Essays on Health Economics

by

Yang Wang

Department of Economics
Duke University

Date: ______________________

Approved:

____
Frank A. Sloan, Advisor

____
Peter Arcidiacono

____
Patrick Bayer

____
Hanming Fang

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2009

Abstract

(Economics)

Essays on Health Economics

by

Yang Wang

Department of Economics
Duke University

Date: ____________________

Approved:

____
Frank A. Sloan, Advisor

____
Peter Arcidiacono

____
Patrick Bayer

____
Hanming Fang

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2009

Abstract

In this dissertation, I discuss two important factors in individuals' decision-making processes: subjective expectation bias and time-inconsistent preferences. In Chapter I, I look at how individuals' own subjective expectations about certain future events are different from what actually happens in the future, even after controlling for individuals' private information. This difference, which is defined as the expectation bias in this paper, is found to have important influence on individuals' choices. Specifically, I look into the relationship between US elderly's subjective longevity expectation biases and their smoking choices. I find that US elderly tend to over-emphasize the importance of their genetic makeup but underestimate the influence of their health-related choices, such as smoking, on their longevity. This finding can partially explain why even though US elderly are found to be more concerned with their health and more forward-looking than we would have concluded using a model which does not allow for subjective expectation bias, we still observe many smokers. The policy simulation further confirms that if certain public policies can be designed to correct individuals' expectation biases about the effects of their genes and health-related choices on their longevity, then the average smoking rate for the age group analyzed in this paper will go down by about 4%.

In Chapter II, my co-author, Hanming Fang, and I look at one possible explanation to the under-utilization of preventive health care in the United States: procrastination. Procrastination, the phenomenon that individuals postpone certain decisions which incur instantaneous costs but bring long-term benefits, is captured in economics by hyperbolic discount factors and the corresponding time-inconsistent preferences. This chapter extends the semi-parametric identification and estimation method for dynamic discrete choice models using Hotz and Miller's (1993) conditional choice probability approach to the setting where individuals may have hyperbolic discounting time preferences and may be naive about their time inconsistency. We implement the proposed estimation method to US adult women's decisions of undertaking mammography tests to evaluate the importance of present bias and naivety in the under-utilization of mammography, controlling for other potentially important explanatory factors such as age, race, household income, and marital status. Preliminary results show evidence for both present bias and naivety in adult women's decisions of undertaking mammography tests. Using the parameters estimated, we further conduct some policy simulations to quantify the effects of the present bias and naivety on the utilization of preventive health care in the US.

To my parents for their unconditional love and unfaltering support.

Contents

List of Tables

List of Figures

Acknowledgements

I am indebted to my committee members, Frank A. Sloan, Peter Arcidiacono, Patrick Bayer, and Hanming Fang for their invaluable support and guidance. I am grateful to Jason Blevins, Ralph Boleslavsky, Han Hong and participants at the 2008 Annual Meeting of American Society of Health Economists, the 6th Annual Meeting of Southeast Health Economics Study Group, the Triangle Health Economics Workshop, and the Applied Microeconometrics Lunch Group at Duke University. All remaining errors are mine. Comments are welcome. Contact information: Department of Economics, Duke University, 213 Social Sciences Building, P.O. Box 90097, Durham, NC 27708-0097. Email: yang.wang@duke.edu.

1

Subjective Expectations: Test for Bias and Implications for Choices

I develop a new method to assess the rational expectations assumption using subjective expectations data. This approach allows the researcher to identify the distributions of the true underlying objective expectations and the expectation bias with respect to unobserved private information. I incorporate subjective longevity expectations into a structural dynamic discrete choice model to explain the smoking behavior of the elderly in the U.S. I find that subjective longevity expectations have important implications for smoking choices. Specifically, the elderly tend to overemphasize the importance of their genetic makeup but underestimate the effects on their longevity of their health-related choices, such as smoking. I also find that the elderly are more concerned with their health and are more forward-looking than under the rational expectations assumption. Policy experiments show that without the individuals' expectation bias the average smoking rate would decrease by 4%.

1.1 Introduction

Expectations about future events play a central role in the decision-making process of individuals who consider dynamic implications of their current choices. How to model these expectations, therefore, is a fundamental issue for economic understanding of individual behavior. Indeed, because observed choice data alone could be consistent with various combinations of expectations and preferences, model predictions and policy implications depend crucially on the underlying assumptions about individual expectations ([68]).

The existing procedures to recover individuals' preferences from the observed choices are typically based on the rational expectations assumption.[1] This assumption states that agents use all relevant information in forming expectations about economic variables, and their expectations do not systematically differ from the realized outcomes. That is, although the future is not fully predictable, people do not make systematic mistakes when predicting the future, and deviations from actual results are due to random errors. In reality, however, individuals' subjective expectations about the future might be biased, and thus cannot be fully captured by the data on observed outcomes only.[2] Assuming rational expectations in these cases can lead to model misspecification and deceptive conclusions about individual preferences, which have important implications for economic analysis of the decision-making processes and public policy outcomes.

In this paper, I propose a new theoretical framework to assess the rational expectations assumption using elicited subjective expectations data. First, I establish the conditions for identification of the distributions of the true underlying objective expectations and the expectation bias. I then incorporate subjective expectations

[1] The rational expectations assumption was proposed by [76] and has been widely used in economics ever since. Recent studies assuming rational expectations include [80], [52], [32], and [3].

[2] See [89] for a brief review of different explanations for the existence of individual bias.

data directly into a structural model to infer individual preferences and explain individual behavior. As an application, I show that subjective longevity expectations have important implications for the smoking choices of the elderly in the U.S.

Using the subjective expectations elicited from the agents in economic models of individual behavior is intuitive and economically appealing. However, this approach faces several important challenges. First, subjective expectations data are not widely available for a range of important variables. Even when subjective expectations data are available, they are difficult to evaluate because of the possible measurement errors and the existence of private information. Further, using subjective expectations data to estimate a model of individual choices can be difficult, as it often requests good knowledge of the formation of subjective expectations; in addition, there might be a mismatch between the elicited subjective expectations and the subjective information required by the model.

Much effort has been taken to address these challenges. Since the 1990's, in addition to some small-scale surveys, various large-scale surveys have started eliciting subjective expectations directly from respondents. Furthermore, economists have provided various explanations and solutions to account for the seemingly unreasonable responses.[3] These developments have spurred a promising new literature on subjective expectations.

One strand of this new literature assesses the elicited subjective expectations, primarily by comparing subjective expectations with historical or actual realizations at the mean or individual level.[4] The main findings have been quite positive about subjective expectations. Specifically, there are strong reasons to believe that individuals

[3] See, e.g., [46], [11], [31], [75] and [8].

[4] For example, [36] and [45] compare the means of elicited subjective longevity expectations with life tables. [21], [92], [93], and [29] compare adults or teens' mean expectations about certain life events with realizations of a different but comparable population. [22], [9], [48], [46], and [84] compare individuals' expectations with their own realizations on weekly earnings, timing of retirement, onsets of health conditions and mortality.

give meaningful answers when their subjective expectations are elicited. Furthermore, the changes in individuals' expectations over time respond in a qualitatively reasonable way to important life events.

The main underlying assumption in these studies is that there is no information in respondents' information sets which is not observed by the econometrician. The question of whether subjective expectations are biased is thus transformed into the question of whether there is any difference between subjective expectations and their "objective" counterparts constructed by the econometrician using only publicly observed information. The conclusions, therefore, depend crucially on the information available to, and used by, the econometrician to form these "objective" counterparts.[5] Such methods, however, cannot deal with the possibility that individuals, when forming their subjective expectations, possess private information that is not observed by the econometrician but affects their subjective expectations in important ways.

In this paper, I show that subjective expectations data combined with observed realizations can be used to identify the key features of the distributions of the true underlying objective expectations and the expectation bias, which is defined as the difference between the subjective expectations and the true underlying objective expectations. In particular, under certain orthogonality assumptions, the whole distributions of the true underlying objective expectations and the expectation bias can be identified with respect to the unobserved private information.

I further show that after bringing subjective expectations data into a classic binary choice model, it is still possible to identify the model parameters and the distributions of the true underlying objective expectations and the expectation bias, under certain conditions. I then use the data on subjective longevity expectations and smoking decisions of the elderly in the U.S. from the Health and Retirement Study

[5] [17] develop an alternative approach to deal with individuals' unobserved information.

(HRS) in a structural and dynamic discrete choice model. In this model, individuals choose from a finite set of actions to maximize their lifetime utilities based on their expectations of future state transitions ([79]). To model the formation of individuals' subjective expectations about events multiple periods ahead, I set up an Expectation Tree, which solves the mismatch problem between the subjective expectations needed in the model and the ones available in the data.

The World Health Organization (World Health Report, 2002) estimates that more than four million people die from the consequences of smoking each year. The Centers for Disease Control and Prevention therefore describe smoking as the "most important preventable risk to human health" and the world's leading cause of preventable premature deaths (See, also, [82]). Due to its profound effects on people's overall welfare, smoking behavior has been the focus of a large and growing literature in economics.[6]

Besides its importance in welfare, smoking behavior is also an excellent testing ground for the influence of subjective expectations on choices. First, the risks to health from smoking have been well documented and are relatively easy to identify and quantify. Second, the smoking decision is made under considerable uncertainty, with significant pecuniary and nonpecuniary consequences. Expectations about the future after the smoking choice is made, therefore, play a crucial role in the decision-making process.

The empirical results show that expectation biases do exist and play an important role in individuals' decision-making processes. Specifically in this application, individuals' genetic makeup, which is found to be statistically insignificant under the rational expectations assumption, is subjectively significant. Additionally, genetic

[6] See [34], [14], [62], or [15] for reviews on smoking behavior in general. See [83] for a review of literature on mature smokers' addictive smoking behavior. Preventive health choices in general are of great importance to individuals' welfare. For an overview of economic issues related to a variety of preventive health decisions, see [53].

makeup has an effect on individuals' longevities that is subjectively as significant as their health-related choices, such as smoking. I also find that individuals are subjectively more concerned with their health and more forward-looking than under the rational expectations assumption. The estimated utility parameter for health is about 50% higher, and the estimated discount factor increases from 0.61 to 0.70 when I directly incorporate subjective longevity expectations in the model.

The policy experiments conducted using the subjective expectations data and the inferred preferences further show that, if individuals had unbiased expectations about the marginal effects of smoking on their longevities, possibly following a more personalized information campaign, smoking rates would go down on average by 4% in this age group.

This paper stands at the intersection of economics literatures on subjective expectations, health behaviors, and dynamic discrete choice models, and makes the following main contributions. First and foremost, I propose a new method to assess the rational expectations assumption by identifying the distributions of the true underlying objective expectations and the expectation bias with respect to unobserved private information using subjective expectations data and observed realizations.

Second, this study belongs to a growing literature on analyzing individual behavior using subjective expectations data, which combines the revealed preferences approach with the stated preferences approach. Few studies have directly employed subjective expectations data to understand and predict choices. [19] analyzes how perceptions about the benefits and costs of different contraceptive methods affect women's birth control choices; [63] investigates the effects of perceptions of the justice system on youth criminal behavior.[7] Both studies use a static structural model.

[7] Other recent studies analyzing relationships between subjective expectations and individual behaviors include [77] and [20], using experimental data; and [47], [49], [92], [93], and [54], using survey data on various decisions such as life-cycle consumption patterns and smoking. [90], on the other hand, treat subjective expectations data as auxiliary data and keep the rational expectations

To the best of my knowledge, this paper is among the first to use subjective expectations data in a structural and *dynamic* discrete choice model of (health-related) behaviors, and therefore also contributes to the large literatures on dynamic discrete choice models and health behaviors. The empirical results are important for the design of public policies aimed at further reducing smoking rates among adult smokers.

The rest of this paper proceeds as follows. Section 1.2 introduces the new theoretical framework to assess the rational expectations assumption and to identify the distributions of the expectation bias. Section 1.3 shows how to identify the preference parameters and the expectation bias when individual choices are introduced into the framework. I describe the data in Section 2.3.2, and then apply the theoretical framework to the data to identify the true underlying objective expectation and the expectation bias for survival in Section 1.5. Section 1.6 explains the estimation strategy and the empirical specifications. Section 2.3.6 presents estimation results and discusses the policy implications. Section 2.4 concludes.

1.2 Identification of Objective Expectations and Expectation Bias

In the absence of subjective expectations data, inferring individual preferences from observed choices typically requires strong assumptions about individuals' expectations, such as the rational expectations assumption. This assumption does not allow for the possibility that respondents' subjective expectations might be systematically different from the observed outcomes. The violation of this assumption has significant implications for economic understanding of the decision-making processes of the agents; hence, it is important to test for the rationality of individual expectations using available data.

The existing methods to assess the rational expectations assumption compare

assumption.

subjective expectations to their objective counterparts based on the public information. In this section, with the same available data, I go one step further and propose a new theoretical framework which can be used to identify the distributions of the true underlying objective expectations and the expectation bias with respect to the unobserved private information. In the next section, I further introduce individual choices to this framework and discuss the identification of the expectation bias and preference parameters.

1.2.1 Continuous Outcome

I start with the case in which the outcome of interest is a continuous random variable, such as income. The available data include the realizations of and individuals' subjective expectations about this random variable, alongside with other public information.

The key factors to explain the observed realizations and agents' expectations are the true underlying objective expectations, also known as "Nature's Law of Motion" ([67]); the expectation bias, which is defined as the deviation in individuals' subjective expectations from the true underlying objective expectations; and the realization shock, which is the difference between the true underlying objective expectation and the observed realization. These three components are generally unobserved to the econometrician due to individuals' private information, and have to be identified using the available data on observed realizations and subjective expectations. I show how to extract information on these three components from the two pieces of available information in a theoretical framework below.

Let S_t denote the state variables observed by the agent at time t. I assume a Markovian structure of these state variables, so that S_t fully describes the information set of the agent at time t. Part of S_t, denoted here as X_t, is observable to both the agent and the econometrician.

The expected values of certain outcomes at time $t+1$ under the true objective underlying state transitions are denoted by $E(O_{t+1}|S_t)$. The agent has his/her own subjective expectations about these outcomes $\hat{E}(O_{t+1}|S_t)$, which might deviate from the true underlying expectations. The difference between these two expectations determines the agent's bias $\lambda_t = \lambda(S_t)$:

$$\lambda_t \equiv \hat{E}(O_{t+1}|S_t) - E(O_{t+1}|S_t), \tag{1.1}$$

while the difference between the realized outcomes O_{t+1} at time $t+1$ and their expectations under the true underlying state transition defines the realization shock ξ_{t+1}:

$$\xi_{t+1} \equiv O_{t+1} - E(O_{t+1}|S_t). \tag{1.2}$$

With these notations, I can now establish sufficient conditions for the identification of the distributions of interest.

Proposition 1. *Assume the following conditions:*

1. *$E(\lambda_t|X_t)$ is finite.*
2. *λ_t, $E(O_{t+1}|S_t)$, and ξ_{t+1} are mutually independent given X_t.*
3. *The random variables possess non-vanishing (a.e.) characteristic functions, conditional on X_t.*

Then the conditional distributions of λ_t, $E(O_{t+1}|S_t)$, and ξ_{t+1} given X_t are identified.

Proof. From the definitions of the expectation bias and the realization shock, it immediately follows that

$$\begin{aligned} O_{t+1} &= E(O_{t+1}|S_t) + \xi_{t+1} \\ \hat{E}(O_{t+1}|S_t) &= E(O_{t+1}|S_t) + \lambda_t, \end{aligned} \tag{1.3}$$

where both O_{t+1} and $\hat{E}(O_{t+1}|S_t)$ are known by the econometrician. In addition, ξ_{t+1} is conditionally mean zero following its definition in Eq. (1.2), while the expected value of the expectation bias given the econometrician's information set, $E(\lambda_t|X_t)$, exists due to Assumption 1. Then, according to the theorem by [57], given Assumptions 1 through 3, the distributions of the three random variables λ_t, $E(O_{t+1}|S_t)$, and ξ_{t+1} conditional on X_t are nonparametrically identified. □

Assumptions 1 and 3 are technical conditions under which the existence of the conditional moments of the expectation bias and the true underlying objective expectations, and the characteristic functions of their distributions given the econometrician's information set is guaranteed. Assumption 2 states that the expectation bias is independent of the true underlying expectation and the realization shock, and weakens the dependence between the underlying expectation and the realization shock to conditional independence. Under these assumptions, all the moments of the distribution of the true underlying objective expectations can be obtained by exploring different moments of the joint distribution of the two left-hand side variables in Eq. (1.3), except for the first moment which will follow directly the fact that ξ_{t+1} has zero conditional mean. Once the distribution of the true underlying expectation is identified, the identification of the distributions of the other two unknowns is straightforward.

For an illustration, suppose we are interested in the true underlying objective expectation and the expectation bias for individuals' income levels in the next year, so O_{t+1} are next year's incomes which can be observed ex post in the data. We elicit from individuals their own subjective expectations about future income levels, so $\hat{E}(O_{t+1}|S_t)$ are also known. After controlling for all the observed individual characteristics which affect individual incomes, such as gender, tenure, education, and occupation, there might still be private information in individuals' information set

(S_t) about their future incomes, such as their ability. The complete information set determines individuals' true underlying objective expectations about future incomes ($E(O_{t+1}|S_t)$), which may differ from the realized income levels due to certain random shocks (ξ_{t+1}) such as unexpected absence from work for family reasons. Overly optimistic or pessimistic views about the economy may bias individuals' subjective expectations about future income (λ_t). According to Proposition 1, the econometrician can identify the distributions of the true underlying objective expectations of next year's income, the expectation bias, and the realization shock in the presence of the private information.

Intuitively, suppose we want to know the variance of the distribution of the true underlying objective expectations of next year's income. The covariance of the realized and individuals' subjective expectations about next year's income, O_{t+1} and $\hat{E}(O_{t+1}|S_t)$, according to Eq. (1.3), can be written as the covariance of the two terms on the right hand side. Since the three components of these two terms are assumed to be mutually independent, the covariance of O_{t+1} and $\hat{E}(O_{t+1}|S_t)$ is exactly the variance of $E(O_{t+1}|S_t)$, the true underlying objective expectations of next year's income. Based on [57] theorem, this argument can be extended to identify the whole distributions of the true underlying objective expectation and the expectation bias about future income levels, with respect to the unobserved private information, such as workers' ability.

The identification approach based on [57] theorem has been applied in various fields including measurement errors (e.g., [61]), auctions (e.g., [58]) and education investment (e.g., [17]). The novel dimension of this paper is to adopt these methods into a framework to uncover unobserved distributions of underlying true expectations and expectation biases in the presence of private information.

1.2.2 Binary Outcome

In many cases of interest, such as individuals' survival, workers' job losses, and the arrival of an economic recession, the outcome is binary. In such situations, respondents report subjective probabilities of the occurrences of these events, while the econometrician observes only the binary outcome, e.g., alive or not. This binary case does not fall into the continuous case setup, since now Assumption 2 of Proposition 1 will no longer hold, as the realization shock will have a heteroskedastic variance and therefore cannot be independent of the true underlying objective expectation.[8] In addition, binary outcomes provide far less information about the underlying distributions than continuous ones. Hence, the binary outcome case requires a special treatment relative to the continuous case outlined in the previous subsection.

Consider a binary realization O_{t+1} which takes two values 0 and 1. Call S_t the objective information observed by the agent, and ξ_{t+1} the realization shock. Then, the realization, O_{t+1}, can be written down in the following way:

$$O_{t+1} = I(S_t - \xi_{t+1} > 0). \tag{1.4}$$

I assume that ξ_{t+1} and S_t are independent, and the distribution of ξ_{t+1}, denoted as $\Phi(\xi_{t+1})$, is known to the agent and the econometrician.

Equation (1.4) implies that S_t determines the objective expectation of O_{t+1} given the agent's full information set. That is,

$$E(O_{t+1}|S_t) \equiv Pr(O_{t+1} = 1|S_t) = Pr(\xi_{t+1} < S_t|S_t) = \Phi(S_t). \tag{1.5}$$

The subjective expectation of O_{t+1} can differ from the objective one because of the agent's expectation bias. Denote by $\hat{E}(O_{t+1}|S_t)$ the subjective expectation of the agent given the full information set. Then, I can implicitly define the subjective

[8] See, e.g., [65].

expectation bias λ_t in the following way:

$$\hat{E}(O_{t+1}|S_t) = \Phi(S_t + \lambda_t). \tag{1.6}$$

Indeed, if the bias is zero ($\lambda_t = 0$), then the agent reports the true objective expectations of his/her survival:

$$\hat{E}(O_{t+1}|S_t, \lambda_t = 0) = \Phi(S_t) = E(O_{t+1}|S_t). \tag{1.7}$$

On the other hand, a positive bias $\lambda_t > 0$ implies that the agent overestimates the objective expectations, while a negative bias implies that s/he underestimates them. The following proposition establishes identification results for the distributions of S_t and λ_t.

Proposition 2. *Assume the following conditions,*

1. *O_{t+1} and $\hat{E}(O_{t+1}|S_t)$ are observed, while S_t and λ_t are unobserved by the econometrician.*

2. *S_t, ξ_{t+1}, and λ_t are mutually independent.*

3. *ξ_{t+1} has a known distribution.*

Then the distributions of S_t and λ_t can be identified.

Proof of this proposition and the corresponding estimation strategy is provided in Appendix A.1.[9]

For example, let O_{t+1} denote individuals' survival status, so that the expected value of this outcome equals the probability of being alive next period. Figure 1.1 provides a graphical illustration of the relationship between the mean survival status (the vertical line) and the subjective survival expectations (the horizontal

[9] This proof draws insights from the literature on measurement errors. See, e.g., [60] and [61]. [16] provides a survey of nonlinear models of measurement errors. See, also, [39].

line). Without expectation bias or any private information, the relationship between the realization of survival status and the elicited subjective survival expectations should follow a 45° line. Indeed, if individuals have rational expectations and the econometrician observes all the information individuals use to form their subjective survival expectations, those who report, say, 80% probability of survival should be alive at the next period with an 80% chance.

The methods in existing literature on subjective expectations allow for expectation bias but assume away individuals' private information, which means that the difference between the actual relationship between the mean realizations and the subjective expectations will be interpreted as the expectation bias. For example, if this actual relationship can be summarized by the dashed curve, then when this curve lies to the right of the 45° line, agents are found to be are overly optimistic about their survival and vice versa. The method proposed in this paper, however, goes one step further than the existing literature by relaxing the assumption that the econometrician has complete knowledge of individuals' information set. Instead, this method allows individuals to have both expectation bias and private information.

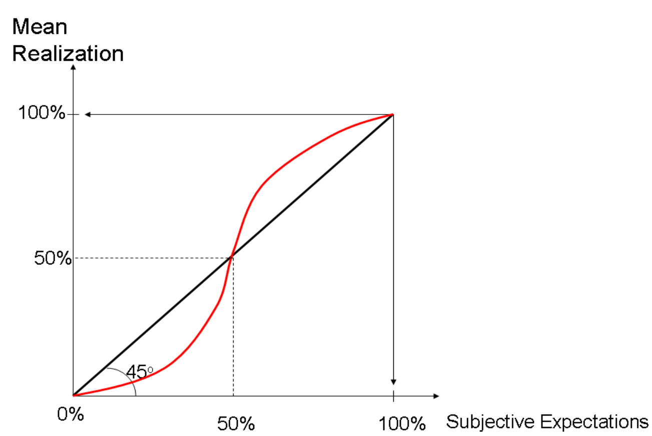


Figure 1.1: Realizations and Subjective Expectations

Graphical illustration of the relationship between mean realizations and subjective expectations for a univariate binary outcome.

1.3 Subjective Expectations and Individual Choices

In the previous section, I introduced a new theoretical framework to assess the rational expectations assumption using the objective realizations and subjective expectations data. In particular, I provided the conditions for the identification of the distributions of the true underlying objective expectation and the expectation bias with respect to the unobserved private information. The next step is to examine the implications of subjective expectations for individual choices without assuming rational expectations. In this section, I discuss the identification of preference parameters and the expectation bias using individual choices in addition to subjective expectations and observed realizations data. I start with a two-period binary choice model, and then consider a full-blown dynamic discrete choice model.

1.3.1 Subjective Expectations in a Binary Choice Model

In the current subsection, I consider individual behavior and extend the classic binary choice model to allow for subjective expectations, which permits the identifications of preference parameters, the true underlying objective expectations and the expectation bias without relying on the rational expectations assumption.

I consider a two-period $(t, t+1)$ binary choice $(A_t \in \{0, 1\})$ model with subjective expectations. I assume that choice 1 is associated with a higher instantaneous utility and higher probabilities of reaching less favorable states in the next period. Therefore, while the agent is faced with the tradeoff between short-term benefits and long-term costs when s/he makes a choice at time t, s/he will surely choose action 1 once reaching time $t+1$. The econometrician observes the agent's optimal choice, although to simplify the exposition, I do not include the agent's choice in the vectors X_t or S_t.

I normalize the instantaneous utility associated with choice 0 to be 0, and assume that the instantaneous utility associated with choice 1 is linear in the observed variables X_t,[10]

$$u_{1,t} = \theta' X_t. \tag{1.8}$$

Denote the exogenous exponential discount factor between periods by β, and define the additively separable choice-specific utility shocks by ε_0 and ε_1, respectively.[11] The difference in these shocks is given by ε $(= \varepsilon_1 - \varepsilon_0)$, and for simplicity, the difference between the utility shocks in the second period is assumed to be 0. The difference in utility shocks (ε) exhibits the appropriate lack of dependence on observable information.

Given the structure of the model, the inter-temporal utilities associated with these two choices at time t satisfy:

$$U_1 = \theta' X_t + \varepsilon_1 + \beta\theta' \hat{E}(X_{t+1}|S_t, 1) \quad \text{and} \quad U_0 = \varepsilon_0 + \beta\theta' \hat{E}(X_{t+1}|S_t, 0), \tag{1.9}$$

where the expectations, $\hat{E}(X_{t+1}|S_t, A_t)$, are the agent's subjective beliefs about future states given his/her current information S_t, possible actions A_t and time preference β. The optimal choice of the agent, A_t^*, is determined by comparing the two utilities in Eq. (1.9):

$$A_t^* = I(\theta' X_t + \varepsilon + \beta\theta' \hat{E}(X_{t+1}|S_t, 1) - \beta\theta' \hat{E}(X_{t+1}|S_t, 0) > 0), \tag{1.10}$$

where the indicator function $I(.)$ takes the value of 1 if the term inside the brackets, the difference in inter-temporal utilities between these two choices, is positive.

While the agent takes into account his/her subjective expectations for *both* possible actions when s/he makes the decision, the *factual* subjective expectations reported by the agent only correspond to the optimal choice A_t^* :

$$\hat{E}(X_{t+1}|S_t, A_t^*) = \hat{E}(X_{t+1}|S_t, 1) \times A_t^* + \hat{E}(X_{t+1}|S_t, 0) \times (1 - A_t^*). \tag{1.11}$$

[10] This linearity assumption is not critical and can be relaxed following arguments in [69], [70], and [71].

[11] This additive separability assumption can also be relaxed following [72].

The relationship between subjective expectations, realization, true underlying expectations and expectation bias can be expressed in a way similar to that in the previous section, recognizing that the econometrician can now observe both X_{t+1} and A_t^* :

$$\begin{aligned} X_{t+1} &= E(X_{t+1}|S_t, A_t^*) + \xi_{t+1}, \\ \hat{E}(X_{t+1}|S_t, A_t^*) &= E(X_{t+1}|S_t, A_t^*) + \lambda_t. \end{aligned} \tag{1.12}$$

Using the same arguments as in Section 1.2, the econometrician can identify the distributions of ξ_{t+1}, $E(X_{t+1}|S_t, A_t^*)$, and λ_t, conditional on X_t and A_t^*. A more interesting question, however, is to characterize the distributions of these variables conditional on each of the two choices individually, not just upon the optimal one. For example, one goal of the model is to learn about individuals' expectations and biases for potential choice 1, which could depend on whether choice 1 is optimal. Another concern is the identification of the utility parameters (θ) and the discount factor (β). I make the following alternative assumptions on the structure of the model to address these issues.

Assumption 1.3.1. ***(Observable Counterfactuals (CO))*** *The econometrician can observe both* $\hat{E}(x_{t+1}|S_t, 1)$ *and* $\hat{E}(x_{t+1}|S_t, 0)$.

Assumption 1.3.2. ***(Choice Independent Bias (CI))*** *Individuals' expectation biases are choice-independent:* $\lambda_t(S_t, 1) = \lambda_t(S_t, 0)$.

This first assumption, that individuals' expectation biases depend only on their states, not on their possible actions, is not as restrictive as it may seem. For example, if certain individuals are innately optimistic, then they will always have a rosy view about their probabilities of getting into a good state, no matter what choices they make.

Assumption 1.3.3. *About the true underlying objective expectations:*

1. ***(Strong Irrelevance of Private Information (SIR))*** *The true underlying expectations depend only on observed public information and choices, not on unobserved private information:* $E(X_{t+1}|S_t, A_t) = E(X_{t+1}|X_t, A_t)$.

2. ***(Weak Irrelevance of Private Information (WIR))*** *The dependence of the true underlying expectations on private information is the same for both possible choices individuals can make.*

The first statement in the second assumption that the true underlying objective expectations do not depend on private information can potentially be strong. However, one nice feature of this assumption is that it is empirically testable. Since the distribution of the expectation biases for certain choices can be obtained using two different methods as explained in Appendix A.2 significantly different distributions of the expectation biases resulting from these two methods, imply that the validity of Assumption SIR is in question. The second statement in Assumption 3.2 says that SIR can be relaxed to the extent that private information can affect the true underlying objective expectations, but its effects are the same under both choices. For example, an extremely noisy working environment might be an individual's private information, which will most likely affect this person's health transitions, but its impact can be the same regardless of his choices.

In an ideal but extremely rare case, the econometrician can observe individuals' subjective expectations under both choices. If it is true, then the following proposition establishes the identification results for preference parameters and the conditional distributions of the objective expectations and the expectation biases:

Proposition 3. *With both factual and counterfactual subjective expectations, and subject to appropriate lack of dependence of the difference in utility shock (ε) on the conditioning variables, conditional independence of the expectation bias and the* *underlying objective expectations given X_t, and existence of necessary moments, supports and characteristic functions,*

1. *Preference parameters (θ, β) and the distribution of ε conditional on X_t and elicited subjective expectations are identified.*

2. *Given Assumption CI or SIR, the conditional marginal distributions of true underlying expectations and expectation biases for both possible actions can also be identified.*

In a more realistic situation, only individuals' factual subjective expectations corresponding to the optimal choices are elicited, then,

Proposition 4. *Subject to appropriate lack of dependence of the difference in utility shock (ε) on the conditioning variables, conditional independence of expectation bias and underlying expectations given X_t, and existence of necessary moments, supports and characteristic functions,*

1. *Assumptions CI and SIR together identify conditional marginal distributions of true underlying expectations and expectation biases for both possible actions.*

2. *Assumptions CI and SIR together identify preference parameters (θ, β) and the distribution of ε conditional on X_t and elicited subjective expectations.*

3. *The conditions for the identification of preference parameters in 2 can be weakened, i.e., Assumption SIR can be replaced by Assumption WIR, at the expense of the identification of the difference in utility shocks, ε.*

Proofs of the above two propositions are provided in Appendices A.2 and A.3, respectively.

Table 1.1 summarizes different assumptions for subjective expectations, expectation bias, and private information in different models. To analyze behaviors assuming rational expectations effectively assumes away any possible expectation bias and

Table 1.1: Identifying Assumptions for Subjective Expectations

		Under Rational	With Subjective Expectations				
Components	Assumptions	Expectations Assumption	Without Behaviors	With Behaviors			
Subjective	Factual	X	X			X	X
Expectations	Factual + Counterfactual			X	X		
Expectation	None	X					
Bias	Choice Independent			X		X	X
	Flexible		X				
Private	Strong Irrelevance	X			X	X	
Information	Weak Irrelevance						X
	Flexible		X				

Necessary identifying assumptions on subjective expectations, expectations bias, and private information in models assuming rational expectations (Column 1), allowing for subjective expectations without behaviors (Column 2), and with behaviors (Column 3).

private information in individuals' subjective expectations. Once the rational expectations assumption is relaxed, the expectation bias can be dealt with conditional upon available data. With behaviors, certain assumptions about the expectation bias and the private information are required, but these assumptions are still weaker than the rational expectations assumption.

1.3.2 Subjective Expectations in a Dynamic Discrete Choice Model

This subsection extends the two-period binary choice model and incorporates subjective expectations data into a finite-horizon single-agent dynamic discrete choice model.[12] In this class of models, agents choose from a finite set of actions to maximize their expected lifetime utilities based on their expectations of future state transitions. This is also the model adopted in the following empirical analysis. Since the data used in this study, like most available data sets with subjective beliefs, contain only factual subjective expectations corresponding to the optimal choices, I follow Proposition 4 and assume that the expectation bias is choice independent (Assumption CI) and that the true underlying objective expectations do not depend upon private information (Assumption SIR). In future work, I plan to further relax Assumption SIR by allowing the private information to affect the true underlying

[12] For detailed reviews of the literature on dynamic discrete choice models, see, for example, [23], [79], and [1].

objective expectations.

This model is set up with the following components:

- A time index, $t \in \{0, 1, 2, ..., T\}$.
- A state space, S, consisting of both observable and unobservable parts. Specifically, $S = (x, \varepsilon)$, where x is observable to all, while ε is observable to the agent but not to the econometrician.
- A choice space, A, with a finite number of discrete choices.
- Agents' subjective expectations about state transitions, $\hat{p}(s_{t+1}|s_t, a_t)$.
- A long-run exponential discount factor, $\beta \in [0, 1]$.[13]
- An instantaneous period utility function, $u(s_t, a_t)$.

Agents have the following inter-temporal utility function at time t:

$$U_t = u(s_t, a_t) + \Sigma_{\tau=t+1}^{T} \beta^{\tau-t} u(s_\tau, a_\tau), \tag{1.13}$$

where $u(s_t, a_t)$ is the instantaneous utility at time t, while the second term on the right-hand side is the summation of discounted future utilities from the time $(t + 1)$ forward.

The agent's value function at the time of choice can, in turn, be expressed as the maximum value of the inter-temporal utility function, with the maximization taken over the action space:

$$V(s_t) = max_{a_t \in A}[u(s_t, a_t) + \beta \int V(s_{t+1}) \hat{p}(s_{t+1}|s_t, a_t) ds_{t+1}].$$

[13] An alternative approach models individuals as hyperbolic discounters who put less weight on future than in the model considered here and therefore make time inconsistent choices. See [59], [38], [26], [35], and [27].

Intuitively, in a dynamic discrete choice model, agents are faced with a set of discrete choices, each of which is associated with a total amount of utilities, as shown by Eq. (1.13). As future states are uncertain, the agents need to calculate these utilities based on their expectations about the future given their current choices and states. So, to recover individuals' preferences (u) from observed choices (a), we should use individuals' *own* subjective expectations ($\hat{p}$) about the future states following their current choices.

Following Rust (1987), I make the following assumptions on the unobservable part in the preferences:

A1. Additive Separability: $u(s_t, a_t) = u(x_t, a_t) + \varepsilon_t(a_t)$.

A2. Conditional Independence:

$$\hat{p}(x_{t+1}, \varepsilon_{t+1}(a_{t+1})|x_t, \varepsilon_t(a_t), a_t) = q(\varepsilon_{t+1}(a_{t+1})|x_{t+1})\hat{p}(x_{t+1}|x_t, a_t)$$

$$q(\varepsilon_{t+1}(a_{t+1})|x_{t+1}) = q(\varepsilon).$$

A3. Extreme Value Error Distribution: ε is i.i.d with extreme value distribution.

Assumption A1 shows that the period utility function consists of two parts: $u(x_t, a_t)$, which depends only on the observed components – state variable x and choice a at time t, and the unobserved state variable ε.

Assumption A2 places simplifying restrictions on the transition probabilities by assuming that the transition of the unobserved state variable is independent of the observed state variables and the agent's choice, and that the transition probabilities of the whole state space are multiplicatively separable in the observed and unobserved state variables, conditional upon the agent's lagged choice and observed state variables. Furthermore, lagged unobserved state variables have no implications for the evolution of future state variables.

The distribution of the unobserved state variable is difficult to identify without making strong parametric assumptions about its functional form. This variable is therefore assumed to be known, with little loss of generality. The particular extreme value distribution in Assumption A3 guarantees a closed-form solution to the ex ante value function, also known in the literature as social surplus function ([79]), and a convenient logistic functional form for the conditional choice probabilities.

The expected maximum value of inter-temporal utilities is expressed as the ex ante value function through the following relationship:

$$V(x_t) = E_\varepsilon max_{a_t \in A}[u(x_t, a_t) + \varepsilon_t(a_t) + \beta \int V(x_{t+1})\hat{p}(x_{t+1}|x_t, a_t)dx_{t+1}]. \quad (1.14)$$

The maximization above is again taken over the choice space, and the expectation is taken over the whole distribution of the unobserved state variable.

Note that the only unobserved part inside the big bracket on the right-hand side of Eq. (1.14) is ε. The remaining observed part is therefore collected and named the *choice-specific value function*:

$$V(x_t, a_t) = u(x_t, a_t) + \beta \int V(x_{t+1})\hat{p}(x_{t+1}|x_t, a_t)dx_{t+1}, \quad (1.15)$$

which is a function of only the observed state variables (x) and the observed choice (a).

Under Assumptions A1 - A3, Eqs. (1.14) and (1.15) together imply that the ex ante value function is related to the choice-specific value functions through the following expression:

$$\begin{aligned} V(x_t) = & \quad E_\epsilon max_{a_t}\{V(x_t, a_t) + \varepsilon\} \\ = & \quad G(V(x_t, a_t), a_t = 0, ..., \#A) = ln\{\textstyle\sum_{a \in A} exp[V(x_t, a_t)]\}, \end{aligned} \quad (1.16)$$

where the first equality in Eq. (1.16) is obtained by replacing the first and third terms inside the big bracket on the right-hand side of Eq. (1.14) using the definition given in Eq. (1.15); while the second equality directly follows the extreme value error distribution of Assumption A3.[14]

Without loss of generality, consider a two-choice case where $A = \{0, 1\}$.[15] The literature has made it clear that what can be identified from the data is the difference in the choice-specific value functions ([43]). In particular, under Assumption A3, the difference in the logarithms of the conditional choice probabilities is equal to the difference in the choice-specific value functions, denoted here as $d(x)$. That is,

$$d(x_t) \equiv V(x_t, 1) - V(x_t, 0) = \ln P(1|x_t) - \ln P(0|x_t). \tag{1.17}$$

If we further assume that the instantaneous utility functions take on the following linear form,

$$u(x_t, 1) = x_{1t}'\theta_1 \quad \text{and} \quad u(x_t, 0) = x_{0t}'\theta_0,$$

then the utility parameters can be identified through the following relationship, using the difference in the choice-specific value functions, $d(x)$, and the state transition probabilities, $\hat{p}(x_{t+1}|x_t, a_t)$, both of which can be obtained from the data:

$$\begin{aligned} d(x_t) - \textstyle\sum_{s=t+1}^{T} \beta^{s-t} \hat{E}\left[\log(1 + e^{-d(x_s)})|x_t = x, a = 1\right] \\ + \textstyle\sum_{s=t+1}^{T} \beta^{s-t} \hat{E}\left[\log(1 + e^{d(x_s)})|x_t = x, a = 0\right] \\ = \textstyle\sum_{s=t}^{T} \beta^{s-t} \hat{E}\left[x_{1s}|x_t = x, a = 1\right]' \theta_1 - \textstyle\sum_{s=t}^{T} \beta^{s-t} \hat{E}\left[x_{0s}|x_t = x, a = 0\right]' \theta_0. \end{aligned} \tag{1.18}$$

The derivation of Eq. (1.18) is provided in Appendix A.4.

We can identify the utility parameters, as described above, for a given discount factor.[16] [64] discuss the nonparametric identification of dynamic discrete choice

[14] For a more detailed derivation of Eq. (1.16), see Rust (1987).

[15] Generalization to a case with multiple choices is straightforward.

[16] This study, as most studies in the literature, treats time preference as exogenous. [6], on the other hand, show a model where time preference can change endogenously as a result of individuals' investment. This possibility is not considered here.

models using a method based on the insights from [43], who view Bellman equations as moment conditions. One of the conclusions Magnac and Thesmar have reached is that the discount factor can be identified if there exist certain exclusive restrictions that shift the expected future utilities (through, say, the transition of the state variables), but do not enter the agents' instantaneous utility functions.

This somewhat abstract argument has an intuitive appeal. Suppose two individuals are identical in all ways except for one. This difference affects their future state transitions but not their instantaneous utilities. That is, this one variable is exogenous to the utility function but still relevant to the state transitions, which makes this variable the exclusive restriction by definition. If these two individuals are completely myopic with a zero discount factor, they should make exactly the same choices because their choices depend only on their identical instantaneous utilities. Therefore, systematic differences in their behaviors should not be expected.

If these two individuals are not completely myopic, then we should expect some differences in their behaviors because their state transitions and therefore the total utilities associated with each choice are different due to the exclusive restriction. The more forward-looking they are, the greater the magnitude of this difference in their behavior. This relationship helps pin down the discount factor.

1.4 Data

This study uses the data from the Health and Retirement Study. The HRS is a nationally representative biennial panel study, in which the baseline interviews were conducted in 1992 (wave 1) with birth cohorts 1931 through 1941 and their spouses, if married. New birth cohorts have been added to the initial sample of 12,652 persons in 7,702 households, and the most recent available data are for year 2006 (wave 8).[17]

[17] The survey history and design are described in more details in [50]. Data flow and other information are also available at http://hrsonline.isr.umich.edu.

A. Subjective Longevity Expectations and Smoking Behaviors

To measure the subjective longevity expectations, which are the main focus of this paper, I use the survey answers to the following questions: 1) "*What is the percent chance that you will live to be 75 or more?*" and 2) "*What is the percent chance that you will live to be 85 or more?*". These questions are asked only of those under age 65 and 75 at the time of the interview, respectively.[18] For consistency, responses across all waves were re-scaled to fall within $[0, 1]$.

The validity of the elicited subjective longevity expectations might be in question in the following two cases: 1) the response is 0% or 100%, or 2) the elicited probabilities of living to age 75 do not exceed those of living to age 85. These two cases are not mutually exclusive.

The first case might exist because people round their answers to the closest integers.[19] Following the literature (e.g., [54]), I add 10% (1%) to the response if 0% is the answer to the first (second) question, and subtract 1% (10%) from the response if 100% is the answer to the first (second) question.

The second case usually implies certain mistakes by the respondents or their misunderstanding of the questions. It might be due to rounding again if the responses to the two questions are the same, which I deal with by subtracting 1% from the elicited subjective percent chance of living to age 75 and then using the resulting number as the subjective percent chance of living to age 85, so the probability of living to age 75 is always at least 1% higher than that of living to age 85.[20]

[18] The questions in 1992 were slightly different from those in the following waves, "Using any number from 0 to 10 where 0 equals *absolutely no chance* and 10 equals *absolutely certain*, what do you think are the chances that you will live to be 75(85) or more?". In addition, from the 5th wave in 2000, the target age in the second longevity question has been based on respondents' ages at the time of interview.

[19] See [75] and [8] for an analysis of the possible rounding problem with subjective probabilities.

[20] As part of the sensitivity test not shown here, I tried adding 1% to the elicited subjective

A response of 50% may also suggest the incapability or carelessness of the respondents to answer those questions, or it may be an expression of uncertainty rather than a quantitative probability (see, e.g., [28], and [13]). However, other studies argue that respondents who reported a value of 50% were drawing from nearby probability points (e.g., [46]), and that 50% could be merely round-off errors instead of cognitive dissonance ([11], p. 306). Furthermore, the responses generally are not excessively more bunched at 50% than at adjacent round values (see, e.g., [68]). I, therefore, do not consider any further adjustment to those responses of 50%.

From the very first survey in 1992, the HRS has asked the respondents about their smoking behaviors using the questions: *"Do you smoke cigarettes now?"* and *"Have you ever smoked before?"* By smoking, the HRS refers to the consumption of more than 100 cigarettes in a respondent's lifetime, excluding pipes and cigars. Using the answers to these two questions, I categorize the respondents into current smokers, former smokers, and never smokers. Because never smokers might abstain from smoking for various reasons, such as religious beliefs and dislike of the flavor, while the HRS does not provide any information on the exact causes of abstinence, I exclude never smokers from the study.

B. Additional Explanatory Variables

In addition to subjective longevity expectations and smoking choices, I also collect information on individuals' health, genetic makeup, and other demographic characteristics.

Self-reported health status is measured on a 5-point Likert scale. Respondents

percent chance of living to age 85 and then using the resulting number as the subjective percent chance of living to age 75. I also tried estimating the whole model excluding those observations with any of these two cases of special responses. Results are robust. Appendix Table A.2 shows the percentages of observations in the final analysis sample with any of the special responses. Appendix Figure A.1 shows the distributions of both the original and the transformed ('corrected') subjective probabilities of living to age 75 and 85.

were asked, *"What do you think is your current health status: 1. excellent; 2. very good; 3. good; 4. fair, 5. poor?"* To avoid possible measurement errors and to simplify the estimation procedure, I summarize the information on self-reported health status by a binary health indicator that is set to 1 if the respondent is in bad health (fair or poor) and 0 otherwise (excellent, very good, or good).[21]

I also control for a number of demographic characteristics, including the respondents' age, gender (female or not), race (non-Hispanic White or not), real household income in "1992 dollars" (calculated using the Consumer Price Index), and the longevity of the respondent's same-gender parent which is summarized by a binary variable set to 1 if the parent is still alive or died at an age greater than 70, and 0 otherwise.[22] This last variable is used to capture the respondents' private information regarding their expected longevity.

The effect of a parent's death on self-assessed survival probabilities is potentially both biological and psychological. For example, if the parent died of a type of disease known to have a genetic link, the child might reassess his/her own longevity expectancy accordingly. In addition, a parent's death may also affect the respondent's subjective longevity expectations because it reminds the respondent of his or her own mortality. It would admittedly be helpful to know the cause of the same-gender parent's death, e.g., due to accident, some choice-related sources, or genetic reasons. Unfortunately, the HRS does not provide detailed information on this matter.

Previous research has focused largely on smoking behaviors of young adults (see, e.g., [62]), yet one important characteristic of cigarette consumption is the long

[21] See [12] for a discussion of self-reported and objective measures of health.

[22] I also tried using only mothers' or fathers' longevity information for all the observations and mothers' longevity information for males and fathers' longevity information for females. Results are robust.

latency period between the time of initiation and the onset of adverse health shocks.[23] And, even for smokers who quit at age 65, [86] show that men gained on average 1.4 to 2.0 years of life, and women 2.7 to 3.7 years. Furthermore, Mendez and Warner ([73], [74]) show that the goal of cutting the smoking prevalence among U.S. adults to 12% by 2010 (*Healthy People 2010*, U.S. Department of Health and Human Services, 2000) cannot be achieved unless the rate of smoking cessation among adults increases. I therefore focus on the smoking behaviors of respondents aged 51 to 61.

Our final analysis sample excludes observations with missing values for any of the variables mentioned above.[24]

Table 1.2: Summary Statistics

Variables	Mean	Std. Dev.
Age	56.7	2.98
Female	0.53	0.50
White	0.82	0.38
Current smoker at period 1	0.38	0.49
Former smoker at period 1	0.62	0.49
Same gender parent alive or died at age > 70	0.67	0.47
Self-rated health at period 1	2.65	1.16
Bad health at period 1	0.23	0.42
Household income at period 1 ('$k)	53.4	65.6
Self-rated health at period 2	2.63	1.25
Bad health at period 2	0.24	0.43
Household income at period 2 ('$k)	51.6	68.4
Subjective probability of living to 75	0.63	0.29
Subjective probability of living to 85	0.42	0.30
Observed deaths in two years	0.021	0.14

Summary statistics of variables for the final analysis sample. HRS panel data from 1992 to 2006.

Table 2.1 provides the summary statistics for our final analysis sample. Approximately 38% of the observations in the sample are current smokers, and the other 62%

[23] 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) ([42]).

[24] Table A.1 compares the final analysis sample with the sample with missing values on subjective longevity expectations and shows that these two samples are not significantly different in important variables such as smoking behaviors and two-year mortality rates.

have a history of cigarette smoking. About 53% of our sample observations are female, with an average age of 56.7. Around 82% of our observations are Non-Hispanic Whites. 67% of our respondents' same-gender parents are still alive or died after age 70. At period one, the average self-rated health level is 2.65, which is between *very good* and *good* health. The average household income is about \$53.4K. At period two, the average self-rated health level drops to 2.63, with the average household income decreasing to \$51.6K. The average subjective probabilities of living to ages 75 and 85 are 63% and 42%, respectively. About 2.1% of the observations do not survive to period 2.

1.5 Survival Expectations and Bias

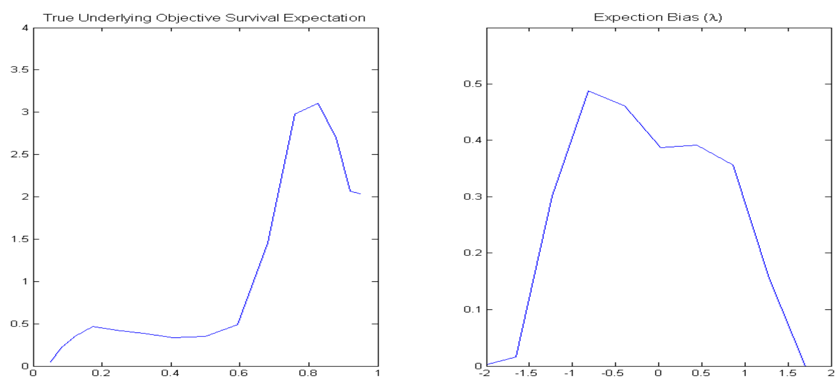


FIGURE 1.2: Objective Survival Expectation and Expectation Bias, Whole Sample

Left panel shows the objective survival expectations $E(O_{t+1}|S_t)$. Right panel shows the distribution of the expectation bias λ_t. Whole sample.

Recall in the binary outcome specification in Section 1.2, an individual information set S_t determines objective expectations, $E(X_{t+1}|S_t) = \Phi(S_t)$; while the expectation bias λ_t is defined in $\hat{E}(X_{t+1}|S_t) = \Phi(S_t + \lambda_t)$. In this section, I consider the binary outcome of survival status at age 75 for those who were 61 to 65 years old and reported their subjective expectations of surviving to age 75 in 1992. This

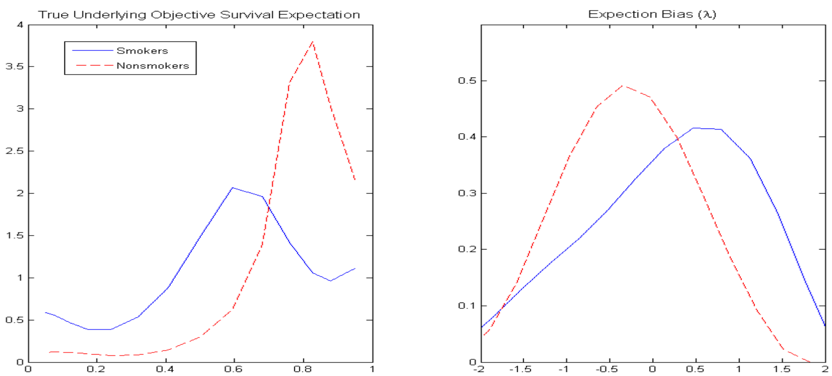


FIGURE 1.3: Objective Expected Survival and Bias, By Smoking Status

Left panel shows the objective survival expectations $E(O_{t+1}|S_t)$ for smokers (solid line) and nonsmokers (dashed line). Right panel shows the distribution of the expectation bias λ_t for smokers (solid line) and nonsmokers (dashed line).

particular analysis sample is used in this subsection, due to the availability of information on both the observations' survival outcomes and their subjective survival probabilities by age 75. Younger age groups have not reached the target age, so their survival outcomes cannot be observed yet; while older age groups' subjective survival probabilities to age 75 were not elicited. For the estimation of the structural dynamic discrete choice model explained in the following sections, the analysis sample is the one described in Subsection 2.3.2.

Figures 1.2 and 1.3 show the distributions of the true underlying objective survival expectations (left panels) and the expectation bias (right panels) for the whole sample, and by smoking status. The shapes of the distributions are qualitatively similar for all samples. However, overall, smokers are objectively less likely to survive to age 75 than nonsmokers, as evident from the heavier left tail of the dashed line in the left panel of Figure 1.3. On the other hand, as shown in the right panel of Figure 1.3, the expectation bias for smokers is positively skewed with a large mass above zero, implying that the majority of smokers are optimistic about their survival; while the opposite is true for nonsmokers.

Table 1.3: Distribution of Objective Expectation and Expectation Bias

	$E(O_{t+1}\|S_t)$	$E(S_t)$	$Std(S_t)$	$E(\lambda_t)$	$Std(\lambda_t)$	$\Phi(E(S_t)+E(\lambda_t))-\Phi(E(S_t))$
All	0.72	0.63	0.67	-0.13	0.78	-0.04
Smoker	0.58	0.26	0.75	0.24	0.99	0.09
Nonsmoker	0.76	0.72	0.56	-0.22	0.73	-0.07

$E(O_{t+1}|S_t)$ is average survival rate in the sample. $E(.)$ and $Std(.)$ are the estimated mean and variance of S_t and λ_t, respectively; while the last column $\Phi(E(S_t)+E(\lambda_t))-\Phi(E(S_t))$ calculates the average difference between subjective and objective survival expectation. Data are from the 1992 HRS.

I quantify the differences in the means and variances of the true underlying objective expectations and expectation bias in Table 1.3. As shown in this table, for the whole sample, the mean objective survival expectations is 0.72, with a small negative average expectation bias implying that this group slightly underestimates its survival probabilities by 0.04. Smokers, on the other hand, have a lower average objective survival probability, but tend to overestimate their survival probabilities. On average, the difference between subjective and objective survival probabilities are 0.09 for smokers, relative to -0.07 for nonsmokers who underestimate their survival probabilities. Smokers are also found to have more dispersed distribution of the expectation bias, which has a standard deviation of 0.90 compared to 0.73 for nonsmokers and 0.78 for the whole sample.

Similar analysis can be conducted based on other individual characteristics, e.g., self-reported health status. Respondents who report having bad health on average have lower objective expectations of survival (0.51) than their healthier counterparts (0.79). Individuals with bad health are also found to be subjectively more pessimistic about their survival, with substantially more dispersed expectation bias (standard deviation of 0.88 versus 0.35 for those with good health). Distributions of objective expectations and expectation bias are qualitatively similar in shape to those depicted in Figures 1.2-1.3, and are omitted in the interest of space.

1.6 Empirical Estimation

1.6.1 Estimation of Survival Probabilities

The dynamic discrete choice model requires knowledge of one-period ahead subjective expectations of state transitions. The available subjective longevity expectations, however, are about survival probabilities at a certain target age which is at least 14 years in the future for the observations in the sample. To solve this mismatch problem, I propose a new method, an Expectation Tree, to analyze the formation of individuals' subjective expectations of certain events multiple periods ahead. The main characteristic of this new method is that it puts dynamics into individuals' formation of such subjective expectations by explicitly taking into consideration forward-looking agents' uncertainty about future state transitions.

Specifically, according to the Expectation Tree illustrated in Figure 1.4, when asked at the time of interview (period 0) about their subjective longevity expectations to certain target age, say 75, respondents first think about the probabilities that they will die before the next period (period 1); then, if alive at period 1, the probabilities that they will be in any of the possible (alive) states at period 1. Now, given that they actually reach one of those possible states at period 1, say state 2, again, they think about the probabilities that they will not survive to period 2; and if still alive, the probabilities that they will reach all the different possible states. They keep thinking forward in this way until reaching the target age.

Therefore, the subjective expectations elicited from the respondents are the weighted summations of probabilities of getting into all the possible states at the target age, taking into account the probabilities of dying along the way (from the age at period 0 to the target age), conditional upon their states at the time of interview. That is, the elicited subjective expectations are results of forward-looking individuals' careful calculations with all possible situations in the relevant future taken into consideration.

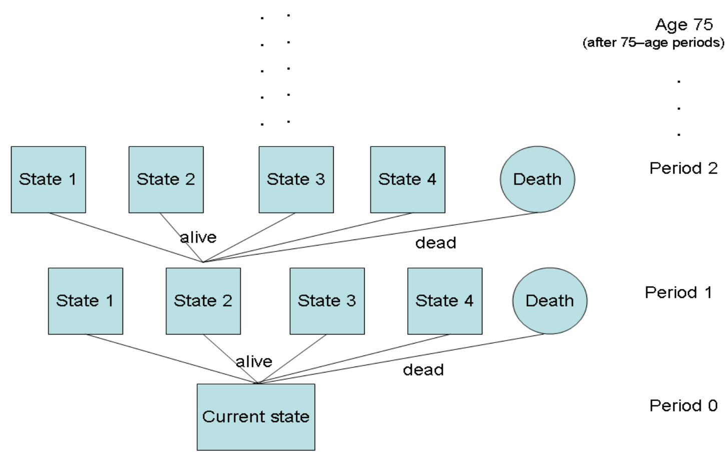


FIGURE 1.4: Expectation Tree: Formation of Longevity Expectations

Graphical illustration of Expectation Tree for survival outcome.

This new dynamic method has the following features. First, this Expectation Tree provides a new way of understanding and analyzing individuals' formation of subjective expectations, which is consistent with the decision-making processes of agents in a dynamic discrete choice model. Specifically, this Expectation Tree can also be viewed as the decision tree forward-looking individuals use to make discrete choices. That is, when faced with options A and B, individuals first consider the probabilities of getting into each and every one of all the possible states following either choice. Once in one of the possible states after their choices, they again need to think about the probabilities of getting into all the possible states following options A and B in the next period. They keep thinking forward in this manner and finally make the optimal decisions after comparing the net benefits of the two options.

Note that, in the decision tree for a dynamic discrete choice model, there are utilities attached to all of the possible states, which individuals use to make the optimal choices which have the highest expected total utilities. When individuals form their subjective expectations using the Expectation Tree, however, they only need to do the forward-looking calculations, no utility maximization needed. This new method provides a connection between individuals' subjective expectations and their discrete choices in a dynamic context where both outcomes – expectations and choices – are generated in the same manner and therefore offer consistent information for the recovery of individual preferences.

Second, this method provides an alternative to the two main methods currently used in the literature in dealing with subjective expectations data: subjective hazard model and Bayesian-style information updating mechanism.[25] Compared with the subjective hazard model, this new method does not require any assumption on the functional form of the underlying subjective hazard, but can still be used to infer subjective (longevity) expectations between any two points in time. Unlike the Bayesian-style information updating mechanism, this new method does not require multiple measures of subjective expectations to infer their formation. Notwithstanding these differences, all three methods can complement each other in dealing with subjective expectations data in different contexts.

To estimate the Expectation Tree, first denote by $F(xa, T)$ the probability of surviving another T periods to the target age under current status xa. Recursively, it follows that

$$F(xa, T) = F(xa, 1) * \sum_{xa'} p(xa'|xa)F(xa', T-1), \tag{1.19}$$

where $F(xa, 1)$ and $F(xa', T-1)$, by definition, refer to the probabilities of surviving another 1 and $T-1$ periods, given the current status xa or xa', respectively. $P(xa'|xa)$ denotes the probability of reaching status xa', given current status, xa.

Equation (1.19) shows that the probability of surviving T periods is the product of the probability of surviving to the next period ($F(xa, 1)$) and the weighted

[25] For subjective hazard model, see, e.g., [36], [54], [55], [78], [30], and [31]; for Bayesian-style information updating mechanism, see, e.g., [91], [95], [94], [85], and [63]. [18] proposes a new metric to measure revisions to subjective expectations.

summation of the probabilities of surviving the remaining $T-1$ periods, where the weights are the probabilities of reaching different statuses in the next period given the current status. Note that *status* here includes both the agents' states (x) excluding their survival status (alive or not), and their choices a. That is,

$$p(xa'|xa) \equiv p_a(a'|x') * p_x(x'|x, a, \text{alive}') \tag{1.20}$$

where p_a and p_x are the conditional choice and transition probabilities, respectively. Variables with $'$ are those in the next period.

When $T = 2$, we can write

$$F(xa, 2) = F(xa, 1) * \sum_{xa'} p(xa'|xa) F(xa', 1). \tag{1.21}$$

Once we assume a certain functional form for $F(xa, 1)$ with a set of parameters, θ, we can write $F(xa, 2)$ according to Eq. (1.21), which can in turn be used to write $F(xa, 3)$ in a similar way by replacing the second $F(xa, 1)$ on the right-hand side of Eq. (1.21) with $F(xa, 2)$, and eventually lead us to $F(xa, 75 - \text{age})$. We can then obtain parameter estimates ($\hat{\theta}$) in $F(xa, 1)$ using optimization methods by matching the elicited subjective expectations to the predicted ones.

Eqs. (1.19) and (1.21) show that the elicited longevity expectations are results of subjective expectations of future state transitions and choices. An ideal investigation of subjective expectations would obtain individual assessments of all possible events in different contexts. For many reasons, it is currently impractical to gather such complex information from survey respondents. Therefore, with only the elicited subjective longevity expectations, we need to make certain compromises. That is, to estimate the subjective probability of surviving to the next period ($F(xa, 1)$), which is an essential part of the dynamic discrete choice model of individuals' smoking behavior, we need to estimate the transition probabilities ($p(xa'|xa)$), including the conditional choice probabilities and the conditional state transitions, using only objectively observed data.

1.6.2 Estimation of Utility Parameters and Discount Factors

For each value of the discount factor β within the reasonable range, we can obtain the parameter estimates, $\hat{\theta}$, from Eq. (1.18) using certain optimization methods. One of the simplest ways is to regress the left-hand side dependent variable which contains only information directly obtained from the data:

$$d(x_t) - \textstyle\sum_{s=t+1}^{T} \beta^{s-t} E\left[\log(1 + e^{-d(x_s)})|x_t = x, a = 1\right] \\ + \textstyle\sum_{s=t+1}^{T} \beta^{s-t} E\left[\log(1 + e^{d(x_s)})|x_t = x, a = 0\right],$$

on the right-hand side generated regressors:

$$\sum_{s=t}^{T} \beta^{s-t} E\left[x_{1s}|x_t = x, a = 1\right]' \theta_1 \quad \text{and} \quad \sum_{s=t}^{T} \beta^{s-t} E\left[x_{0s}|x_t = x, a = 0\right]' \theta_0,$$

where the only unknowns are θ.[26]

Given the utility parameter estimates at each possible value of the discount factor, β, we can then use a linear line search to locate the discount factor with the maximum likelihood, where the likelihood is evaluated at both the discount factor and the corresponding utility parameter estimates.

Conditional choice probabilities are flexibly estimated using the Logit model, with up to fourth order polynomials and interaction terms of the state variables. Conditional transition probabilities of the state variables (excluding the vital status) are estimated non-parametrically.[27] For comparison purposes, I also estimate

[26] Appendix A.5 shows the steps to obtain $\sum_{s=t}^{T} \beta^{s-t} E\left[x_a^s|x_t = x, a\right]'$.

[27] For estimation purposes, I discretize the continuous variable – the logarithm of household income – into a finite number of discrete values. Furthermore, due to the large number of discrete states and the relatively sparse data for each value of the discrete state variables, I kernel smooth over the discretized income levels when I nonparametrically estimate the conditional transition probabilities of the state variables. For details on kernel smoothing, see Appendix A.6.

the model by assuming rational expectations. The only difference in this objective estimation (since there are no subjective expectations data) is that it uses a Logit model with the observed mortality as the dependent variable for the estimation of survival probabilities.

The estimation steps are:

Step 1 Flexibly estimate the dependent variable and the regressors in Eq. (1.18). This estimation requires estimations of conditional choice probabilities, conditional health and income transitions, and survival probabilities.

Step 2 For each discount factor β inside the reasonable range $[0, 1]$, recover $\hat{\theta}$ using Eq. (1.18).

Step 3 Construct the sample likelihood at β and $\hat{\theta}(\beta)$.

Step 4 Repeat Steps 2 and 3 for each discount factor β along a line search to find the β with the maximum likelihood.

The estimation procedure described above assumes there are no persistent unobserved differences in the transition of the state variables and in the preferences for smoking and health. It is also possible to account for unobserved heterogeneity following the insights of [41], [52], [51], and [44], and the methods of empirical implementation suggested by [3] and [2].

1.6.3 Empirical Specifications of States and Utilities

The empirical specification follows the framework of the demand for health introduced by [33].[28] In each period, individuals choose whether to smoke ($a = 1$) or not ($a = 0$), after they observe all the state variables, including the utility shocks, which are unobserved to the econometrician. The complete set of state variables includes

[28] See, also, [7], [4], and [5].

whether the individuals are alive or not, whether they are in bad health or not, (the logarithm of) their real household income,[29] age, gender, race (Non-Hispanic White or not), and same-gender parents' longevity. This last variable, as mentioned in Subsection 2.3.2, is created to control for the differences in the agents' expected longevity due to (unobserved) familial and genetic reasons. This variable also serves as the exclusive restriction for the identification of the discount factor (see Section 1.3.2). Specifically, I assume that same-gender parents' longevities affect the agents' own expectations about future survival and health, without affecting the agents' instantaneous utilities associated with their smoking choices in a different way.

The instantaneous period utility when the agents choose not to smoke ($a = 0$), depends upon whether they have bad health and (the logarithm of) their household income which is used here to measure the composite good. If the agents decide to smoke ($a = 1$), the period utility depends only on the (logarithm of) their household income. The utility of a deceased individual is normalized to be zero. That is,

$$u_0 = \alpha_0 + \alpha_1 * \text{bad health} + \alpha_2 * \log(\text{household income})$$

$$u_1 = \log(\text{household income}).$$

Agents' uncertainty about future comes from uncertainty about future survival status, whether they will be alive or not, and if alive, future health and income states. The survival probabilities and the conditional probabilities of having bad health and a certain amount of household income in the next period depend upon individuals' choices (whether to smoke or not), the state variables that reflect individuals' natural and biological initial conditions (age, race, gender, same-gender parents' longevity), and the state variables that are linked to the decisions that have been made up to the current period (alive or not, current health status, and household income).[30]

[29] Household income per capita is also tried here as a robustness check.

[30] See [24] and [25] for economic justification for the variables included in this specification.

1.7 Estimation Results

1.7.1 Conditional Transition Probabilities

Figure 1.5 shows the estimation results for conditional health transitions. As expected, individuals who are in bad health at period one or do not have long-lived same-gender parents are much more likely to have bad health in the next period (Figure 1.5, (a) and (b)). Smoking today increases one's chance of having bad health tomorrow, regardless of the current health state (Figure 1.5, (c) and (d)). Higher household income, as is clear from every panel in Figure 1.5, predicts a lower probability of having bad health in the next period.

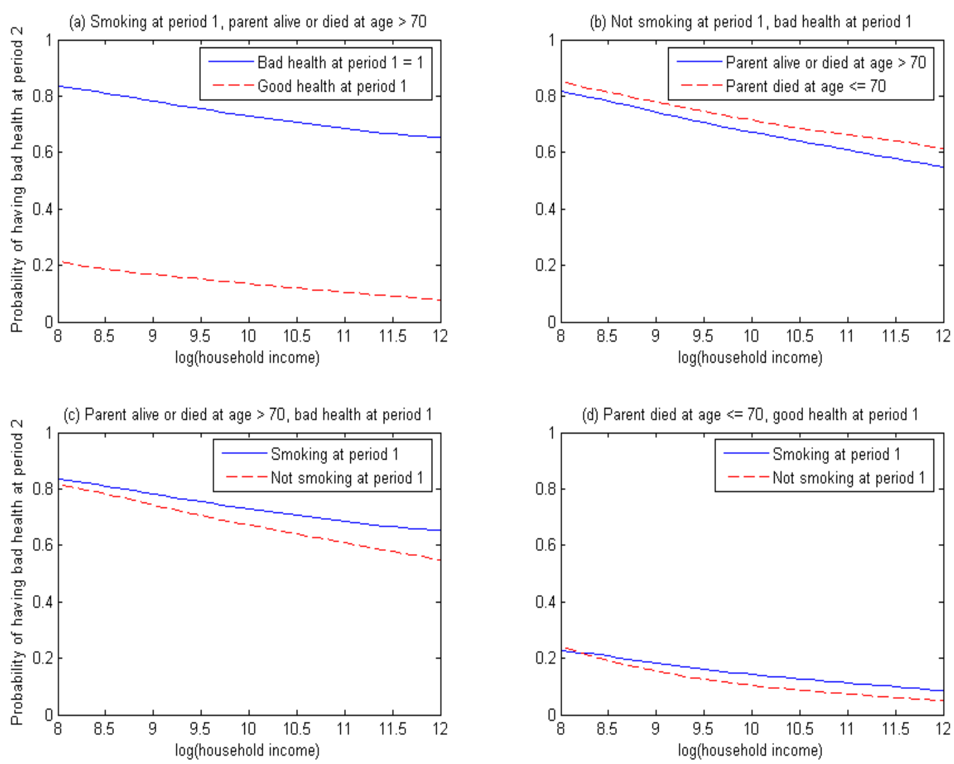


FIGURE 1.5: Health Transition Conditional Upon Survival

Probabilities of having bad health in two years as a function of household income by current health status, same-gender parents' longevity, and smoking status.

1.7.2 Survival Probabilities

Table 1.4 reports the estimation results for the probabilities of dying before the next period (in 2 years in this case). The first two columns are the results of the objective estimation based on the rational expectations assumption, with observed deaths before the next period as the dependent variable. The results are as expected: currently smoking, having had health, and aging increase one's probability of dying in 2 years, while White females with higher household incomes and long-lived same-gender parents are less likely to die in 2 years. All estimates are statistically significant at the 1% level, except for the variable measuring same-gender parents' longevity, which, although with the expected sign, is neither economically nor statistically as significant as other factors.

Table 1.4: Estimation Results for Probability of Dying in Two Years

Variable	Obj. Coef.	Std. Err.	Subj. Coef.	Std. Err.
Currently smoke	0.42**	0.11	0.39**	0.19
Same-gender parent's longevity	-0.10	0.11	-0.33**	0.04
Bad health	1.65**	0.13	3.19**	0.66
Log(household income)	-0.25**	0.05	-0.13**	0.08
White	-0.25	0.13	0.34**	0.04
Female	-0.68**	0.12	-0.25**	0.03
Age	0.08**	0.02	0.07**	0.02
Constant	-6.13**	1.30	-8.01**	1.53

Probabilities are parameterized using Logistic specification. Objective estimation uses a Logit model with observed mortality as the dependent variable. Subjective estimation relies on elicited multiple periods ahead subjective longevity expectations using the Expectation Tree method.
*Statistically significant at 5% level; **statistically significant at 1% level.

The last two columns of Table 1.4 report the estimation results of the probabilities of dying in 2 years using subjective longevity expectations data. First, note that each parameter estimate has the same sign as its objective counterpart, except for "being White". That is, people understand correctly the general effects of different determinants on their survival probabilities. For example, having bad health right

Table 1.5: Marginal Effects on Probabilities of Dying in Two Years

Variable	Obj. Coef.	Std. Err.	Subj. Coef.	Std. Err.
Currently smoke	0.006*	0.002	0.004**	0.001
Same-gender parent's longevity	-0.001	0.002	-0.004**	0.001
Bad health	0.040**	0.004	0.034**	0.002
Log(household income)	-0.004**	0.001	-0.001**	0.0004
White	-0.004	0.002	0.004**	0.001
Female	-0.010**	0.002	-0.003**	0.002
Age	0.001**	0.0002	0.001**	0.0003

Objective estimation uses a Logit model with observed mortality as the dependent variable. Subjective estimation relies on elicited multiple periods ahead subjective longevity expectations using the Expectation Tree method. Marginal effects are calculated holding other variables constant at mean levels. Discrete changes for binary variables from 0 to 1 are reported.
*Statistically significant at 5% level; **statistically significant at 1% level.

now will increase their chances of dying in 2 years, and having a long-lived same-gender parent is positively associated with survival. However, Whites in our sample, as mentioned in the previous paragraph, are objectively estimated to be less likely to die, which is consistent with the life tables,[31] but subjectively, our respondents think that being White lowers ones' chances of survival. The result that individuals have the right idea about the directions of the effects of different determinants is also found by others using different methods for different data sets (see, e.g., [84], [48], and [46]).

Because both sets of survival parameters are estimated using nonlinear methods, the marginal effects are also reported (Table 1.5). The magnitudes of the estimates show that currently being in bad health has the greatest effect on the probability of dying in 2 years, objectively or subjectively. Being a smoker is almost equivalent to adding 4 to 6 years to one's age in terms of its effects on mortality.

One main difference in the relative magnitudes of objective and subjective parameter estimates is noteworthy. Genetic information, summarized by the same-gender parents' longevity, is given a disproportionately heavy weight subjectively, making

[31] See, for example, http://www.cdc.gov/nchs/products/pubs/pubd/lftbls/decenn/1991-89.htm for the life table constructed by CDC.

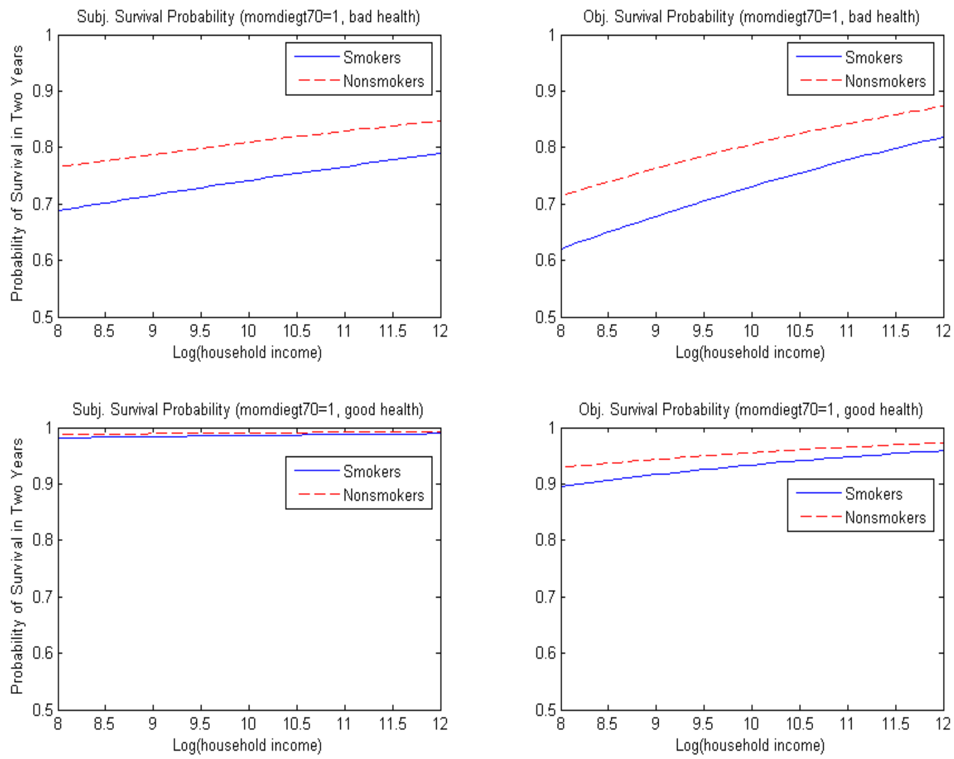


Figure 1.6: Two-Year Survival Probabilities for Non-White Males

Probabilities of surviving another two years as a function of household income by smoking status for Non-White males. Left panels: subjective estimations based on the Expectation Tree. Right panels: objective estimations using observed mortality data.

it almost as important as current smoking status. One interpretation is that, when forming their longevity expectations, people think that having "good" genes is as important as the health-related choices they make. Alternatively put, people believe that as long as they have good genes, the negative effects of their bad health behaviors such as smoking can be canceled out.

That people tend to overestimate the importance of their parents' longevity for their own life expectancy was also found by [36] using a smaller sample of younger age groups. One explanation to this overestimation might be what [88] call the "availability heuristic," an over-reliance on apparently relevant information. However, the fact that individuals form expectations differently than described by the rational expectations assumptions, regardless whether they are justified to think this way, is the main message here.

As a graphical example, Figure 1.6 shows the 2-year survival probabilities for Non-White males with long-lived parents. The right two figures are recovered from subjective longevity expectations, while the left two are based on objective estimation assuming rational expectations. We can make several observations from this figure. First, subjective and objective estimations both show that, survival probabilities are lower for smokers and those with bad health but increase with household income. Second, subjective survival probabilities are almost everywhere above their objective counterparts, showing that individuals in this group are optimistic about their longevity. Third, the difference in survival probabilities by smoking status estimated using subjective survival expectations is smaller than that estimated under rational expectations assumption, implying that individuals in this group tend to underestimate the negative health effects of smoking on their longevity.

1.7.3 Utility and Time Preferences

Table 1.6 reports the objective and subjective estimates of utility and time preferences. The first similarity between these two sets of estimation results is the negative sign associated with having bad health, implying that bad health hurts nonsmokers more than smokers. Alternatively put, current smokers are less concerned with having bad health than nonsmokers, which could potentially explain their smoking choices in the first place. This result is consistent with what is found in the literature. For example, [56], using the stated preferences approach, also find that current smokers have substantially lower willingness-to-pay than nonsmokers to be rid of the highly costly and mostly smoking-related disease: Chronic Obstructive Pulmonary Disease (COPD).

The second similarity lies in the sign and magnitude of parameter estimates for the

Table 1.6: Utility Parameters and Discount Factor

Variables	Obj.Estimates	Subj.Estimates
Bad health	-0.46	-0.77
Log(household income)	1.35	1.33
Constant	-3.48	-3.42
β	0.61	0.70

Objective estimation is based on the rational expectations assumption (Column 1). Subjective estimation uses subjective longevity expectations (Column 2).

logarithm of real household income: both are positive and greater than 1. Since the utility of current smokers is normalized to be the logarithm of their household income, these estimates show that: 1) marginal utility of income is higher for nonsmokers than for current smokers; and 2) the difference in individuals' instantaneous utilities by smoking status increases with (logarithm of) household income, so individuals with higher household incomes will experience more utility loss if they choose to smoke than their lower-income counterparts. These findings concur with previous studies' results that smoking tends to be more prevalent among low-income consumers,[32] and are consistent with the idea of state-dependent utilities introduced by [96].

Two differences in the estimated utility parameters and discount factors between objective and subjective estimations are economically and statistically significant. First, although they share the same sign, the subjective estimation of the parameter for having bad health is more than 60% greater than the objective estimation. This result implies that the difference in how much they care about having bad health between current smokers and nonsmokers is much more significant than we can estimate under the rational expectations assumption. Without subjective expectations data, we would have underestimated this difference.

The second main difference lies in the estimated discount factors. The (two-year) discount factors are estimated to be 0.61 and 0.70 for objective and subjective estima-

[32] See, e.g., [97], [87], and [15].

tions, respectively. Both are within the reasonable range suggested by the literature, implying that agents in this analysis sample are rather forward-looking.[33] However, the subjective discount factor is much greater than the objective one, showing that individuals are much more forward-looking or patient with time than we would have concluded under the rational expectations assumption using only objective data. Without subjective expectations data, we would have underestimated individuals' patience.

Combining the estimation results on survival probabilities with the estimated utility parameters and discount factors, we can reach two important conclusions. First, individuals care greatly about their health, much more than we would have concluded. Second, individuals are also more patient and pay more attention to their future than under the rational expectations assumption. Both conclusions have important implications as we evaluate the effectiveness of various public policies. One possible explanation of the fact that we can still find so many smokers is that they attach disproportionately large weight to their genetic makeup when they form their longevity expectations. They even think that "good" genes can protect them against the detrimental effects of harmful health behaviors, such as smoking.

This finding that there is still gap between individuals' subjective beliefs in the relative effects of their smoking choices and other characteristics on their survival and health and those objectively identified calls for further effort in policy intervention such as information campaign. Specifically, more personalized anti-smoking messages which highlight both the absolute and the relative importance of quitting smoking to mortality and morbidity may further shrink the gap between subjective and objective expectations and lead to lower smoking rates.[34]

[33] See Appendix Table A.3 for examples of discount factors estimated in the literature.

[34] Estimation results with unobserved heterogeneity are work in progress.

1.7.4 Goodness of Fit

Figure 1.7 shows two measures of within sample goodness of fit of the estimated parameters. Specifically, I compare the predicted smoking rates using both subjectively and objectively estimated parameters with those from the data at different ages and (logarithm of) household income levels. These two graphs show that both estimation methods have done a great job in matching smoking rates with the data: the predicted smoking rates are right around the true smoking rates from the data, and both methods capture the decreasing trend in smoking rates over age and income levels.

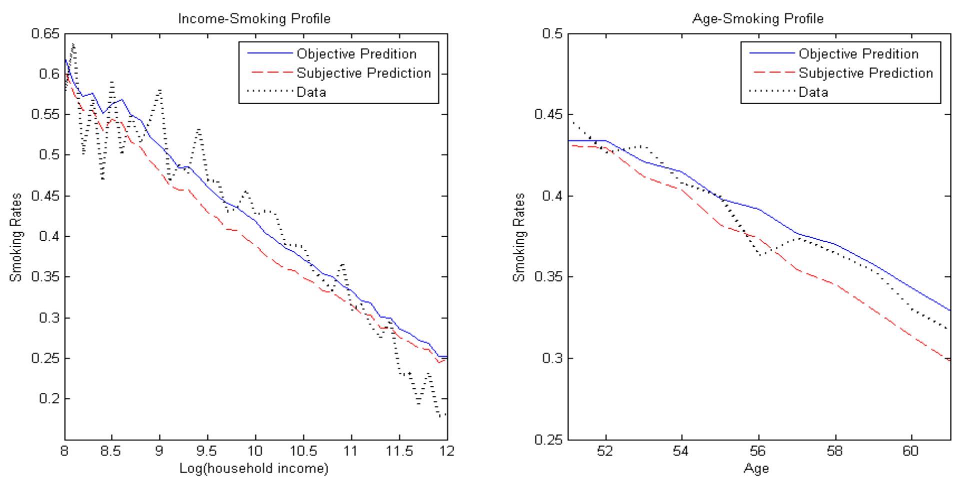


FIGURE 1.7: Within-Sample Goodness of Fit

Model implied smoking probabilities based on the rational expectations assumption, subjective expectations data, and from the data. Left panel depicts the smoking rates with respect to household income; right panel depicts the smoking rates with respect to age.

The finding that estimation using subjective longevity expectations data alone, without any objectively observed information on mortality, can still perform very well in fitting the data is consistent with the literature. [30], using the data from Asset and Health Dynamics Among the Oldest Old, also find that parameter estimates using subjective mortality risk perform better in predicting out-of-sample wealth levels than estimates using life table mortality risks. Similarly, [63] shows that the perceived crime rate among youth can better predict their criminal behavior than the official neighborhood crime rate.

1.7.5 Policy Implications

Subjective expectations data and the estimated utility and time preferences can be used to assess by simulation precisely what the impact would be of different policy interventions on smoking choices. As an illustration, I consider below two specific public policies. The first one is an information campaign, which has been proven to be effective in informing individuals of the harmful effects of smoking and consequently reducing the smoking rate. The case we study here is when the information campaign is so effective that subjective marginal effects of smoking on survival are exactly the same as those objectively estimated under the rational expectations assumption. This policy experiment aims at showing the effects of (at least partially) reducing the expectation bias.

The second policy we consider is the effects of advances in medical technology. We focus on changes in the age-smoking profile given anticipated or unanticipated advances in medical technology, which are assumed to make smoking less detrimental to health by lowering the harmful effects of smoking on survival by 50 percent. We focus on people aged 51 and assume that the advances in medical technology occur when they are at age 54.

The left panel of Figure 1.8 shows that, if individuals' subjective expectations about the marginal effects of smoking on their survival are exactly like the objective ones, the smoking rate for the age group studied in this paper will decrease. This drop in smoking rate ranges from 2% to 6% with an average of 4%.

The right panel of Figure 1.8 shows the results of the second policy experiment. As expected, if individuals anticipate the advances in medical technology, the smoking rate will go up even before the advances actually occur. This is because individuals

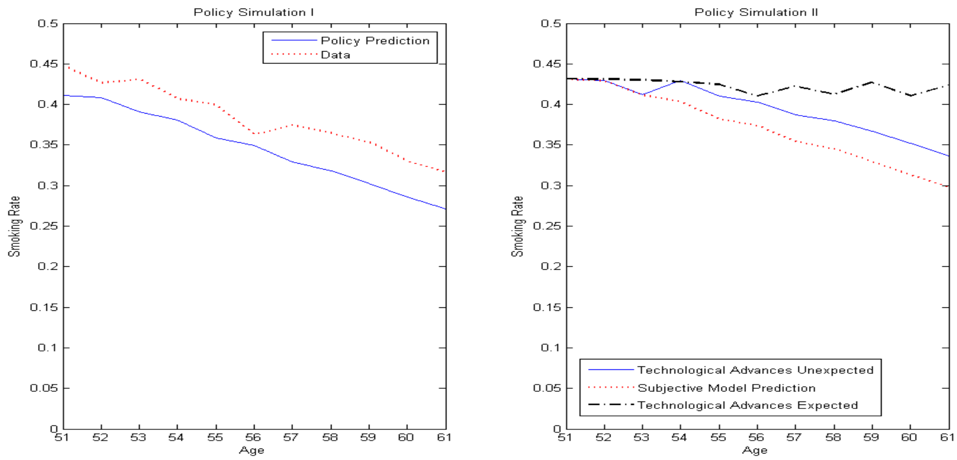


FIGURE 1.8: Policy Experiments

Left panel: smoking rates from the data and those predicted by policy experiment which sets the subjective marginal effects of smoking at the objectively estimated levels. Right panel: smoking rates from the data and those predicted by policy experiments with expected and unexpected technological advances which cut down the negative effects of smoking by half.

expect the health costs of smoking to be smaller in the future, which makes smoking a more attractive option now. If individuals do not anticipate the advances in medical technology, then the smoking rate will not jump before the advances occur. Anticipated or not, the medical advances will increase the smoking rate, but the smoking rate will still decrease with age.

1.8 Conclusion and Discussion

In this paper, I present a new approach to assess the rational expectations assumption using subjective expectations data. I also show how to relax the rational expectations assumption by bringing subjective expectations data into discrete choice models to improve the inference on behavior. As one of the few studies using subjective expectations data in a dynamic structure model, this paper shows that these data are critical in conducting analysis under weaker assumptions than what is usually imposed in the literature.

My central empirical finding is that individuals care more about their health and are more forward-looking than we would have concluded under the rational expectations assumption. To recover these true preferences, I analyze individuals' own longevity expectations and find that the expectation bias does exist. Specifically, individuals put disproportionately heavier weights on genetic information, yet pay less attention to the effects of their health related behaviors. Policy experiments further show that if individuals do not have biased expectations about the marginal effects of smoking on their mortality, the average smoking rate will be around 4% lower than the current value. Therefore, at least for the empirical question asked in this paper, it is not innocuous to make the simplifying yet strong rational expectations assumption. Subjective expectations data should be used to conduct the analysis in order to allow for the expectation biases which have important implications for choice predictions and policy evaluations.

Individuals in the current model are assumed to have possibly biased yet mature opinions about the effects of different determinants on certain outcomes. This assumption can be justified in the current sample of the elderly in the U.S. who one can argue have passed the initial information-gathering stage. An interesting extension of this model is to allow for learning or belief revision which could play an important role in explaining adolescent behaviors.[35]

The empirical results are specific to the empirical question asked and the analysis sample used in this paper. The method proposed and the model used here, however, can be generalized to other decision-making processes such as consumption and saving patterns, decisions on labor supply and education investment choices.

Even though this paper is able to relax certain strong assumptions about individuals' expectations, it still requires other assumptions in its estimation mainly due to the available subjective expectations data. For example, the availability of counterfactual subjective expectations data would allow for more flexible forms of

[35] See, e.g., [95], [91], [94], [10], and [18] for analyses of belief revision.

expectation bias or a greater role of private information in ones' decision-making processes.

2

Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions

This paper extends the semi-parametric identification and estimation method for dynamic discrete choice models using Hotz and Miller's (1993) conditional choice probability (CCP) approach to the setting where individuals may have hyperbolic discounting time preferences and may be naive about their time inconsistency. We implement the proposed estimation method to the decisions of undertaking mammography to evaluate the importance of present bias and naivety in the underutilization of mammography. Preliminary results show evidence for both present bias and naivety.

2.1 Introduction

Dynamic discrete choice models have been used to understand a wide range of economic behavior. The early dynamic discrete choice models that are empirically implemented tend to be parametric;[1] but recently, a growing list of authors have addressed the non- or semi-parametric identification of dynamic discrete choice models. The earliest attempt in this regard is Hotz and Miller (1993) which pioneered the approach of using conditional choice probabilities to infer about choice-specific continuation values. Rust (1994a, 1994b) showed that the discount factor in standard dynamic discrete choice models are generically not identified; Magnac and Thesmar (2002) expanded Rust's non-identification results, and proposed exclusive restrictions that lead to the identification of the standard discount factor (see more below).

All of the above-mentioned literature model the impatience of the decision makers by assuming that agents discount future streams of utility or profits *exponentially* over time. As is now well known, Strotz (1956, p.172) showed that exponential discounting is not just an analytically convenient assumption; without this assumption, intertemporal marginal rates of substitution will change as time passes, and preferences will be time-inconsistent. A recent theoretical literature has built on the work of Strotz (1956) and others to explore the consequences of relaxing the standard assumption of exponential discounting. Drawing both on experimental research and on common intuition, economists have built models of quasi-hyperbolic discounting to capture the tendency of decision makers to seize short term rewards at the expense of long-term preferences.[2] This literature studies the implications of time-inconsistent preferences, and their associated problems of self-control, for a variety of economic

[1] The earliest formulation and estimation of parametric dynamic discrete choice models include Wolpin (1984) for fertility choice, Miller (1984) for occupational choice, Pakes (1986) for patent renewal, and Rust (1987) for bus engine replacement.

[2] A body of experimental research, reviewed in Ainslie (1992) and Loewenstein and Elster (1992), indicates that hyperbolic time discounting may parsimoniously explain some basic features of the intertemporal decision making that are inconsistent with simple models with exponential discounting. Specifically, standard decision models with exponential discounting are not easy to reconcile with commonly observed preference reversals: subjects choose the larger and later of two prizes when both are distant in time, but prefer the smaller but earlier one as both prizes draw nearer to the present (see Rubinstein 2003, however, for an alternative explanation of preference reversals).

choices and environments.[3]

A small list of empirical papers that attempted to estimate dynamic models with hyperbolic discounting time preferences have followed the parametric approach (Fang and Silverman 2007, Laibson, Repetto and Tobacman 2007 and Paserman 2007).[4] Fang and Silverman (2007) empirically implement a dynamic structural model of labor supply and welfare program participation for never-married mothers with potentially time-inconsistent preferences. Using panel data on the choices of single women with children from the National Longitudinal Survey of Youth (NLSY 1979), they provide estimates of the degree of time-inconsistency, and of its influence on the welfare take-up decision. For the particular population of single mothers with dependent children, they estimate the present bias factor and the standard discount factor to be 0.338 and 0.88 respectively, implying a one-year ahead discount rate of 238%. Laibson, Repetto and Tobacman (2007) use Method of Simulated Moments (MSM) to estimate time preferences - both short and long run discount rates – from a structural buffer stock consumption model that includes many realistic features such as stochastic labor income, liquidity constraints, child and adult dependents, liquid and illiquid assets, revolving credit and retirement. Under parametric assumptions on the model, the model is identified from matching the model's predictions of retirement wealth accumulation, credit card borrowing and consumption-income co-movement with those observed in the data. Their benchmark estimates imply a 48.5% short-term annualized discount rate and a 4.3% long-term annualized discount rate. Paserman (2007) estimates the structural parameters of a model of job search with hyperbolic discounting and endogenous search effort, using data on duration of

[3] For example, models of time-inconsistent preferences have been applied by Laibson (1997) and O'Donoghue and Rabin (1999a,b) to consumption and savings; by Barro (1999) to growth; by Gruber and Koszegi (2001) to smoking decisions; by Krusell, Kuruşçu, and Smith (2002) to optimal tax policy; by Carrillo and Mariotti (2000) to belief formation; and by Della Vigna and Paserman (2001) to job search.

[4] Also related, Arcidiacono, Sieg and Sloan (2007) estimate a parametric forward-looking dynamic discrete choice model of smoking and heavy drinking for late-middle age men in the Health and Retirement Studey (HRS) and find that a forward-looking model fits the data (mainly the age profile for heavy-drinking and smoking) better than a myopic model. Their model assumes exponential discounting and thus does not incorporate the possibility that time inconsistent preferences may play a role in the consumption of alcohol and cigarettes.

unemployment spells and accepted wages from the NLSY 1979. Under parametric assumptions of the model, identification of the hyperbolic discounting parameters comes from the variation in the relative magnitude of unemployment duration and accepted wages. Indeed he finds that the results are sensitive to the specific structure of the model and on the functional form assumption for the distribution of offered wages. For low-wage workers, he rejects the exponential discounting model and estimates a one-year discount rate of about 149%.[5]

None of the above papers allow for the possibility that a hyperbolic discounting decision-maker may also be naive. Most importantly, the identification of the present bias and standard discount factors in these papers are often based on parametric assumptions imposed on the model. To the best of our knowledge, it is not known whether dynamic discrete choice models with hyperbolic discounting preferences can be semi-parametrically identified using standard short-panel data that are typically used in these papers.[6]

In this paper, we examine the conditions under which dynamic discrete choice models with hyperbolic discounting time preference can be partially identified using short-panel (two periods) data. We show that, if there exist *exclusion variables* that affect the transition probabilities of states over time but do not affect the decision-makers' static payoff functions, a condition that is similar to that in Magnac and Thesmar (2002) necessary for the identification of dynamic discrete choice models with standard exponential discounting, then we can potentially identify all three discount factors $\beta, \tilde{\beta}$ and δ.

The intuition for why exclusion variables that affect the transition of state variables but not static payoffs can provide source of identification for the discount factors can be simply described as follows. Consider two decision-makers who share the same

[5] There are other inferential studies about discount rates that exploit specific clear-cut intertemporal trade-offs. For example, Hausman (1979), and Warner and Pleeter (2001) estimate discount rates ranging from 0 to 89% depending on the characteristics of the individual and intertemporal trade-offs at stake.

[6] The exception is Fang and Silverman (2006), which argued that exponential discounting and hyperbolic discounting models are distinguishable, using an argument based on observed choice probabilities.

period-payoff relevant state variables but differ only in the exclusion variables. Because their exclusion variables only affect the transition of the payoff-relevant state variables, their effects on the choices in the current period will inform us about the degree to which the agents discount the future. The intuition for why $\beta, \tilde{\beta}$ and δ can be separately identified will be provided later in Section 2.2.4.

We provide two estimation approaches that are intimately related to our identification arguments. One approach is based on maximizing a pseudo-likelihood function and the other is based on minimizing the estimated variance of the utility functions. Monte Carlo experiments show that both estimators perform well in large samples, but in relatively small samples, the maximum pseudo-likelihood based estimator performs better. We thus use maximum pseudo-likelihood based estimator in our empirical application.

Our paper also represents an interesting intermediate case between the literature on estimating dynamic discrete choice single-agent decision problems (see Miller 1984, Wolpin 1984, Pakes 1986, Rust 1987, Hotz and Miller 1993 for early contributions and Rust 1994a, 1994b for surveys) and the more recent literature on estimating dynamic games (Pakes and McGuire 1994, Pakes, Ostrovsky and Berry 2007, Bajari, Benkard and Levin 2007). As is well-known, if an agent has hyperbolic discounting time preferences, the outcome of her decision process can be considered as the equilibrium outcome of an *intra-personal game* with the players being *the selves at different periods.* There are two crucial differences, however, between the intra-personal games we analyze for agents with time-inconsistent time preferences and those in the existing dynamic games literature. The first key difference is that in our case, we *do not observe* the actions of all the players. More specifically, the outcomes – choices and the evolutions of the state variables – we observe in the data are affected only by the current selves, even though the current selves' choices are impacted by their perception of future selves' actions. Secondly, the dynamic games literature (e.g. Bajari, Benkard and Levin 2007) may allow for players to have different period payoff functions, in our setting, however, the payoffs for the players – the current self and the future selves – differ only in time preferences; moreover, under hyperbolic discounting, we are assuming a rather restricted form of time preference differences between the players.

We also apply our identification and estimation method to investigate the role of time inconsistent preferences in women's mammography decisions. We consider a simple model where mammography can potentially lower the probability of death in the next two years, and it may lower the probability of bad health conditional on surviving in two years, but undertaking mammography may involve immediate costs (most of which we would like to interpret as psychological and physical costs instead of financial costs). In particular, we use the indicator for either the woman's mother is still alive or she died at age greater than 70 as the exclusive variable that does not enter the instantaneous utility but affects the transition probability of other state variables that enter the instantaneous utility function. Our preliminary estimates indicate that individuals exhibit both present bias and naivety as $\tilde{\beta}/\beta$ is estimated to be about 1.80, suggesting both $\beta < 1$ (present bias) and $\tilde{\beta} > \beta$ (naivety). We also estimate $\tilde{\beta}\delta$ to be about 0.86. These suggest that both present bias and naivety might have played an important role in the fact that nearly 25% of the women do not undertake mammography as advised by American Cancer Association, which is universally regarded as a very cost effective way for early detection of breast cancer.

The remainder of the paper is structured as follows. In Section 2.2 we describe a general dynamic discrete choice model with hyperbolic discounting time preferences and provide detailed analysis for identification and estimation of the model, particularly the discount factors; we also evaluate the performance of our proposed estimation methods from Monte Carlo experiments. In Section 2.3 we provide the background information for mammography, which is the decision we examine in our empirical application; we also describe the data set used in our study and provide some basic descriptive statistics of the samples; we then provide details about the empirical specification of our model of the decision for undertaking mammography and present the preliminary main estimation results. Finally, Section 2.4 concludes.

2.2 Dynamic Discrete Choice Model with Hyperbolic Discounting Time Preferences

2.2.1 Basic Model Setup

Consider a decision maker whose intertemporal utility is additively time separable. The agent's instantaneous preferences are defined over the action she chooses from a discrete set of alternatives $i \in \mathcal{I} = \{0, 1, ..., I\}$, and a list of state variables denoted by $h \equiv (x, \varepsilon)$ where x, which for notational simplicity includes time t, are observed by the researcher, and $\varepsilon \equiv (\varepsilon_1, ..., \varepsilon_I)$ are the vector of random preference shocks for each of the I alternatives. We make the following assumption about the instant utility from taking action i, $u_i^* (h) \equiv u_i^* (x, \varepsilon)$:

Assumption 1. ***(Additive Separability)*** *The instantaneous utilities are given by, for each* $i \in \mathcal{I}$,

$$u_i^* (x, \varepsilon) = u_i (x) + \varepsilon_i,$$

where $(\varepsilon_1, ..., \varepsilon_I)$ *has a joint distribution* G, *which is absolutely continuous with respect to the Lebesgue measure in* R^I.

We assume that the *time horizon is infinite* with time denoted by $t = 1, 2, ...$ The decision-maker's intertemporal preferences are represented by a simple and now commonly used formulation of agents' potentially time-inconsistent preferences: (β, δ)-preferences (Phelps and Pollak 1968, Laibson 1997, and O'Donoghue and Rabin 1999a):

Definition 1. (β, δ)*-preferences are intertemporal preferences represented by*

$$U_t (u_t, u_{t+1}, ...) \equiv u_t + \beta \sum_{k=t+1}^{\infty} \delta^{k-t} u_k$$

where $\beta \in (0, 1], \delta \in (0, 1]$.

Following the terminology of O'Donoghue and Rabin (1999a), the parameter δ is called the *standard discount factor* and captures long-run, time-consistent discounting; the parameter β is called the *present-bias factor* and captures short-term impatience. The standard model is nested as a special case of (β, δ)-preferences when $\beta = 1$. When $\beta \in (0, 1)$, (β, δ)-preferences capture "quasi-hyperbolic" time discounting (Laibson, 1997). We say that an agent's preferences are time-consistent if $\beta = 1$, and are present-biased if $\beta \in (0, 1)$.

The literature on time-inconsistent preferences distinguishes between *naive* and *sophisticated* agents (Strotz 1956, Pollak 1968, O'Donoghue and Rabin 1999a,b). An agent is *partially naive* if the self in every period t underestimates the present-bias of her future selves, believing that her future selves' present bias is $\tilde{\beta} \in (\beta, 1)$; in the extreme, if the present self believes that her future selves are time-consistent, i.e. $\tilde{\beta} = 1$, she is said to be *completely naive.* On the other hand, an agent is *sophisticated* if the self in every period t correctly knows her future selves' present-bias β and anticipates their behavior when making her period-t decision.

Following previous studies of time-inconsistent preferences, we will analyze the behavior of an agent by thinking of the single individual as consisting of many autonomous *selves,* one for each period. Each period-t self chooses her current behavior to maximize her current utility $U_t(u_t, u_{t+1}, ...)$, while her future selves control her subsequent decisions.

More specifically, let the observable state variable in period t be $x_t \in \mathcal{X}$ where $\mathcal{X}$ denotes the support of the state variables and the unobservable choice-specific shock $\varepsilon_{it} \in \Im$, and $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{It})$. We assume that $\mathcal{X}$ is a finite set and denote $X = \#\mathcal{X}$ to be the size of the state space.

A *strategy profile* for all selves is $\sigma \equiv \{\sigma_t\}_{t=1}^{\infty}$ where $\sigma_t : \mathcal{X} \times \Im^I \to \mathcal{I}$ for all t. It specifies for each self her action in all possible states and under all possible realizations of shock vectors. For any strategy profile σ, write $\sigma_t^+ \equiv \{\sigma_k\}_{k=t}^{\infty}$ as the *continuation strategy profile* from period t on.

To define and characterize the equilibrium of the intra-personal game of an agent with potentially time-inconsistent preferences, we first introduce a useful concept: write $V_t\left(x_t, \varepsilon_t; \sigma_t^+\right)$ as the agent's period-t expected continuation utility when the state variable is x_t and the shock vector is ε_t under her *long-run* time preference for a given continuation strategy profile σ_t^+. We can think of $V_t\left(x_t, \varepsilon_t; \sigma_t^+\right)$ as representing (hypothetically) her intertemporal preferences from some prior perspective when her own present-bias is irrelevant. Specifically, $V_t\left(x_t, \varepsilon_t; \sigma_t^+\right)$ must satisfy:

$$V_t\left(x_t, \varepsilon_t; \sigma_t^+\right) = u^*_{\sigma_t(x_t,\varepsilon_t)}\left(x_t, \varepsilon_{\sigma_t(x_t,\varepsilon_t)t}\right) + \delta \mathrm{E}\left[V_{t+1}\left(x_{t+1}, \varepsilon_{t+1}; \sigma_{t+1}^+\right)|x_t, \sigma_t\left(x_t, \varepsilon_t\right)\right], \tag{2.1}$$

where $\sigma_t\left(x_t, \varepsilon_t\right) \in \mathcal{I}$ is the choice specified by strategy σ_t, and the expectation is taken over both the future state x_{t+1} and ε_{t+1}.

We will define the equilibrium for a partially naive agent whose period-t self believes that, beginning next period, her future selves will behave optimally with a present-bias factor of $\tilde{\beta} \in [\beta, 1]$. Following O'Donoghue and Rabin (1999b, 2001), we first define the concept of an agent's *perceived continuation strategy profile* by her future selves.

Definition 2. *The perceived continuation strategy profile for a partially naive agent is a strategy profile* $\tilde{\sigma} \equiv \{\tilde{\sigma}_t\}_{t=1}^{\infty}$ *such that for all* $t = 1, 2, ...,$ *all* $x_t \in \mathcal{X}$, *and all* $\varepsilon_t \in \Im^I$,

$$\tilde{\sigma}_t\left(x_t, \varepsilon_t\right) = \arg\max_{i \in \mathcal{I}} \left\{u_i^*\left(x_t, \epsilon_{it}\right) + \tilde{\beta}\delta \mathrm{E}\left[V_{t+1}\left(x_{t+1}, \varepsilon_{t+1}; \tilde{\sigma}_{t+1}^+\right)|x_t, i\right]\right\}.$$

That is, if an agent is partially naive with perceived present-bias by future selves at $\tilde{\beta}$, then her period-t self will anticipate that her future selves will follow strategies $\tilde{\sigma}_{t+1}^+ \equiv \{\tilde{\sigma}_k\}_{k=t+1}^{\infty}$. Given this perception, the period-t self's best response is called *perception-perfect strategy profile.*

Note, importantly, what the strategy profile $\tilde{\sigma} \equiv \{\tilde{\sigma}_t\}_{t=1}^{\infty}$ describes is the *perception of the partially naive agent regarding what her future selves will play.* It is *not* what will generate the *actual play* that we observe in the data. What we actually observe is generated from the perception-perfect strategy profile that we define below.

Definition 3. *A perception-perfect strategy profile for a partially naive agent is a strategy profile* $\sigma^* \equiv \{\sigma_t^*\}_{t=1}^{\infty}$ *such that, for all* $t = 1, 2, ...,$ *all* $x_t \in \mathcal{X}$, *and all* $\varepsilon_t \in \Im^I$,

$$\sigma_t^*\left(x_t, \varepsilon_t\right) = \arg\max_{i \in \mathcal{I}} \left\{u_i^*\left(x_t, \varepsilon_{it}\right) + \beta\delta \mathrm{E}\left[V_{t+1}\left(x_{t+1}, \varepsilon_{t+1}; \tilde{\sigma}_{t+1}^{+}\right) | x_t, i\right]\right\}.$$

It is key to note the *difference and connection* between $\tilde{\sigma}$ and σ^*. $\tilde{\sigma}$ is the unobserved *perception* of the partially naive agent regarding what her future selves will do, under the partially naive assumption that her future selves does not suffer from the present bias as described by the parameter β, but instead is governed by present bias parameter $\tilde{\beta}$ that may differ from β. σ^* is what the self in each period will optimally choose to do and that is what will be observed in the data. Note also that when β and $\tilde{\beta}$ coincide, that is, when the agent is sophisticated, then $\sigma^* = \tilde{\sigma}$.

Assumption 2. ***(Stationarity)*** *We assume that the observed choices are generated under the stationary perception-perfect strategy profile of the infinite horizon dynamic game played among different selves of the decision makers.*

2.2.2 Decision Process

Now we describe the decision process of the decision maker. First, define the current choice-specific value function (deterministic component), $W_i\left(x\right)$, as follows:

$$W_i(x) = u_i(x) + \beta\delta \int V(x')\pi(x'|x, i)dx', \tag{2.2}$$

where $\pi(x'|x, i)$ denotes the transition probabilities for state variables x when action i is taken; and $V\left(\cdot\right)$ is the *perceived long-run* value function defined according to (2.1) under perception perfect strategy profile for a partially naive agent $\tilde{\sigma}$ as defined in Definition 3).

We can also use $V\left(\cdot\right)$ as defined in (2.1) to define the *choice-specific value function of the next-period self as perceived by the current self*, $Z_i\left(x\right)$, as follows:

$$Z_i\left(x\right) = u_i\left(x\right) + \tilde{\beta}\delta \int V(x')\pi(x'|x, i)dx'. \tag{2.3}$$

Note that there are two key difference between $W_i(x)$ and $Z_i(x)$. The first difference is in how they discount the future streams of payoffs: in $W_i(x)$ the payoff t periods from the current period is discounted by $\beta\delta^t$, while in $Z_i(x)$ the payoff t periods from now is discounted by $\tilde{\beta}\delta^t$. The second difference is interpretational: $W_i(x)$ represents how the current-period self evaluates the deterministic component of the payoff from choosing alternative i, while $Z_i(x)$ is how the current-period self perceives how her future self would evaluate the deterministic component of the payoff from choosing alternative i. it is obvious but important to note that $W_i(x)$ will regulate the current self's optimal choice, but $Z_i(x)$ will regulate the perception of the current self regarding the choices of her future selves.

Given $Z_i(x)$, we know that the current self's perception of her future self's choice, i.e., $\tilde{\sigma}$ as defined in Definition 2 is simply

$$\begin{aligned}\tilde{\sigma}(x,\varepsilon) &= \max_{i\in\mathcal{I}}\left[u_i(x)+\varepsilon_i+\tilde{\beta}\delta\int V(x')\pi(x'|x,i)dx'\right] \\ &= \max_{i\in\mathcal{I}}\left[Z_i(x)+\varepsilon_i\right]. \qquad (2.4)\end{aligned}$$

Let us define the probability of choosing alternative j by the the next period self as perceived by the current period self, $\tilde{P}_j(x)$ when next period state is x :

$$\begin{aligned}\tilde{P}_j(x) &= \Pr\left[\tilde{\sigma}(x,\varepsilon)=j\right] \qquad (2.5) \\ &= \Pr\left[Z_j(x)+\varepsilon_j \geq Z_{j'}+\varepsilon_{j'} \text{ for all } j'\neq j\right].\end{aligned}$$

With the characterization of $\tilde{\sigma}(x,\varepsilon)$, now we can provide a characterization of $V(\cdot)$. For this purpose, further denote

$$V_i(x) = u_i(x)+\delta\int V(x')\pi(x'|x,i)dx'. \qquad (2.6)$$

According to the definition of $V(\cdot)$ as given by (2.1), $V(x)$ is simply the expected value of $[V_i(x)+\varepsilon_i]$ where i is the chosen alternative according to $\tilde{\sigma}(x,\varepsilon)$. Thus it must satisfy the following relationship:

$$V(x) = \mathrm{E}_{\varepsilon}\left[V_{\tilde{\sigma}(x,\varepsilon)}(x) + \varepsilon_{\tilde{\sigma}(x,\varepsilon)}\right]. \tag{2.7}$$

Now note from (2.3) and (2.6), we have

$$V_i(x) = Z_i(x) + \left(1-\tilde{\beta}\right)\delta\int V(x')\pi(x'|x,i)dx'. \tag{2.8}$$

The relationship (2.8) is crucial as it allows us to rewrite (2.7) as:

$$\begin{aligned}
V(x) &= \mathrm{E}_{\varepsilon}\left[V_{\tilde{\sigma}(x,\varepsilon)}(x) + \varepsilon_{\tilde{\sigma}(x,\varepsilon)}\right] \\
&= \mathrm{E}_{\varepsilon}\left[Z_{\tilde{\sigma}(x,\varepsilon)}(x) + \varepsilon_{\tilde{\sigma}(x,\varepsilon)} + \left(1-\tilde{\beta}\right)\delta\int V(x')\pi(x'|x,\tilde{\sigma}(x,\varepsilon))dx'\right] \\
&= \mathrm{E}_{\varepsilon}\max_{i\in\mathcal{I}}\left[Z_i(x)+\varepsilon_i\right] + \left(1-\tilde{\beta}\right)\delta\mathrm{E}_{\varepsilon}\int V(x')\pi(x'|x,\tilde{\sigma}(x,\varepsilon))dx' \\
&= \mathrm{E}_{\varepsilon}\max_{i\in\mathcal{I}}\left[Z_i(x)+\varepsilon_i\right] + \left(1-\tilde{\beta}\right)\delta\sum_{j\in\mathcal{I}}\tilde{P}_j(x)\int V(x')\pi(x'|x,j)dx' \quad (2.9)
\end{aligned}$$

where the second equality just follows from (2.8); and the third equality follows from (2.4) and thus

$$\mathrm{E}_{\varepsilon}\left[Z_{\tilde{\sigma}(x,\varepsilon)}(x) + \varepsilon_{\tilde{\sigma}(x,\varepsilon)}\right] = \mathrm{E}_{\varepsilon}\max_{i\in\mathcal{I}}\left[Z_i(x)+\varepsilon_i\right];$$

and the fourth equality follows from the fact that

$$\begin{aligned}
\mathrm{E}_{\varepsilon}\int V(x')\pi(x'|x,\tilde{\sigma}(x,\varepsilon))dx' &= \sum_{j\in\mathcal{I}}\Pr\left(\tilde{\sigma}(x,\varepsilon)=j\right)\int V(x')\pi(x'|x,j)dx' \\
&= \sum_{j\in\mathcal{I}}\tilde{p}_j(x)\int V(x')\pi(x'|x,j)dx'.
\end{aligned}$$

Now we make two additional assumptions about the transition of the state variables and the distribution of the shocks.

Assumption 3. ***(Conditional Independence):***

$$\pi(x_{t+1},\varepsilon_{t+1}|x_t,\varepsilon_t,d_t) = q(\varepsilon_{t+1}|x_{t+1})\pi(x_{t+1}|x_t,d_t)$$
$$q(\varepsilon_{t+1}|x_{t+1}) = q(\varepsilon).$$

Assumption 4. ***(Extreme Value Distribution):*** *ε_t is i.i.d extreme value distributed.*

Remark 1. *It is well-known that the distribution of the choice-specific shocks to payoffs in discrete choice models are not non-parametrically identified (see Magnac and Thesmar 2002, for example). Thus one has to make an assumption about the distribution of ε. We make the extreme value distribution assumption for simplicity, but it could be replaced by any other distribution G .*

With the above preliminary notations, now we can describe how the agent will make the choices when the state variables are given by x. From Definition 3 for perception perfect strategy profile and Equation (2.2), we know that the current period decision maker will choose i if and only if

$$i \in \arg\max_{i\in\mathcal{I}} \{W_i(x) + \varepsilon_i\}.$$

That is, the perception-perfect strategy profile $\sigma^*(x,\varepsilon)$ is:

$$\sigma^*(x,\varepsilon) = \arg\max_{i\in\mathcal{I}} \{W_i(x) + \varepsilon_i\}.$$

Under Assumption 3, the probability of observing action i being chosen at a given state variable x is:

$$P_i(x) = \Pr\left[W_i(x) + \varepsilon_i > \max_{j\in\mathcal{I}\setminus\{i\}} \{W_j(x) + \varepsilon_j\}\right] = \frac{\exp[W_i(x)]}{\sum_{j=0}^{I}\exp[W_j(x)]}. \tag{2.10}$$

$P_i(x)$ is the current-period self's equilibrium choice probabilities and will be observed in the data.

Now we derive some important relationships that will be used in our identification exercise below. First, note that by combining (2.2) and (2.3), we have that:

$$Z_i(x) - u_i(x) = \frac{\tilde{\beta}}{\beta}\left[W_i(x) - u_i(x)\right]. \tag{2.11}$$

Since both $Z_i(\cdot)$ and $W_i(\cdot)$ depends on $V(\cdot)$, we would like to use (2.9) to derive a characterization of $V(\cdot)$. Note that under Assumptions 3, we have:

$$\mathrm{E}_{\varepsilon}\max_{i\in\mathcal{I}}\{Z_i(x)+\varepsilon_i\} = \ln\left\{\sum_{i\in\mathcal{I}}\exp\left[Z_i(x)\right]\right\}. \tag{2.12}$$

Moreover, from (2.5), we have that

$$\tilde{P}_j(x) = \frac{\exp\left[Z_j(x)\right]}{\sum_{i=0}^{I}\exp\left[Z_i(x)\right]}. \tag{2.13}$$

Using (2.12) and (2.13), we can rewrite (2.9) as

$$V(x) = \ln\left\{\sum_{i\in\mathcal{I}}\exp\left[Z_i(x)\right]\right\} + \left(1-\tilde{\beta}\right)\delta\sum_{j\in\mathcal{I}}\frac{\exp\left[Z_j(x)\right]}{\sum_{i=0}^{I}\exp\left[Z_i(x)\right]}\int V(x')\pi(x'|x,j)dx'. \tag{2.14}$$

The three set of equations (2.3), (2.11) and (2.14) will form the basis of our identification argument below. Let us first make a few useful remarks.

Remark 2. *We have three value functions $\{W_i(x), Z_i(x), V_i(x) : x \in \mathcal{X}\}$ as defined respectively in (2.2), (2.3) and (2.6). Both $W_i(\cdot)$ and $Z_i(\cdot)$ are related to V_i. It is worth emphasizing that $W_i(x)$ will regulate the current self's choice behavior as demonstrated by (2.10); and $Z_i(x)$ will regulate the current self's perception of future selves choices as demonstrated by (2.13). $V_i(x)$ is an auxiliary value function that simply uses the long-run discount factor δ to evaluate the payoffs from the choices that the current self perceives that will be made by her future selves.*

Remark 3. *If $\tilde{\beta} = 1$, i.e., if the decision maker is completely naive, we can see from (2.8) that $V_i(x) = Z_i(x)$ for all x. This makes sense because when $\tilde{\beta} = 1$, the current self perceives her future selves to be time consistent. Thus the current self is already perceiving her future selves to be behaving according to the long run discount factor δ only.*

Remark 4. *If $\tilde{\beta} = \beta$, i.e., when an agent is sophisticated, we have that $W_i(x) = Z_i(x)$. if the decision maker is sophisticated, then the current self's own choice rule will be identical to what she perceives to be her future self's choice rule.*

Remark 5. *When the decision maker is partially naive, there are two distinct value functions $W_i(x)$ and $Z_i(x)$ that separately regulate the choice of the current self and the perceived choice of her future selves. Equation (2.11) clarifies that it is the fact that we allow for potential* naivety *in the hyperbolic model that is creating the wedge between $W_i(x)$ and $Z_i(x)$: if $\tilde{\beta} = \beta$, i.e., if agents are sophisticated (even when they suffer from present bias), it would be true that $V_i(x) = W_i(x)$. This is an important point because, as we see below in (2.10), the observed choice probabilities (our data) would provide direct information about $W_i(x)$, without needing any information about the discount factors. When $\tilde{\beta} = \beta$, it also provides direct information about $Z_i(x)$; but when $\tilde{\beta}$ and β are potentially not equal, we can no longer learn about $Z_i(x)$ directly from the observed choice probabilities.*

Relationship with the Dynamic Games Literature. We analyze the observed outcome of the dynamic discrete choice problem of a hyperbolic discounting decision process as the equilibrium outcome of an *intra-personal game* with the players being *the selves at different periods.* Thus our paper represents an interesting intermediate case between the literature on classical estimating single-agent dynamic discrete choice decision problems and the more recent literature on estimating dynamic games. It is worth pointing out that there are two crucial differences between the intra-personal games we analyze for agents with time-inconsistent time preferences and those in the existing dynamic games literature.

The first key difference is that in our case, we *do not observe* the actions of all the players. More specifically, the outcomes – choices and the evolutions of the state variables – we observe in the data are affected only by the current selves, even though the current selves' choices are impacted by their perception of future selves' actions. The current self's perception of how her future selves will play has to be inferred by the researcher using the equilibrium restriction imposed by the theory. As can be seen from the above discussion, $\tilde{P}_j(x)$ as defined in (2.13), captures the current self's perception of how her future selves will play, which is crucial for us to understand the current self's actual choices. However, as a researcher, we do not observe $\tilde{P}_j$, only observe P_j, the choice probabilities by the current self. In the standard dynamic games literature, it is always assumed that the action of all the players are observed.

Secondly, the dynamic games literature (e.g. Bajari, Benkard and Levin 2007) may allow for players to have different contemporaneous payoff functions, in our setting, however, the payoffs for the players – the current self and the future selves – differ only in their time preferences; moreover, under hyperbolic discounting, we are assuming a rather restricted form of difference in time preferences between the selves in different periods.

2.2.3 Data and Preliminaries

Before we describe our results on identification, let us assume that we the data provides us with the following information:

DATA:

- **(Conditional Choice Probabilities)** For all $x \in \mathcal{X}$, we observe the choice probabilities $P_i(x)$ for all $i \in \mathcal{I}$;
- **(Transitional Probabilities for Observable State Variables)** For all $(x, x') \in \mathcal{X}^2$, all $i \in \mathcal{I}$, we observe the transition probabilities $\pi(x'|x, i)$; we denote
$$\pi \equiv \left\{\pi(x'|x, i) : (x, x') \in \mathcal{X}^2, i \in \mathcal{I}\right\};$$

- **(Short Panels)** We have access to at least two periods of the above data, even though the data results from a stationary infinite horizon model.

Because we assume that our data is short panel as in Magnac and Thesmar (2002), we assume that the *structure* of the model, denoted by b, is defined by parameters

$$b = \left\{ \left(\beta, \tilde{\beta}, \delta\right), G, \langle \{u_i(x), Z_i(x'), V_i(x') : i \in \mathcal{I}, x \in \mathcal{X}, x' \in \mathcal{X}\} \rangle \right\}.$$

Note that the elements in b in our setting differs from those in Magnac and Thesmar (2002) in that we have two additional parameters β and $\tilde{\beta}$ that measure present bias and naivety; moreover, the interpretation of $V_i(x')$ in our paper differs from Magnac and Thesmar (2002). In their paper $V_i(x')$ directly informs about the *actual* choice probabilities of the decision maker in the second period, namely, $\Pr(i|x') = \Pr(V_i(x') + \varepsilon_i > V_j(x') + \varepsilon_j \text{ for all } j \neq i)$. In our paper, $Z_i(x')$ captures the current self's *perception* of the choice probability of the next period's self, which is *never actually* observed in the data; while $V_i(x')$ is just an auxiliary value function to account for the exponentially discounted payoff streams from the perceived choices made according to $\tilde{\sigma}(x, \varepsilon)$. Another difference is that in Magnac and Thesmar (2002), the vector $\{V_i(x') : x' \in \mathcal{X}\}$ are completely free parameters; in our setting, however, neither $Z_i(x')$ and $V_i(x')$ are completely free parameters as they are subject to the restriction that they have to satisfy (2.3), (2.11) and (2.14).

We denote by $\mathcal{B}$ the set of all permissible structures. The set $\mathcal{B}$ requires that the structure satisfies the assumptions we adopted in the model, as well as the restrictions (2.3),(2.11) and (2.14).

Given any structure $b \in \mathcal{B}$, the model predicts the probability that an agent will choose alternative $i \in \mathcal{I}$ in state $x \in \mathcal{X}$, which we denote by $\hat{P}_i(x; b)$ and is given by

$$\hat{P}_i(x; b) = \Pr$$
$$\left\{ u_i(x) + \varepsilon_i + \beta\delta \int V(x')\pi(x'|x, i)dx' = \max_{j \in \mathcal{I}} \left[u_j(x) + \varepsilon_j + \beta\delta \int V(x')\pi(x'|x, j)dx' \right] \middle| x, b \right\}.$$

As is standard in the identification literature, we call the predicted choice probabilities $\hat{P}_i(x;b)$ as the *reduced form* of structure $b \in \mathcal{B}$. We say that two structures $b, b' \in \mathcal{B}$ are *observationally equivalent* if

$$\hat{P}_i(x;b) = \hat{P}_i(x;b') \, \forall i \in \mathcal{I} \text{ and } x \in \mathcal{X}.$$

A model is said to be *identified* if and only if for any $b, b' \in \mathcal{B}$, $b = b'$ if they are observationally equivalent.

2.2.4 Identification Results

We first describe identification of $\langle \{u_i(x), Z_i(x'), V_i(x') : i \in \mathcal{I}, x \in \mathcal{X}, x' \in \mathcal{X}\} \rangle$ with for a given set of discount factors $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$. Then we provide conditions pertinent to the identification of $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$.

For any given joint distribution G of $\tilde{\varepsilon} \equiv (\varepsilon_1, ..., \varepsilon_I)$, the choice probability vector $\mathbf{P}(x) = (P_1(x), ..., P_I(x))$ is a mapping Q of $\mathbf{W}(x) = (W_0(x), W_1(x), ..., W_I(x))$. Hotz and Miller (1993) showed that the mapping Q can be inverted and one of the $W_i(x)$ has to be normalized. That is, one can find

$$D_i(x) \equiv W_i(x) - W_0(x) = Q_i(\mathbf{P}(x); G)$$

where Q_i is the i^{th} component of the inverse of Q. Under our Assumption 3 (that ε_i is iid extreme value distributed), the mapping Q_i is especially simple [following from (2.10)]:

$$D_i(x) = W_i(x) - W_0(x) = \ln \frac{P_i(x)}{P_0(x)}. \tag{2.15}$$

Since we observe $P_i(x)$ and $P_0(x)$ from the data, we immediately learn about $D_i(x)$. We thus proceed as if $D_i(x)$ is observable.

From (2.11), we have, for all $i \in \mathcal{I}$,

$$Z_i(x) = \frac{\tilde{\beta}}{\beta} W_i(x) + \left(1 - \frac{\tilde{\beta}}{\beta}\right) u_i(x) \tag{2.16}$$

Together with (2.15), we have, for all $i \in \mathcal{I} \backslash \{0\}$ and $x \in \mathcal{X}$,

$$Z_i(x) - Z_0(x) = \frac{\tilde{\beta}}{\beta} D_i(x) + \left(1 - \frac{\tilde{\beta}}{\beta}\right) [u_i(x) - u_0(x)]. \tag{2.17}$$

This allows us to rewrite $\tilde{P}_j(x)$ as follows:

$$\begin{aligned}
\tilde{P}_0(x) &= \frac{1}{1 + \sum_{i=1}^{I} \exp\left[\frac{\tilde{\beta}}{\beta} D_i(x) + \left(1 - \frac{\tilde{\beta}}{\beta}\right)[u_i(x) - u_0(x)]\right]} \\
\tilde{P}_j(x) &= \frac{\exp\left[\frac{\tilde{\beta}}{\beta} D_j(x) + \left(1 - \frac{\tilde{\beta}}{\beta}\right)[u_j(x) - u_0(x)]\right]}{1 + \sum_{i=1}^{I} \exp\left[\frac{\tilde{\beta}}{\beta} D_i(x) + \left(1 - \frac{\tilde{\beta}}{\beta}\right)[u_i(x) - u_0(x)]\right]} \text{ for } j \neq 0.
\end{aligned}$$

Assuming that the set of states $\mathcal{X}$ is finite and contains X elements, Equation (2.14) can be written as:

$$\begin{aligned}
V(x) &= Z_0(x) + \ln\left\{\sum_{i \in \mathcal{I}} \exp[Z_i(x) - Z_0(x)]\right\} \\
&+ \left(1 - \tilde{\beta}\right) \delta \sum_{j \in \mathcal{I}} \frac{\exp[Z_j(x) - Z_0(x)]}{\sum_{i=0}^{I} \exp[Z_i(x) - Z_0(x)]} \sum_{x' \in \mathcal{X}} V(x') \pi(x'|x, j), \qquad (2.18) \\
&\text{for all } x \in \mathcal{X}.
\end{aligned}$$

Note that equation (2.18) is simply a system of X linear equation in $\{V(x) : x \in \mathcal{X}\}$. In fact, if we write $\mathbf{V}$ as the $X \times 1$ column vector $[V(1), ..., V(X)]^T$, $\mathbf{A}$ as the $X \times 1$ column vector $\left[Z_0(x) + \ln\left\{\sum_{i \in \mathcal{I}} \exp[Z_i(x) - Z_0(x)]\right\}\right]_{x \in \mathcal{X}}$. If we further write

$$\tilde{\mathbf{P}} = \begin{bmatrix} \tilde{\mathbf{P}}_0 & \cdots & \tilde{\mathbf{P}}_I \end{bmatrix}_{X \times [(I+1)X]},$$

where

$$\tilde{\mathbf{P}}_j = \begin{bmatrix} \tilde{P}_j(0) & 0 & \cdots & 0 \\ 0 & \tilde{P}_j(1) & \cdots & 0 \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ 0 & 0 & \cdots & \tilde{P}_j(X) \end{bmatrix}_{X \times X}.$$

Also properly stack up the transition matrices $\pi(x'|x,j)$ into an $[(I+1)X] \times X$ matrix as follows:

$$\mathbf{\Pi} = \begin{bmatrix} \mathbf{\Pi}_0 \\ \vdots \\ \mathbf{\Pi}_I \end{bmatrix}_{[(I+1)X]\times X}$$

where

$$\mathbf{\Pi}_j = \begin{bmatrix} \Pi_j(1) \\ \Pi_j(2) \\ \vdots \\ \Pi_j(X) \end{bmatrix}_{X\times X}$$

where

$$\Pi_j(x) = \begin{bmatrix} \pi(1|x,j) & \ldots & \pi(X|x,j) \end{bmatrix}$$

is an $1 \times X$ row vector.

Thus we can write (2.18) as

$$\begin{aligned} \mathbf{V} &= \mathbf{A} + \left(1-\tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\mathbf{V} \\ \mathbf{V} &= \left[\mathbf{I} - \left(1-\tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1}\mathbf{A} \qquad (2.19) \end{aligned}$$

Thus for fixed values of $\left(\beta, \tilde{\beta}, \delta\right)$, we can plug (2.19) into (2.3) and obtain, for all $x \in \mathcal{X}$,

$$\begin{aligned} Z_0(x) &= u_0(x) + \tilde{\beta}\delta \sum_{x' \in \mathcal{X}} V(x')\,\pi(x'|x,0) \\ &= u_0(x) + \tilde{\beta}\delta\mathbf{\Pi}_0(x)\,\mathbf{V} \qquad (2.20) \\ &= u_0(x) + \tilde{\beta}\delta\mathbf{\Pi}_0(x)\left[\mathbf{I} - \left(1-\tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1}\mathbf{A} \end{aligned}$$

where $\mathbf{\Pi}_0(x) = [\pi(1|x,0), ..., \pi(X|x,0)]$ is an $X \times 1$ vector. Similarly, for any

$i \in \{1, ..., I\}$ and for all $x \in \mathcal{X}$, we can obtain the following analogously:

$$
\begin{aligned}
Z_i(x) &= u_i(x) + \tilde{\beta}\delta \sum_{x' \in \mathcal{X}} V(x')\,\pi(x'|x,i) \\
&= u_i(x) + \tilde{\beta}\delta \mathbf{\Pi}_i(x)\,\mathbf{V} \qquad (2.21) \\
&= u_i(x) + \tilde{\beta}\delta \mathbf{\Pi}_i(x)\left[\mathbf{I} - \left(1-\tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1}\mathbf{A}
\end{aligned}
$$

Now consider the system of equations given by (A.2), (2.20) and (2.21). We know that we have to normalize the utility for the reference alternative 0, without loss of generality set $u_0(x) = 0$ for all $x \in \mathcal{X}$. The unknowns contained in the equation system include $(I+1) \times X$ values for $\{Z_i(x) : i \in \mathcal{I}, x \in \mathcal{X}\}$ and $I \times X$ values for $\{u_i(x) : i \in \mathcal{I}/\{0\}, x \in \mathcal{X}\}$, thus the total unknowns is $(2I+1) \times X$. It is also easy to see that the total number of equations in the system is also equal to $(2I+1) \times X$.

Relationship to the Standard Exponential Discounting Special Case: $\beta = \tilde{\beta} = 1$. Before we discuss whether the non-linear equation system (A.2), (2.20) and (2.21) has a unique solution, let us show how it is related to the existing case in the literature for the identification of dynamic discrete choice models with exponential discounting (i.e., the case with $\tilde{\beta} = \beta = 1$). In that case, (A.2) is reduced to the well-known relationship

$$
Z_i(x) - Z_0(x) = \ln P_i(x) - \ln P_0(x); \qquad (2.22)
$$

that is, in standard models the difference in the choice probabilities for alternative i and the reference alternative 0 informs us about the difference in the value from choosing i relative to the value from choosing 0. This is of course also true when $\tilde{\beta} = \beta < 1$. The potential naivety we allow in our setup breaks this direct relationship between $P_i(x)/P_0(x)$ and $Z_i(x) - Z_0(x)$.

Moreover, when $\tilde{\beta} = \beta = 1$, (2.20) is reduced to (using the normalization that

$u_0(x) = 0)$:

$$Z_0(x) = \delta \sum_{x' \in \mathcal{X}} Z_0(x') \pi(x'|x,0) + \delta \sum_{x' \in \mathcal{X}} \ln \left[\sum_{i \in \mathcal{I}} \frac{P_i(x')}{P_0(x')} \right] \pi(x'|x,0).$$

For simplicity, denote the $X \times 1$ vector $\{Z_0(x)\}_{x \in \mathcal{X}}$ as $\mathbf{Z}_0$; write the $X \times X$ matrix $\pi(x'|x,0)$ as $\mathbf{\Pi}_0$, and write the $X \times 1$ vector $\left\{ \ln \left[\sum_{i \in \mathcal{I}} \frac{P_i(x')}{P_0(x')} \right] \right\}_{x \in \mathcal{X}}$ as $\mathbf{m}$. The above equation can be written as

$$\mathbf{Z}_0 = \delta \mathbf{\Pi}_0 (\mathbf{Z}_0 + \mathbf{m}).$$

Thus,

$$\mathbf{Z}_0 = (\mathbf{I} - \delta \mathbf{\Pi}_0)^{-1} \delta \mathbf{m}.$$

Given this unique solution of $\mathbf{Z}_0$, (2.22) immediately provides $Z_i(x)$ for all $i \in \mathcal{I} / \{0\}$ and all $x \in \mathcal{X}$. To obtain $u_i(x)$ for $i \in \mathcal{I} \backslash \{0\}$, note that (2.21) implies that

$$\mathbf{u}_i = \mathbf{Z}_i - \delta \mathbf{\Pi}_i \mathbf{Z}_0 - \delta \mathbf{\Pi}_i \mathbf{m}, \tag{2.23}$$

where $\mathbf{Z}_i$ and $\mathbf{u}_i$ are $X \times 1$ vectors of $\{Z_i(x)\}_{x \in \mathcal{X}}$ and $\{u_i(x)\}_{x \in \mathcal{X}}$ respectively, $\mathbf{\Pi}_i$ is the $X \times X$ matrix $\pi(x'|x,i)$. Recall that in the standard exponential discounting model we have $Z_i(x) = V_i(x)$, thus we can conclude that , $\{\mathbf{u}_i\}_{i \in \mathcal{I} \backslash \{0\}}$ and $\{\mathbf{V}_i\}_{i \in \mathcal{I}}$ are identified once δ, G and $\{u_0(x)\}_{x \in \mathcal{X}}$ are fixed. This replicates the proof of Proposition 2 in Magnac and Thesmar (2002).[7]

The system of equations is no longer linear in the more general case we consider there. But we can also reduce the system of equations (A.2), (2.20) and (2.21) into a single Denote by $\mathbf{c}$ the $X \times 1$ vector

$$\ln \left\{ \sum_{i \in \mathcal{I}} \exp \left[\frac{\tilde{\beta}}{\beta} \ln \frac{P_i(x')}{P_0(x')} + \left(1 - \frac{\tilde{\beta}}{\beta} \right) u_i(x') \right] \right\}_{x' \in \mathcal{X}}. \tag{2.24}$$

[7] The only difference is that our argument above indicates that $\mathbf{V}_0$ does not have to be fixed. It can be identified from the model.

Note that there is a crucial difference between the vector $\mathbf{m}$ we defined earlier for the case $\tilde{\beta} = \beta$ and $\mathbf{c}$: $\mathbf{m}$ contains only observables and thus can be treated as known; in contrast $\mathbf{c}$ here depends on unknown current payoffs $\{u_i(x) : i \in \mathcal{I} \setminus \{0\}\}$, thus cannot be considered as an observable. With this in mind, we can write (2.20) in matrix form as (since $\mathbf{A} = \mathbf{Z}_0 + \mathbf{c}$) :

$$\mathbf{Z}_0 = \tilde{\beta}\delta\mathbf{\Pi}_0 \left[\mathbf{I} - \left(1 - \tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1} (\mathbf{Z}_0 + \mathbf{c}) .$$

Thus,

$$\mathbf{Z}_0 = \tilde{\beta}\delta\mathbf{Q}_{\mathbf{A}}\mathbf{c}$$

where

$$\mathbf{Q}_{\mathbf{A}} = \left(\mathbf{I} - \tilde{\beta}\delta\mathbf{\Pi}_0 \left[\mathbf{I} - \left(1 - \tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1}\right)^{-1} \mathbf{\Pi}_0 \left[\mathbf{I} - \left(1 - \tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1} .$$

Now (2.21) can be written in matrix form as

$$\mathbf{Z}_i = \mathbf{u}_i + \tilde{\beta}\delta\mathbf{\Pi}_i \left[\mathbf{I} - \left(1 - \tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1} (\mathbf{Z}_0 + \mathbf{c}) \tag{2.25}$$

$$= \mathbf{u}_i + \tilde{\beta}\delta\mathbf{\Pi}_i\mathbf{Q}_{\mathbf{B}}\mathbf{c} \tag{2.26}$$

where

$$\mathbf{Q}_{\mathbf{B}} = \left[\mathbf{I} - \left(1 - \tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1}$$
$$\left(\mathbf{I} + \left(\mathbf{I} - \tilde{\beta}\delta\mathbf{\Pi}_0 \left[\mathbf{I} - \left(1 - \tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1}\right)^{-1} \tilde{\beta}\delta\mathbf{\Pi}_0 \left[\mathbf{I} - \left(1 - \tilde{\beta}\right)\delta\tilde{\mathbf{P}}\mathbf{\Pi}\right]^{-1}\right) .$$

Finally, note that (A.2) can be written in matrix form as

$$\mathbf{Z}_i - \mathbf{Z}_0 = \frac{\tilde{\beta}}{\beta}\mathbf{D}_i + \left(1 - \frac{\tilde{\beta}}{\beta}\right)\mathbf{u}_i. \tag{2.27}$$

where $\mathbf{D}_i$ denotes the $X \times 1$ vector $\{\ln P_i(x) - \ln P_0(x)\}_{x \in \mathcal{X}}$. Thus, we have

$$\begin{aligned} \mathbf{u}_i &= \mathbf{D}_i + \beta\delta \mathbf{Q}_{\mathbf{A}}\mathbf{c} - \beta\delta \mathbf{\Pi}_i \mathbf{Q}_{\mathbf{B}}\mathbf{c} \\ &= \mathbf{D}_i + \beta\delta \left(\mathbf{Q}_{\mathbf{A}} - \mathbf{\Pi}_i \mathbf{Q}_{\mathbf{B}}\right)\mathbf{c} \end{aligned}$$

Let

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_I \end{bmatrix}, \mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_I \end{bmatrix}$$

both be $(I \times X) \times 1$ vector with $\mathbf{u}_i$ and $\mathbf{D}_i$ stacked together; and let

$$\mathbf{E} = \begin{bmatrix} \mathbf{Q}_{\mathbf{A}} - \mathbf{\Pi}_1 \mathbf{Q}_{\mathbf{B}} \\ \vdots \\ \mathbf{Q}_{\mathbf{A}} - \mathbf{\Pi}_I \mathbf{Q}_{\mathbf{B}} \end{bmatrix}$$

be a $(I \times X) \times X$ matrix. Then we have

$$\mathbf{U} = \mathbf{D} + \beta\delta \mathbf{E}\mathbf{c}. \tag{2.28}$$

Note that in equation (2.28), the three parameters $\left(\beta, \tilde{\beta}, \delta\right)$ appears in three different combinations: $\beta/\tilde{\beta}$, $\left(1-\tilde{\beta}\right)\delta$ and $\tilde{\beta}\delta$ [Note that $\beta\delta = \tilde{\beta}\delta \times \left(\beta/\tilde{\beta}\right)$]. Despite its seeming simplicity, Equation (2.28) is a highly nonlinear system of equations in $\mathbf{U}$, which is an $(I \times X) \times 1$ column vector. In order to proceed with our identification arguments, we have to make the following additional assumption:additional assumptions:

Assumption 5. ***(Unique Solution)*** *For any values of* $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$ *and* $\{u_0(x) : x \in \mathcal{X}\}$, *(2.28) has a unique solution.*

Under the above assumption, we can uniquely solve $u_i(x)$ for all $i \in \mathcal{I} \backslash \{0\}$ and $x \in \mathcal{X}$ and $V_i(x)$ for all $i \in \mathcal{I}$ and $x \in \mathcal{X}$ for any given $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$.

Now we discuss conditions for identification related to $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$. This discussion is closely related to that in Magnac and Thesmar (2002). We impose the following *exclusion restriction assumption:*

Assumption 6. ***(Exclusive Restriction)*** *Suppose that there exists variables* $(x_1, x_2) \in \mathcal{X}^2$ *with* $x_1 \neq x_2$, *but:*

- *for* all $i \in \mathcal{I}$, $u_i(x_1) = u_i(x_2)$;
- *for* some $i \in \mathcal{I}$, $\pi(x'|x_1, i) \neq \pi(x'|x_2, i)$.

More specifically, to satisfy the exclusion restriction assumption, there must be a variable that does not directly affect the *contemporaneous utility function* u_i for all $i \in \mathcal{I}$ but the variable may matter for choices because it affects the transition of state variables. The extent to which individuals' choice probabilities differ at state x_1 and x_2 reveals information about the discount factors. This is the key intuition from Magnac and Thesmar's (2002) result where they are interested in identifying a single long-term discount factor δ. In their setting, if $\delta = 0$, i.e., if individuals are completely myopic, then the choice probabilities would have been the same under x_1 and x_2; to the extent that choice probabilities differ at x_1 and x_2, it reveals information about the degree of time discounting. Their intuition, however, can be easily extended to the hyperbolic discounting case, as we will exploit in the proposed estimation strategy below. For notational simplicity, we will divide the state variables into two groups (x_r, x_e) where x_r refers to the state variables that directly enter the contemporaneous payoff function $u_i(x_r)$ and x_e refers to the state variables that satisfy the exclusive restriction assumption (i.e., they do not enter the contemporaneous payoff function but affects the state transition probabilities).

Before we show how we can use the exclusion restriction to construct estimators for the three discount factors $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$, it is useful to provide some intuition as to how $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$ come to affect the observed choice behavior by the current self differently.

β vs. δ. It may seem counterintuitive that β and δ could be separately identified in a short two-period panel data set. To provide some intuition, let us consider the case that $\tilde{\beta} = \beta$. The question is: "Can we distinguish the behavior of an agent with exponential discounting rate $\hat{\delta} = \beta\delta$ from the behavior of a sophisticated time-inconsistent agent with preference (β, δ)?" Under stationarity assumption, if an agent has time consistent exponential discounting rate $\hat{\delta} = \beta\delta$, her expected continuation utilities is completely determined by the observed choice probabilities. To see this, observe that in equation (2.18), if one replaces $\tilde{\beta}$ by 1 and δ by $\hat{\delta}$, we will have

$$V(x) = Z_0(x) + \ln\left\{\sum_{i\in\mathcal{I}} \exp\left[Z_i(x) - Z_0(x)\right]\right\},$$

which only depends on $D_i(x)$ when $\beta = \tilde{\beta}$.

However, for a sophisticated time-inconsistent agent with preference (β, δ), there is an incongruence between current self and her perceived future self regarding how they evaluate the future stream of payoffs. Though the current self has to defer to her next-period self in terms of the actual next-period choice that will be chosen, they disagree on how much weight to put on payoffs two-periods from now. It is this incongruence that leads to the last term in Equation (2.18), which in turn breaks the tight link between observed choice probabilities and the continuation utilities.

β vs. $\tilde{\beta}$. To help provide intuition for why β could be distinguished from $\tilde{\beta}$, let us suppose that $\delta = 1$. First note that the ratio $\tilde{\beta}/\beta$ appears in term $Z_i(x)$ [see Eq. (2.16)]. This ratio regulates the incongruence between the current self's own behavior and her perception of the behavior of her future selves. Eq. (A.2) shows that if $\tilde{\beta}/\beta = 1$, then $Z_i(x) - Z_0(x)$ is uniquely determined by the observed $D_i(x)$; thus the current self's perception about her future self's action is identical to her own action. Since we do not directly observe the current self's perceptions, this insight could not directly be used to test whether $\tilde{\beta}/\beta$ is equal to 1. Only with the exclusion restriction can we use the above intuition to help identify $\tilde{\beta}/\beta$.

For notational simplicity, we will divide the state variables into two groups (x_r, x_e) where x_r refers to the state variables that directly enter the contemporaneous payoff function $u_i(x_r)$ and x_e refers to the state variables that satisfy the exclusive restriction assumption (i.e., they do not enter the contemporaneous payoff function but affects the state transition probabilities).

If is also useful to note that so far, our discussion has focused on short-panel (two period) data sets under stationarity assumption. Having two-period data allows one to non-parametrically estimate the transition probabilities $\pi(x'|x, i)$; stationarity ensures that looking at a two-period slice of a potentially long panel is sufficient.[8]

2.2.5 Estimation Strategies

We propose two related two-step estimation strategies. The first step for the two estimation strategies are the same: estimate from the data the choice probabilities $P_i(x)$ for all $i \in \mathcal{I}$ and all $x \in \mathcal{X}$, as well as the state transition probabilities $\pi(x'|x, i)$ for all $i \in \mathcal{I}$ and all $(x', x) \in \mathcal{X}^2$.

The second step of both estimation strategies involves two loops. (1) In the inner loop, we solve equation (2.28) for $u_i(x)$ for a given triple of values for $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$ for all $i \in \mathcal{I} \backslash \{0\}$ and all $x \in \mathcal{X}$ where we normalize $u_0(x) = 0$ for all $x \in \mathcal{X}$. Note the role played by Assumption 5. (2). In the outer loop, we maximize an objective function (to be stated below) over values of $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$. The two estimation strategies differ in the objective function used in the outer loop.

In the first estimation strategy, we take the $\hat{u}_i(x)$ solved from the inner loop and impose (by Assumption 6) the restriction that

$$\hat{u}_i(x_r) = \sum_{\{(x_r, \tilde{x}_e): \tilde{x}_e \in \mathcal{X}_e\}} \hat{u}_i(x_r, \tilde{x}_e) \tag{2.29}$$

[8] Fang and Silverman (2006) considered a case without stationarity (specifically a finite horizon model) and showed that β and δ could be potentially identified without exclusion restriction if the researcher has access to at least three-period panel data.

where $\mathcal{X}_e$ is the set of possible values for the payoff irrelevant state variables x_e we discussed in Assumption 6. We then use the above utility function $\hat{u}_i(x_r)$ to predict the pseudo-choice probabilities for the individuals in a given state $x \in \mathcal{X}$. Denote the pseudo-choice probabilities as $\hat{P}_i(x)$. The first estimation strategy will maximize the pseudo-likelihood by iterating over combinations of $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$. Under the maintained model, at the true values of $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$, $\hat{u}_i(x_r, \tilde{x}_e)$ as solved from the inner loop should be independent of $\tilde{x}_e$, the payoff irrelevant state variables, and thus $\hat{u}_i(x_r)$ as calculated in (2.29) would be exactly equal to $\hat{u}_i(x_r, \tilde{x}_e)$ for all $\tilde{x}_e \in \mathcal{X}_e$. Thus the pair of $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$ that maximizes the pseudo-likelihood of the observed choices is a consistent estimator.

In the second estimation strategy, we exploit the identifying assumption 6 in a somewhat different manner. We know that at the true values of $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$, for all $\tilde{x}_e \in \mathcal{X}_e$,the standard deviation of $\hat{u}_i(x_r, \tilde{x}_e)$ with respect to $\tilde{x}_e$ should be 0, because Assumption 6 requires that $u_i(x_r, \tilde{x}_e)$ does not depend on $\tilde{x}_e$. Therefore the estimator for $\left\langle \beta, \tilde{\beta}, \delta \right\rangle$ in our second estimation strategy is

$$\begin{aligned} \left\langle \beta, \tilde{\beta}, \delta \right\rangle &= \arg \min_{\left\{\beta, \tilde{\beta}, \delta\delta\right\}} f(\beta, \tilde{\beta}, \delta) \\ &= \arg \min_{\left\{\beta, \tilde{\beta}, \delta\right\}} \sum_{x_r \in \mathcal{X}_r} \left| \text{std}\left[\hat{u}_i\left(x_r, \tilde{x}_e; \left\langle \beta, \tilde{\beta}, \delta \right\rangle \right) \right] \right|. \end{aligned}$$

2.2.6 Monte Carlo Experiments

In this section we provide Monte Carlo evidence for the identification of discount factors in a dynamic discrete choice model using the two estimation methods described above. In this simple Monte Carlo exercise, we consider a 2-choice decision problem facing an agent with infinite horizon and stationary state transition. There are two state variables x_1 and x_2. The state variable $x_1 \in \{0, 1, 2, 3, 4, 5\}$, affects both instantaneous utility and state transition; while state variable $x_2 \in \{0, 1, 2, 3\}$, affects

Table 2.1: Monte Carlo Results Under the Two Proposed Estimation Methods.

Estimation Method	Maximizing Pseudo-Loglikelihood				Minimizing Variance of Utilities			
Parameters	$\tilde{\beta}\delta$	$\tilde{\beta}/\beta$	α_0	α_1	$\tilde{\beta}\delta$	$\tilde{\beta}/\beta$	α_0	α_1
True Values	0.8	1.2	-0.1	0.5	0.8	1.2	-0.1	0.5
Sample Size: $240,000$								
Mean	0.828	1.246	-0.100	0.500	0.824	1.240	-0.100	0.500
Std. Dev.	0.076	0.124	0.008	0.003	0.084	0.121	0.009	0.003
Sample Size: $120,000$								
Mean	0.823	1.244	-0.100	0.500	0.817	1.235	-0.100	0.500
Std. Dev.	0.086	0.146	0.012	0.005	0.096	0.136	0.012	0.005
Sample Size: $12,500$								
Mean	0.766	1.243	-0.104	0.503	0.765	1.253	-0.105	0.503
Std. Dev.	0.172	0.234	0.037	0.016	0.186	0.234	0.038	0.016

For each sample size, we generate 5000 random simulation samples. The Mean and Standard Deviations of the estimated parameters are with respect to the 5000 samples.

only the state transition.[9] When $a = 1$, instantaneous utility is $u_1(x_1) = \alpha_0 + \alpha_1 x_1$; when $a = 0$, instantaneous utility is simply $u_0(x_1) = 0$ for all x_1. The true parameters are $\alpha_0 = -0.1$, $\alpha_1 = 0.5$, $\tilde{\beta}/\beta = 1.2$, and $\tilde{\beta}\delta = 0.8$. We estimate the discount factors and corresponding utility parameters $5,000$ times for various sample sizes, to show the differences in performance of these two estimation methods. Results are shown in Table 2.1. Table 2.1 shows that both estimation methods do an excellent job in recovering the true parameter values of in large samples. For the Monte Carlo sample size analogous to our actual estimation sample (12,500), the maximum pseudo-loglikelihood method seems to perform better. As a result, we will use the maximum pseudo-loglikelihood method in our empirical exercise.

[9] The state transition matrices are generated as follows. We first generate two random matrices M_0 and M_1 each with dimension $\#X_1 \times \#X_2 = 6 \times 4 = 24$, with each entry a random number generated from a uniform $[0, 1]$ distribution. We then normalize the entry in each row by its row sum to ensure a proper probability matrix. The resulting matrices are denoted Π_0 and Π_1. The (j, k)-th entry of matrix Π_i where $(j, k) \in \{1, ..., 24\}^2$ is the probability that (x_1', x_2') takes on k-th combination condition on (x_1, x_2) taking on j-th combination in this period and action chosen is $i \in \{0, 1\}$. The matrices Π_1 and Π_0 are assumed to be known by the decision maker; and are directly taken to be the state transition probabilities in our Monte Carlo exercise reported in Table 2.1.

2.3 Application: Empirical Application: Mammography Decisions

2.3.1 Background on Mammography

Breast cancer is the third most common cause of death, and the second leading cause of cancer death, among American women. From birth to age 39, one woman in 231 will get breast cancer ($< 0.5\%$ risk); from age 40-59, the chance is 1 in 25 (4% risk); from age 60-79, the chance is 1 in 15 (nearly 7%). Assuming that a woman lives to age 90, the chance of getting breast cancer over the course of an entire lifetime is 1 in 7, with an overall lifetime risk of 14.3%.[10]

Breast cancer takes years to develop. Early in the disease, most breast cancers cause no symptoms. When breast cancer is detected at a localized stage (it hasn't spread to the lymph nodes), the 5-year survival rate is 98%. If the cancer has spread to nearby lymph nodes (regional disease), the rate drops to 81%. If the cancer has spread (metastasized) to distant organs such as the lungs, bone marrow, liver, or brain, the 5-year survival rate is 26%.

A screening – mammography – is the best tool available to find breast cancer before symptoms appear. Mammography can often detect a breast lump before it can be felt and therefore save lives by finding breast cancer as early as possible. For women over the age of 50, mammography have been shown to lower the chance of dying from breast cancer by 35%.[11] Leading experts, the National Cancer Institute, the American Cancer Society, and the American College of Radiology recommend annual mammography for women over 40. The U.S. Preventive Services Task Force recommends mammography screening for women beginning at age 50 every 12-24 months in order to reduce the risk of death from breast cancer (DHHS 2002).

---

[10] As a useful comparison, breast cancer has a higher incidence rate than lung cancer in the US. In 2004, there were 217,440 new cases for breast cancer in US (American Cancer Society).

[11] Source: American Cancer Society. Also note that finding breast cancers early with mammography has also meant that many more women being treated for breast cancer are able to keep their breasts. When caught early, localized cancers can be removed without resorting to breast removal.

2.3.2 Data

The data used in this analysis are from the Health and Retirement Study (HRS). The HRS is a nationally representative biennial panel study of birth cohorts 1931 through 1941 and their spouses as of 1992. The initial sample includes 12,652 persons in about 7,600 households who have been interviewed every two years since 1992. The most recent available data are for year 2006 (wave 8). The survey history and design are described in more details in Juster and Suzman (1995). Since the HRS started asking women questions about their usage of mammography in 1996, our sample is limited to women interviewed in the HRS from 1996 to 2006. We focus on the age group 51 to 64, and exclude those observations with missing values for any of the critical variables.[12] We also exclude those who have ever been diagnosed of (breast) cancer,[13] since those who are diagnosed of cancer might be of a different group who do not make decisions on mammography or any other preventive health care the same way as do others. Our final sample consists of 12,506 observations (each observation is a two-period short panel pooled from two consecutive waves) for 7,067 individuals.

2.3.3 Descriptive Statistics

Table 2.2 provides summary statistics of the key variables for the sample we use in our empirical analysis. The sample of women we select are aged from 51 to 64 (note that we are combining two-period panels using individuals that appear in two consecutive samples), with an average age of 57.8. A large majority of the sample are non-Hispanic white (78%) and married (70.5%). The average household income is about 50K dollars. 23% of the sample has a self-reported bad health and about 76% of the women undertook mammogram in the survey year. About 1.5% of the women who were surveyed in a wave died within two years. Finally, about 75% of

[12] The key variables are marital status, non-Hispanic white, self-rated health, household income, and mammography usage (see the summary statistics table).

[13] For those who entered in 1992 (i.e. the first wave), we know whether they have been diagnosed of breast cancer as of 1992. But for those who entered HRS in later waves, we only know whether cancer has been diagnosed of since the survey from 1994 did not ask about breast cancer specifically.

Table 2.2: Summary Statistics of Key Variables in the Estimation Sample.

Variable	Mean	Std. Dev.	Min	Max	Obs.
Age	57.81	3.95	51	64	12506
White (Non-Hispanic)	.781	.413	0	1	12506
Married	.705	.456	0	1	12506
Household Income ($1000)	50.04	69.12	0.101	2,140	12506
Log of Household Income	10.32	1.07	4.61	14.57	12506
Bad Health	.230	.421	0	1	12506
Mammogram	.758	.428	0	1	12506
Death	.015	.121	0	1	12506
Mother Still Alive or Died After Age 70	.749	.433	0	1	12506
Bad Health $(t+1)$	.249	.432	0	1	12319
Household Income $(t+1)$ ($1000)	48.98	177.52	0.103	17,600	12319

The last two variables in the table are observed only for those who survive to the second period.

the mothers of the women in our sample are either still alive or died at age greater than 70 at the time of the interview.

2.3.4 Decision Time Line

Figure 2.1 depicts the time line for mammography decisions for women in our sample. As we mentioned earlier, we only consider women who are alive and have not yet been diagnosed with any cancer (thus not breast cancer) in the first period. Given her period-1 state variables, she makes the decision of whether to undertake mammography. Mammography detects breast cancer with very high probability, though not for certain, if the woman has breast cancer. In the event that the woman has breast cancer, early detection of breast cancer will lead to higher survival probability.

To fully capture the diagnostic nature of mammography, we would need to have information about whether the woman has breast cancer at any period, and estimate the probability of detecting breast cancer with (P1 in Figure 2.1) and without (P2) mammogram. However, we do not have access to such data. In HRS, even though we have information on women's mammography choices from 1996 on and we know whether their doctors have told them that they have *any* cancer, we do not have

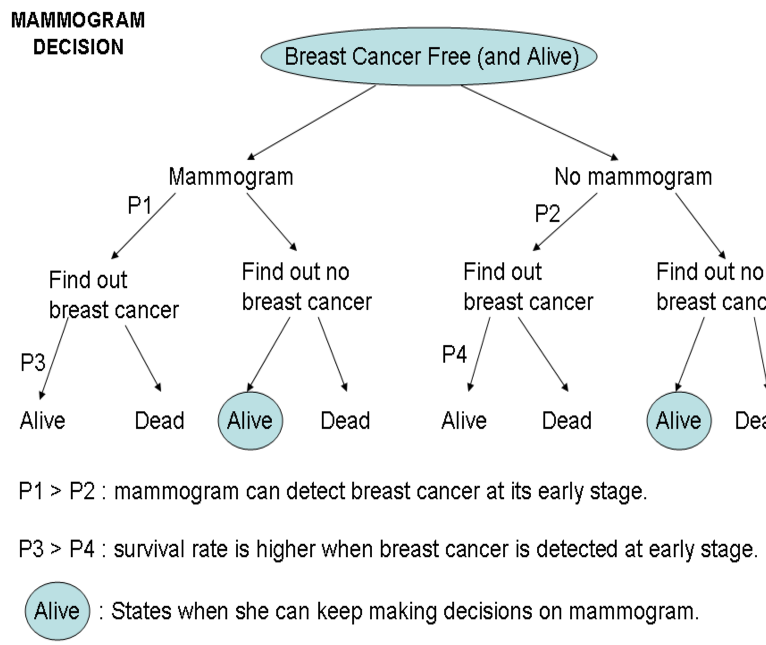


FIGURE 2.1: The Timeline for Mammography Decisions

information on which kind of cancer they have been diagnosed of.[14]

Due to these data issues, we decide to go directly from the mammography decision to the live/death outcome and health status if alive (see our empirical specification below), without going through the intermediate step (having breast cancer or not). That is, we simply capture the ultimate effect of undertaking mammography as to lower the probability of dying, and to lower the probability of being in bad health status if alive.

2.3.5 Empirical Specification

In this dynamic discrete choice model, each agent decides whether to get mammography or not. It is clear from the previous section on identification, we need to impose some normalization of the contemporaneous utilities.

[14] Even for those over age 65 with matching Medicare claim data, the number of observations is not big enough for us to get the needed probabilities of being diagnosed with breast cancer with and without mammography stratified by all the state variables we want to control in our model such as age, race, health status, income, and marital status.

Let us first be specific about the state variables we use in our empirical specification. They include: Age (Age); Marital Status (Married); Bad Health (BadHealth) which indicates whether the individual self reports bad health; Log of Household Income (LogIncome), Death (Death) and whether her mother is still alive or died at age greater than 70 (Mother70)

Each woman decides whether to undertake mammography ($i = 1$ if she does and $i = 0$ if she does not). We normalize the individual's instantaneous utility at the death state to be zero. Note that no decision is necessary if one reaches the death state.

If an individual stays alive, then we specify her utility from taking mammogram relative to not taking mammogram as:

$$u_1(x) - u_0(x) = \alpha_0 + \alpha_1 \text{BadHealth} + \alpha_2 \text{LogIncome} + \varepsilon_t, \qquad (2.30)$$

where "BadHealth" is a binary variable indicating whether the agent is in bad health or not at time t; "LogIncome" denotes the logarithm of household income of the agent at time t, α_0 and α_1 are the utility parameters, and ε_t denotes difference in the choice-specific utility shocks at time t. It is important to remark that even though the other state variables Age, White and Married do not show up in the above specification, it does not mean that these variables do not affect the instantaneous utility of the individual; what it means is that these variables affect the instantaneous utility under action 1 (mammogram) and action 0 (no mammogram) in exactly the same way.

The agents make their decisions about whether to get mammography by comparing the expected summations of current and discounted future utilities from each choice. Individuals are uncertain about their future survival probabilities, and if alive, the transition probabilities of future health and household income. These probabilities depend on their choices about whether or not to get mammography; time-variant state variables including their lagged health status, their lagged household income, and their age, denoted as Age; and time-invariant state variables including their race, denoted by a binary variable White, their marital status Married, and the longevity of their mothers, denoted by the binary variable MOTHER70 which takes value 1 if the mother is still alive or died at the age older than 70 and 0 otherwise.

The exclusion restriction variable for our empirical analysis is MOTHER70. We make the plausible identifying assumption that Mother70 affects the transition matrix of the key state variables (BADHEALTH and DEATH) that directly enter into the instantaneous utilities, but MOTHER70 itself does not directly affect the instantaneous utilities. In the preliminary results below, we will show evidence for this assumption.

2.3.6 Preliminary Results

First Step Estimates

As we noted earlier, our estimation strategy has two steps. In the first step, we need to use the data to estimate choice probabilities, and the state transitions. Here we report these first-step estimation results. The choice probabilities and the death probability are estimated using Logit regressions; but the transition of BADHEALTH and LOGINCOME are estimated non-parametrically.

Reduced Form Result for the Determinant of Mammography Table 2.3 produces the reduced form Logit regression results for the determinants of whether a woman will undertake mammogram in a given year. There are two statistically significant coefficient estimates: high income women are more likely and whites are less likely to undertake mammograms with both coefficient estimates significant at 1%. While the rest of the coefficient estimates are not significant, their signs are plausible: women whose mothers are still alive or died after age 70 are less likely, women with bad health are less likely, and married women are more likely, to undertake mammograms.[15]

[15] Table 2.3 is a shorter version of the actual regressions we used to measure the choice probabilities where we also include many interactions among the variables included in Table 2.3: MOTHER70×AGE, MOTHER70×AGE^2, BADHEALTH×AGE, BADHEALTH×AGE^2, WHITE×MOTHER70, WHITE×HLOGINCOME, WHITE×AGE, WHITE×AGE^2, MARRIED×MOTHER70, MARRIED×LOGINCOME, MARRIED×AGE, MARRIED×AGE^2, WHITE×MARRIED, WHITE×MARRIED×LOGINCOME. The results from this flexible Logit regression are harder to interpret than those in Table 2.3.

Table 2.3: Determinants of Mammography Decisions: Reduced Form Logit Regression Results.

Variable	Coeff. Est.	Std. Err.	t-stat.
MOTHER70	-.028	.050	-0.57
BADHEALTH	-.0067	.0531	-0.13
LOGINCOME	.378***	.024	15.57
AGE	.172	.174	0.98
AGE^2	-.0014	.00151	-0.95
WHITE	-.209***	.0547	-3.82
MARRIED	.0427	.052	0.82
Constant	-7.65	5.03	-1.52

Table 2.4: Determinants of Probability of Dying in Two Years.

Variable	Coeff. Est.	Std. Err.	t-stat.
MAMMOGRAM	1.94	2.550	0.76
MAMMOGRAM×AGE	-0.04	0.043	-0.92
MAMMOGRAM×BADHEALTH	-0.04	0.333	-0.12
MOTHER70	-0.25*	0.160	-1.54
BADHEALTH	1.72***	0.275	6.25
LOGINCOME	-0.183***	0.078	-2.35
AGE	0.294	0.667	0.44
AGE^2	-0.0016	0.006	-0.29
WHITE	-0.126	0.167	-0.75
MARRIED	-0.100	.174	-0.57
Constant	-14.1	19.6	-0.72

Determinants of the Probability of Dying in Two Years Table 2.4 reports the Logit regression results for the probability of dying in two years. It shows that on net, undertaking mammogram lowers the probability of death (notice that the average age of the sample is about 58 years). Not surprisingly, women with bad health are more likely to die, but mammogram reduces the probability of dying conditional on bad health. Also note that women whose mothers are either still alive or died after age 70 are less likely to die, suggesting a genetic link of longevity between mothers and daughters.

Evolution of Bad Health in Two Years Figure 2.2 depicts a subset of the results from the non-parametric estimation of the evolution of BadHealth for a selective combinations of the other state variables and the mammogram choice. For example, Panel (a) shows that probability of having bad health in period 2 is higher for women with bad health in period 1 conditional on undertaking mammogram in period 1 and their mothers are either still alive or died after age 70. Panel (b) showed that women with longer living mothers are less likely to experience bad health. Panel (c) and (d) showed that undertaking mammogram lowers the probability of bad health in period 2. Also note from all panels that the probability of bad health decreases with household income.

Utility Parameters and Discount Factors

Here we report the preliminary estimation results for the parameters in the utility function specification (2.30) and the identified discount factor combinations $\tilde{\beta}/\beta$ and $\tilde{\beta}\delta$. We found that having bad health lowers the utility of undertaking mammography relative to not undertaking mammography, consistent with our earlier finding in Table 2.3 that women with bad health are less likely to undertake mammography. We found that the relative utility of undertaking mammography increases with household income, consistent with the finding in Table 2.3 that women with higher household incomes are more likely to undertake mammography.

More interestingly, we estimate $\tilde{\beta}/\beta$ to be 1.80 and $\tilde{\beta}\delta$ to be around 0.86. We still need to calculate the standard error for these estimates. But these point estimates shows that women exhibit substantial present bias ($\beta < 1$) as well as naivety about their present bias ($\tilde{\beta} > \beta$) when making mammography decisions.

Our estimates for $\beta\delta = \tilde{\beta}\delta \times \beta/\tilde{\beta} = 0.86/1.80 \approx 0.48$. This can be compared with the estimate in Fang and Silverman (2007) where they estimate β to be 0.338 and δ to be 0.88, with $\beta\delta \approx 0.30$ for a group of single mothers with dependent children. It is important to note that our sample period is two years, while Fang and Silverman's sample period is one year; also the sample of women here are older and have very

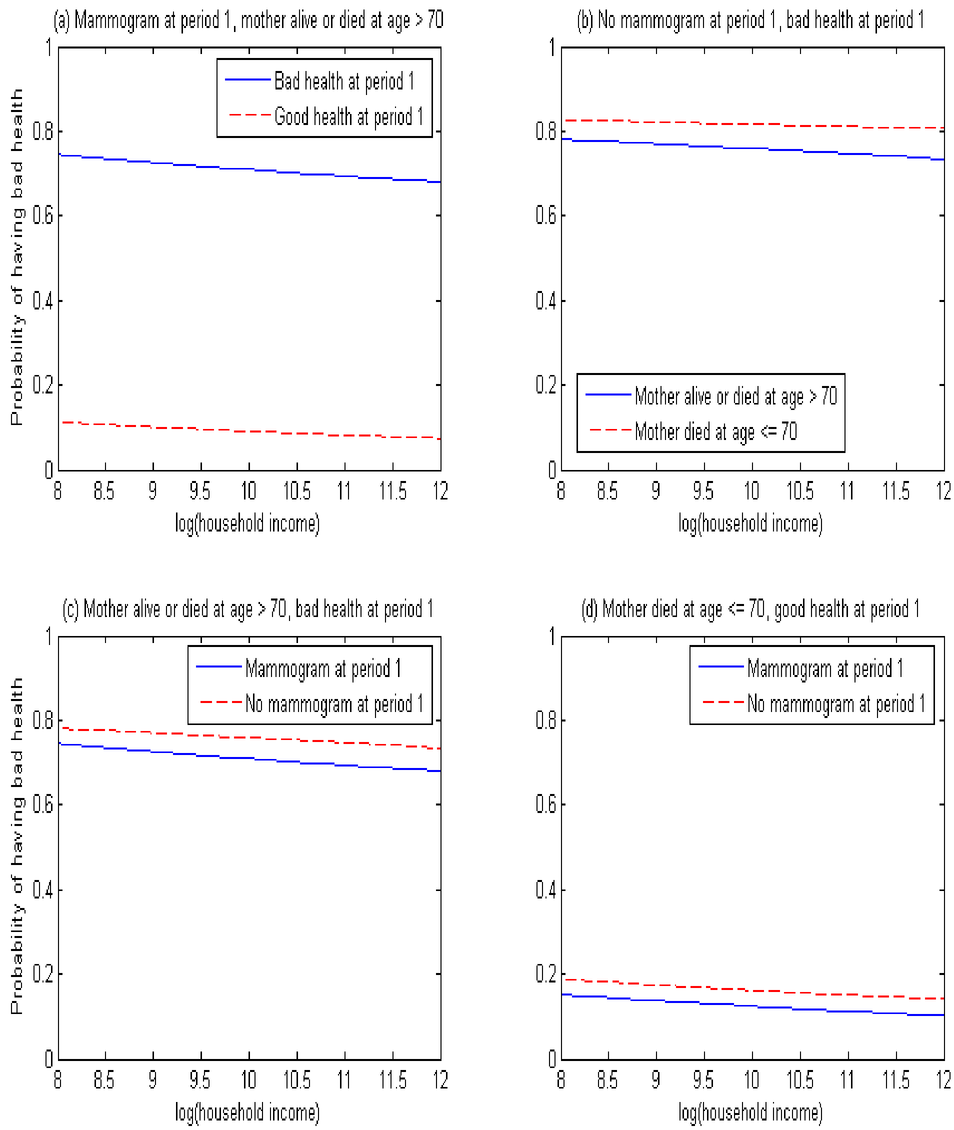


FIGURE 2.2: Non-parametric Estimate of the Determinants of Bad Health in Two Years

Table 2.5: Preliminary Parameter Estimates for Utility Function and Combinations of Discount Factors.

Variable	Coeff. Est.	Std. Err.
Utility Function Parameters		
BadHealth	-0.29	
LogIncome	0.51	
Constant	-4.66	
Discount Factor Combinations		
$\tilde{\beta}/\beta$	1.80	
$\tilde{\beta}\delta$	0.86	

different social economic status (e.g. education, income, marital status) from the sample in Fang and Silverman.

2.4 Conclusion

This paper extends the semi-parametric identification and estimation method for dynamic discrete choice models using Hotz and Miller's (1993) conditional choice probability (CCP) approach to the setting where individuals may have hyperbolic discounting time preferences and may be naive about their time inconsistency.

Our analysis showed that the three discount factors, the present bias factor β, the standard discount factor δ and the perceived present bias factor $\tilde{\beta}$ for naive agents can not be separately identified; however, with identifying exclusive restrictions, we can identify two combinations of the above three parameters, namely, $\tilde{\beta}/\beta$ and $\tilde{\beta}\delta$. The key identifying restriction is that there exists variables that do not directly enter the instantaneous utility function but affect the transition of other payoff relevant state variables.

Our discussion on identification also makes it clear that, with variables that satisfy the identifying exclusive restrictions, β and δ can not be separately identified in a fully sophisticated model with hyperbolic discounting time preferences (i.e. $\tilde{\beta} = \beta$) with two-period short panel data; but they could be separately when the agents are assumed to be naive about their present bias (i.e. $\tilde{\beta} = 1$).

We proposed two estimation strategies based on the identification argument, and implement the proposed estimation method to the decisions of undertaking mammography to evaluate the importance of present bias and naivety in the underutilization of mammography. Preliminary results show evidence for both present bias and naivety. The preliminary estimate of $\tilde{\beta}/\beta$ is about 1.80 and $\tilde{\beta}\delta$ is about 0.86.

Appendix A

Supplemental Materials

A.1 Proof of Proposition 2

Proof. For simplicity, I will drop time subscripts t and $t+1$ in this proof. Denote $\hat{z} = \Phi^{-1}(\hat{E})$. Then, as $\Phi(.)$ is monotonic and known, $\hat{z}$ is effectively observable, so we can rewrite the equation (1.6) in the following way.

$$\hat{z} = S + \lambda. \tag{A.1}$$

Due to monotonicity of $\Phi(.)$, the information in $\hat{E}$ is the same as in $\hat{z}$. It is, however, easier to work with the transformed variable due to linearity.

Then, from equation (1.4),

$$P(O|\hat{z}) = P(S - \xi > 0|\hat{z}) = P(\varepsilon < S|\hat{z}) = \int P(\xi < S|S, \hat{z}) f(S|\hat{z}) dS$$

$$= \int \Phi(S) f(S|\hat{z}) dS = \int \Phi(S) \frac{f(S,\hat{z})}{f(\hat{z})} dS = \int \Phi(S) f(S) f_\lambda(\hat{z} - S) dS \frac{1}{f(\hat{z})}$$

So,

$$P(O|\hat{z}) f(\hat{z}) = \int \Phi(S) f(S) f_\lambda(\hat{z} - S) dS$$

Take Fourier Transform of both sides. Since the left-hand side is observed, $\int P(O|\hat{z})f(\hat{z})e^{it\hat{z}}d\hat{z}$ is observed and is equal to the convolution between $\Phi(S)f(S)$ and $f(\lambda)$:

$$\int P(O|\hat{z})f(\hat{z})e^{it\hat{z}}d\hat{z} = \int \Phi(S)f(S)e^{itS}dS \int f_\lambda(\lambda)e^{it\lambda}d\lambda \tag{A.2}$$

where $\int \Phi(S)f(S)e^{itS}dS$ is defined as $Ee^{itS}\Phi(S)$, and $\int f_\lambda(\lambda)e^{it\lambda}$ is defined as $Ee^{it\lambda}$. That is, Fourier Transform of convolution is equal to the product of Fourier transforms.

Now, from Eq. (A.1), $Ee^{it\hat{z}} = Ee^{itS}Ee^{it\lambda}$. So, define

$$\Phi_{\hat{z}}(t) = \frac{\int P(O|\hat{z})f(\hat{z})e^{it\hat{z}}d\hat{z}}{Ee^{it\hat{z}}}, \tag{A.3}$$

where $\Phi_{\hat{z}}(t)$ is observable from the data. From (A.2) it follows that

$$\Phi_{\hat{z}}(t) = \frac{\int \Phi(S)f(S)e^{itS}dS}{\int f(S)e^{itS}dS}, \quad \forall t,$$

which leads to

$$\int (\Phi(S) - \Phi_{\hat{z}}(t))f(S)e^{itS}dS = 0, \quad \forall t. \tag{A.4}$$

Eq. (A.4) can be considered a "linear functional relation" of $f(S)$, which must be satisfied for all t, subject to the integrability constraint on $f(S)$. Once $f(S)$ is identified, identification of the distribution of λ is straightforward. □

Equation (A.4) provides a basis for empirical identification of the unknown distribution $f(S)$ in the data. In the empirical analysis, I take $\Phi(\xi)$ to be standard Normal distribution. I discretize the support of $f(S)$ into N equidistant points $\{S_1, S_2, \ldots S_N\}$ with increment ΔS, and consider M possible values for t $\{t_1, t_2, \ldots t_M\}$.

Denote the vector of values of $f(S)$ at the selected points $f(\mathbf{S})$:

$$f(\mathbf{S}) = \begin{bmatrix} f(S_1) & f(S_2) & \cdots & f(S_N) \end{bmatrix}'. \tag{A.5}$$

Let us construct a complex-valued $M \times N$ matrix H, whose (j, k) element is

$$H(j, k) = (\Phi(S_k) - \Phi_{\hat{z}}(t_j))e^{it_j S_k} \Delta S. \tag{A.6}$$

This matrix can be constructed from the data; further, given equation (A.4), for $N \to \infty$ it should be true that

$$Hf(\mathbf{S}) \to 0. \tag{A.7}$$

For a fixed N, I minimize the sum of squared products of real and imaginary parts of rows of H and the unknown function $f(\mathbf{S})$, subject to the integrability constraint on $f(\mathbf{S})$.

Denote $W = \begin{bmatrix} \text{Real}(H) \\ \text{Imag}(H) \end{bmatrix}$, and let W_k denote the kth row of W. I then select $f(\mathbf{S})$ according to the following objective function

$$\min_{f(\mathbf{S})} \sum_k (W_k f(\mathbf{S}))^2, \tag{A.8}$$

subject to the integrability constraint $\sum_n f(S_n)\Delta S = 1$. Taking first-order conditions, the optimal solution to $f(\mathbf{S})$ is then given by

$$f(\mathbf{S}) \propto (\sum_k W_k' W_k)^{-1} \times \mathbf{1}, \tag{A.9}$$

where the proportionality coefficient is chosen to ensure integrability of $f(\mathbf{S})$ to 1.

In principle, this approach permits the estimation of the unknown distribution function of S, $f(\mathbf{S})$, for an arbitrarily large grid N. In practice, I find that this approach works quite well for N around $10 - 30$, and for a $30 - 50$ discrete values of t ranging from 0 to about $10 - 20$ (negative values of t are redundant in W).

After I identify the distribution of S, it is straightforward to obtain the distribution of the expectation bias (λ) using inverse Fourier methods. Indeed, I can estimate the characteristic function of λ using

$$Ee^{it\lambda} = Ee^{it\hat{z}}/Ee^{itS}, \tag{A.10}$$

and then recover the probability distribution function $f_\lambda(\lambda)$ using inverse Fourier transform:

$$f_\lambda(\lambda) = \frac{1}{2\pi}\int e^{-it\lambda} Ee^{it\lambda} dt \tag{A.11}$$

A.2 Proof of Proposition 3

Proof of Statement 1: In the ideal case that econometrician observes both factual and counterfactual subjective expectations, the only unobserved variable in the choice equation (1.10) is ε, and the binary choice problem falls into a standard framework analyzed by [66], [69] and [17], among others. The conditions provided in these works (e.g., independence of ε from observed state variables) can be used for identification of model parameters θ and β, as well as the conditional distribution of ε in Eq. (1.10) given X_t, $\hat{E}(x_{t+1}|S_t, 1)$ and $\hat{E}(x_{t+1}|S_t, 0)$.[1]

Proof of Statement 2: When both factual and counterfactual subjective expectations are elicited, the conditional distributions of the true underlying state transition, the expectation bias, and the realization shock given X_t and A_t are identified for $A_t = (0, 1)$, following the arguments in Section 1.2.

Let us first add Assumption 1.3.2 on choice-independent expectation bias. Then,

$$\begin{aligned} \hat{E}(X_{t+1}|S_t, 0) &= E(X_{t+1}|S_t, 0) + \lambda_t \\ \hat{E}(X_{t+1}|S_t, 1) &= E(X_{t+1}|S_t, 1) + \lambda_t. \end{aligned} \tag{A.12}$$

[1] See Appendix A in [40] for a complete list of Matzkin's conditions.

The conditional mean of λ_t given X_t can be identified using the optimal choice and expectations data in the following way,

$$E(\lambda_t|X_t) = E(\lambda_t|X_t, A_t^* = 1) \times Pr(A_t^* = 1) + E(\lambda_t|X_t, A_t^* = 0) \times Pr(A_t^* = 0).$$

As the conditional distribution of λ_t given X_t and optimal choice A_t^* is identified (see Section 1.3.1) and the optimal actions are observable, $E(\lambda_t|X_t)$ is identified as well. Hence, the means of the conditional distributions of $E(X_{t+1}|S_t, 0)$, $E(X_{t+1}|S_t, 1)$ and λ_t in Eq. (A.12) are identified. Applying the Kotlarski theorem, the marginal distributions of these variables can be identified given that $E(X_{t+1}|S_t, 0)$, $E(X_{t+1}|S_t, 1)$ and λ_t are mutually independent (conditional on X_t) and characteristic functions exist.

Now, let us use Assumption SIR instead. Then, the underlying expectation $E(X_{t+1}|S_t, A_t)$ is equal to $E(X_{t+1}|X_t, A_t)$, and can be directly identified from the objective data by looking at the optimal choices of the agents. The difference between objective and subjective expectations for the two choices allows to identify the distribution of the expectation bias.

A.3 Proof of Proposition 4

Proof of Statement 1: Without counterfactual subjective expectations, Assumptions CI and SIR can be combined to identify the underlying true expectations $E(X_{t+1}|S_t, A_t)$ $(= E(X_{t+1}|X_t, A_t))$ from the objective data and agents' optimal choices (Assumption SIR). Matching factual subjective and objective expectations, we can further identify the expectation bias, which by Assumption CI is independent of the choice.

Proof of Statement 2: Under Assumptions CI and SIR, the counterfactual subjective expectations can be obtained using the choice-independent bias and the identifiable underlying true expectations. Once the counterfactual subjective expectations are known, we can bring them back to Eq. (1.10) and the identification of the model parameters and the distribution of ε proceed as discussed above.

Proof of Statement 3: Without Assumptions CI and SIR, non-parametric identification of the distributions of the expectation bias, the realization shock, and the true underlying expectations conditional upon *optimal* choices (A_t^*) are still possible based on the arguments in Section 1.2. However, identification of model parameters (θ, β) and the distribution of ε in Eq. (1.10) are in general only possible to a certain extent.

One alternative is to keep Assumption CI and to relax Assumption SIR in the following way. Writethe true underlying objective expectation as:

$$E(X_{t+1}|S_t, A_t) = E(X_{t+1}|X_t, A_t) + \nu(S_t, A_t),$$

where $\nu(.)$ denotes the only part in the true underlying objective expectation that depends on the unobserved private information in S_t. Then, the difference in the true underlying expectations by choices, $\Delta_{1-0}E(X_{t+1}|S_t)$, is:

$$\Delta_{1-0}E(X_{t+1}|S_t) = \Delta_{1-0}E(X_{t+1}|X_t) + \Delta_{1-0}\nu(S_t, A_t). \tag{A.13}$$

Given Assumption CI, we can identify the model parameters (θ, β), as well as the distribution of the term ($\varepsilon + \beta\theta'\Delta_{1-0}\nu(S_t, A_t)$), assuming $\varepsilon + \beta\theta'\Delta_{1-0}\nu(S_t, A_t)$ has zero median conditional upon X_t.

Indeed, since Assumption CI still hold, using Eq. (A.13) in which the first term on the right-hand side can be obtained from the data, we have:

$$\Delta_{1-0}\hat{E}(X_{t+1}|S_t) = \Delta_{1-0}E(X_{t+1}|S_t) = \Delta_{1-0}E(X_{t+1}|X_t) + \Delta_{1-0}\nu(S_t),$$

which can be used in Eq. (1.10) to give us:

$$A_t^* = I(\theta' X_t + \beta\theta'\Delta_{1-0}E(X_{t+1}|X_t) + \varepsilon + \beta\theta'\Delta_{1-0}\nu(S_t)) > 0).$$

According to [66], [69] and [17], if we know that $\varepsilon + \beta\theta'\Delta_{1-0}\nu(S_t)$ has zero median conditional upon X_t, we can identify θ, β, and the distribution of $\varepsilon + \beta\theta'\Delta_{1-0}\nu(S_t)$.

A.4 Derivation of Eq. (1.18)

Eq. (1.16) implies that the ex ante value function can be expressed as a function of choice specific value functions, which can be further written as a function of the differences in choice specific value functions ($d(x) \equiv V(x,1) - V(x,0)$) and one baseline choice specific value function.

With two choices, we can write Eq. (1.16) in the following way

$$\begin{aligned} V(x_t) = & \log\Big\{\exp(V(x_t,1)) + \exp(V(x_t,0))\Big\} \\ = & \log\Big\{\exp(V(x_t,1) - V(x_t,0)) + 1\Big\} + V(x_t,0) \\ = & \log\Big\{\exp(V(x_t,0) - V(x_t,1)) + 1\Big\} + V(x_t,1) \end{aligned} \tag{A.14}$$

where the first equality is from the two-choice assumption, and the second and third equalities are obtained by adding and subtracting $\ln(\exp(V(x_t,0)))$ and $\ln(\exp(V(x_t,1)))$, respectively.

By inserting the second line of Eq. (A.14) into Eq. (1.15) and rearranging the terms, we can have the following expression with $d(x_t)$

$$\begin{aligned} & V(x_t,0) - \beta \textstyle\int V(x_{t+1},0)\hat{p}(x_{t+1}|x_t,0)dx_{t+1} \\ = \ & u(x_t,0) + \beta \textstyle\int \log(1 + e^{d(x_{t+1})})\hat{p}(x_{t+1}|x_t,0)dx_{t+1} \end{aligned} \tag{A.15}$$

where the second term on the right hand side can be obtained directly from the data. Taking this term as given, the left hand side subsequently defines a backward induction relation that recursively expresses $V(x_t,0)$ for all t as functions of $u(x_t,0)$ only.

Specifically, in the last period, the continuation value is equal to the instantaneous utility, for all x and a

$$V(x_T, a_T) = u(x_T, a_T) \tag{A.16}$$

For any period $t < T$, we can rewrite Eq. (A.15) and apply it recursively by moving the time index forward

$$\begin{aligned} V(x_t, 0) = u(x_t, 0) + \beta \textstyle\int \log(1 + e^{d(x_{t+1})})\hat{p}(x_{t+1}|x_t, 0)dx_{t+1} \\ + \beta \textstyle\int V(x_{t+1}, 0)\hat{p}(x_{t+1}|x_t, 0)dx_{t+1} \\ = \quad u(x_t, 0) + \beta \textstyle\int \log(1 + e^{d(x_{t+1})})\hat{p}(x_{t+1}|x_t, 0)dx_{t+1} \\ + \beta \textstyle\int u(x_{t+1}, 0)\hat{p}(x_{t+1}|x_t, 0)dx_{t+1} \\ + \beta^2 \textstyle\int\int \log(1 + e^{d(x_{t+2})})\hat{p}_{t+1,t+2}(x_{t+2}|x_{t+1}, 0)\hat{p}_{t,t+1}(x_{t+1}|x_t, 0)dx_{t+2}dx_{t+1} \\ + \beta^2 \textstyle\int\int V(x_{t+2}, 0)\hat{p}_{t+1,t+2}(x_{t+2}|x_{t+1}, 0)\hat{p}_{t,t+1}(x_{t+1}|x_t, 0)dx_{t+2}dx_{t+1} \\ = \quad \ldots \end{aligned}$$

which can be (relatively more succinctly) written as:

$$\begin{aligned} V(x_t, 0) = u(x_t, 0) + \beta \textstyle\int \log(1 + e^{d(x_{t+1})})\hat{p}(x_{t+1}|x_t, 0)dx_{t+1} \\ + \beta \textstyle\int V(x_{t+1}, 0)\hat{p}(x_{t+1}|x_t, 0)dx_{t+1} \\ = \quad u(x_t, 0) + \beta \textstyle\int \log(1 + e^{d(x_{t+1})})\hat{p}(x_{t+1}|x_t, 0)dx_{t+1} \\ + \beta \textstyle\int u(x_{t+1}, 0)\hat{p}(x_{t+1}|x_t, 0)dx_{t+1} + \beta^2 \textstyle\int \log(1 + e^{d(x_{t+2})})\hat{p}_{t,t+2}(x_{t+2}|x_t, 0)dx_{t+2} \\ + \beta^2 \textstyle\int V(x_{t+2}, 0)\hat{p}_{t,t+2}(x_{t+2}|x_t, 0)dx_{t+2} \\ = \quad \ldots \end{aligned}$$

which leads to the following backward induction relation

$$\begin{aligned} V(x_t, 0) = \quad & \textstyle\sum_{s=t}^{T} \beta^{s-t} E\left[u(x_s, 0)|x_t = x, a = 0\right] \\ & + \textstyle\sum_{s=t+1}^{T} \beta^{s-t} E\left[\log(1 + e^{d(x_s)})|x_t = x, a = 0\right] \end{aligned} \tag{A.17}$$

In the expectation here, the distribution of x_s given $x_t = x, a = 0$ is induced by the transition of the state variables x from t to s as if action $a = 0$ is always taken between periods t to $s-1$.

Analogously, if we instead insert the third line of Eq. (A.14) into Eq. 1.15, then we have

$$\begin{aligned} & V(x_t, 1) - \beta \int V(x_{t+1}, 1)\hat{p}(x_{t+1}|x_t, 1)dx_{t+1} \\ = \quad & u(x_t, 1) + \beta \int \log(1 + e^{-d(x_{t+1})})\hat{p}(x_{t+1}|x_t, 1)dx_{t+1} \end{aligned} \tag{A.18}$$

And for any period $t < T$, Eq. (A.18) translates into the following backward induction relation

$$\begin{aligned} V(x_t, 1) = \quad & \textstyle\sum_{s=t}^{T} \beta^{s-t} E\left[u(x_s, 1)|x_t = x, a = 1\right] \\ & + \textstyle\sum_{s=t+1}^{T} \beta^{s-t} E\left[\log(1 + e^{-d(x_s)})|x_t = x, a = 1\right] \end{aligned} \tag{A.19}$$

In the above conditional expectations, the distribution of x_s given $x_t = x, a = 1$ is induced by the transition of the state variables x from t to s as if action $a = 1$ is always taken between periods t to $s-1$.

Given that the data identify $d(x_t)$ ($\equiv V(x_t, 1) - V(x_t, 0)$), taking the difference between Eqs. (A.17) and (A.19), and rearranging terms shows that the difference

$$\sum_{s=t}^{T} \beta^{s-t} E\left[u(x_s, 1)|x_t = x, a = 1\right] - \sum_{s=t}^{T} \beta^{s-t} E\left[u(x_s, 0)|x_t = x, a = 0\right]$$

is identified through the data as

$$\begin{aligned} & d(x_t) - \textstyle\sum_{s=t+1}^{T} \beta^{s-t} E\left[\log(1 + e^{-d(x_s)})|x_t = x, a = 1\right] \\ & \qquad + \textstyle\sum_{s=t+1}^{T} \beta^{s-t} E\left[\log(1 + e^{d(x_s)})|x_t = x, a = 0\right]. \end{aligned}$$

Given the linear form of the instantaneous period utility functions, difference between Eq. (A.18) and Eq. (A.15) can be written as

$$\begin{aligned} & d(x_t) - \sum_{s=t+1}^{T} \beta^{s-t} E\left[\log(1+e^{-d(x_s)})|x_t = x, a = 1\right] \\ & \quad + \sum_{s=t+1}^{T} \beta^{s-t} E\left[\log(1+e^{d(x_s)})|x_t = x, a = 0\right] \\ = & \sum_{s=t}^{T} \beta^{s-t} E\left[x_{1s}|x_t = x, a = 1\right]' \theta_1 - \sum_{s=t}^{T} \beta^{s-t} E[x_{0s}|x_t = x, a = 0]' \theta_0 \quad \text{(A.20)} \end{aligned}$$

which is exactly Eq. (1.18).

Note that, this identification method is an improvement over the current literature, because it requires fewer estimation steps.

A.5 Derivation of $\sum_{s=t}^{T} \beta^{s-t} E\left[x_a^s|x_t = x, a\right]'$

With discrete state variables, we can write the "regression" relation in Eq. (A.20) as

$$\begin{aligned} & d(x_t) - \sum_{s=t+1}^{T} \beta^{s-t} P_1^{t,s} \log(1+e^{-d(x_s)}) + \sum_{s=t+1}^{T} \beta^{s-t} P_0^{t,s} \log(1+e^{d(x_s)}) \\ = & \left[\sum_{s=t}^{T} \beta^{s-t} P_1^{t,s} x_1\right]' \theta_1 - \left[\sum_{s=t}^{T} \beta^{s-t} P_0^{t,s} x_0\right]' \theta_0 \quad \text{(A.21)} \end{aligned}$$

where $d(x)$ denotes the vector of differences in the choice specific value functions with length equal to the number of discrete states, and $P_1^{t,s}$ and $P_0^{t,s}$ are defined as

$$\begin{aligned} P_1^{t,s} &= P_1^{t,t+1} P_1^{t+1,t+2} \cdots P_1^{s-1,s} \\ P_0^{t,s} &= P_0^{t,t+1} P_0^{t+1,t+2} \cdots P_0^{s-1,s} \end{aligned}$$

with

$$\begin{aligned} P_1^{t,t} &= I \\ P_0^{t,t} &= I \end{aligned}$$

More precisely, we can write

$$\begin{aligned} \sum_{s=t}^{T} \beta^{s-t} P_1^{t,s} = & \quad I + \beta P^{t,t+1} + \beta^2 P^{t,t+2} + \ldots + \beta^{T-t} P^{t,T} \\ = & \quad I + \beta P^{t,t+1} + \beta P^{t,t+1} \beta P^{t+1,t+2} + \ldots + \beta P^{t,t+1} \times \cdots \times \beta P^{T-1,T} \end{aligned}$$

It is worth noting that when death is allowed and the states in P_1 and P_0 do not include death, P_1 and P_0 only have to be "sub" stochastic matrices, in the sense that their rows sum to < 1 instead of 1.

If we define

$$Q_t = \sum_{s=t}^{T} \beta^{s-t} P_1^{t,s}$$

then we can compute Q_t through recursive relations.

First of all,

$$Q_T = I$$

Secondly, for all $t < T$

$$Q_t = \beta P^{t,t+1} Q_{t+1} + I$$

It is also worth noting that there is a certain relation between the matrices on the left hand side and the right hand side of Eq. (A.21). In particular, the left hand side matrices are $\beta P_1^{t,t+1}$ and $\beta P_0^{t,t+1}$ multiplied by the left hand side matrices shifted forward by one period

$$\sum_{s=t+1}^{T} \beta^{s-t} P_1^{t,s} = \beta P_1^{t,t+1} (\sum_{s=(t+1)}^{T} \beta^{s-(t+1)} P_1^{(t+1),s}),$$

and

$$\sum_{s=t+1}^{T} \beta^{s-t} P_0^{t,s} = \beta P_0^{t,t+1} (\sum_{s=(t+1)}^{T} \beta^{s-(t+1)} P_0^{(t+1),s}).$$

A.6 Kernel Smoothing

For estimation purposes, I discretize the continuous variable – the logarithm of household income – into a finite number of discrete values. Furthermore, due to the large number of discrete states and the relatively sparse data for each value of the discrete state variables, I kernel smooth over the discretized income levels when I nonparametrically estimate the conditional transition probabilities of the state variables. Kernel weights are calculated as follows

$$w(b_2, l_2|b_1, l_1, s_1, m) = \frac{p(b_2, l_2|b_1, l_1, s_1, m)}{\sum_{i=1}^{n} p(b_{2i}, l_{2i}|b_{1i}, l_{1i}, s_{1i}, m_i)}$$

where

$$p(b_2, l_2|b_1, l_1, s_1, m) =$$
$$\frac{\frac{1}{nh^2}\sum_{i=1}^{n} k(\frac{l_{2i}-l_2}{h})k(\frac{l_{1i}-l_1}{h})I(b_{2i}=b_2)I(b_{1i}=b_1)I(s_{1i}=s_1)I(m_i=m)}{\frac{1}{nh}\sum_{i=1}^{n} k(\frac{l_{1i}-l_1}{h})I(b_{1i}=b_1)I(s_{1i}=s_1)I(m_i=m)}$$

is the kernel estimate of the conditional joint density of the log(household income) and health status in period 2 conditional on log(household income), health status, and smoking status in period 1 as well as the indicator of same-gender parent's longevity. b denotes bad health, l denotes log(household income), s denotes smoking status, m denotes the indicator of whether the same-gender parent is still alive or died at an age greater than 70. $k(\cdot)$ is the kernel function and h is the bandwidth used in the nonparametric probability estimation and adapted from "Silverman's Rule of Thumb" ([81]).[2]

[2] See [37] for a discussion of bandwidth choice.

A.7 Additional Tables and Figures

Table A.1: Comparing Final Sample and the Sample with Missing Subjective Probabilities

	Final Sample		Missing Subj. Prob.	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Age	56.7*	2.98	56.7	2.97
Female	0.53**	0.50	0.51	0.50
White	0.82**	0.38	0.81	0.39
Current smoker at period 1	0.38	0.49	0.38	0.49
Former smoker at period 1	0.62	0.49	0.62	0.49
Long-lived same gender parent	0.67	0.47	0.67	0.47
Self-rated health at period 1	2.65**	1.16	2.69	1.17
Bad health at period 1	0.23**	0.42	0.24	0.43
Household income at period 1 ('$k)	53.4	65.6	53.0	78.3
Self-rated health at period 2	2.63*	1.25	2.65	1.27
Bad health at period 2	0.24**	0.43	0.29	0.50
Household income at period 2 ('$k)	51.6	68.4	50.8	68.2
Observed deaths in two years	0.021	0.14	0.023	0.15

Differences in means between the final analysis sample and the sample with missing subjective probabilities.
*Statistically significant at 5% level; **statistically significant at 1% level.

Table A.2: Special Cases of Subjective Probabilities of living to Ages 75 and 85

Cases	%
Live to 75 = 0	0.064
Live to 85 = 0	0.170
Live to 75 = 1	0.200
Live to 85 = 1	0.087
P75 > P85	0.019
Both probabilities = 0	0.064
Both probabilities = 1	0.085
P75 >= P85	0.290
Any of the special case, with P75 > P85	0.380
Any of the special case, with P75 >= P85	0.500

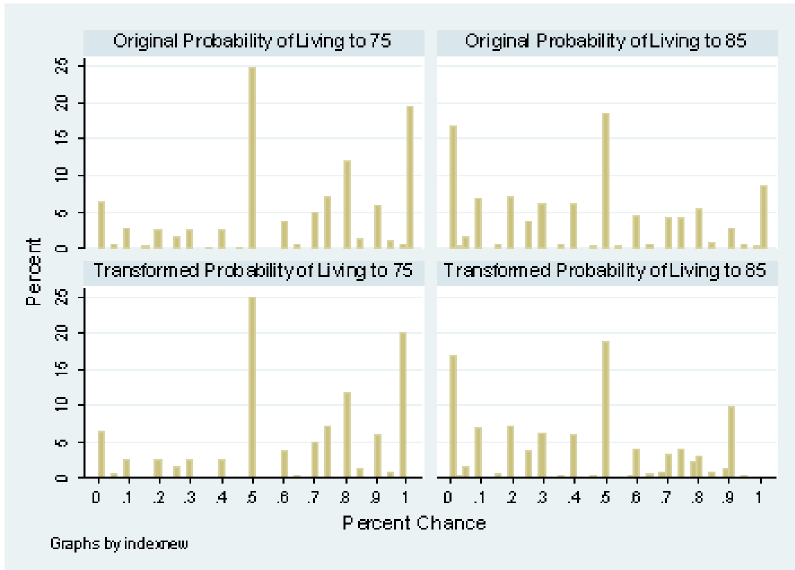
Frequencies of special elicited subjective expectations in the data.

Table A.3: Discount Factors in the Literature

Sources	Estimated Discount Factor
Moore and Viscusi (1990)	0.88 to 0.99
Moore and Viscusi (1988)	0.89 to 0.91
Arcidiacono, Sieg, and Sloan (2007)	0.91
Dreyfus and Viscusi (1995)	0.85, 0.88, 0.90
Hausman(1979)	0.83
Warner and Pleeter (2001)	0.77 to 1

Examples of discount factors estimated in the literature.

Figure A.1: Original and Transformed Subjective Probabilities of Living to Ages 75 and 85



Distributions of original elicited subjective responses and transformed answers.

Bibliography

[1] Daniel Ackerberg, Lanier Benkard, Steven Berry, and Ariel Pakes. Econometric tools for analyzing market outcomes. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume 6, Amsterdam, 2005. North Holland.

[2] Peter Arcidiacono and Robert Miller. Ccp estimation of dynamic discrete choice models with unobserved heterogeneity. Revise and Resubmit, 2008.

[3] Peter Arcidiacono, Holger Sieg, and Frank A. Sloan. Living rationally under the volcano? an empirical analysis of heavy drinking and smoking. *International Economic Review*, 48(1):37 – 65, 2007.

[4] G. Becker, M. Grossman, and K. Murphy. Rational addiction and the effect of price on consumption. *American Economic Review*, 71(3):933 – 946, 1991.

[5] G. Becker, M. Grossman, and K. Murphy. An empirical analysis of cigarette addiction. *American Economic Review*, 84(3):396 – 418, 1994.

[6] Gary Becker and Casey B. Mulligan. The endogenous determination of time preference. *The Quarterly Journal of Economics*, pages 729 – 758, 1997.

[7] Gary Becker and Kevin Murphy. A theory of rational addiction. *Journal of Political Economy*, 96(4):675 – 701, 1988.

[8] Arie Beresteanu and Francesca Molinari. Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763 – 814, 2008.

[9] B. Douglas Bernheim. The timing of retirement: A comparison of expectations and realizations. In David Wise, editor, *The Economics of Aging*, pages 335 – 355. Chicago: The University of Chicago Press, 1989.

[10] B. Douglas Bernheim. How do the elderly form expectations: An analysis of responses to new information. In David Wise, editor, *Issues in the Economics of Aging*, pages 259 – 283. Chicago: The University of Chicago Press, 1990.

[11] A. Borsch-Supan. Comment on 'subjective survival curves and life cycle behavior'. In D. A. Wise, editor, *Inquiries in the Economics of Aging*, pages 306 – 309. Chicago: University of Chicago Press, 1998.

[12] J. Bound. Self-reported and objective measures of health in retirement models. *Journal of Human Resources*, 24:106–138, 1991.

[13] W. Bruine de Bruin, B. Fischhoff, B. Halpern-Felsher, and S. Millstein. Expressing epistemic uncertainty: It's a fifty-fifty chance. *Organizational Behavior and Human Decision Processes*, 81:115 – 131, 2000.

[14] F. Chaloupka, J. A. Tauras, and M. Grossman. *Tobacco Control in Developing Countries*, chapter The Economics of Addiction, pages 107 – 1290. Oxford: Oxford University Press., 2000.

[15] F. Chaloupka and K. Warner. *Handbook of Health Economics*, chapter The Economics of Smoking, pages 1539 – 1627. New York: North Holland, 2000.

[16] Xiaohong Chen, Han Hong, and Denis Nekipelov. Nonlinear models of measurement errors. Under Review, 2008.

[17] Flavio Cunha, James J. Heckman, and Salvador Navarro. Separating uncertainty from hetergeneity in life cycle earnings. *Oxford Economic Papers*, 57(2):191 – 261, April 2005. Oxford University Press.

[18] Adeline Delavande. Measuring revisions to subjective expectations. *Journal of Risk and Uncertainty*, 36(1), 2008.

[19] Adeline Delavande. Pill, patch or shot? subjective expectations and birth control choice. *International Economic Review*, 49(3), 2008.

[20] J. Dominitz and A. Hung. Homogeneous actions and heterogeneous beliefs: Experimental evidence on the formation of information cascades. working paper, Heinz School of Policy and Management, Carnegie Mellon University, 2003.

[21] J. Dominitz and C. Manski. Perceptions of economic insecurity: Evidence from the survey of economic expectations. *Public Opinion Quarterly*, 61:261 – 287, 1997.

[22] Jeff Dominitz. Earnings expectations, revisions, and realizations. *Review of Economics and Statistics*, 80:374 – 388, 1998.

[23] Zvi Eckstein and Kenneth I. Wolpin. The specification and estimation of dynamic stochastic discrete choice models: A survey. *The Journal of Human Resources*, 24:562 – 598, 1989.

[24] I. Ehrlich. Uncertain lifetime, life protection, and the value of life saving. *Journal of Health Economics*, 19:341 – 367, 2000.

[25] I. Ehrlich and Y. Yin. Explaining diversities in age-specific life expectations and values of life saving: A numerical analysis. *Journal of Risk and Uncertainty*, 31(2):129 – 162, 2005.

[26] Hanming Fang and Dan Silverman. Time-inconsistency and welfare program participation: Evidence from the nlsy. International Economic Review, forthcoming, 2007.

[27] Hanming Fang and Yang Wang. Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. Working Paper, 2008.

[28] B. Fischhoff and W. Bruine de Bruin. Fifty-fitfy = 50? *Journal of Behavioral Decision Making*, 12:149 – 163, 1999.

[29] Baruch Fischhoff, Andrew M. Parker, Wandi Bruine De Bruin, Julie Downs, Claire Palmgren, Robyn Dawes, and Charles F. Manski. Teen expectations for significant life events. *Public Opinion Quarterly*, 64:189 – 205, 2000.

[30] Li Gan, Guan Gong, Michael Hurd, and Daniel McFadden. Subjective mortality risk and bequests. 2004. working paper.

[31] Li Gan, Michael Hurd, and Daniel McFadden. Individual subjective survival curves. NBER Working Paper 9480, 2003.

[32] D. Gilleskie. A dynamic stochastic model of medical care use and work absence. *Econometrica*, 66:1 – 46, 1998.

[33] Michael Grossman. On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223 – 255, 1972.

[34] Michael Grossman, F. Chaloupka, and R. Anderson. A survey of economic models of addictive behavior. *Journal of Drug Issues*, 28:631 – 643, 1998.

[35] Jonathan Gruber and Botond Koszegi. Is addiction “rational”? theory and evidence. *Quarterly Journal of Economics*, 116(4):935 – 958, 2001.

[36] Daniel Hamermesh. Expectations, life expectancy, and economic behavior. *Quarterly Journal of Economics*, 95:537 – 557, 1985.

[37] Wolfgang Hardle. *Applied Nonparametric Regression*. Cambridge University Press, 1990.

[38] C. Harris and D. Laibson. Dynamic choice of hyperbolic consumers. *Econometrica*, 69(4):935 – 958, 2001.

[39] James J. Heckman and Edward E. Leamer. *Handbook of Econometrics*, volume 6B. Elsevier, 2007.

[40] James J. Heckman and Salvador Navarro. Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*, 136:341 – 396, 2006.

[41] James J. Heckman and Burton Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271 – 320, 1984.

[42] T. Hodgson. cigarette smoking and lifetime medical expenditures. *The Milbank Quarterly*, 70(1):85 – 125, 1992.

[43] J. Hotz and R. Miller. Conditional choice probabilities and estimation of dynamic models. *Review of Economic Studies*, 60:497 – 529, 1993.

[44] Yingyao Hu and Mattew Shum. Nonparametric identification of dynamic models with unobserved state variables. Working Paper, April 2008.

[45] M. D. Hurd and K. McGarry. Evaluation of the subjective probabilities of survival in the health and retirement study. *Journal of Human Resources*, 30:S268 – S292, 1995.

[46] M. D. Hurd and K. McGarry. The predictive validity of subjective probabilities of survival. *The Economic Journal*, 112:996 – 985, 2002.

[47] Michael D. Hurd, Daniel McFadden, and Li Gan. *Inquiries in the Economics of Aging*, chapter Subjective Survival Curves and Life Cycle Behavior, pages 353 – 387. University of Chicago Press, 1998.

[48] Michael D. Hurd, Daniel McFadden, and Angela Merrill. *Themes in the Economics of Aging*, chapter Predictors of Mortality Among the Elderly, pages 171 – 197. University of Chicago Press, 2001.

[49] Michael D. Hurd, James P. Smith, and Julie M. Zissimopoulos. The effects of subjective survival on retirement and social security claiming. *Journal of Applied Econometrics*, 19:761 – 775, 2004.

[50] F. Thomas Juster and Richard Suzman. An overview of the health and retirement study. *Journal of Human Resources*, 40:934 – 940, 1995.

[51] Hiroyuki Kasahara and Katsumi Shimotsu. Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica forthcoming*, 2008.

[52] Michael P Keane and Kenneth I Wolpin. The career decisions of young men. *Journal of Political Economy*, 105(3):473 – 522, 1997.

[53] Donald S. Kenkel. Prevention. In A.J. Culyer and J.P. Newhouse, editors, *Handbook of Health Economics*, volume 1, pages 1675 – 1720. North Holland, 2000.

[54] Ahmed Khwaja, Frank A. Sloan, and Sukyung Chung. The relationship between individual expectations and behaviors: Mortality expectations and smoking decisions. *Journal of Risk and Uncertainty*, 35(2):179 – 201, October 2007.

[55] Ahmed Khwaja, Frank A. Sloan, and Martin Salm. Evidence on preferences and subjective beliefs of risk takers: The case of smokers. *International Journal of Industrial Organization*, 24:667 – 682, 2006.

[56] Ahmed Khwaja, Frank A. Sloan, and Yang Wang. Do smokers value their health and longevity less? forthcoming, Journal of Law and Economics, 2008.

[57] I. Kotlarski. On characterizing the gamma and normal distribution. *Pacific Journal of Mathematics*, 20:729 – 738, 1967.

[58] Elena Krasnokutskaya. Identification and estimation in highway procurement auctions under unobserved auction heterogeneity. Revise and resubmit in Review of Economic Studies, 2003.

[59] D. Laibson. Gloden eggs and hyperblic discounting. *Quarterly Journal of Economics*, 112(2):443 – 477, 1997.

[60] Tong Li. Robust and consistent estimation of nonlinear errors-in-variables models. *Journal of Econometrics*, 110:1 – 26, 2002.

[61] Tong Li and Q. Vuong. Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65:139 – 165, 1998.

[62] L. Liang, F. Chaloupka, M. Nichter, and R. Clayton. Prices, policies and youth smoking, may 2001. *Addiction*, 98:105 – 122, 2003.

[63] Lance Lochner. Individual perceptions of the criminal justice system. *American Economic Review*, 97:444 – 460, 2007.

[64] Thierry Magnac and David Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 20(2):801 – 816, 2002.

[65] Aprajit Mahajan. Identification and estimation of regression models with misclassification. *Econometrica*, 74(3):631 – 665, 2006.

[66] Charles F. Manski. Identification of binary response models. *Journal of the American Statistical Association*, 83(403):729 – 728, September 1988.

[67] Charles F. Manski. Dynamic choice in social settings. *Journal of Econometrics*, 58:121 – 136, 1993.

[68] Charles F. Manski. Measuring expectation. *Econometrica*, 72(5):1329 – 1376, 2004.

[69] Rosa L. Matzkin. Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica*, 60:239 – 270, 1992.

[70] Rosa L. Matzkin. Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics*, 58:137 – 168, 1993.

[71] Rosa L. Matzkin. Restrictions of economic theory in nonparametric methods. In R. Engle and D. McFadden, editors, *Handbook of Econometrics*, pages 2523 – 2558. North-Holland, New York, 1994.

[72] Rosa L. Matzkin. Nonparametric estimation of nonadditive random functions. *Econometrica*, 71(5):1339 – 1375, 2003.

[73] David Mendez and Kenneth E. Warner. Smoking prevalence in 2010: Why the healthy people goal is unattainable. *American Journal of Public Health*, 90:401 – 403, 2000.

[74] David Mendez and Kenneth E. Warner. Adult cigarette smoking prevalence: Declining as expected (not as desired). *American Journal of Public Health*, 94:251 – 252, 2004.

[75] Francesca Molinari and Chuck Manski. Rounding probabilistic expectations in surveys. Working Paper, 2008.

[76] John F. Muth. Rational expectations and the theory of price movements. *Econometrica*, 29(3):315 – 335, July 1961.

[77] Y. Nyarko and A. Schotter. An experimental study of belief learning using elicited beliefs. *Econometrica*, 70:971 – 1005, 2002.

[78] Gabriel Picone, Frank A. Sloan, and Donald Taylor. Effects of risk and time preference and expected longevity on demand for medical tests. 28:39 – 53, 2004.

[79] John Rust. Structural estimation of markov decision processes. In R. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, pages 3082 – 3139. North Holland, 1994.

[80] John Rust and Christopher Phelan. How social security and medicare affect retirement behavior in a world of incomplete markets. *Econometrica*, 65(4):781 – 831, 1997.

[81] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.

[82] F. A. Sloan, J. Ostermann, G. Picone, C. Conover, and D.H. Taylor. *The Price of Smoking*. MIT Press, 2004.

[83] Frank A. Sloan and Yang Wang. Economic theory and evidence on smoking behavior of adults. *Addiction*, 103:1777 – 1785, November 2008.

[84] V. Kerry Smith, Donald H. Taylor, and Frank A. Sloan. Longevity expectations and death: Can people predict their own demise? *American Economic Review*, 91(4):1126 – 1134, 2001.

[85] V. Kerry Smith, Donald H. Taylor, Frank A. Sloan, F. Reed Johnson, and William H. Desvousges. Do smokers respond to health shocks. *Review of Economics and Statistics*, 83(4):675 – 687, 2001.

[86] Donald H Taylor, Vic Hasselblad, S Jane Henley, Michael J Thun, and Frank A Sloan. Benefits of smoking cessation for longevity. *American Journal of Public Health*, 92(6):990 – 996, 2002.

[87] J. Townsend, P. Roderick, and J. Cooper. Cigarette smoking by socioeconomic group, sex, and age: Effects of price, income, and health publicity. *British Medical Journal*, 309:923 – 927, 1994.

[88] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124 – 1131, 1974.

[89] Eric Van den Steen. Rational overoptimism (and other biases). *The American Economic Review*, 94(4):1141 – 1151, September 2004.

[90] W. Van der Klaauw and K. Wolpin. Social security and the savings and retirement behavior of low income households. *Journal of Econometrics*, 145(1-2):21 – 42, 2008.

[91] W. K. Viscusi. A bayesian perspective on biases in risk perception. *Economics Letters*, 17:59 – 62, 1985.

[92] W. K. Viscusi. Do smokers underestimate risks? *Journal of Political Economy*, 98(6):1253 – 1269, 1990.

[93] W. K. Viscusi. Age variations in risk perceptions and smoking decisions. *The Review of Economics and Statistics*, 73(4):577 – 588, November 1991.

[94] W. Kip Viscusi and W. A. Magat. *Learning About Risk: Consumer and Worker Responses to Hazard Information*. Cambridge, MA: Harvard University Press, 1987.

[95] W. Kip Viscusi and Charles J. O'Conner. Adaptive responses to chemical labeling: Are workers bayesian decision makers? *American Economic Review*, 74:942 – 956, 1984.

[96] W.K. Viscusi and W. Evans. Utility functions that depend on health status: Estimation and economic implications. *American Economic Review*, 80:353 – 375, 1990.

[97] Jeffrey Wasserman, Willard G. Manning, Joseph P. Newhouse, and John D. Winkler. The effects of excise taxes and regulations on cigarette smoking. *Journal of Health Economics*, 10(1):43 – 64, May 1991.

Biography

Yang Wang was born in Dalian, Liaoning Province, P. R. China on May 16th, 1977. She earned her B.A. in International Relations from Beijing (Peking) University in May 1999 and her M.P.P in Public Policy from University of Wisconsin, Madison in May 2003. Yang began her graduate studies in Economics at Duke University in August 2003.