cigarettes, liquor, beer and wine collected by American Chamber of Commerce Researchers' Association (ACCRA) for a large number different localities. A price index for alcohol is constructed by weighting the prices for liquor, beer and wine using the expenditure shares in the Consumer Price Index. We assign alcohol and cigarette prices to individuals in the HRS sample using prices of the nearest locality for which we have ACCRA data.[19]

[17] For a discussion of self-reported versus objective measures of health see Bound (1991).

[18] Admittedly it would be helpful to know the causes that led to the mother's death. In that case one could construct more informative variables which would distinguish between accidental sources of death, choice related sources of early death, and differences in death due to genetic differences. Unfortunately the HRS does include sufficiently specific information.

[19] For computational reasons, we need to discretized prices for alcohol and tobacco. We split the sample into 16 subsamples and use the mean prices in the subsamples as points in the grid.

3 The Framework

Previous analysis of the rational addiction hypothesis has been typically based on versions of the Becker-Murphy (1988) model which rests on the assumption that individuals make continuous choices. We are mostly interested in non-marginal changes in smoking and heavy drinking which are caused by changes in income and health status. We therefore adapt an alternative modelling framework based on the dynamic discrete choice literature.

3.1 A Model of Forward Looking Behavior

Consider the following model of individual decision making. Let J denote the number of alternatives in the choice set (C_t) available to an individual in a given period. In our model, individuals choose whether or not to smoke and whether or not to engage in heavy drinking or not, $J = 4$.[20] Hence there are four mutually exclusive elements in the choice set at each point of time. Let $c_{jt} \in \{0, 1\}$ denote an indicator variable which equals one if the individual chooses alternative j at time t and is zero otherwise. Let the vector $c_t = (c_{1t}, ..., c_{Jt})$ characterize choices of an individual at t. Since the alternatives are mutually exclusive,

$$\sum_{j=1}^{J} c_{jt} = 1 \tag{1}$$

Besides the choice variables, c_t, there is a vector of state variables, X_t. State variables can be decomposed into an unobserved component, ϵ_t, and an observed component, x_t. Individuals have beliefs about uncertain future states of the world. These beliefs are captured by a Markov transition density of the state variables $q_t(X_{t+1}|X_t, c_t)$, which satisfies the following assumption:

$$q_t(X_{t+1}|X_t, c_t) = g_t(\epsilon_{t+1} \,|\epsilon_t, c_t)\; f_t(x_{t+1} \,|x_t, c_t) \tag{2}$$

[20] As noted in the previous section, heavy drinking and smoking are strongly correlated. Failing to control for either heavy alcohol use or smoking is likely to bias the findings, both regarding the effects of alcohol and tobacco consumption on health, earnings and other outcomes.

In our model, individuals face two types of uncertainty: uncertainty over future health and over future income. First, and most importantly, an individual does not know how his health status will evolve over time. In particular, the relationship between smoking and drinking habits and mortality and morbidity status is stochastic. An individual who engages in heavy drinking and smoking does not necessarily experience bad health outcomes, but rather has a higher probability of experiencing negative health shocks in the future. As individuals experience negative health shocks they will update their believes about the remaining life expectancy and may change the behavior. Thus our model differs significantly from the previous literature that has tried to identify forward looking behavior based on the response of individuals to price changes. Here we are mostly interested in the response of individuals to changes in health status.

More formally, individuals face uncertainty about the evolution of their health status, h_t. We assume that that health status can take on three different values, good, bad and dead. Thus uncertainty about health status also includes uncertainty about death. We assume that the transition probabilities for being in bad health (bh_{t+1}) and mortality status follow a multinomial logit model:[21]

$$\begin{aligned} Pr(death_{t+1}) &= \frac{\exp(x_{ht}\theta^d)}{1+\exp(x_{ht}\theta^d)+\exp(x_{ht}\theta^b)} \\ Pr(bh_{t+1}) &= \frac{\exp(x_{ht}\theta^b)}{1+\exp(x_{ht}\theta^d)+\exp(x_{ht}\theta^b)} \end{aligned} \tag{3}$$

where x_{ht} includes health status at time t, age, and the smoking and drinking choices at time t. Note that consistent estimates of the θ^d's and the θ^b's relies upon no unobserved variables that may affect health status and are correlated with the elements of x_{ht}. We relax this assumption in the next section when we discuss initial conditions and unobserved heterogeneity.

We assume that individuals also face income uncertainty. They must, therefore, forecast

[21] We experimented with ordered logit models of health status which assume that there is one underlying health variable with cut points at bad health and death. This model was too restrictive and did not allow enough individuals to move from good health to death. We also experimented with nested logit models of health status and could not reject the simpler multinomial logit model.

the evolution of income. The transitions for future income are implicitly given by a log-normal regression of income at time $t+1$ on income, health status, age, and smoking and drinking choices at time t.[22] We index the parameters estimated here as θ^y.

We assume that individuals have preferences which can be represented by a time separable (expected) utility function with discount factor β. Let $u_t(c_t, X_t)$ denote the single period utility function which depends on the state and control variables in t. We assume that

$$u_t(c_t, X_t) \;=\; \sum_{j=1}^{J} c_{jt}[u_j(x_t) + \epsilon_{jt}] \tag{4}$$

We use the following specification for part of the utility function which depends on observed state variables:[23]

$$\begin{aligned} u_j(x_t) \;=\; & \alpha_{0j} \;+\; \alpha_{1j}\, \ln(y_t \;-\; e(a_{jt}, c_{jt})) \;+\; \alpha_{2j}\, bh_t \;+\; \alpha_3\, 1\{\text{quit drinking}\} \\ & + \alpha_4\, 1\{\text{quit smoking}\} \end{aligned} \tag{5}$$

where y_t and h_t refer to income and health at time t and $e(d_{jt}, s_{jt})$ refer to expenditures on drinking and smoking given choice j. $1\{\cdot\}$ denotes an indicator function which is equal to one if the event inside the brackets happens and is zero otherwise. This specification implies that each choice has a fixed benefit (α_{0j}). The utility derived from income, α_{1j}, and health, α_{2j}, varies across choices. Utility depends on net income. Unfortunately, the HRS does not contain data on expenditures on alcohol and tobacco; we therefore impute expenditures using the following expression $e(d_{jt}, s_{jt}) = \min\{d_{jt}p_t^d + s_{jt}p_t^s, \phi y_t\}$.[24] Using the minimum of mean expenditures and a fraction of income, ϕy_t, guarantees that imputed expenditures on alcohol and tobacco are not excessively high for low income people who are

[22] To simplify the analysis, we assume that prices evolve deterministically and that individuals have correct point expectations about current and future prices. For a time series analysis of tobacco prices and an investigation of decision making in the presence of price uncertainty, see Coppejans, Gilleskie, Sieg, and Strumpf (2004).

[23] For a discussion of utility functions that depend on health status, see Viscusi and Evans (1990).

[24] d_{jt} (s_{jt}) are the mean quantity of alcohol (tobacco) consumption in the sample for individuals who consume alcohol given choice j. It is zero if choice j is abstaining.

likely to consume cheaper products than the average individual in our sample.

The last two parameters, α_3 and α_4, are designed to capture habit persistence of smoking and heavy drinking. We expect that α_3 (α_4) is negative indicating that quitting heavy drinking (smoking) is costly to individuals because of withdrawal effects. Since it is hard to assign quitting costs to the choices, we prefer to think of these parameters of the utility function as measures of habit persistence and estimate them as part of the parameters of the utility function.[25]

Individuals are rational and forward looking and behave according to an optimal decision rule $\delta_t(X_t) = c_t$ which solves the following maximization problem:

$$\max_{\delta=(\delta_1,\ldots,\delta_T)} E_\delta \left\{ \sum_{t=0}^{T} \beta^t \; u_t(c_t, X_t) \;\middle|\; X_0 = X \right\} \tag{6}$$

where E_δ denotes the expectation with respect to the controlled stochastic process $\{X_t, c_t\}$ induced by the decision rule, δ. This notation is thus consistent with the key feature of the model that individuals do not have perfect foresight and face uncertainty about the impacts of smoking and heavy drinking on health outcomes and mortality.

3.2 Estimation

The parameters of our model consist of the parameters of the utility function and the transition probabilities. Since a Markov decision process yields deterministic decision rules, we need to rely on unobserved state variables to generate a properly defined econometric model. Rust (1987) shows that if the unobserved state variables satisfy the assumptions of additive separability (AS) and conditional independence (CI), conditional choice probabilities are well defined. If we additionally assume that the unobservables in preferences follow a Type

[25]Since discrete choices only depend on relative utility levels, we can normalize a set of coefficients. We set $\alpha_{11} = 1$ and $\alpha_{i1} = 0$ for $i \neq 1$.

I extreme value distribution, we obtain Rust's multinomial dynamic logit specification:

$$P_t(c_{jt} = 1 \mid x_t) \;\; = \;\; \frac{\exp(v_{jt}(x_t, \alpha, \beta, \theta))}{\sum_{i=1}^{J} \exp(v_{it}(x_t, \alpha, \beta, \theta))} \tag{7}$$

where the conditional value function $v_{jt}(x, \theta)$ is recursively defined as:

$$\begin{aligned} v_{jT}(x_T, \alpha, \beta, \theta) \;\; &= \;\; u_{jT}(x_T, \alpha) \\ v_{jt}(x_t, \alpha, \beta, \theta) \;\; &= \;\; u_{jt}(x_t, \alpha) \; + \; \beta \int \left(\log \left[\sum_{j=1}^{J} \exp\{v_{jt+1}(x_{t+1}, \alpha, \beta, \theta)\} \right] \right) \; f(dx_{t+1} | x_t, c_t, \theta) \end{aligned} \tag{8}$$

Substituting these value functions into equation (7) yields the conditional choice probabilities of the dynamic logit model. We observe a panel of N individuals over T periods. The log-likelihood for the full model is then given by:

$$L(\alpha, \beta, \theta) \;\; = \;\; \sum_{n=1}^{N} \sum_{t=1}^{T} L_{cnt}(\alpha, \beta, \theta) \; + L_{hnt}(\theta^d, \theta^b) \; + L_{ynt}(\theta^y) \tag{9}$$

where L_{cnt} is the log-likelihood of the individual n's choices on smoking and heavy drinking at time t, with L_{hnt} and L_{ynt} the corresponding contributions of the health and income transitions. Note that since the log likelihood is additively separable, it is possible to estimate the model in steps.[26] In particular, we could estimate the θ's using only the second two terms of the log likelihood above. Taking these values of the θ's as given, we could then use the first term in the log likelihood only to estimate the α's and β.

3.3 Initial Conditions and Unobserved Heterogeneity

The above likelihood function assumes there are no persistent unobserved variables in the health and income transitions and no unobserved persistent tastes for drinking and smoking. An individual with bad health endowments and strong tastes for smoking and drinking is much more likely to be classified as problem drinker or problem smoker at the beginning of the sample period. Similarly, a highly educated person may be more strongly aware of

[26] See Rust and Phelan (1997) for an example of sequential estimation of dynamic discrete choice problems.

the negative health effects of smoking and drinking or may discount the future less. As a consequence he would be less likely to enter the panel with a strong smoking or drinking history. These examples illustrate that it is problematic to ignore unobserved heterogeneity, especially if it is correlated with initial conditions that characterize previous pattern of habit formation in the sample. Ignoring these problems is likely to cause inconsistent estimators of the parameters of the model.

We therefore assume individuals differ by unobserved characteristics at the start of the panel. Following Heckman and Singer (1984) and Keane and Wolpin (1997), we account for unobserved state variables using a semi-nonparametric approach which allows for a finite mixture of types, each comprising a fixed proportion of the population. These type probabilities depend on all time invariant state variables observed in the beginning of the panel including observed measures of past smoking and drinking histories. More formally, we assume that there are M discrete types of individuals. The probability that an individual in our sample with observed time invariant characteristics z_n is of type m is given by:

$$\pi_n^m = \frac{\exp(\gamma_m z_n)}{1 + \exp(\gamma_m z_n)} \qquad\qquad m = 1, ...M \tag{10}$$

In our application, z_n contains the problem drinker and problem smoker variables, as well as age at wave 1 interacted with lagged smoking and drink choices. It also includes education and a measure of the mother's longevity, whether one's mother died by the age of 70.

An individual's type affects both the health and income transitions as well as his tastes for drinking and smoking. For income, we assume that the intercept term in the log income regression varies across type. For health, we could allow type to have an independent effect on death and bad health. However, since once an individual dies we observe no future health transitions, we instead assume that each additional type adds only one parameter to the health transitions. Namely, we restrict the effect of type on the death and bad health probabilities to be proportional to the corresponding coefficients on lagged bad health. The probabilities of being in death and bad health are thus given by:

$$
\begin{aligned}
Pr(death_{t+1}) &= \frac{\exp(x_{1ht}\theta_1^d + bh_t\theta_2^d + I_m\theta_m\theta_2^d)}{1+\exp(x_{1ht}\theta_1^d + bh_t\theta_2^d + I_m\theta_m\theta_2^d) + \exp(x_{1ht}\theta_1^h + bh_t\theta_2^h + I_m\theta_m\theta_2^h)} \\
Pr(bh_{t+1}) &= \frac{\exp(x_{1ht}\theta_1^h + bh_t\theta_2^h + I_m\theta_m\theta_2^h)}{1+\exp(x_{1ht}\theta_1^d + bh_t\theta_2^d + I_m\theta_m\theta_2^d) + \exp(x_{1ht}\theta_1^h + bh_t\theta_2^h + I_m\theta_m\theta_2^h)}
\end{aligned}
$$

where I_m takes on a value of one if the individual is a member of the mth type and x_{1ht} refers to x_{ht} without the lagged bad health status.

We also allow the value of type to affect the utility of heavy drinking and smoking directly in two ways. First, the intercept terms for smoking and heavy drinking are allowed to vary by type. Second, we allow the discount factor, β, to vary across type in some versions of the model. Adding unobserved heterogeneity then changes the log likelihood function to:

$$L(\alpha,\beta,\theta,\gamma) = \sum_{n=1}^{N} \ln\left(\sum_{m=1}^{M} \pi_m(\gamma) \prod_{t=1}^{T} l_{cntm}(\alpha,\beta,\theta)\; l_{hntm}(\theta^d,\theta^b)\; l_{yntm}(\theta^y)\right) \tag{11}$$

where the likelihoods, l's, are type-specific. Note that the log likelihood is no longer additively separable and hence a two-step estimator is not feasible. Estimation can still be accomplished by maximizing the objective function with respect to all parameters at once. However, this approach is computationally intensive. We pursue, therefore, in this paper an iterative estimator proposed by Arcidiacono and Jones (2003). This estimator is consistent and significantly decreases the required estimation time.[27]

Arcidiacono and Jones (2003) note that one of the techniques for solving mixture models, the Expectation-Maximization (EM) Algorithm, reintroduces the additive separability of the log likelihood function at the maximization step. The EM algorithm works as follows. First, given initial values of the parameters we calculate the conditional probability that an individual is a particular type. Using these conditional probabilities as weights, we treat type as observed and maximize the now additively separable log likelihood function. Given

[27] This algorithm has been used in similar situations by Arcidiacono (2004) and Arcidiacono (2005).

the new parameter estimates, we then update the conditional probabilities of being each of the types and iterate until convergence.

More formally, given values for the α's, β's, θ's, and γ's, the conditional probability of individual n being the mth type follows from Bayes' rule:

$$P_{nm} = \frac{\pi_m(\gamma) \prod_{t=1}^{T} l_{cntm}(\alpha, \beta, \theta)\ l_{hntm}(\theta^d, \theta^b)\ l_{yntm}(\theta^y)}{\sum_{m=1}^{M} \pi_m(\gamma) \prod_{t=1}^{T} l_{cntm}(\alpha, \beta, \theta)\ l_{hntm}(\theta^d, \theta^b)\ l_{yntm}(\theta^y)} \tag{12}$$

We then maximize the expected log likelihood function:

$$\sum_{n=1}^{N} \sum_{m=1}^{M} \sum_{t=1}^{T} P_{nm} \ \ (L_{cntm}(\alpha, \beta, \theta) + L_{hntm}(\theta^d, \theta^b) + L_{yntm}(\theta^y)) \tag{13}$$

Taking the P_{nm}'s as given, we can estimate the θ's using only the last two terms. Given values for the θ's, we can then estimate the α's and the β's. With the estimates of the α's, β's, and θ's, we can estimate the γ's by maximizing (11) with respect to the γ's, taking the l's (the likelihoods) as given. Once we have the γ's, we update the conditional type probabilities and again proceed with the sequential maximization, iterating on these steps until convergence.

3.4 Myopic Behavior

As discussed above, heavy drinking and smoking can also stem from myopic behavior which implies that individuals ignore future risks of harmful consumption goods. The myopic model is nested in the forward looking model described above. In the limiting case of the myopic model, individuals ignore the future.[28] We can obtain this version of the myopic model by setting the discount factor β equal to zero. The choice probabilities are then given by the simple (static) logit model:

$$P_t(c_{jt} = 1 \mid x_t) \quad = \quad \frac{\exp(u_{jt}(x_t, \theta))}{\sum_{i=1}^{J} \exp(u_{it}(x_t, \theta))} \tag{14}$$

[28]We condition our analysis on behavior prior to age 50. Hence, our analysis does not allow us to make statements about behavior of younger adults who may be more likely to engage in myopic behavior.

More generally, we can thus distinguish between myopic and forward looking models based on the value of β. Low estimated values of β suggest myopia; large values of the two-period discount factor β in the magnitude of 0.9 are an indication of "fully rational" behavior.[29]

3.5 Identification of the Discount Factor and Model Selection

Before we proceed and report the main empirical findings of this paper it is useful to discuss the main differences between myopic and forward looking behavior in our context. In a forward looking model individuals will trade-off the short term benefits of the consumption of alcohol and smoking – measured by the per period utility function – with the long term costs – measured by the expected future value function. These long-term costs result because excessive smoking and alcohol consumption make it more likely that an individual will be in bad health. Smoking and heavy drinking also increases the mortality risks and hence lowers life expectancies. The costs largely depend on the evolution of the health status which is stochastic. Thus in a forward looking model smoking and heavy drinking may be quite attractive activities. They may yield a higher instantaneous utility than abstaining. However, individuals will primarily stay away from these activities because they recognize that the greater health risks significantly lower the expected future value associated with these activities.

In contrast, myopic individuals will largely abstain from smoking and heavy drinking because these options are unattractive to them. They result in low values of the instantaneous utility function. As a consequence, if we fit forward looking and myopic models to the same data, we expect to obtain quite different parameter estimates. To explain the same observed pattern in the sample, myopic and forward looking models will have to be based on different parameter estimates. The discount factor is thus primarily identified in our paper from the observed smoking-age and heavy drinking-age profiles in the data. By

[29] The utility function in the simple logit model can also be interpreted as a linear approximation of the conditional value function. Following that line of reasoning, we are then comparing dynamic models based on full solution algorithms with those based on linear approximations of value functions. It is hard to distinguish between the two alternative interpretations of the simple logit model.

excluding age as an explanatory variable in the instantaneous utility function, we let the data determine the level of forward looking behavior which largely determines the relative values of each option in the choice set. There is thus a close link the smoking-age and drinking age-profiles and the rate of time preference if one does not place age directly into the utility function.

The discussion also implies that myopic and forward looking models, evaluated at their parameter estimates, will have different out-of-sample predictions. We can also evaluate the validity of each behavioral hypothesis by conducting a horse race and analyzing which model has better out-of-sample predictive properties.

4 Estimation Results

We estimated a variety of models with different number of types. Here we only report the estimates of models with four types.[30] Tables 3 through 6 summarize the main findings for four different model specifications. The first model is a myopic model that is obtained by setting the discount factor equal to 0.0. The fully rational model uses a two-period discount rate of 0.9. We also estimate an unconstrained dynamic model by searching over a grid for β. The last model that we consider also allows for heterogeneity in discount factors among the four types of individuals. Tables 3 and 4 report the results for the transition probabilities. Table 5 summarizes the parameter estimates for the utility function. Finally, Table 6 reports the estimates for the type probabilities.

The behavior predicted by the forward looking model depends on beliefs that individuals hold about the evolution of the main state variables. In our model, individuals have beliefs about transitions of their health status, mortality, and income. Estimating these transitions is important because smoking and heavy drinking are likely to reduce the life span of an individual, and thus reduce expected lifetime utility. Out-of-sample predictions of survival rates are needed to solve the computational dynamic programming model.

[30] Results for one and two type models are available upon request from the authors.

Table 3: Parameter Estimates of the Health Transitions

	β=0		β=0.9		Estimated β		Heterogeneous β	
	Coeff	Std. Err	Coeff	Std. Err	Coeff	Std. Err	Coeff	Std. Err
Bad Health								
Constant	-2.867	0.0997	-2.8429	0.0991	-2.8526	0.0994	-2.8541	0.0995
Lagged Bad Health	2.7923	0.0709	2.7928	0.0709	2.793	0.0709	2.7926	0.0709
Age	0.0871	0.0129	0.0862	0.0129	0.0865	0.0129	0.0864	0.0129
Lagged Drinking	-0.0241	0.1333	-0.0171	0.1322	-0.0187	0.1328	-0.0183	0.1326
Lagged Smoking	0.1432	0.0809	0.1347	0.0815	0.1395	0.0812	0.1386	0.0812
Death								
Constant	-6.3415	0.2426	-6.3167	0.242	-6.3264	0.2423	-6.3274	0.2423
Lagged Bad Health	2.8554	0.1477	2.8555	0.1479	2.8556	0.1478	2.8546	0.1478
Age	0.2562	0.0269	0.2554	0.0269	0.2556	0.0269	0.2555	0.0269
Lagged Drinking	-0.0691	0.2545	-0.0613	0.2538	-0.0632	0.2542	-0.0628	0.2540
Lagged Smoking	0.7357	0.1609	0.7258	0.1613	0.7311	0.1611	0.7302	0.1611
Type=2	0.1482	0.0453	0.1402	0.0447	0.1431	0.0451	0.1438	0.0449
Type=3	0.2478	0.0310	0.2419	0.0310	0.2439	0.0310	0.2448	0.0310
Type=4	-0.2193	0.0426	-0.2303	0.0427	-0.2233	0.0425	-0.2247	0.0427

Accurate estimation of these transitions is therefore an important component of the overall modelling strategy. We assume that individuals subjective beliefs correspond to probability measures that can be estimated based on observed data.[31]

Table 3 reports the results for the health and death transitions. Individuals who are in bad health are much more likely to face bad health outcomes in the future. Similarly, as individuals age they are more likely to experience bad health outcomes or die. The same holds true for smoking as smoking today makes bad health more likely in the future, with even stronger effects on the probability of dying. Since there is a large negative intercept term for death, significant death probabilities only result when an individual is very old or in bad health. Hence, the marginal effect of smoking on the probability of dying is much higher at seventy than at fifty. This will then lead to a downward sloping age-smoking profile, at least at ages below ninety. The type effects indicate that type 1 and 4 have better health transitions than types 2 and 3. As we will see in the discussion of type probabilities, type 4 individuals tend to be those who graduate from college while type 1's find smoking and heavy drinking unattractive in general. The parameter estimates are robust across the four model specifications.

The one puzzling result is the lack of an effect of heavy drinking on either bad health or death. Previous medical studies have documented the potential negative health effects of heavy drinking among the elderly. Equal doses of alcohol produce higher blood alcohol levels among the elderly compared to the young. Heavy alcohol use increase the risks of injuries and illnesses, including falls, depression, and cognitive impairment, and mortality (Andreasson, 1998). The elderly are also more susceptible to the ill effects of alcoholism because of medication and drug-alcohol interactions (Adams, Gary, Rhyne, Hunt, and Goodwin, 1990). We do not see these effects here even though these would be consistent with the downward sloping age-drinking profiles in the data. However, the standard errors are large and economically significant effects can be found within the ninety-five percent confidence intervals of the estimates .

[31] Sloan, Smith, and Taylor (2001) provide some evidence that individuals correctly update their beliefs in response to health shocks.

Table 4: Parameter Estimates of the Income Function

	β=0		β=0.9		Estimated β		Heterogeneous β	
	Coeff	Std. Err	Coeff	Std. Err	Coeff	Std. Err	Coeff	Std. Err
Constant	5.4447	0.1185	5.4487	0.1185	5.4459	0.1185	5.4476	0.1185
Lagged Income	0.5026	0.0104	0.5021	0.0104	0.5024	0.0104	0.5023	0.0104
Lagged Bad Health	-0.2432	0.0247	-0.243	0.0246	-0.2432	0.0246	-0.243	0.0246
Age	-0.0587	0.0037	-0.0586	0.0037	-0.0586	0.0037	-0.0586	0.0037
Lagged Drinking	-0.0677	0.0396	-0.0623	0.0393	-0.0643	0.0395	-0.0631	0.0394
Lagged Smoking	-0.0476	0.024	-0.044	0.0242	-0.0457	0.0241	-0.0452	0.0240
Type=2	0.0789	0.0371	0.0717	0.0367	0.0749	0.037	0.0727	0.0369
Type=3	-0.0213	0.0247	-0.0214	0.0249	-0.0208	0.0248	-0.0214	0.0248
Type=4	0.3332	0.0264	0.3368	0.0264	0.3334	0.0263	0.3349	0.0264

Estimating beliefs that individuals hold about income transitions is also complex. Current income of individuals obviously is largely affected by current and past labor market participation, retirement and saving decisions. Modelling these decisions within a well-defined dynamic programming model is exceedingly complicated as documented, for example, by Rust and Phelan (1997). We do not seek to improve upon these efforts. Instead, we analyze interactions between consumption of various harmful goods and their health effects. We therefore adapt a reduced form approach for modelling the income process. The main approach for modelling income transitions is based on log-normal income regressions. Estimates of the transition parameters of the income equation are given in Table 4.

In general, our findings are quite reasonable. The parameters have the expected sign in almost all cases and are often estimated precisely. Not surprisingly, income follows a strong autoregressive process.[32] Health status and age are significant in all three models and have the expected negative sign. Heavy drinking and smoking negatively affect future income. Types 2 and 4 have higher labor market skills than types 1 and 3, with particularly high incomes for Type 4's who are primarily college graduates.

We estimate the preference parameters of the utility function specified in equation (5) using myopic and forward looking model specifications. Most of the coefficients in the utility functions have the expected sign and are statistically significant. Our results suggest that, on average, lower income individuals are more likely to engage in smoking and heavy drinking.[33]

[32]Income is measured as total household income and therefore one would expect a lower coefficient than estimates typically found for labor income.

[33]The maximum fraction of income that an individual can spend on alcohol and tobacco is set at 15%.

Table 5: Parameter Estimates of the Utility Function

		β=0		β=0.9		Estimated β		Heterogeneous β	
		Coeff	Std. Err	Coeff	Std. Err	Coeff	Std. Err	Coeff	Std. Err
	Two-year β	0		0.9		0.8192	0.05152	0.9933	0.0292
	Type 2 Two-year β							0.6851	0.1456
	Type 3 Two-year β							0.7354	0.0838
	Type 4 Two-year β							0.8208	0.0681
	Constant	-13.6817	5.8207	-4.7333	0.987	-4.7101	1.0242	-3.5658	1.2025
d=1,s=1	Ln Income	0.7563	0.0824	0.807	0.0684	0.8034	0.0698	0.7934	0.0715
	Bad Health	-0.2907	0.1954	0.2121	0.1745	0.0557	0.1942	0.0092	0.1913
	Constant	-2.6736	0.6527	-1.2057	0.5793	-1.2795	0.5913	-1.0814	0.6004
d=1,s=0	Ln Income	0.8062	0.0606	0.825	0.0535	0.823	0.0543	0.8233	0.0553
	Bad Health	-0.3677	0.1608	-0.4184	0.1384	-0.4035	0.141	-0.3991	0.1468
	Constant	-9.5438	5.7733	-1.899	0.8009	-1.8512	0.8157	-0.9353	1.0204
d=0,s=1	Ln Income	0.823	0.0505	0.8414	0.0375	0.8427	0.0391	0.8351	0.0401
	Bad Health	-0.3011	0.1179	0.364	0.116	0.1824	0.1533	0.1286	0.1454
	Quitting Cost-Drinking	-1.6785	0.1212	-1.7369	0.1197	-1.7148	0.1205	-1.7232	0.1204
	Quitting Cost-Smoking	-4.5368	0.1033	-4.4754	0.1028	-4.5024	0.1029	-4.5072	0.1029
	Type 2-Abstain Alc	-8.8634	5.7477	-3.6128	0.5626	-3.2602	0.5773	-2.0123	1.014
	Type 2-Abstain Smoke	-3.5922	0.2253	-2.8462	0.21	-2.8594	0.2074	-2.5073	0.2808
	Type 3-Abstain Alc	-8.6261	5.7473	-3.651	0.5606	-3.2662	0.5774	-2.1591	0.9681
	Type 3-Abstain Smoke	-1.6559	0.225	-1.4104	0.1983	-1.3525	0.1963	-1.0267	0.2513
	Type 4-Abstain Alc	-7.9795	5.7472	-3.3455	0.5817	-2.9373	0.5962	-2.1336	0.9501
	Type 4-Abstain Smoke	0.3256	0.3902	0.4311	0.3702	0.5865	0.3856	0.7379	0.4125
	Log-likelihood	17199		17195		17194		17192	

Including variables for habit formation improves the fit of the model considerably. There are substantial costs for quitting smoking or heavy drinking. Quitting smoking, however, is much harder than quitting heavy drinking. The parameter estimates of the utility function differ significantly across model specifications. The parameter estimates of the instantaneous utility function indicate that smoking or heavy drinking are more attractive alternatives in forward looking models. This finding is due to the fact that myopic models do not account for the fact that the future expected utility of abstaining is higher due to improved health and longer life expectancies.

The finding that there are only small differences in the estimated quitting costs between myopic and the forward looking models may at first seem puzzling. One potential explanation for this result is as follows. If the true data generating process involves forward looking behavior, estimating a myopic model will still be approximating the results from a forward looking model. We find much lower intercept terms for drinking and smoking in the utility function for the myopic model. When a myopic model is used, the differences in expected future utility across the choices are largely captured by the intercepts. Hence, the myopic model shows that individuals generally do not have a preference for drinking and smoking. In contrast, the forward looking models indicate that individuals enjoy drinking and smoking. But they do not engage in these activities because of the future health risks. The lagged variables themselves pick up the same level of persistence in the drinking and smoking decisions that are revealed in the data regardless of whether we estimate a myopic model or a forward looking model.

The values of the likelihood functions indicate that the forward looking model ($\beta = 0.9$) fits the data slightly better than the myopic model ($\beta = 0.0$). The unconstrained estimate of the model yields a point estimate for the two-period discount factor β of approximately 0.82, which translates into an annual discount factor of approximately 0.91.[34] Based on the likelihood values, we conclude that forward looking models provide a slightly better fit than

[34] Discount rates both higher and lower have been reported in the literature. These studies use different approaches for estimating discount rates than we do. See Hausman (1979), Dreyfus and Viscusi (1995), Moore and Viscusi (1988, 1990), and Warner and Pleeter (2001).

myopic models.

We have argued before that controlling for unobserved heterogeneity and differences in initial conditions is important for obtaining reliable parameter estimates. The results in Tables 3 - 5 clearly indicate that there are substantial differences among the four types. Moreover, we also find that the parameter estimates for the discount factor may be seriously biased downward if one does not account for unobserved heterogeneity. For example, the point estimate in the one-type model for the two-year discount factor is equal to 0.59 which is almost 25% lower than the point estimate obtained in the four type model. Finally, we also estimated a model which allowed for heterogeneity in the discount factors among the four types. The last two columns in Table 5 indicate that the two-year discount factors vary from 0.685 to 0.993 which suggests individuals may have different time preferences.

Finally, we consider the estimation results which measure the impact of observed time invariant characteristics on the type probabilities. These results are summarized in Table 6 for all four model specifications. We find that there are many commonalities among the specifications. Types 2, 3, and 4 are all more likely to enter the panel with significant histories of smoking and heavy drinking.

Educated individuals have typically high levels of income and thus have access to better health care. As a consequence they have longer life expectancies, and thus smoking or heavy drinking may be less attractive for them. More educated individuals may also be more aware of the negative health effects of smoking. We explore this hypothesis and include education in the type probabilities. We find that education has a strong impact on the type probabilities. Hence there is some evidence that smoking patterns among older individuals differ by educational backgrounds. The type probabilities depend to a lesser degree on the longevity of the mother thus picking up differences in health endowments. These differences are also reflected in the point estimates reported in Table 3.

In summary, we find that forward looking models fit the data slightly better than myopic models. The models also yield different estimates for key parameters of the model. In particular, estimates of myopic models imply that the instantaneous utility of smoking and

Table 6: Parameter Estimates: Type Probabilities

Model	Variable	Type=2	Type=3	Type=4
Two-Year β=0	Constant	-2.618	-1.9969	-19.9151
	Problem Drink	1.0132	2.4038	1.2658
	Problem Smoke	11.0168	19.1057	10.2435
	Mother Dead by 70	-1.2098	1.0836	-0.7443
	College	0.9138	-25.00	21.361
	Drink Wave 1	6.0235	1.3318	1.0363
	Smoke Wave 1	20.1204	8.563	19.2468
	Drink Wave 1×Age Wave 1	-0.3455	-2.8825	-0.2868
	Smoke Wave 1×Age Wave 1	-1.2077	1.9818	-1.3763
	Mean Prob.	0.1339	0.3633	0.1758
Two-year β=.9	Constant	-2.6269	-2.1111	-19.8829
	Problem Drink	0.9877	2.3918	1.3047
	Problem Smoke	10.7948	19.5158	10.0378
	Mother Dead by 70	-1.2497	1.0989	-0.7655
	College	1.0251	-25.00	21.3974
	Drink Wave 1	6.169	0.9259	1.2546
	Smoke Wave 1	20.2845	7.9617	19.6808
	Drink Wave 1×Age Wave 1	-0.3713	-2.899	-0.3134
	Smoke Wave 1×Age Wave 1 e	-1.1878	2.1078	-1.4888
	Mean Prob.	0.1362	0.3566	0.1756
Two-year β Estimated	Constant	-2.6669	-2.0512	-19.8732
	Problem Drink	1.0565	2.4166	1.3124
	Problem Smoke	10.8846	19.3771	10.0877
	Mother Dead by 70	-1.2672	1.083	-0.7546
	College	1.0773	-25.00	21.4057
	Drink Wave 1	6.2174	1.115	1.2783
	Smoke Wave 1	20.2076	8.0965	19.6299
	Drink Wave 1×Age Wave 1	-0.3736	-2.8989	-0.3102
	Smoke Wave 1×Age Wave 1	-1.1963	2.0689	-1.4589
	Mean Prob.	0.1345	0.3603	0.177
Two-year β Heterogeneous	Constant	-2.6385	-2.0386	-19.8848
	Problem Drink	1.0473	2.4057	1.3267
	Problem Smoke	10.8674	19.3722	10.1171
	Mother Dead by 70	-1.2614	1.1019	-0.7633
	College	1.0514	-25.00	21.3973
	Drink Wave 1	6.2029	1.114	1.2989
	Smoke Wave 1	20.209	8.105	19.6232
	Drink Wave 1×Age Wave 1	-0.3707	-2.8963	-0.3118
	Smoke Wave 1×Age Wave 1	-1.1823	2.0729	-1.4595
	Mean Prob.	0.1355	0.3612	0.1763

heavy drinking is lower than the one in forward looking models. As a consequence, we expect these models to have different predictive properties. We investigate these issues in detail in the next section.

5 Goodness of Fit and Out-of-Sample Predictions

We consider both within sample fit and out-of sample predictions of the models estimated above. Table 7 reports the predicted probabilities of smoking and heavy drinking for the observations from waves 2-4 of the HRS and compares these predictions to the actual data. Table 7 suggests that there are only small differences in the models' within-sample predictions. All of the models match the overall trend in the data reasonably well.

Table 7: Within Sample Fit

		Wave 2	Wave 3	Wave 4
Proportion Drinking	Myopic	.085	.075	.068
	Fully Rational	.084	.075	.069
	Estimated Beta	.084	.075	.068
	Hetero. Beta	.084	.075	.068
	Data	.091	.069	.066
Proportion Smoking	Myopic	.253	.222	.205
	Fully Rational	.257	.222	.202
	Estimated Beta	.256	.222	.202
	Hetero. Beta	.256	.222	.202
	Data	.256	.230	.203

A more interesting comparison of the different models examines out-of sample predictions. In this study, we focus on predicted smoking profiles as a function of age. This allows us to evaluate the models since we do not include age as an explanatory variable in the preferences. We run the following experiment. We take the set of individuals in our data and forecast out the probabilities of smoking at each age. We condition on being alive by dividing the unconditional probabilities of smoking by the mean probability of being alive at a particular age. Individuals with a high probability of dying are weighted less in later years than those who have a low probability of dying.

We plot in Figure 1 the age-smoking profiles for HRS data as well as data from the Asset and Health Dynamics Among the Oldest Old (AHEAD). AHEAD is a companion data set to HRS which contains individuals that were aged 70 or over in 1993, and their spouses who could be any age. In particular, the HRS data are used for ages 52 to 72, and then the first 3 waves of the AHEAD data are attached for ages 72 to 90. One caution is necessary when examining the AHEAD data: the model was estimated with permanent abstainers removed from the HRS; this was not possible for the AHEAD data.

Large differences across the models are apparent when examining the age-smoking profiles. The myopic model predicts fairly flat profiles. As individuals age, they become more unhealthy which makes smoking less attractive. However, they also become poorer which makes smoking more attractive. In contrast, the fully rational model predicts sharp declines in smoking rates. Here, individuals smoke more than in the myopic model up until the age of 62. The decline in smoking rates occurs because the marginal adverse health effects of smoking increase as individuals age. After the age of 80 the myopic model shows decreases in smoking rates, the fully rational model shows *increases* in smoking rates. This occurs because of an end-of-life effect. Individuals expect to die soon. Hence they are more likely to ignore the negative health effects of smoking and heavy drinking. The model in which the discount factor is estimated yields profiles in between the fully rational and myopic model for ages 60 through 85. After age 85, the model predicts lower smoking rates than either of the myopic or the fully rational model. Finally, the model with heterogeneous discount factors predicts an even steeper profile. This model clearly matches the trends in the data best, both in and out of sample.

Finally, we consider heavy drinking pattern as a function of age. The results are plotted in Figure 2. Our findings suggest that all of our models do not match the trend in the

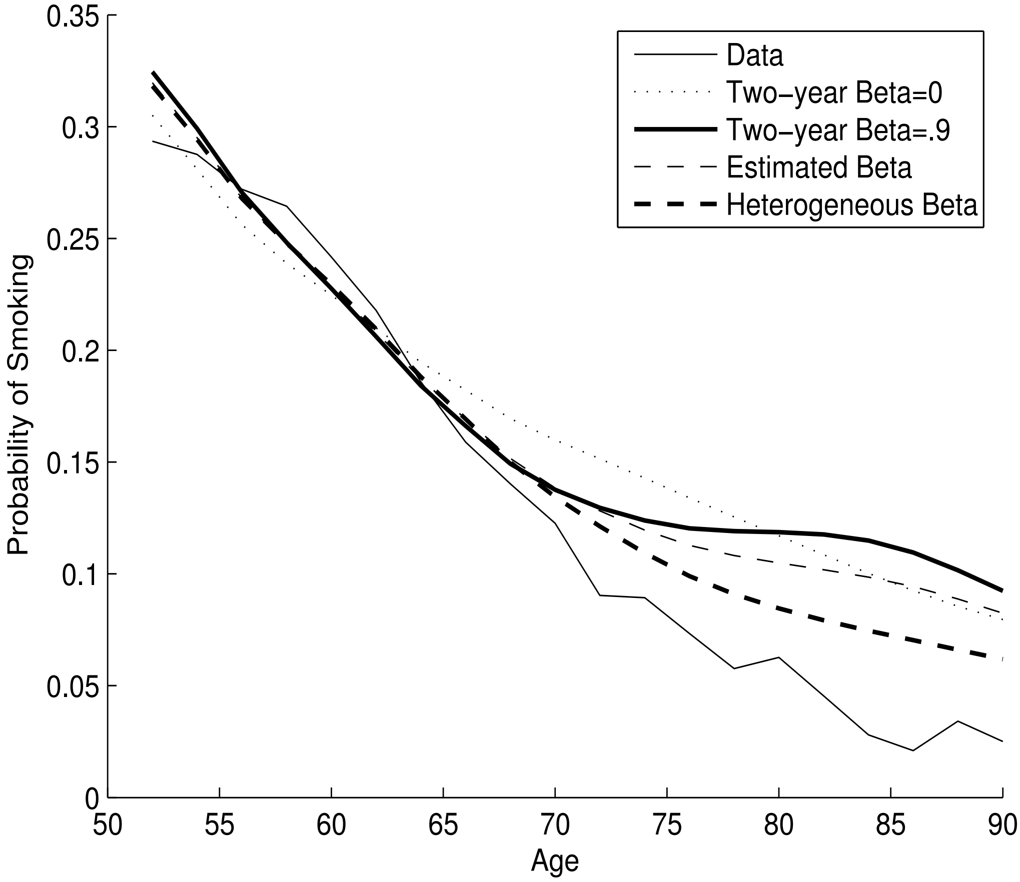


Figure 1: Age-smoking profiles under different discount factors

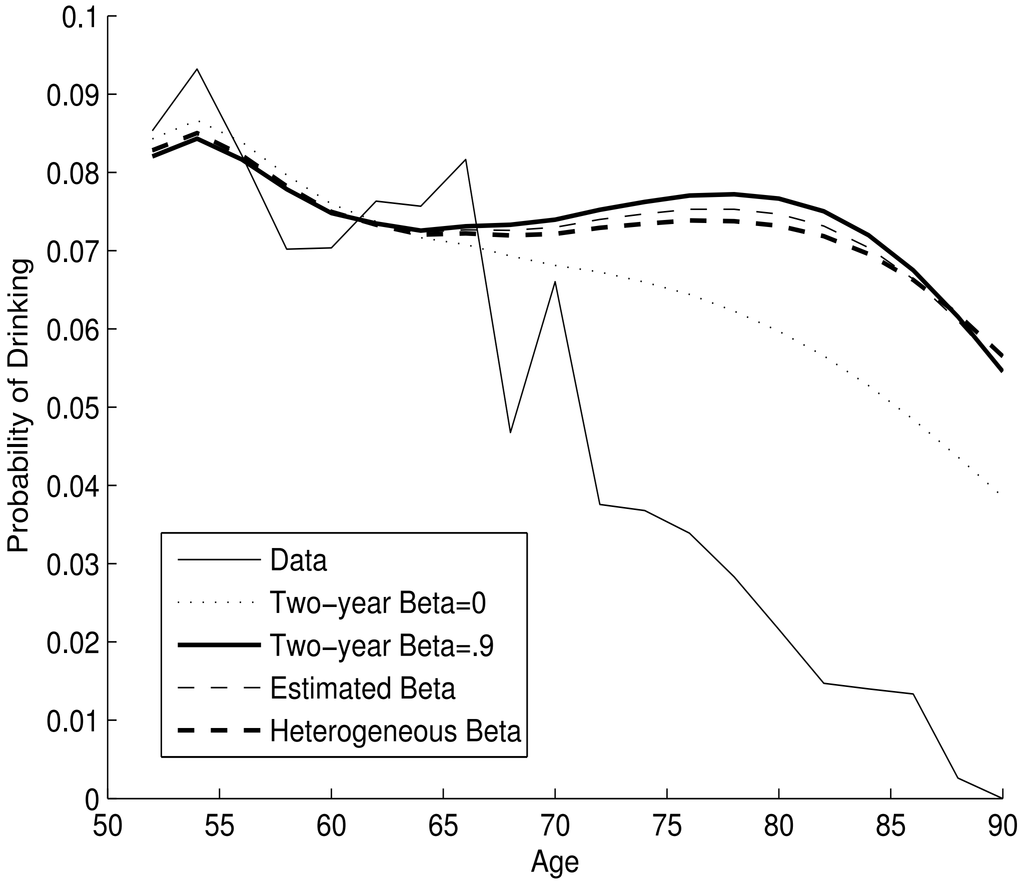


Figure 2: Age-drinking profiles under different discount factors

AHEAD well. We find that both the myopic and the estimated discount factor model predict very flat age-drinking profiles. This corresponds to the drinking pattern found in the HRS. The fully rational model predicts declining alcohol consumption until the age 78 and increasing consumption after that. The data show sharp declines in alcohol consumption later in the life-cycle.

There are two plausible explanations for the lack of predictive power of the alcohol-age profiles of all our models. First, our estimates imply that heavy drinking has only a small impact on health outcomes. As a consequence all of our models have a hard time explaining why individuals quit as they get older. If we used higher parameter estimates in the health transitions for the impact of heavy drinking, the forward-looking models would fit the trend in the data much better. Second, the lack of fit may also be due to a deficiency in the construction of the data. AHEAD did not have the information needed to remove permanent abstainers. This would obviously affect the profiles for heavy drinking more than for smoking. It is thus likely that the observed patterns in AHEAD sample are not good predictors for future drinking for the HRS sample used in the estimation of our model.

6 Implications of the Model

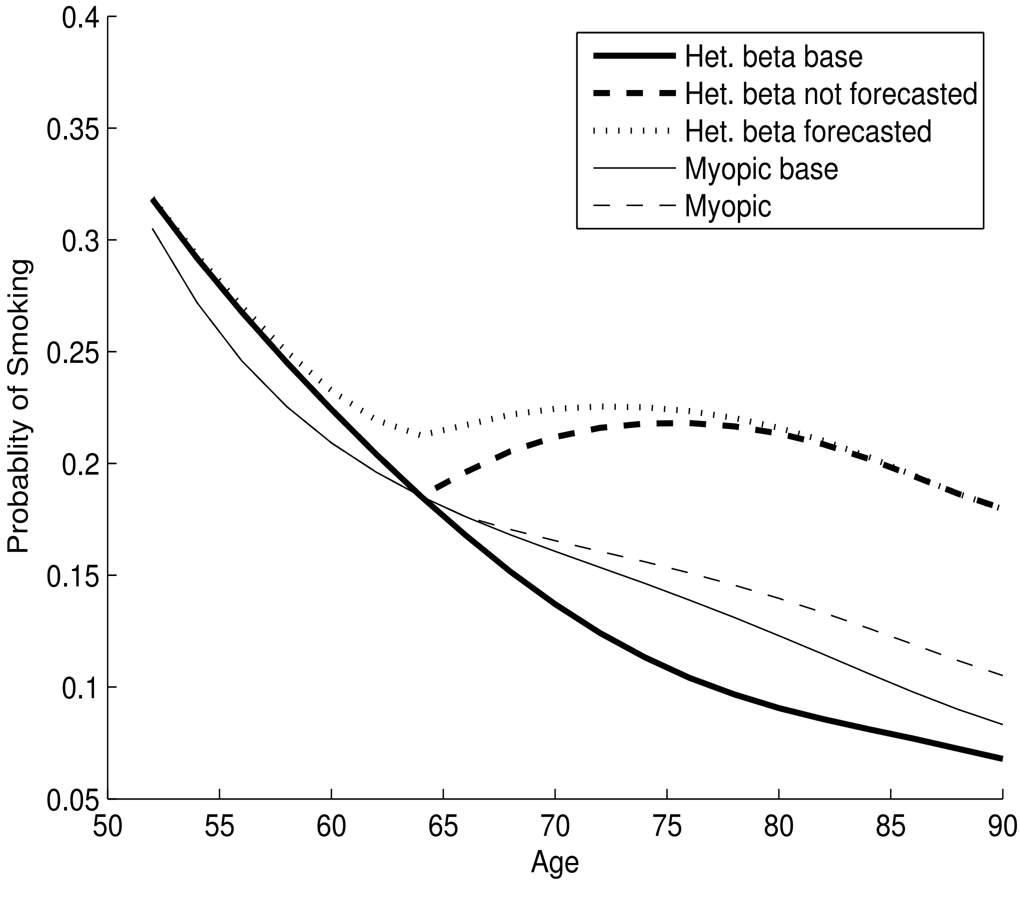
The different models also lead to differential responses to either public policy or advances in medical technology. Here we focus on how the age-smoking profile change given advances in medical technology. We consider a case in which smoking is less taxing on one's health.[35] In our experiment we focus on 52 year old individuals and assume that at age 65 a medical advance occurs such that the coefficients on smoking in the health transitions fall by fifty percent.

We plot the predicted age-smoking profiles of five models in Figure 3. The baseline forward looking model with heterogeneous discount factors is given by the bold line. If

[35] We also investigated the effects of price changes on smoking and heavy drinking behavior. Increasing the price of either alcohol or cigarettes leads to increases in the demand for the other suggesting that heavy drinking and smoking serve as substitutes.

individuals anticipate the advance in technology, the results are given by the dotted line. If the advance is a surprise, the results are given by the bold dashed line. The myopic baseline model is given by the solid line. The myopic model after the change in technology is given by the dashed line. Note that the myopic model has the same predictions regardless of whether the medical advance is anticipated or not.

Figure 3: Age-smoking profiles with a medical advance



First consider the case in which the medical advance comes as a surprise and is thus not anticipated. Not surprisingly, Figure 3 shows that the behavior until age 65 is not affected by an unanticipated medical advance in both myopic and forward-looking models.

However, there are pronounced behavior responses starting at age 65. The forward looking model predicts smoking rates start to increase at age 65. This occurs because individuals in the forward looking model like to smoke but avoid it because of the negative health consequences. Relaxing the negative consequences then increases smoking rates. In contrast, the medical advance has no direct effect on the evaluation of the different alternatives in the myopic model. Smoking rates increase in the myopic model after the medical advance because individuals are less likely to be in bad health. This (indirect) effect is, however, small in comparison to the changes in behavior in the forward looking model.

Figure 3 also shows that smoking rates increase in a forward-looking model before age 65 when the medical advance is anticipated. Individuals know that the expected health costs of smoking will be smaller in the future. They anticipate the future benefits of the medical advance. As a consequence smoking is a more attractive option even before the medical innovation occurs.

Given the large differences in behavioral responses, we also expect differences in the predicted death rates. Figure 4 plots the percent change in the probability of dying from moving from the base case to the case with the medical advance.[36] We find that death rates actually increase before age 65 when the medical advance is anticipated. Individuals forecasting the medical advance increase their smoking rates before age 65. This leads to a higher probability of death. Death rates drop the most in the myopic model since the myopic model induces the smallest change in smoking behavior. More individuals smoke in forward looking models in the post-advance regime than in the myopic model.

The death rates actually increase in the forward looking models after age 75. This occurs in part because people who would have died earlier are now living longer. However, the increased death rates are also due to the higher smoking rates. In fact, a crossing point occurs: the overall probability of living until age 86 is higher with the medical advance but is lower post age 86 because of the higher smoking rates. This does not occur in the myopic

[36]The comparison group for the forward looking models is the model with heterogeneous discount factors without the advance while the comparison group for the myopic model is the myopic model without the advance.

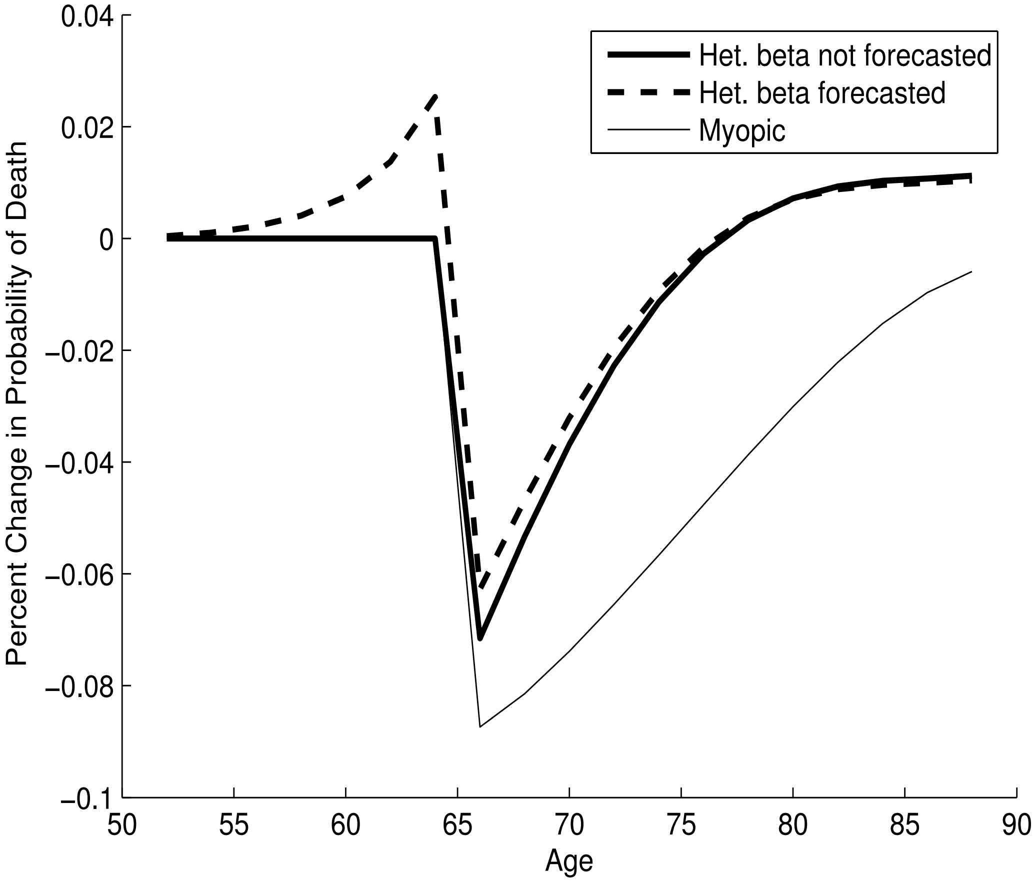


Figure 4: Changes in the probability of dying with a medical advance

model because of the lack of a behavioral response to the medical advance.

7 Conclusions

This study has analyzed smoking and heavy drinking in a sample of elderly individuals drawn from the HRS. The main objective of the analysis has been to explore differences between myopic and forward looking models. We have investigated whether models of forward looking behavior explain the main regularities found in the data better than myopic models. Earlier studies have typically assumed that individuals are either myopic or forward looking. In contrast, we provide an empirical framework which allows us to control for varying degrees of forward looking behavior of individuals and, therefore, nest the competing theories. Estimating the discount factor is desirable and helps us to distinguish between the competing hypotheses which have been put forward to explain the consumption of harmful and addictive goods. Our analysis suggests that forward looking models which control for observed habit formation, unobserved heterogeneity and differences in initial conditions fit the data the best. They also have the best predictive power for a sample of mean above age 70. Our analysis thus provides strong evidence for the hypothesis that older individuals are forward looking and take future risks associated with smoking and heavy drinking into consideration when determining their choices.

Our analysis relies on a number of simplifying assumptions. Models based on hyperbolic discounting present an alternative to the two types of models considered here. Our framework could be extended to estimate a behavioral model with hyperbolic discount factors. Future research, with longer panels, should investigate whether these hybrid models provide better fits of the data than models we analyzed. Although estimating hyperbolic-discounting models are feasible, in principle, additional problems arise because of the need to estimate two separate discount factors. Given the availability of only short panels, it may prove to be quite challenging to precisely estimate these types of models.[37] Nevertheless, our results

[37] Fang and Silverman (2004) provide some evidence which suggests the presence of hyperbolic discounting among welfare recipients.

are quite promising for further work which combines formal decision-theoretic analysis and estimation to address questions regarding the degree of forward looking behavior imposed in modelling the consumption of potentially harmful and addictive goods.

References

Adams, W., Gary, P., Rhyne, R., Hunt, W., and Goodwin, J. (1990). Alcohol Intake in the Healthy Elderly. *JAGS*, *38*, 211–216.

Aguirregabiria, V. and Mira, P. (2002). Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models. *Econometrica*, *70 (4)*, 1519–1544.

Andreasson, S. (1998). Alcohol and J-Shaped Curves. *Alcoholism: Clinical and Experimental Research*, *22*, 359–364.

Arcidiacono, P. (2004). Ability Sorting and the Returns to College Major. *Journal of Econometrics*, *121*, 343–375.

Arcidiacono, P. (2005). Affirmative Action in Higher Education: How do Admission and Financial Aid Rules Affect Future Earnings?. *Econometrica*. forthcoming.

Arcidiacono, P. and Jones, J. (2003). Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm. *Econometrica*, *71 (3)*, 933–946.

Becker, G., Grossman, M., and Murphy, K. (1991). Rational Addiction and the Effect of Price on Consumption. *American Economic Review*, *81 (2)*, 237–242.

Becker, G., Grossman, M., and Murphy, K. (1994). An Empirical Analysis of Cigarette Addiction. *American Economic Review*, *84 (3)*, 396–418.

Becker, G. and Murphy, K. (1988). A Theory of Rational Addiction. *Journal of Political Economy*, *96 (4)*, 675–700.

Bernadt, M., Taylor, J. M. C., Smith, B., and Murray, R. (1982). Comparison of Questionnaire and Laboratory Tests in the Detection of Excessive Drinking and Alcoholism. *Lancet*, 325–328.

Bound, J. (1991). Self-reported and Objective Measures of Health in Retirement Models. *Journal of Human Resources*, *26*, 106–138.

Brien, M., Lillard, L., and Stern, S. (2000). Cohabitation, Marriage, Divorce in a Model of Match Quality. University of Virginia Working Paper.

Chaloupka, F. (1991). Rational Addictive Behavior and Cigarette Smoking. *Journal of Political Economy*, *99 (4)*, 722–742.

Chaloupka, F. and Warner, K. (2000). The Economics of Smoking. In *Handbook of Health Economics*, pp. 1151–1227. North Holland.

Choo, E. (2000). Rational Addiction and Rational Cessation: A Dynamic Structural Model of Cigarette Consumption. Working Paper.

Coppejans, M., Gilleskie, D., Sieg, H., and Strumpf, K. (2004). Consumer Demand in Markets with High Price Volatility. Working Paper.

Dreyfus, M. and Viscusi, K. (1995). Preferences and Consumer Valuations of Automobile Safety and Fuel Efficiency. *Journal of Law and Economics*, *38(1)*, 79–105.

Eckstein, Z. and Wolpin, K. (1999). Why Youth Drop out of High School: The Impact of Preferences, Opportunities and Abilities. *Econometrica*, *67*, 1295–1340.

Edwards, G., Marshall, E., and Cook, C. (1997). *The Treatment of Drinking Problems: A Guide for the Helping Professions*. Cambridge University Press. Cambridge.

Fang, H. and Silverman, D. (2004). Time-inconsistency and Welfare Program Participation: Evidence from the NLSY. Working Paper.

Gilleskie, D. (1998). A Dynamic Stochastic Model of Medical Care Use and Work Absence. *Econometrica*, *66*, 1–46.

Gilleskie, D. and Strumpf, K. (2004). The Behavioral Dynamics of Youth Smoking. Working Paper.

Graham, K. (1985). Identifying and Measuring Alcohol Abuse among the Elderly: Serious Problems with Existing Instrumentation. *Journal of the Studies of Alcohol*, *47*, 322–326.

Gruber, J. and Koszegi, B. (2001). Is Addiction Rational: Theory and Evidence. *Quarterly Journal of Economics*, *116(4)*, 1261–1304.

Harris, C. and Laibson, D. (2001). Dynamic Choice of Hyperbolic Consumers. *Econometrica*, *69 (4)*, 935–958.

Hausman, J. (1979). Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables. *Bell Journal of Economics*, *10(1)*, 33–54.

Heckman, J. and Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data. *Econometrica*, *52 (2)*, 271–320.

Hodgson, T. (1992). Cigarette Smoking and Lifetime Medical Expenditures. *The Milbank Quarterly*, *70(1)*, 85–125.

Hotz, J. and Miller, R. (1993). Conditional Choice Probabilities and Estimation of Dynamic Models. *Review of Economic Studies*, *60*, 497–529.

Keane, M. and Wolpin, K. (1994). The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *Review of Economics and Statistics*, *76*, 648–672.

Keane, M. and Wolpin, K. (1997). The Carrer Decisions of Young Men. *Journal of Political Economy*, *105(3)*, 473–523.

Khwaja, A. (2001). Health Insurance, Habits and health Outcomes: A Dynamic Stochastic Model of Investments in Health. Working Paper.

King, M. (1986). At Risk Drinking Among General Practice Attenders: Validation of the CAGE Questionnaire. *Psychological Medicine*, *16*, 213–217.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics*, *112(2)*, 443–77.

Lowry, M. (1947). *Under the Volcano*. Reynal and Hitchcock. New York.

Miller, R. (1984). Job Matching and Occupational Choice. *Journal of Political Economy*, *92 (6)*, 1086–1120.

Moore, M. and Viscusi, K. (1988). The Quality-Adjusted Value of Life. *Economic Inquiry*, *26(3)*, 369–388.

Moore, M. and Viscusi, K. (1990). Models for Estimating Discount Rates in Long-Term Health Risks Using Labor Market Data. *Journal of Risk and Uncertainty*, *31(4)*, 381–401.

Orphanides, A. and Zervos, D. (1995). Rational Addiction with Learning and Regret. *Journal of Political Economy*, *103 (4)*, 739–758.

Pakes, A. (1986). Patents as Options: Some Estimates of the Value of Holding European Patent Stocks. *Econometrica*, *54*, 755–785.

Palfai, T., Ostafin, B., Monti, P., and Hutchinson, K. (2000). Effects of Nicotine Deprivation on Alcohol-Related Information Processing and Drinking Behavior. *Journal of Abnormal Psychology*, *109*, 96–105.

Rosengren, A., Wilhelmsen, L., and Wedel, H. (1988). Separate and Combined Effects of Smoking and Alcohol Abuse in Middle-aged Men. *Acta Med Scand*, *223*, 111–118.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurchner. *Econometrica*, *55 (5)*, 999–1033.

Rust, J. and Phelan, C. (1997). How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets. *Econometrica*, *65*, 781–832.

Schlecht, N., Franco, E., J. Pintos, J., and et. al. (1999). Interaction between Tobacco and Alcohol Consumption and the Risk of Cancers of the Upper Aero-digestive Tract in Brazil. *American Journal of Epidemiology*, *150*, 1129–37.

Sloan, F., Smith, V., and Taylor, D. (2001). Longevity Expectations and Death: Can People Predict Their Own Demise?. *American Economic Review*, *91 (4)*, 1126–1134.

Sloan, F., Smith, V., and Taylor, D. (2003). *Parsing the Smoking Puzzle: Information, Risk Perception and Choice*. Harvard University Press. Cambridge.

Stigler, G. and Becker, G. (1977). De Gustibus Non Est Disputandum. *American Economic Review*, *67*, 76–90.

Thaler, R. and Sheffrin, H. (1981). An Economic Theory of Self-control. *Journal of Political Economy*, *89(2)*, 392–406.

Viscusi, K. and Evans, W. (1990). Utility Functions that Depend on Health Status: Estimates and Economic Implications. *American Economic Review*, *80*, 353–375.

Warner, J. and Pleeter, S. (2001). The Personal Discount Rate: Evidence from Military Downsizing Programs. *American Economic Review*, *91(1)*, 33–53.

Watson, C., Tilleskjor, C., Hoodecheck-Schow, E., Purcel, J., and Jacobs, L. (1984). Do Alcoholics Give Valid Self-Reports?. *Journal of Studies on Alcohol*, *45*, 344–348.

Winston, G. (1980). Addiction and Backsliding: The Theory of Compulsive Consumption. *Journal of Economic Behavior and Organization*, *1*, 295–324.

Wolpin, K. (1984). An Estimable Dynamic Stochastic Model of Fertility and Child Mortality. *Journal of Political Economy*, *92*, 852–874.