A Dynamic Model of Health Insurance Choices and Health Care Consumption[1]

Jian Ni

Johns Hopkins University

Email: jni@jhu.edu

Nitin Mehta

University of Toronto

Email: nmehta@rotman.utoronto.ca

Kannan Srinivasan

Carnegie Mellon University

Email: kannans@cmu.edu

Baohong Sun

Cheung Kong Graduate School of Business

Email: bhsun@ckgsb.edu.cn.

January 2014

[1] This paper is based on Chapter 2 of Jian Ni's Ph.D. dissertation at Carnegie Mellon University. This paper was funded by the Andrew Mellon doctoral fellowship to Jian Ni and SSHRC grant # 410-2011-0189 to Nitin Mehta and Jian Ni. The authors are grateful to Sunder Kekre for helping us obtain the data and to George-Levi Gayle, Hanming Fang and Avi Goldfarb for helpful comments. The authors also appreciate feedback from seminar participants at Emory, Georgia Tech., Johns Hopkins, HKUST, Minnesota, National University of Singapore, NYU, University of Southern California, Washington University and Yale University. The usual disclaimer applies.

A Dynamic Model of Health Insurance Choices and Health Care Consumption

Abstract

The increase in healthcare costs over the last few decades is particularly relevant for chronic diseases, which account for 75% of healthcare expenditure. Individuals suffering from chronic diseases can consume three types of services: secondary-preventive care which includes diagnostic tests, primary-preventive care which consists of drugs that help prevent the illness from getting worse, and curative care which includes surgeries and more expensive drugs that provide quantum boost to the consumer's health. Although majority of cases can be managed by preventive care, most consumers opt for more expensive curative care that leads to substantial increase in overall costs. To examine these inefficiencies, we build a dynamic model of consumers' annual insurance plan decisions and periodic consumption decisions, and apply it to a panel dataset on these two consumers' decisions. Our results indicate that there exists a sizable segment of consumers who purchase more comprehensive plans than needed because of high uncertainty vis-à-vis their health status; and once in the plan, they opt for curative care even when their illness could be managed through preventive care. We examine how changing cost sharing characteristics of insurance plans and providing more accurate information to consumers via secondary-preventive care can reduce these inefficiencies.

Keywords: Insurance Choice, Healthcare Service, Nested Dynamic Decisions, Primary and Secondary Preventive Care, Curative Care, Informative Effect, Investment Effect, Dynamic Programming

Section 1: Introduction

Healthcare is one of the most important and personal services that consumers buy and it has a pervasive impact on the quality of daily life and the economy. In the U.S., insurance firms are the major health insurance providers. They cover 58.5% of the population through employer-sponsored health insurance, followed by Medicare and Medicaid which cover 29.0% of the population (Economic Policy Institute 2009). During the past decade, health care costs incurred by these providers have more than doubled, resulting in a price tag of $2.6 trillion in 2008 that constitutes 17% of U.S. GDP (New York Times 2009). Such an increase has resulted in insurance firms passing on the costs to consumers, which has caused 119% increase in premiums over the past 10 years. These skyrocketing premiums have not only imposed huge burden on the consumer's wallet, but has also led to a dramatic increase in the number of people who cannot pay such hefty premiums (Price Water House Coopers 2008). Consequently many people are left either uninsured or covered by government-run programs, which has in turn caused an enormous increase in government spending.

Consumer choices and consumption has been the cornerstone of research in marketing for the past several decades. They have examined a number of categories such as ketchup, diapers, soda, yogurt, coffee to name a few (Given the enormous number of studies, we omit citations). Surprisingly, when one shifts away from product choices to services, the literature in marketing becomes incredibly sparse. Consumers' choices of mortgages, life insurance and mutual funds are important and often dwarf the consumer packaged goods that has been so extensively studied. Similarly, we are unaware of any research in marketing focused on purchase and consumption of health insurance and healthcare services, a market that is so large that only a handful of economies in the world exceed its size.

Research in marketing tends to examine the interdependency between choice and consumption. For example, there are studies that examine whether or not purchase of a larger size increases consumption (for example, see Sun 2005). Yet the inter-relationship is far more critical in the area of healthcare. Consumers expectation of health status in future will drive their insurance plan choices and given those choices, their consumption behavior of healthcare services. Thus, this category offers excellent scope for understanding this aspect that has been the cynosure of research in marketing for a longtime.

The need to understand healthcare consumption has assumed greater urgency due to the explosive growth in cost. The increase in costs is particularly relevant in the context of chronic diseases such as heart disease, cancer, diabetes and hypertension, which afflict more than 133 million people, and account for 75% of the overall health care costs in U.S (New York Times, 2009). See Table 1A for some typical examples of chronic diseases. Chronic diseases differ from acute diseases[2] and other ailments in two respects: first, unlike other ailments, individuals suffering from chronic diseases require health care over a much longer time horizon; second, unlike other ailments, individuals suffering from chronic diseases can take two types of healthcare services to manage their illnesses: preventive care and curative care (Kenkel 2000).

Preventive care as defined by Center for Disease Control and Prevention constitutes two types of services: primary prevention and secondary prevention. Primary prevention refers to services that include certain types of drugs, which help prevent the disease from getting worse so that the consumers can lead a normal life. Secondary prevention refers to services that help detect the illness and the extent of its severity. These include diagnostic tests, screening and routine physical exams. Curative care refers to treatments such as surgeries and certain types of drugs that help improve symptoms or cure the medical problems. See Table 1B for some typical examples of services and treatments that fall under primary preventive, secondary preventative and curative care. Although primary preventive and curative cares serve a similar purpose, they differ in two respects. First, curative care is more expensive and provides a greater boost to consumer's health compared to primary prevention. Second, primary prevention is useful only when the illness is detected at an early stage and its severity ranges from low to moderate. On the other hand, curative care is useful when the severity ranges from moderate to high.

The distinction between preventive care and curative care is crucial in the context of chronic diseases since according to experts, a vast majority of cases in chronic diseases can be effectively managed by preventive care (Grossman & Rand 1974; Thorpe 2008). An analysis from RAND (Manning et al. 1987) claimed that a better management of chronic diseases through preventive care could result in twenty million fewer impatient days and nine million fewer office visits based on only four chronic diseases. However the reality is that more than 96% of overall health care expenditure for chronic diseases goes to curative care (Thorpe 2008). For instance, consider prostate cancer which inflicts around 200,000 men every year in the U.S. (Wall Street Journal 2010). An estimated 85% of tumors grow so slowly that they will never cause problems as long as they are effectively managed through preventive care, which mainly includes taking primary preventive care drugs such as statins that control the spread of prostate cancer (along with using secondary preventive care on a

[2] Acute disease refers to disease or disorder that lasts a short time, inflicts rapidly and is accompanied by distinct symptoms such as fractures

periodic basis such as PSA tests and digital rectal exams). However less than 10% of consumers opt to go for preventive care; instead, they undergo expensive curative treatments such as surgeries and radiation therapy, which substantially increase the overall healthcare costs.

It follows from the above discussion that there could be inefficiencies in managing chronic diseases: instead of seeking less expensive primary preventive care, consumers suffering from chronic diseases opt for the more expensive curative care that leads to disproportionally higher healthcare costs. In this paper we seek to investigate one of the reasons for this inefficiency, which is moral hazard. Recall from our earlier discussion that consumers suffering from chronic diseases with moderate levels of severity can be treated by either curative care or primary preventive care – where curative care, although more expensive, provides a greater boost to the consumer's health as compared to primary preventive care. Thus, if the cost of treatment to the consumers (i.e., the out of pocket expenses incurred by consumers when seeking the treatment) was not an issue, such consumers will always opt for curative care. Now since consumers obtain healthcare services through insurance plans, it follows that the more comprehensive the consumer's insurance plan, the lower will be the share of the healthcare expense incurred by the consumer, and consequently the greater will be the incentive for such a consumer to opt for curative care. In summary, moral hazard arises when consumers suffering from chronic diseases with moderate levels of severity purchase comprehensive insurance plans in which a large part of the cost of treatment is shared by the insurance firm; as a result, these consumers opt for the more expensive curative care to manage their illness that could have been managed by primary preventive care.

This preamble leads us to the goal of our paper, which is to examine in the context of chronic diseases, whether and to what extent does moral hazard play a role in creating inefficiencies and the types of policies that can be employed to reduce these inefficiencies. To achieve this goal, we have three important objectives.

1.1 Research Objectives and Main Findings

The *first* objective is to model how consumers make annual insurance plan decisions and periodic healthcare consumption decisions conditional on their chosen insurance plans. The insurance plan decision entails choosing between different types of health insurance plans that differ in terms of annual premiums and cost sharing characteristics. And the healthcare consumption decision entails choosing between primary preventive, secondary preventive, curative cares or the no-consumption option in each period in a policy year. We model the two decisions in a nested dynamic framework with risk-averse and forward looking consumers who are uncertain about their health status.

We model the healthcare consumption decision as a tradeoff between the extent to which curative, primary preventive and secondary preventive cares impact the consumer's health status and the out of pocket expenses incurred by the consumer while consuming each of these services. We model the impact of curative, primary and secondary preventive cares on consumer's health status via two mechanisms: investment and informative effects (Arrow 1963; Grossman 1972). The investment effect refers to the increase in consumer's health after consuming the healthcare service and informative effect refers to the information that the healthcare service provides to the consumer about her current health status that enables her to judiciously decide on future course of treatment. And regarding the annual insurance plan decision, we model it as a tradeoff between the annual premium charged in the insurance plan and consumer's expected future health care consumption – where the health care consumption, as discussed before, depends on the consumer's health status, consumer's uncertainty in her health status and the cost sharing characteristics of the insurance plan.

The *second* objective is to estimate our model on panel data that consists of consumers' insurance plan and consumption choices over time. The panel data has detailed information on insurance plans available to each consumer each year, characteristics of each plan, plan chosen by each consumer each year, healthcare consumption decisions on a weekly level, and the cost incurred by the consumer and insurance company for each consumption decision. Our key empirical results are as follows. *A.* Both preventive and curative cares have significant impact on consumer's health status. Specifically, the informative effect of secondary preventive care is significant; and while both primary preventive and curative cares have a significant investment effect, the investment effect is much higher for curative care. *B.* The variation in insurance plan decisions across consumers stems more from heterogeneity in their health status, their uncertainty vis-à-vis their health status and degradation in their health status than the heterogeneity in their price sensitivities. *C.* The usage of both primary and secondary preventive cares is affected most by changes in copayment, followed by changes in the deductible and coinsurance rate, while the opposite is true for curative care. *D.* moral hazard plays a crucial role in creating inefficiencies. Our results indicate that the underlying reason for inefficiencies is that there exists a sizable segment of 'problem' consumers (13% of consumers) who have moderate health status, high uncertainty about their health status and low price sensitivities. Because of high uncertainty, these consumers choose the comprehensive plan as opposed to the medium plan that would have been a better match for their moderate health status; and once in the plan, they opt for more expensive curative care instead of primary preventive care.

The *third* objective is to investigate how the inefficiencies can be reduced. We examine two routes. The first is the 'immediate route' in which we change cost sharing characteristics of insurance plans at the beginning of the policy year, which would incentivize 'problem' consumers to choose the medium plan in that policy year itself. The second is the 'delayed route' in which we incentivize the 'problem' consumers to consume more secondary preventive care in the first year (and onwards). Doing so will decrease their uncertainty and will induce them to choose medium plans in future. Our results indicate that the immediate route is not feasible since these consumers have low price elasticities in insurance plan decisions. We conduct two counterfactuals in the context of the delayed route in which we examine the impact of the following on health care costs: (a) changing the cost sharing characteristics that encourage use of secondary preventive care, and (b) providing more accurate information through secondary preventive care. In both counterfactuals, we find that health care costs reduce substantially as a result of 'problem' consumers choosing medium plans in future.

1.2 Related Literature

The last five years have seen a growing interest in broad area of healthcare economics with topics ranging from studying the impact of labor market frictions on healthcare expenditures (e.g. Fang and Gavazza 2011), testing for bounded rationality in the healthcare market (e.g. Handel and Kolstad 2013; Fang et al. 2013) to identifying asymmetric information and separating selection from moral hazard. Given the wide range of topics, we restrict our attention to only those papers that have reasonable relevance to our work, and do not discuss the papers that are only tangentially related.

Prior empirical papers that have modeled consumers' healthcare consumption decisions have used the Grossman's health investment framework (Grossman 1972; Grossman and Rand 1974). In this framework, the consumer's utility is modeled as a function of her health status and the out of pocket expenses for consuming a healthcare service. The consumer's health status is assumed to degrade over time, and the consumption of a health care service serves to increase the consumer's health status via the investment effect only. Most empirical papers that have modeled healthcare consumption decisions have treated the health insurance decisions as exogenous. A few empirical papers that have endogenized the insurance plan decisions are Cameron et al. (1988), Cardon and Hendel (2001), Carlin and Town (2008), Bajari et al. (2012), Einav et al. (2013) and Handel (2013). Unlike our paper, they did not model consumption decisions in terms of whether or not consumers should seek preventive or curative care in each period in the policy year; instead, they modeled the consumption decision as a one shot decision, namely, how much money to spend on healthcare services in a policy year. Further, these papers only modeled the investment effects of consumption options, and not the informative effects. Finally, these papers focused on existence of asymmetric information and separating selection from moral hazard. In other words, their objective was to demonstrate the existence of moral hazard by showing that conditional on consumer's health status, the choice of her insurance plan has a significant impact on her annual healthcare expenditure[3].

Our paper builds on these papers by modeling the underlying decision making process of consumers that connects the insurance plan decision to their annual healthcare expenditures. We do so by modeling consumers' periodic consumption decisions within a policy year in terms of choosing between primary preventive, curative and secondary preventive cares. Doing so is important since it allows us to examine as to how changes in policies can influence the extent of moral hazard. We further build on these papers by modeling the informative effects of consumption options which play a crucial role in managing chronic diseases. It allows us to examine as to how policies that help provide cheaper and easier access to information on one's health can reduce the extent of moral hazard. Apart from these differences, our paper differs from the prior papers on two other grounds. *First*, the data used in the prior papers is from a single employer only. In our paper, the data is from multiple employers in which the specifics of insurance plans (premium and cost sharing characteristics) differ across employers. As argued in Cardon and Hendel (2001), this choice set heterogeneity allows us to do a better job of separating selection from moral hazard. *Second*, these papers used cross sectional data which covered the insurance plan and consumption decisions for one policy year only[4]. On the other hand, we use panel data that covers the decisions over multiple years. This has two important implications. Unlike the prior literature, (a) we are able to do a much better job in identifying the unobserved heterogeneity, which as we will discuss later, plays a crucial role in identifying the extent of moral hazard, and (b) we can examine the impact of consumption decisions in one year on insurance plan and consumption decisions in the following year. Point (b) is crucial, since as we discussed before, one of our key findings is that it is the delayed route and not the immediate route that helps decrease the extent of moral hazard. Note that the delayed route can only be investigated if we model the decisions over multiple years.

[3] The exception here is Handel (2013) who assumed away moral hazard and focused on identifying switching costs.

[4] The exception here is the working paper by Khwaja (2001) who proposed a dynamic life-cycle model to study consumers' insurance and healthcare expenditure decisions over time.

Our paper is also related to empirical contract literature starting with Chiappori and Salanie's (2000) positive correlation test of asymmetric information. This test estimates the correlation between insurance coverage and insurance utilization implied by the consumer's risk types. However the difficulty of separating selection from moral hazard leads to mixed evidence with some empirical studies finding strong support for adverse selection while others find weak support (Fang, Keane and Silverman 2006). Finkelstein and McGarry (2006) and Cutler, Finkelstein and McGarry 2008 resolved these mixed findings by positing two dimensions of heterogeneity instead of just one: heterogeneity in risk preferences and heterogeneity in risk type. If individuals differ in more than one dimension, i.e., in their risk type as well as in their risk preferences, then it is possible that these two types of private information offset each other to produce an equilibrium with no correlation between insurance coverage and utilization. Therefore, it is important to take into account heterogeneity in both the risk type and in the risk preference, in order to separate out moral hazard from selection. This finding is crucial since the objective of our paper is to identify the inefficiencies that result from moral hazard. In our model, we incorporate heterogeneity in the risk type and the risk preference by allowing for consumers to differ in terms of their health status and their uncertainty toward their health status.

Section 2: Data

We use a proprietary dataset provided by an anonymous health insurer. This insurer offers employer sponsored Preferred Provider Organization (PPO) plans to consumers through their employers typical of the American health care system. Around 87% of the insurer's group business is in a PPO product category. In our data, consumers come from different employers, but their health plans are offered by the same single insurer from 2005 to 2007.[5] Across these three years, the insurer offered consumers three plans labeled as basic, medium and comprehensive. All consumers in the data are enrolled in one of the three plans. All three plans have the same hospital/physician network, same covered medical services, and the same contractual agreement with healthcare providers (such as discounts for services provided). The three plans differ in terms of cost sharing characteristics and annual premiums which determine the mapping from total medical expenditure from a given treatment to out-of-pocket expense incurred by the consumer. The characteristics of the three plans

[5] Although we do not observe employees' health insurance options outside the firm, according to the Kaiser Annual Survey of Employer Health Benefits, 82% of eligible workers enroll in plans offered by their own employers. This is also verified by the insurer who provided us the data.

vary annually and also across consumers who choose same type of plan. This is because even though the insurer offered a total of three plans to employers each year, the insurer customized these three plans for different employers. Therefore, any two individuals across different employers who choose the same type of a plan in the data would face different premium and cost sharing characteristics.

The cost sharing characteristics of a plan include copayment, deductible, coinsurance rate and out-of-pocket-maximum. Copayment is the specified dollar amount that a consumer must pay out of her own pocket for a specified doctor visit at the time the service is rendered. Deductible is the flat amount that an individual must pay before the insurer will make any benefit payments. Coinsurance rate is the percentage of all remaining eligible health care expenses after the deductible amount has been paid which is a method of cost-sharing in the health insurance setting. Out-of-pocket-maximum is the dollar amount set by the insurer that limits the amount an individual has to pay out of her own pocket for particular healthcare services during a particular time period. See Figure 1 for an illustration of pricing structure and the empirical cost sharing from our data.

We observe every consumer's annual health insurance plan choice from 2005 to 2007 and her detailed claim history over the three years. For each consumer, we have information on demographics and the annual premiums and cost sharing characteristics of each insurance plan in each consumer's choice set. And for each claim, we have information on when the specific healthcare service was consumed, the diagnostic, procedure and therapeutic codes for that specific service and the payment information such as copayment, deductible paid, coinsurance paid, insurer paid and total billed charged. The typical chronic diseases in our data are heart disease, cancer, hypertension, chronic respiratory diseases, diabetes, Alzheimer's disease and kidney disease. We classify consumers' health care consumption services into three types: (i) secondary preventive care, which includes diagnostic tests such as mammograms and Prostate-Specific Antigen (PSA) tests, (ii) primary preventive care such as statins (Lipitor) used for lowering blood cholesterol, and (iii) curative care which includes surgeries such as coronary angioplasty and coronary stent placement and curative drugs. The classification of different treatments into primary preventive, secondary preventive and curative care is based on the definition of American Medical Association (International Classification Ninth Revision – Clinical Modification 2008 and Current Procedural Terminology 2008) and the internal manual provided by the insurance company.

Table 2A reports the number of different treatments that fall under a given consumption option (i.e., secondary preventive, primary preventive and curative cares) for three common chronic diseases. Observe that for each disease, there are multiple treatments within a given consumption option. For instance, there are five different treatments that fall under secondary preventive care for type-2 diabetes. Thus to keep the model tractable, for each disease, we aggregate all treatments that fall under a given consumption option into a single consumption option with a single aggregate cost. The procedure of aggregating the costs is explained in Appendix I of the paper, and is similar to what prior papers have done to operationalize the aggregate price of a category or brand composite when using scanner data sets. Essentially, the aggregate cost of a consumption option for a given disease is calculated as a weighted average cost of all treatments that fall under that consumption option and disease – where the weights are based on frequencies of usage of the different treatments.

To see whether or not this would lead to large aggregation biases, we computed the weighted average cost (and the std. dev. of the costs) of different treatments that fall under each consumption option for the same three diseases. Note that if the standard deviations of the costs are high, it will imply that there is a large variation in costs of treatments that fall under a given consumption option and disease, which will in turn imply that aggregating those costs may lead to large aggregation biases. The weighted average cost (and the corresponding std. dev.) of a given consumption option and disease is calculated based on the costs recorded in our data for all those observations in which the consumers consumed either of the treatments that fall under that given consumption option and disease. The weighted average costs and standard deviations of costs are reported in Table 2A. Observe in Table 2A that standard deviations of all costs are an order of magnitude lower than the average costs. This implies that using a weighted average measure to aggregate the costs across treatments within a given consumption option and disease, although not perfect, is reasonable.

Moving on to the other aspects, around 95% of consumers in our data consume at most one type of service in a given week; and in our estimation sample, this number is 100%. Thus in our model, we assume that consumers make discrete choices of healthcare consumption (whether to choose primary preventive care, secondary preventive care, curative care or neither) on a weekly basis. Further, we only model the consumption decisions that are related to chronic diseases. We do not model consumer's visits that are not related to chronic diseases for two reasons. First, the fraction of such visits (=5.3%) is as such small and the expenditure on these visits is even smaller. Second, as we discussed in Section 1, the consumer's consumption choice set for non-chronic diseases does not consist of primary preventive, secondary preventive and curative care. Thus, the modeling framework that we propose in this paper does not apply to non-chronic diseases.

We focus on individual insurance policy holders who are continuously enrolled for the entire period. In our estimation sample, we only consider those consumers who have purchased individual plans and not family plans. The reason for that is because in our modeling framework, we model the insurance plan and consumption decisions based on the health status of a single individual. This leaves us with 2,833 insurance contract holders for estimation. In our estimation sample, 92% of consumers have taken preventive or curative care prescribed for chronic diseases, and 34.2% of consumers have taken preventive or curative care for more than one chronic disease.

This completes the discussion on the construction of the estimation sample. The dependent variables are the consumers' annual insurance plan choices (i.e., whether to choose basic, medium or comprehensive plan) and their weekly consumption choices (i.e., whether to choose curative care, primary preventive care, secondary preventive care or neither). And the independent variables are the annual premiums and cost sharing characteristics for each insurance plan in the consumer's choice set, and the actual price paid by each consumer for each option. From the total cost for each option and the cost sharing characteristics of the chosen plan, we construct the actual price paid by each consumer for each option. The procedure for doing so is explained later in Section 3.

In our sample, 2.9% of consumers did not purchase any of the three consumption options, 52.8% purchased all three consumption options at least once, and 79.6% purchased secondary preventive care (i.e., diagnostic tests) at least once during their purchase history. And between primary preventive care and curative care, 14.3% of consumers did not use curative care but used primary preventive care at least once, 20.1% did not use primary preventive care but used curative care at least once, and 58.9% used both primary preventive and curative cares at least once during their entire purchase history. Amongst the consumers who have used secondary preventive care at least once, the average number of secondary preventive care visits per consumer is 5.7, and amongst the consumers who have used primary preventive care (curative care) at least once, the average number of primary preventive care (curative care) visits per consumer is 40.8 (11.1). And amongst the 58.9% of consumers who have used both primary preventive and curative cares, the average number of primary preventive care (curative care) visits per consumer is 44.7 (12.3).

Tables 2B and 2C present consumers' insurance plan choices and plan switching behavior over the three years. On an average, 43% of the consumers chose comprehensive plans, 22% chose basic plans and the rest chose the medium plans; and on an average, 7.5% consumers change their insurance plans each year. Table 2D provides the summary statistics of health insurance and claim information across the three types of plans. The premium that consumers pay for the basic plan is only one third of the premium for the comprehensive plan; and the coinsurance rate, deductible and out-of-pocket maximum decrease as we move from the basic plan to the comprehensive plan. On an average, people within comprehensive plan seek health care 22 times a year, which is much higher than that for the consumers enrolled in basic plan. As we move from the basic plan to the comprehensive plan, the number of visits per consumer per year (and healthcare cost per consumer per year) increases for each of the three consumption options; however, this increase is the steepest for curative care and lowest for primary preventive care.

Section 3: Model

We propose a dynamic model of consumers' healthcare decisions in which their periodic (weekly) healthcare consumption decisions are nested within their annual insurance plan purchase decisions. Consumers decide the type of health insurance plans on an annual basis based the premium charged and their expected healthcare consumption in the future. Conditional on the health insurance plan, they then decide their periodic consumption decisions which include taking primary preventive care, secondary preventive care, curative care or the no consumption option. In our analysis, we assume that the physician is a perfect agent and acts in the best interest of the patient. In other words, we assume that the healthcare consumption decisions are made by the consumer based on her beliefs about her health status, and her out of pocket expenses she would incur while consuming the different options. This is a standard assumption in the healthcare literature (Felder and Mayrhofer 2011) and a reasonable one in our context since the level of informational asymmetry between doctor and patient will be small in the context of chronic diseases (Vera-Hernandez 2003).

3.1 Insurance Plan and Health Care Consumption Decisions

Consider i=1.. I consumers who make annual insurance choices from a set of j=1..J health insurance plans at year a=1..A. In our data, the number of available insurance plans for each consumer are J=3, where j=1 indicates the basic plan, j=2 the medium plan, and j=3 the comprehensive plan. The characteristics of the three plans differ in terms of annual premiums, deductibles, co-insurance rates and out of pocket maximums. Since all consumers are enrolled in one of the three plans over the entire period of the data, we only model the insurance plan decisions in terms of which of the three plans to purchase (i.e., we do not allow for a no-purchase option or an outside option in insurance plan decisions). We use d_{ija} to denote consumer's choice of insurance plan j at year a as:

$$d_{ija} = \begin{cases} 1, \text{ if consumer } i \text{ chooses plan } j \text{ at the beginning of year } a \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

Conditional on the insurance plan in year a, the consumer makes her health care consumption decision k in each period (i.e., week) t=1..T. The consumption decision entails choosing either primary preventive care pp, secondary preventive care sp, curative care c or no-consumption option, no. Let another variable c_{ikt} represent the health care consumption decision, where $k \in \{pp, sp, c, no\}$.

$$c_{ikt} = \begin{cases} 1, & \text{if consumer } i \text{ chooses health care service } k \text{ at week } t \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

In a given period, these alternatives are mutually exclusive such that $\sum_k c_{ikt} = 1$. See Figure 2 for an illustration of the time line of the two decisions. In what follows, we discuss the factors that influence the two decisions: consumer's health status, investment and informative effects of the consumption options and the insurance plan's pricing components. For notational ease, we will drop the subscript of i for the consumer in the remainder of this section. We will introduce the subscript for the consumer when we discuss heterogeneity in section 4.

3.2 Health Status and Degradation

The consumer's health status is a key factor that influences her health care consumption and health insurance decisions. Following Grossman's (1972) health production framework, health status is treated as a (latent) human capital stock. Let H_t be the consumer's true health status at the end of period t. Unlike the quality learning literature (Erdem and Keane 1996; Erdem et al. 2005; Zhang 2010; Lin et al. 2014) in which the true product quality is time-invariant, consumers' true health status can evolve over time due to degradation and investment effects of medical interventions such as curative or preventive care. Starting with degradation, the medical literature has established that in absence of any medical intervention, the health status of individuals deteriorates over time because of aging, and more so if they are suffering from chronic diseases (Kenkel 2000). We thus model the degradation in consumer's true health status over time in absence of investment effects as follows:

$$H_t = H_{t-1} - \delta_d \tag{3}$$

where δ_d represents the degradation rate, which is the deterioration in the consumer's true health status over each period. Similar to the prior literature, δ_d is assumed to be known to the consumer.

3.3 Investment Effects

Consumers seek health care to improve their health status. Following Grossman's health production framework, we assume that a healthcare service provides a quantum boost to the consumer's health status. We thus model the evolution of consumer's true health status in the presence of degradation and investment effects as follows

$$H_t = H_{t-1} - \delta_d + \sum_k \delta_k c_{kt} \tag{4}$$

where δ_k is the investment effect of option $k \in \{pp, sp, c, no\}$ consumed in period t. Similar to prior literature, δ_k is assumed to be known to the consumer. We allow for investment effects to differ across consumption options, and we set the investment effect of no-consumption option to zero.

3.4 Informative Effects

Since consumers are not well informed about their true health status, they would have uncertainty about it (Arrow 1963; Hsieh & Lin 1997). Therefore their insurance plan and consumption decisions will be based on their beliefs of their true health status. Consumers can learn about their true health status via the informative effects of the consumption options. To model these informative effects, we assume that consumption options provide noisy signals to the consumers about their true health status, and consumers learn about their true health status from these signals in a Bayesian fashion. To understand the evolution of consumers' beliefs about their health status from the end of period t-1 to the end of period t, see Figure 3 which illustrates the sequence of consumer's periodic degradation, investment and informative effects. In this sequence, we start with the consumer's beliefs about her true health status at the beginning of period t (or end of period t-1) as

$$H_{t-1} \sim N\left(h_{t-1}, \sigma_{t-1}^2\right) \tag{5}$$

Following that, we update the priors in equation (5) based on degradation of the health status in period t and the investment effects from consumption of option k in period t. Since the degradation rate and investment effects are known to the consumer, they will only impact the mean and not the variance of the consumer's health status beliefs. Thus by the end of this step, we get the consumer's beliefs using equations (4) and (5) as

$$H_t^{\Delta} \sim N\left(h_{t-1} - \delta_d + \delta_k c_{kt}, \sigma_{t-1}^2\right) \tag{6}$$

In the next step, we update the priors in equation (6) based on the informative effects from the consumption of option k in period t. [6]Let λ_{kt} be the noisy consumption signal that the consumer receives about her true health status from the consumption of option k in period t. This is given as

$$\lambda_{kt} = H_t + \sigma_k \eta_{kt} \tag{7}$$

[6] It does not matter whether we model the investment or the informative effects in the first step. Either way, we will get the same evolution of the consumer's beliefs of her health status.

The first term on the RHS of equation (7), H_t, represents the consumers true health status at the end of period *t* after she has experienced degradation in her health in period *t* and the investment effect of the option consumed in period *t*. The second term on RHS of equation (7), $\sigma_k\eta_{kt}$, represents the signal noise of option *k*, where η_{kt} is a standard normal random variable and σ_k is the standard deviation of the signal noise which is a measure of the informational inaccuracy of healthcare service *k*. If $\sigma_k=0$, it implies that the healthcare option *k* provides perfect information to the consumer about her heath status. And if $\sigma_k=\infty$, it implies that healthcare option *k* does not provide any information to the consumer about her heath status. Substituting the expression for H_t given in equation (4) into equation (7), we get the following expression for λ_{kt} in terms of the consumer's true health status at the beginning of period *t*:

$$\lambda_{kt} = H_{t-1} - \delta_d + \delta_k c_{kt} + \sigma_k \eta_{kt} \tag{8}$$

Given the prior in equation (6) and the signal in equation (8), the consumer's posterior beliefs at the end of period *t* will be distributed as $H_t \sim N(h_t, \sigma_t^2)$ where

$$h_t = \frac{\dfrac{h_{t-1} - \delta_d + \sum_k \delta_k c_{kt}}{\sigma_{t-1}^2} + \sum_k \dfrac{c_{kt}\lambda_{kt}}{\sigma_k^2}}{\dfrac{1}{\sigma_{t-1}^2} + \sum_k \dfrac{c_{kt}}{\sigma_k^2}} \tag{9a}$$

$$\frac{1}{\sigma_t^2} = \frac{1}{\sigma_{t-1}^2} + \sum_k \frac{c_{kt}}{\sigma_k^2} \tag{9b}$$

Equations (9a) and (9b) represent the evolution of the mean and variance of health status from the consumer's perspective. To complete the specification of this evolution, we represent the consumer's prior beliefs at the beginning of period *t*=1 in year *a*=1 by $H_0 \sim N(h_0, \sigma_0^2)$. We assume that at the beginning of {*a*=1, *t*=1}, the consumer has rational expectations about her health status (Crawford and Shum 2005), i.e., the mean and variance of her prior beliefs, h_0 and σ_0^2, are also the mean and the variance of the distribution of the true health status across the consumer population at the beginning of {*a*=1, *t*=1}.

3.5 Out-of-Pocket Expenses

The pricing structure of the health plans offered by the insurer affects consumers' health insurance and consumption decisions through the out of pocket expenses incurred by the consumer when purchasing the insurance plan and when consuming the healthcare services. If a consumer chooses a plan *j*, her out of pocket expense will be the premium ($Prem_{j,a}$) that she pays at the beginning of year *a*. Once enrolled in plan *j*, her out of pocket expense when choosing a health care consumption option will depend on the actual price of the chosen option and the cost-sharing features of the plan.

In what follows, we discuss the construction of consumer's out of pocket expenses for consumption decisions. Let p_{k,j,t_a} be the consumer's out of pocket expense if she were to consume healthcare option *k* in period *t* of year *a* if she were enrolled in insurance plan *j*. We represent p_{k,j,t_a} as $p_{k,j,t_a} = p'_{k,j,t_a} + cop_{k,j,a}$ where p'_{k,j,t_a} is consumer *i*'s marginal effective price for option *k* in period *t*, and $cop_{k,j,a}$ is the copayment amount incurred by consumer *i* for using option *k* in year *a* if she were enrolled in plan *j*. The consumer's marginal effective price is given by

$$p'_{k,j,t_a} = \begin{cases} cost_{k,j,t_a} \ \text{ if } \ cost_{k,j,t_a} \leq r_{j,t_a} \text{ and } \sum_{\tau=1}^{t_a-1} p_{k,\tau} < opm_{j,a} \\ r_{j,t_a} + \left(cost_{k,j,t_a} - r_{j,t_a}\right) ci_{k,j,a} \ \text{ if } \ cost_{k,j,t_a} > r_{j,t_a} > 0 \text{ and } \sum_{\tau=1}^{t_a-1} p_{k,\tau} \geq opm_{j,a} \\ 0 \ \text{ if } \ \sum_{\tau=1}^{t_a-1} p_{k,\tau} \geq opm_{j,a} \end{cases} \tag{10}$$

where $opm_{j,a}$ denotes the annual out-of-pocket maximum of plan *j* chosen by consumer i in year *a* and $ci_{k,j,a}$ is the coinsurance rate faced by consumer i in year *a* for option *k*. Further, $cost_{k,j,t_a}$ is the overall cost (the sum of healthcare costs incurred by the consumer and insurer) of healthcare option k in period t of year *a*, whose operationalization is discussed in Appendix I. Finally, r_{j,t_a} is the dollar amount remaining in the consumer's annual deductible in period *t*, which is defined as

$$r_{j,t_a} = \begin{cases} ded_{j,a} \quad \text{ if } \ t = 1 \\ \max\left\{0, r_{j,t-1_a} - p_{k,j,t-1_a} c_{k,t-1_a}\right\} \ \text{ if } \ t = 2..T_a \end{cases} \tag{11}$$

In equation (11), T_a denotes the last period of year *a* and $ded_{j,a}$ denotes the total deductible amount paid by consumer *i* for the chosen plan *j* in year *a* before the insurer makes any benefit payments. Although we do not model the hospital/doctor visits related to non-chronic diseases, we take into account the expenditures incurred by the consumer on such visits when calculating r_{j,t_a} in equation (11). For tractability, we assume that consumers know the actual costs of treatments in each period. This assumption is a reasonable one in our context since the cost information for chronic diseases is more widely available compared to other diseases.

Notice in equations (10) and (11) the dynamics in prices induced by the deductible: as long as the deductible is not met, the cost of treatment in each period influences the out-of-pocket expenses in that period which in turn influences the dollars remaining in the deductible in the future time period, and thereby influences the out of pocket expenses in the future period till the out of pocket maximum is met (Aron-Dine et al. 2012). However, once the deductible is met, the cost of treatment has no impact on future out of pocket expenses, since the out of pocket expenses in any period will simply be the cost of the treatment in that period scaled by the coinsurance rate.

3.6 Per period (Weekly) Consumption Utility

We define the consumer's period t utility for choosing the healthcare option $k \in \{pp, sp, c, no\}$ as a function of her out of pocket expenses p_{k,j,t_a} and her true health status H_t that would result if she were to consume that option k in period t. We assume that the utility satisfies the basic property of concavity with respect to the health status (i.e., $u''(H_t)<0$) which is synonymous with risk-aversion. This is a standard assumption in health economics literature (Felder and Mayrhofer 2011) and is needed in order for the model to conform to consumer behavior observed in the healthcare market (more on this later). We consider the CARA functional form for consumption utility for choosing option k, which satisfies this assumption[7] (Chan and Hamilton 2006), and is given as

$$u_{kt} = -\exp(-rH_t) - \alpha p_{k,j,t_a} + \varepsilon_{k,t} \qquad (12)$$

In equation (12), r (>0) is the consumer's degree of absolute risk aversion, α is the consumer's price sensitivity and ε_{kt} is the econometrician's error that is assumed to be IID Type 1 extreme value distributed. It captures the unobserved factors (which are observed by the consumer in period t, but not by the econometrician) that can influence consumer's consumption decisions, such as health shocks that are not captured by degradation/investment/informative effects. For each consumption option, the consumer would face different out of pocket expenses, p_{k,j,t_a}, and a different evolution of her health status in the future. The consumer's information set when making a consumption decision at the beginning of period t consists of (i) her beliefs of her health status at the end of period t-1, which are distributed as $H_{t-1} \sim N(b_{t-1}, \sigma_{t-1}^2)$) (ii) her out of pocket expenses for all consumption options in period t and onwards, (iii) and unobserved factors that influence her

[7] Note that our objective in the paper is not to test whether or not consumers are risk averse; we a priori assume that consumers are risk averse. This can be seen in equation (12) in which the parameter r is restricted to be positive, which necessarily implies that consumers are risk averse (the reason why r is restricted to be positive is because a value of r<0 implies that the consumer's utility is decreasing in her health status, which is incorrect).

consumption decision in period t, $\{\varepsilon_{kt}\}$. Conditional on this information set, her period t expected utility for choosing option k follows from equation (12) as

$$E_t u_{kt} = -E\left(\exp\left(-rh_t + \frac{r^2\sigma_t^2}{2}\right)\middle| h_{t-1}, \sigma_{t-1}, c_{kt} = 1\right) - \alpha p_{k,j,t_a} + \varepsilon_{k,t} \tag{13}$$

In equation (13), the term $r^2\sigma_t^2/2$ denotes consumer's the risk premium in period t, in which the specification of σ_t^2 conditional on $\{h_{t\text{-}1}, \sigma_{t\text{-}1}, c_{kt}\text{=}1\}$ follows from equation (9b) as

$$\sigma_t^2 = \frac{\sigma_{t-1}^2\sigma_k^2}{\sigma_{t-1}^2 + \sigma_k^2} \tag{14a}$$

Equation (14a) shows that the informative effects serve to decrease the risk premium in the expected utility by decreasing the uncertainty in perceived health status. Next, in equation (13), the specification of h_t conditional on $\{h_{t\text{-}1}, \sigma_{t\text{-}1}, c_{kt}\text{=}1\}$ can be derived using equation (9a), which specifies h_t as a function of $h_{t\text{-}1}$, δ_d, δ_k, $\sigma_{t\text{-}1}$, σ_k and the signal λ_{kt}. Note that at the beginning of period t, although the consumer knows the values of $h_{t\text{-}1}$, δ_d, δ_k, $\sigma_{t\text{-}1}$, σ_k, she does not know the value of the signal λ_{kt}. This is because the signal from consuming option k is only realized by the consumer at the end of period t. Thus from the consumer's perspective at the beginning of period t, λ_{kt} will be a random variable that will be distributed as per equation (8). Substituting equation (8) into equation (9a), we get the specification of h_t conditional on $\{h_{t\text{-}1}, \sigma_{t\text{-}1}, c_{kt}\text{=}1\}$ from the consumer's perspective at the beginning of period t as

$$h_t = h_{t-1} + \left(\delta_k c_{kt} - \delta_d\right) + \left(c_{kt}\frac{\sigma_{t-1}^2}{\sqrt{\sigma_k^2 + \sigma_{t-1}^2}}\upsilon_{kt}\right) \tag{14b}$$

where the term υ_{kt} is a standard normal random variable. In equation (14b), the second term on the RHS, $\delta_k c_{kt}\text{-}\delta_d$, represents the change in the mean health status which stems from degradation and the investment effect of option k; and the third term, $c_{kt}\frac{\sigma_{t-1}^2}{\sqrt{\sigma_k^2 + \sigma_{t-1}^2}}\upsilon_{kt}$, represents the change in the mean health status that stems from the informative effect of option k. Since the third term is a normal random variable, it implies that informative effects impact the evolution of the mean health status through the "learning draw" only – where the expected value of the learning draw that stems from consumption of option k in period t (i.e., $\frac{\sigma_{t-1}^2}{\sqrt{\sigma_k^2 + \sigma_{t-1}^2}}\upsilon_{kt}$) is zero, from the consumer's perspective at the beginning of period t. The reason for that is because at the beginning of period t when the consumer is making the decision on whether or not to consume option k, the consumer does not know what the informative effect of that option k will reveal to her about her health status at the end of period t – it could reveal no change, a positive or a negative change.

Substituting equations (14a) and (14b) into equation (13), we get the consumer's period t expected utility conditional on $\{h_{t-1}, \sigma_{t-1}, c_{kt}=1\}$ if she were to choose option k in period t as

$$E_t u_{kt} = -E_{\upsilon_{kt}} \exp\left(-r\left(h_{t-1} - \delta_d + \delta_k + \frac{\sigma_{t-1}^2}{\sqrt{\sigma_k^2 + \sigma_{t-1}^2}} \upsilon_{kt} \right) + \frac{r^2 \sigma_{t-1}^2 \sigma_k^2}{2\left(\sigma_{t-1}^2 + \sigma_k^2\right)} \right) - \alpha p_{k,j,t_a} + \varepsilon_{k,t} \quad (15)$$

This completes the discussion on the per-period utility. In the next section, we discuss the key implications of what we have discussed so far.

3.7 Implications

We discuss two implications that follow from sections 3.1-3.6. These implications hold true regardless of the functional form of utility; the only stipulation required is that the utility be increasing and concave in the health status. The first implication pertains to concavity of the utility and the usage of an investment treatment (i.e., a treatment that only has investment effects and not informative effects). The second implication pertains to consumer's forward looking behavior, concavity of the utility and usage of an informative service (i.e., a service such as a diagnostic test that only has informative effects).

<u>1. Concavity of the Utility and the Overall Frequency of Usage of Investment Treatments</u>: Recall from section 3.6 that the concavity of the utility implies that the marginal utility of health decreases with the increase in health status. Since the marginal utility of health is directly related to the consumer's incentive for seeking an investment treatment, it implies the following: (a) the sicker the consumer, the greater will be her incentive to choose an investment treatment over the no consumption option (and vice versa). This makes sense since consumers undergo treatments when they are sick and not when they are healthy. (b) The greater the consumer's degradation rate, the greater will be her incentive to seek an investment treatment, and consequently the greater will be the frequency of usage of the investment treatment. (c) The greater the investment effect of an option k, the lesser will be the frequency of usage of that option after its first usage. This is because once a consumer consumes an option with a large investment effect, her health status gets a significant boost which then decreases her incentive to repurchase such an option in future. This

makes sense since as discussed in section 2, curative care treatments are not used frequently. (d) As a result of degradation in health, the consumer's propensity to undergo an investment treatment increases with the time lapsed since its last purchase; and the greater the degradation rate, the greater is the increase in the consumer's propensity to repurchase the treatment over time. Thus if there is no degradation, her propensity to repurchase the treatment in any given period does not depend on how long it has been since she last purchased the treatment.

2. Concavity, Forward Looking Behavior and Usage of Informative Services: Recall from section 3.6 that when a consumer consumes an informative service (such as diagnostic tests) at the beginning of period t, she receives a learning draw about her health status at the end of period t, $\frac{\sigma_{t-1}^2}{\sqrt{\sigma_k^2 + \sigma_{t-1}^2}} \upsilon_{kt}$.

This learning draw provides an option value to the consumer since it enables her to make judicious consumption decisions (in terms of whether or not to go for investment treatments) in periods $t+1$ and onwards. For instance, if the consumer receives a very negative value of the learning draw at the end of period t, she would opt for a treatment with large investment effects in period $t+1$ which would increase het optimal utility in period $t+1$; similarly, if she receives a positive learning draw, she would opt for the no consumption option in period $t+1$. This point showcases the importance of modeling the consumers' forward looking behavior and for the utility to be concave. If consumers are not forward looking, there will be no incentive for them to seek information from diagnostic tests that can help them make judicious decisions in future. If the utility is not concave and is instead linear in the health status, there will be no incentive for the consumer to seek information from diagnostic tests that would help her make judicious decisions in the future. This is because there will be no option value of the learning draw since the probabilities of purchasing the investment treatment in period $t+1$ will be the same regardless of whether or not the consumer receives a positive or a negative learning draw from the diagnostic test.

3.8 Weekly Consumption and Annual Insurance Plan Purchase Decisions

We model the weekly consumption and annual insurance plan decisions in a nested dynamic framework with forward looking consumers. We start with the consumer's weekly consumption decision in period t given her insurance plan coverage and her perceived health status at the beginning of period t, h_{t-1}. Following that, we specify her annual insurance choice decisions assuming optimal weekly consumption decisions.

Consumption decisions in period t: The consumer chooses the healthcare option k which maximizes her discounted lifetime expected utility in period t in year a derived from both consumption and subsequent insurance purchases. Specifically, the value function at period $t<T$ (where T is the last period of the insurance year) conditional on insurance plan j is given as

$$V\left(S_t^a \middle| j\right) = \max_k \left\{ -\exp\left(-r\left(h_{t-1} - \delta_d + \delta_k\right) + \frac{r^2\sigma_{t-1}^2}{2} \right) - \alpha p_{k,j,t_a} + \varepsilon_{kt} + \beta E_t V\left(S_{t+1}^a \middle| S_t^a, k, j\right) \right\} \quad \text{(16a)}$$

In equation (16a), β is the weekly discount factor which is set at 0.999, and S_t^a denotes the state variables at the beginning of period t of year a. These state variables are $S_t \equiv \{h_{t-1}, \sigma_{t-1}, \{p_{kt}\}_k, \{\varepsilon_{kt}\}_k\}$, where the state transition processes of $\{h_s, \sigma_s\}$ for the future periods from the consumer's perspective are described in equations (14a) and (14b), and the state transition of prices are described in equation (10) and (11). In the last time period $t=T$ of insurance year a, when an insurance decision will be made in the next time period, the value function is given by

$$V\left(S_T^a \middle| j\right) = \max_k \left\{ -\exp\left(-r\left(h_{T-1} - \delta_d + \delta_k\right) + \frac{r^2\sigma_{T-1}^2}{2} \right) - \alpha p_{k,j,T_a} + \varepsilon_{kT} + \beta E_T W\left(S_1^{a+1} \middle| S_T^a, k\right) \right\} \quad \text{(16b)}$$

where $W\left(S_1^{a+1} \middle| S_T^a, k\right)$ represents the value of the future utilities at the beginning of the next year when the consumer makes another insurance decision, which we define in a moment. To compute the value functions, we employ a variant of Keane and Wolpin (1994) approximation method, which we have explained in Appendix II.

Insurance plan choice at the beginning of year a: At the beginning of each year, the consumer chooses one of the J=3 insurance plans in her choice set based on premium ($Prem_{ij,a}$) and her expected utilization of health care for the coming year. Thus the value of utility of a particular insurance plan j in period t=1 of year a will be[8]

$$VI_j\left(S_t^a\right) = -\alpha Prem_{j,a} + V\left(S_1^a \middle| j\right) + \varphi_{j,a} \quad \text{(17)}$$

[8] We model risk preference in consumption stage because consumers have uncertainty about their health status, but we did not incorporate preference for risk in the insurance plan stage because of the following reasons: first, Cardon & Hendel (2001) noted the CARA risk coefficient, r, is insignificant in the plan choice stage and preference toward risk are determined by both r and concavity of consumption stage utility which we model here; Second, given the nature of employer-sponsored health insurance we studied here, individuals have more uncertainty about their health status, not their plans; Third, Vera-Hernadez (2003) pointed out that risk preference is more about health care consumption.

where the coefficient for the premium, α, measures how much the consumer i values premium and is same as the parameter that captures the out of pocket expenses in consumption decisions, and $\varphi_{j,a}$ is an additive idiosyncratic error that is independent across time and consumers and follows Type-1 extreme value distribution. $V(S_1^a|j)$ is the continuation value at the beginning of the year from the consumption. As stated in equation (16b), the expected lifetime utility at the beginning of year a is

$$W(S_1^a) = E_{T_{a-1}+1} \max VI_j(S_1^a) \tag{18}$$

Section 4: Econometric Specification

4.1 Choice Probabilities

The choice probabilities of insurance plan and consumption decisions follow from equations (16)-(18). However, note that the period t value functions in equations (16)-(18) are defined from the consumer's perspective, where the consumer's health status at the beginning of period t, h_{t-1}, is known to the consumer. On the other hand, h_{t-1} is not known to the econometrician. This is because h_{t-1} is a function of the past signals received by the consumer from periods s=1 to s=t-1 (see equation 9a), and unlike the consumer, the econometrician does not observe the values of these past signals. Therefore before we use equations (16)-(18) to specify the choice probabilities in period t, we need to first specify h_{t-1} from the econometrician's perspective.

From the econometrician's perspective, the past signals will be random variables as specified in equation (8). Thus substituting the expression for the signal given in equation (8) into equation (9a), we get the consumer's mean health status at the end of period t-1 as

$$h_{t-1} = \left(h_{t-2} - \delta_d + \sum_k \delta_k c_{kt-1}\right) + \left(\frac{1}{\sigma_{t-2}^2} + \sum_k \frac{c_{kt-1}}{\sigma_k^2}\right)^{-1} \left(\sum_k \frac{c_{kt-1}}{\sigma_k^2}\left(H_{t-2} - h_{t-2} + \sigma_k \eta_{kt-1}\right)\right) \tag{19}$$

where δ_d is the degradation rate, δ_k is the investment effect of consumption option $k \in \{pp, sp, c, no\}$, σ_k is the noise the in consumption signal of consumption option k, and η_{ks} is a standard normal random variable. Defining $\Delta h_{t-1} \equiv h_{t-1} - H_{t-1}$, we get the evolution of Δh_s from equations (4) and (19) as

$$\Delta h_{t-1} = \Delta h_{t-2} + \left(\frac{1}{\sigma_{t-2}^2} + \sum_k \frac{c_{kt-1}}{\sigma_k^2}\right)^{-1} \left(\sum_k \frac{c_{kt-1}}{\sigma_k^2}\left(-\Delta h_{t-2} + \sigma_k \eta_{kt-1}\right)\right) \tag{20}$$

Using equation (20) recursively, we get $\Delta h_{t-1} \equiv h_{t-1} - H_{t-1}$ in terms of $\Delta h_0 \equiv h_0 - H_0$ as

$$h_{t-1} - H_{t-1} = \left(\frac{1}{\sigma_0^2} + \sum_k \frac{N_{kt-1}}{\sigma_k^2} \right)^{-1} \left(\frac{h_0 - H_0}{\sigma_0^2} + \sum_k \frac{\sum_{s=1}^{t-1} c_{ks} \eta_{ks}}{\sigma_k} \right) \tag{21}$$

where $N_{kt\text{-}1}$ is the total number of times the consumer has chosen healthcare option k from period s=1 to the end of period s=t-1. Next note that the true health status in period t-1 can be recursively derived using equation (4) as

$$H_{t-1} = H_0 - (t-1)\delta_d + \sum_k \delta_k N_{kt-1} \tag{22}$$

Thus substituting equation (22) into equation (21), we get

$$h_{t-1} = H_0 - (t-1)\delta_d + \sum_k \delta_k N_{kt-1} + \left(\frac{1}{\sigma_0^2} + \sum_k \frac{N_{kt-1}}{\sigma_k^2} \right)^{-1} \left(\frac{h_0 - H_0}{\sigma_0^2} + \sum_k \frac{\sum_{s=1}^{t-1} c_{ks} \eta_{ks}}{\sigma_k} \right) \tag{23}$$

Finally, recall that we had assumed rational expectations as per which the consumer's prior beliefs at the beginning of {t=1, a=1}, i.e. h_0 and σ_0^2, are the mean and the variance of the distribution of the true health status across the consumer population at the beginning of {t=1, a=1}. Consequently, the initial Δh_0 at the beginning of {t=1, a=1} can be written as $H_0 = h_0 + \sigma_0 \eta_0$, where η_0 is a standard normal random variable. Thus substituting $H_0 = h_0 + \sigma_0 \eta_0$ into equation (22), we get

$$h_{t-1} = h_0 - (t-1)\delta_d + \sum_k \delta_k N_{kt-1} + \left(\frac{1}{\sigma_0^2} + \sum_k \frac{N_{kt-1}}{\sigma_k^2} \right)^{-1} \left(\sum_k \frac{\sum_{s=1}^{t-1} c_{ks} \left(\sigma_k \eta_{ks} + \sigma_0 \eta_0 \right)}{\sigma_k^2} \right) \tag{24}$$

Equation (24) represents the specification of h_{t-1} from the econometrician's perspective. Note that h_{t-1} is a function of the random errors, $\omega_t \equiv \left\{ \eta_0, \left\{ \eta_{ks} \right\}_{s=1..t-1,k} \right\}$, which reflect the econometrician's uncertainty in knowing the true value of h_{t-1}. Substituting h_{t-1} as given in equation (24) into the value functions in equations (16a)-(16b) will yield the period t value functions from the econometrician's perspective. Given the period t value functions, the relevant probabilities follow immediately. The relevant probabilities are the probability of choosing a particular insurance plan and the probability of seeking health care consumption during the policy year respectively. Conditional on the insurance plan chosen by an individual and the errors ω_t, the probability of choosing no consumption, primary preventive, secondary preventive or curative care is given by

$$\Pr\left(c_{kt_a}=1\middle|d_{j,a},\omega_{t_a}\right)=\frac{\exp\left(V_k\left(S_t^a\left(\omega_{t_a}\right)\middle|d_{j,a}\right)\right)}{\sum_{k'=\{pp,sp,c,no\}}\exp\left(V_{k'}\left(S_t^a\left(\omega_{t_a}\right)\middle|d_{j,a}\right)\right)} \quad (25)$$

where $V_k\left(S_t^a\left(\omega_{t_a}\right)\middle|d_{j,a}\right)$ denotes the choice-specific value function for primary preventive care, secondary preventive care and the no consumption option (conditional on the errors ω_t and the insurance plan choice) as defined in equation (16) after substituting the expression for h_{t-1} given in equation (24). The choice probabilities of alternative insurance plans at the beginning of year a conditional on the errors are given as

$$\Pr\left(d_{ja}=1\middle|\omega_{1_a}\right)=\frac{\exp\left(W_j\left(S_1^a\left(\omega_{1_a}\right)\right)\right)}{\sum_{j=1\ldots J}\exp\left(W_{j'}\left(S_1^a\left(\omega_{1_a}\right)\right)\right)} \quad (26)$$

4.2 Parameters and Heterogeneity

Until now we have introduced the following parameters: (i) degree of absolute risk aversion, r, (ii) price sensitivity, α, (iii) weekly degradation in consumer's health status, δ_d, (iv) investment effects of primary preventive, curative and secondary preventive care options, $\{\delta_{pp}, \delta_c, \delta_{sp}\}$ (v) noise in primary preventive, curative and secondary preventive care signals, $\{\sigma_{pp}, \sigma_c, \sigma_{sp}\}$ and (vi) mean and variance of consumer's prior beliefs in the beginning of first year in the estimation sample, $\{h_0, \sigma_0\}$.

Recall that secondary preventive care comprises of diagnostic tests and screening exams. Since these services do not provide any boost to the health status, we set the investment effect of secondary preventive care to zero (i.e., we set δ_{sp}=0) and only estimate the investment effects of primary preventive and curative care. Furthermore, across the three health care services, it is only the secondary preventive care that provides information to consumers about their current health status and not the curative or primary preventive care. We thus set the informative effects of curative and primary preventive care to zero[9] (i.e., we set $\sigma_c=\sigma_{pp}=\infty$) and only estimate σ_{sp}, which is the noise in the consumption signal for secondary preventive care.

Next note that the risk preference parameter r and the other parameters, $\{\delta_d, \delta_{pp}, \delta_c, \sigma_{sp}, h_0, \sigma_0\}$, always enter the utility in a multiplicative fashion as $\{r\delta_d, r\delta_{pp}, r\delta_c, r\sigma_{sp}, rh_0, r\sigma_0\}$. This can be seen in equation (27) where we have derived the period t value function from the econometrician's perspective using equations (16a) and (24) in terms of the parameters $\{r, \alpha, \delta_d, \delta_{pp}, \delta_c, \sigma_{sp}, h_0, \sigma_0\}$.

[9] Even though we set the informative effects of primary preventive and curative cares to zero, we do allow for consumers to be aware that their health status has been boosted by the investment amount δ_{pp} (δ_c) once they consume primary preventive (curative) care.

$$V\left(S_t^a\right)=\max_k\left\{\begin{array}{l}-\exp\left(\begin{array}{l}-\left(rh_0\right)+\left(r\delta_d\right)t-\left(r\delta_c\right)N_{c,t}-\left(r\delta_{pp}\right)N_{pp,t}\\+\dfrac{1}{\beta_0+N_{sp,t-1}}\left(\dfrac{\left(r\sigma_{sp}\right)^2}{2}+\left(r\sigma_0\right)N_{sp,t-1}\eta_0+\sum_{s=1}^{t}c_{sp,s}\left(r\sigma_{sp}\right)\eta_{sp,s}\right)\end{array}\right)\\-\alpha p_{k,j,t_a}+\varepsilon_{k,t}+\beta E_t V\left(S_{t+1}^a\middle|S_t^a,k\right)\end{array}\right\} \quad (27)$$

where $c_{k,s}$ is a 0-1 indicator variable that takes a value of 1 if the consumer consumes option k ($\forall$ $k\in\{sp, pp, c, no\}$) in period s, the terms η_0 and $\{\eta_{sp,s}\}_{s=1}^{t}$ are standard normal random variables, the terms $\{N_{sp,t}, N_{c,t}, N_{pp,t}\}$ respectively represent the number of times the consumer has consumed secondary preventive care, curative care and primary preventive care from periods s=1 to s=t, and the term $\beta_0=\left(r\sigma_{sp}\right)^2/\left(r\sigma_0\right)^2$. Equation (27) clearly shows that r cannot be identified from the other parameters, $\{\delta_d, \delta_{pp}, \delta_c, \sigma_{sp}, h_0, \sigma_0\}$. Accordingly we normalize r=1.[10]

To capture unobserved heterogeneity in consumers' insurance plan and consumption decisions, we use latent class segmentation in which we allow for the following parameters to differ across segments: (i) the prior mean and variance of consumer's health status, $\{h_0, \sigma_0\}$, (ii) price sensitivity, α, and (iii) weekly degradation rate, δ_d. Since we normalize r=1, we capture heterogeneity in consumers' risk premiums by heterogeneity in their uncertainty about their health status.

To estimate the parameters, let Θ denote the vector of parameters to be estimated. The likelihood for consumer i is

$$L_i(\Theta)=\int\ldots\int\prod_a\prod_{j=1}^{J}\Pr(d_{ijt_a}=1)^{d_{ijt_a}}((\prod_t\prod_k\Pr(c_{ikt_a}=1\mid d_{ijt_a})^{c_{ikt_a}}\Phi(dh_{ik};\Theta))G(d\omega_i) \quad (28)$$

Integrating over the joint distribution of the errors is difficult because of the large dimensionally. We thus use Monte Carlo simulations with R=500 draws to compute the likelihood.

4.3 Identification

We start with the identification of distribution of price sensitivities across the population. The price sensitivity parameter, α, enters as the coefficient of the out of pocket expenses in both the insurance plan and health care consumption decisions. Thus its identification follows from the inter-temporal

[10] Some prior papers that have used the CARA utility have estimated r as a separate parameter. These papers include Chan and Hamilton (2006) and Ching and Ishihara (2010). The reason why Chan and Hamilton (2006) were able to identify r is because in their data, they observed the values of the signals. And the reason why Ching and Ishihara (2010) were able to identify r is because they instead normalized the true quality of one of the alternatives to one. We thank Andrew Ching for pointing this out.

and cross-sectional variation in the out of pocket expenses and the impact of this variation on both the insurance plan and consumption decisions. In insurance plan decisions, the out of pocket expenses vary over time for a given consumer since the insurance premiums vary from year to year; and they also vary across consumers in any given year since the premiums for the same plan are different across different employers. In consumption decisions, the out of pocket expenses vary over time for a given consumer because of non-linear pricing induced by the deductibles and the out of pocket maximum limits in each plan, and also because these cost sharing characteristics vary from year to year; and they vary across consumers with the same plan since the cost sharing characteristics of the same plan are different across different employers. This cross sectional and longitudinal variation of out of pocket expenses allows us to identify the distribution of price sensitivity.[11]

Given the price sensitivity parameter, we next discuss the identification of parameters that determine the usage of a healthcare option with only informative effects (secondary preventive care), and the parameters that determine the usage of an option with only investment effects (curative and primary preventive cares). The usage of option k with only investment effects depends on three parameters: (i) investment effect of that option, δ_k, (ii) consumer's prior mean health status at the beginning of $\{t=1, a=1\}$, h_0, and (iii) her degradation rate, δ_d. And the usage of consumption option k with only informative effects depends on two parameters[12]: (i) noise in option k's signal, σ_k, and (ii) consumer's uncertainty about her health status at the beginning of $\{t=1, a=1\}$, σ_0.

Before we discuss the identification of these parameters, it is important to note that the identification arguments that we provide for parameters that determine the usage of an informative service k, i.e., $\{\sigma_k, \sigma_0\}$, do not apply to consumers who have either never purchased the informative service k throughout their observed history in the data or have repeatedly purchased the informative service k across all observations in their observed history. Similarly, the identification arguments that we provide for parameters that determine the usage of an investment treatment k, i.e., $\{h_0, \delta_k, \delta_d\}$, do not apply to consumers who have either never purchased the investment treatment k throughout their observed history or have repeatedly purchased the investment treatment k across all

[11] This explains how we can identify whether a consumer, who is in a comprehensive plan and consumes costly curative care, has low price sensitivity & moderate health or moderate price sensitivity & poor health - unlike a consumer with low price sensitivity, a consumer with high price sensitivity will show differences in consumption behavior before and after the deductible is reached and before and after the out of pocket maximum is reached.

[12] Note that while we allow for degradation in consumer's mean health status over time (through δ_d), we do not model the increase in consumer's uncertainty over time, where the increase in uncertainty over time stems from periodic health shocks that persist over time and are not observed by the consumer. We have discussed this issue in detail in section 7.

observations in their observed history.[13] This is analogous to the identification of parameters in brand choice models in which the parameters cannot be identified from purchase histories of those consumers who have chosen the same alternative across all observations in their purchase histories.

We start with the parameters σ_{sp} and σ_0 that determine the usage of secondary preventive care. Their identification stems from two features in the data: (a) number of weeks lapsed from {t=1, a=1} till the period when the consumer first purchases secondary preventive care in the data, and (b) the consumer's purchase frequency of secondary preventive care over the three years after her first purchase of secondary preventive care in the data. Table 3A reports the distribution of consumers in the data, who have used secondary preventive care at least once, along (a) and (b). To see how (a) and (b) help identify these parameters, note that all else being the same, the greater the prior uncertainty in health status, σ_0, the greater will be the probability that the consumer will use secondary preventive care early on in the data, and the greater will be her subsequent of purchase of secondary preventive care (i.e., frequency after her first purchase in the data) over the three years in the data; on the other hand, all else being the same, the smaller the signal noise, σ_{sp}, the greater will be the probability that the consumer will use secondary preventive care early on in the data, but lesser will be her subsequent frequency of purchase of secondary preventive care over the three years in the data. The reason for that is because if secondary preventive care provides accurate information about the health status, then there is little incentive for the consumer to repurchase it.

The aforementioned differences allow us to identify σ_{sp} and σ_0. For instance, even though 'small values of both prior uncertainty and signal noise' will yield a similar value of (a) as 'large values of both prior uncertainty and signal noise', however, 'small values of both prior uncertainty and signal noise' will yield a smaller value of (b) as compared to 'large values of both prior uncertainty and signal noise'. Similarly, even though a 'large value of prior uncertainty and small value of signal noise' will yield a similar value of (b) as a 'small value of prior uncertainty and large value of signal noise', however, a 'large value of prior uncertainty and small value of signal noise' will yield a much smaller value of (a) compared to a 'small value of prior uncertainty and large value of signal noise'.

We next move on to the parameters δ_k, h_0, and δ_d that determine the usage of an investment treatment k. Similar to what we did for identifying the informative effects, we start with the following two features in our data vis-à-vis the investment treatments: (i) number of periods lapsed from {t=1, a=1} till the period when the consumer first consumes an investment treatment k in the

[13] Such consumers are less than 3% in our data.

data, (ii) consumer's purchase frequency of the investment treatment *k* over the three years after her first usage of that treatment *k* in the data. Note that although these two features can help identify the prior mean health status from the investment effect (the logic of which we will explain shortly afterwards), they are not sufficient to help separately identify the two from the degradation rate. Why? This is because the prior mean health status and the (inverse of) degradation rate impact (i) and (ii) in a directionally similar manner: the greater the prior mean health status (or the lesser the degradation rate), the larger is the value of (i) and the smaller is the value of (ii). This implies that before we can use (i) and (ii) to separately identify the prior mean health status and the investment effect, we need to first identify the degradation rate. This is discussed as follows.

Recall from section 3.7 that degradation in consumer's health over time leads to a specific type of time dependence, i.e., it results in the consumer's propensity to repurchase an investment treatment to increase with the time elapsed since its last purchase. In other words, if there is no degradation, the consumer's propensity to repurchase an investment treatment in any period after its last purchase will not depend on how long it has been since she last purchased that treatment. And as the value of degradation rate increases, the greater will be this time dependence, i.e., the greater will be the increase in the consumer's propensity to repurchase the treatment with the time elapsed since its last purchase. Next, note that it is only degradation in health that leads to this specific type of time dependence, and not the investment effects or the prior mean health status. This is because while the investment effect and prior mean health status influence the consumer's propensity to purchase the investment treatment at $\{t=1, a=1\}$ as well as how this propensity will change after every successive purchase of the treatment, they do not influence how the consumer's propensity to purchase the treatment will change with the time elapsed since its last purchase. Finally, note that this time dependence is nothing but the slope of the hazard function (with respect to the time elapsed since the last purchase) of inter-purchase times of investment treatments. This implies that it is only the degradation rate that will impact the slope of hazard function of inter-purchase times (the investment effect and the prior mean health status, as we will discuss shortly afterwards, will instead impact the intercept of the hazard function). If degradation rate is zero, the slope of the hazard function will be zero, which implies a constant hazard rate. And as degradation rate increases, the slope of the hazard function will also increase.

It follows from the above discussion that the degradation rate will be identified by the slope of the hazard function of inter-purchase times. Given the degradation rate, the identification of prior mean health status h_0 and the investment effect of option *k*, δ_k, follows immediately from (i) and (ii) discussed before. Table 3B (3C) reports the distribution of consumers in the data, who have used primary preventive care (curative care) at least once, along (i) and (ii). To see how (i) and (ii) help identify the parameters h_0 and δ_k, note that all else being the same, the lower the value of the prior mean health status h_0, the greater will be the probability of using the investment treatment k early on in the data, and the greater will be the subsequent frequency of purchase of treatment k over the three years in the data. On the other hand, all else being the same, the greater the investment effect δ_k, the greater will be the probability of using the treatment k early on in the data, but the lesser will be the subsequent frequency of purchase of treatment k over the three years in the data. To see why that is so, consider two treatments k and k', where k has a larger investment effect compared to k'. Since option k provides a larger boost to consumer's health (and makes her completely healthy), the incentive for the consumer to repurchase k will be smaller than her incentive for the consumer to repurchase k'. This difference allows us to identify prior mean health status from investment effect.

We next present some reduced form analysis to support the aforementioned identification arguments for the parameters δ_k, h_0 and δ_d. We estimate a hazard model of inter-purchase times of consumers' purchases of primary preventive care. To keep it simple, we only consider those consumers (N=171) who were enrolled in the medium plan throughout the three years and used primary preventive care but not curative care during their entire purchase histories. Consider the following proportional hazard model that captures consumer's propensity to purchase primary preventive care treatment in week τ after her last purchase of primary preventive care:

$$h(\tau|t, N_t) = \exp(\gamma_0 + \gamma_1\tau)\exp(\gamma_2 t + \gamma_3 N_t)\exp(c\xi) \qquad (28)$$

The first term on the RHS of equation (28), $\exp(\gamma_0+\gamma_1\tau)$, is the Gompertz baseline hazard (Vilcassim and Jain 1991) which is a function of the weeks lapsed since the last purchase, τ. The second term on the RHS captures the impact of the other covariates on the hazard function. In this term, (i) t represents the number of weeks lapsed from the first period in the data to the period in which the primary preventive care treatment was last purchased (thus $t+\tau$ represents the total number of weeks lapsed from the first period in the data to the focal period considered in the hazard function in equation 28), (ii) N_t is the number of primary preventive care treatments purchased by the consumer in the first t weeks of the data (which includes the purchase in week t). Finally, the third term on the RHS controls for unobserved heterogeneity, where ξ is a standard normal random error.

In equation (28), the parameter γ_0 captures the baseline effect, which is the consumer's propensity to purchase primary preventive care in the first period of the data. Following our earlier discussion, this baseline effect will be a function of all three factors: it will be inversely related to prior mean health status, h_0, and positively related to both the degradation rate, δ_d, and the investment effect, δ_{pp}. The parameter γ_1 represents the slope of the hazard function with respect to the time elapsed since the last purchase of the treatment. As discussed before, this will be a function of only the degradation rate. Thus a value of $\gamma_1>0$ will imply that there is degradation; and the larger the degradation rate, the larger will be the value of γ_1.

The parameter γ_2 captures the impact of t (time lapsed from the first period in the data to the period in which the primary preventive care treatment was last purchased) on the consumer's propensity to purchase primary preventive care in the focal period, $t+\tau$. Since time dependence stems only from degradation, the parameter γ_2 will be a function of the average weekly degradation rate in the first t periods of the data. Similar to γ_1, a value of $\gamma_2>0$ will imply that there is degradation in the first t periods. The parameter γ_3 represents the impact of N_t (number of past purchases of the treatment) on the consumer's propensity to repurchase the treatment in the focal period, $t+\tau$. This will be a function of only the investment effect δ_{pp}, and not the degradation rate δ_d, or the prior mean health status, h_0[14]. Recall from our earlier discussion that the larger the investment effect, the lesser and lesser will be the consumer's propensity to purchase the investment treatment after its every successive repurchase. This implies that $\gamma_3<0$ will indicate a positive investment effect; and the lower the value of γ_3, the greater will be the investment effect. Thus in summary, the degradation rate is identified by the slope of the hazard function, γ_1, and the investment effect is identified by the coefficient of number of past purchases of the investment treatment, γ_3.[15]

We account for both left and right censoring while estimating the reduced form model. For primary preventive care, we get the following estimates of the parameters: γ_0 =1.471 (std. err.=0.389), γ_1=0.202 (std. err.=0.089), γ_2=0.182 (std. err.= 0.054), γ_3=-0.124 (std. err.=0.051) and c=0.531 (std. err.=0.127). Note that the estimates of γ_1 and γ_2 are positive and significant. This implies that there is degradation. The estimate of γ_3 is also negative and significant, which implies positive investment effects for primary preventive care. Finally, we do the same reduced form analysis for curative care. Once again for simplicity, we only consider those consumers who were enrolled in the medium plan

[14] This is because we have controlled for degradation in health in periods prior to the focal period through the terms $\gamma_1\tau$ and $\gamma_2 t$. Further since γ_0 captures the consumer's propensity to purchase primary preventive care in the first period of the data, the mean health status in the first period of the data, h_0, will impact the purchase probability in the focal period, $t+\tau$, only through the baseline effect γ_0.

[15] Given the degradation rate and investment effect, the prior mean health status can then be imputed from the baseline effect, γ_0. This can be done if we know the baseline probability in terms of degradation rate, investment effects and the prior mean health status, which we do in our model.

and used curative care but not primary preventive care during their entire purchase histories (N=185). We get the following estimates for curative care: γ_0=1.027 (std. err.=0.474), γ_1=0.299 (std. err.=0.106), γ_2=0.369 (std. err.=0.112), γ_3=-0.227 (std. err.= 0.095), c=0.516 (std. err.= 0.211). Note that the estimates of γ_1 and γ_2 for curative care are both positive and significant. This implies that there is degradation. Further, these values are larger in magnitude than their corresponding values for primary preventive care. This implies that consumers who use curative care have higher degradation rates than consumers who only use primary preventive care. The estimate of γ_3 for curative care is also negative and significant, which implies positive investment effects. Further, this value is larger in magnitude compared to its corresponding value for primary preventive care. This implies that curative care has a larger investment effect compared to primary preventive care.

Section 5: Empirical Results

5.1 Model Fit and Comparison

As discussed in section 1, the prior literature has not (i) modeled informative effects of consumption options, and (ii) distinguished between curative and preventive cares while estimating the investment effects of healthcare services. We thus compare the goodness of fit of our proposed model (Model I) with two competing models which showcases the importance of modeling (i) and (ii). The first competing model (Model II) is similar to Model I except that it does not allow for informative effects. It aggregates primary preventive and secondary preventive care into one option and only allows for investment effects for that option. The second competing model (Model III) is similar to Model I except that it does not allow for different investment effects of primary preventive care and curative care. It aggregates primary preventive and curative care into one option and estimates a joint investment effect. The goodness of fit results are reported in Table 4. The comparison of goodness of fit statistics shows that our proposed model performs the best. This illustrates that importance of modeling the informational role of preventive care and the importance of distinguishing between preventive and curative care when modeling consumption decisions.

5.2 Parameter Estimates and Discussion

Table 5 reports the parameter estimates of our proposed model. We first discuss the distribution of the parameters that vary across the four segments. As per our estimates, segment A accounts for 22.5%, segment B accounts for 31.5%, segment C accounts for 12.9% and segment D accounts for 33.1% of the consumer population. The estimates of the prior mean health status across the four segments are $h_{0,A}$=2.732 , $h_{0,B}$=1.632, $h_{0,C}$=1.612 and $h_{0,D}$=0.892. This shows that segment A is the healthiest at {t=1, a=1}, followed by segments B and C that are similar in their prior mean health status, and finally followed by segment D. The estimates of degradation rates for the four segments are $\delta_{d,A}$=0.0005, $\delta_{d,B}$=0.0012, $\delta_{d,C}$=0.0012 and $\delta_{d,D}$=0.0053. Note that across the four segments, the degradation rates are inversely correlated with prior mean health status. This implies that segment A is the healthiest segment (in terms of needing least amount of investment treatments over time), followed by segments B and C and finally followed by segment D. We thus refer to segment A as the healthy segment, B and C as the moderately healthy segments and D as the least healthy segment.

The estimates of price sensitivity (which is the sensitivity for out-of-pocket expenses and premium) across the four segments are α_A=0.168, α_B=0.188, α_C=0.131 and α_D=0.099. This shows that the healthy segment A is more price sensitive compared to the least healthy segment D; and the moderately healthy segment B is more price sensitive compared to moderately healthy segment C. The estimates of the prior uncertainty in the health status across the four segments are $\sigma_{0,A}$=2.724, $\sigma_{0,B}$=2.812, $\sigma_{0,C}$=3.986 and $\sigma_{0,D}$=2.887, which shows that the prior uncertainty of segments A, B and D are similar and smaller compared to the prior uncertainty of segment C.

These results imply the following regarding the insurance plan decisions of consumers. If we only consider heterogeneity in the prior mean health status and degradation rate, we would predict that in the first year, consumers in the healthy segment A would primarily prefer the basic plan, consumers in the moderately healthy segments B and C would primarily prefer the medium plan and consumers in the least healthy segment D would prefer the comprehensive plan. However in our data, 22.2%, 33.9% and 43.9% of consumers choose basic, medium and comprehensive plans respectively in the first year. Comparing these percentages with the percentages implied by heterogeneity in the prior mean health status and degradation rate, we can see a misalignment: only 33.1% consumers are in the weakest segment D, but a much higher percentage (=43.9%) of consumers in the data choose the comprehensive plan; and a total of 44.4% consumers are in the moderate health segments B and C, but a much lower percentage (=33.9%) of consumers choose the medium plan. This issue gets resolved once we also consider heterogeneity in the prior uncertainty. Since segments A, B and D have low uncertainties, their insurance plan decisions are dictated more by their mean health status – and as a result, consumers in healthy segment A primarily choose the basic plan, in moderately health segment B primarily choose the medium plan and in the least healthy segment D primarily choose the comprehensive plan. On the other hand, since consumers in the moderately healthy segment C have a high risk premium because of high uncertainty in their health status, they choose the comprehensive plan instead of the medium plan.

An important point that follows from above is that segment C is a potential source of moral hazard. Note that segment C consumers have moderate health status, which implies that they can manage their well-being by either curative or primary preventive care. However since they purchase the comprehensive plan and not the medium plan, they choose to consume the more expensive curative which increases the overall healthcare costs. We will revisit this issue in section 6 where we discuss the policies that can be put in place to decrease the extent of moral hazard.

We next discuss the investment and informative effects. The estimates of investment effects of curative and primary preventive cares are δ_c=0.063 and δ_{pp}=0.014. Both estimates are much greater than the degradation rate of any segment, which implies that consumers in all four segments will improve their health if they take any of the two treatments. Further, the estimate of the investment effect for curative care is significantly greater than the investment effect for primary preventive care. This makes intuitive sense since curative care consists of surgeries and drugs that provide a quantum boost to the consumer's health, while primary preventive care consists of drugs that prevent the disease from getting worse. Moving on to the informative effects, the estimate of the noise in the signal of secondary preventive care is σ_{sp}=7.143, which yields the noise to information ratio (i.e., the ratio of the noise in the signal, σ_{sp}, to the prior uncertainty in the health status for segment j, $\sigma_{0,j}$) for segments A, B, C and D as 2.622, 2.540, 1.792 and 2.474 respectively. A small value of the noise to information ratio means that there will be large reduction in uncertainty after consuming secondary preventive care. The aforementioned values of the noise to information ratios imply that all four segments can reduce their uncertainty by consuming secondary preventive care, but segment C will reduce it the most.

Finally, we discuss the impact of pricing components on insurance choice and consumption decisions in each of the four segments. The pricing components include copayment, coinsurance, deductible, out-of-pocket maximum and premium. Tables 6A-6C report the elasticities of each of the pricing components of each of the three plans on the insurance plan decisions of consumers in the four segments. These elasticities represent the percentage change in the average probabilities of each choice following a 1% change in the pricing component. Further, these are contemporaneous elasticities, i.e., they capture the impact of the change in the pricing component at the beginning of a given year on the insurance plan choices in that year only. Observe that the elasticities are smaller for segments C and D as compared to those for segments A and B. This result stems from the fact that segments C and D have the lowest price sensitivities. Moreover, although all the pricing component elasticities are statistically significant, they are small in magnitude. This implies that the consumer's health status, her uncertainty in her health status and her degradation rate play a more important role in her insurance plan decision as compared to the pricing components.

Table 7 reports the contemporaneous elasticities of each of the cost sharing characteristics (copayment, coinsurance and deductible) on the consumption choices of the four segments. Similar to Table 6, the elasticities in Table 7 are higher for segments A and B compared to segments C and D. The reason for that once again is because segments C and D have lower price sensitivities. And unlike Table 6, the elasticities in Table 7 are larger in magnitude, which implies that the cost sharing characteristics have a greater impact on consumption choice decisions as compared to insurance choice decisions. Across the four segments, the usage of both primary and secondary preventive cares is influenced most by changes in the copayment, followed by changes in the deductible and coinsurance rate. The opposite is true for curative care, which makes intuitive sense - since curative care is expensive, a decrease in the coinsurance rate translates to greater savings for the consumer as compared to a similar percentage decrease in the deductible or the copayment.

Section 6: Policy Simulations

The structural parameters of an individual's optimization problem regarding insurance plan choice and health care consumption decisions enable us to simulate alternative policies and predict the change in consumer behavior and the overall healthcare costs. In this section, we examine how the moral hazard can be decreased by incentivizing consumers in segment C to choose the medium plan. This can be done via two routes. The first route is the 'immediate route' in which we change the cost sharing characteristics of insurance plans at the beginning of the policy year, which will incentivize consumers in segment C to choose the medium plan in that policy year itself. The second route is the 'delayed route' in which we provide incentives for consumers in segment C to consume more secondary preventive care in the first year (and onwards). Recall from Section 2 that the usage of secondary preventive was as such low across all consumers in the data. This implies that the additional usage of secondary preventive care beyond that in the data will help substantially reduce the uncertainty of segment C consumers vis-à-vis their health status; and this will in turn reduce their risk premium, which will then induce them to choose the medium plan in the future.

Regarding the immediate route, our results in Section 5.2 indicate that this may not be feasible. This is because the contemporaneous elasticities for insurance choice decisions are small. Thus changing the cost sharing characteristics of any of the three plans in a given year will not be sufficient to induce segment C consumers to switch to the medium plan in that year itself. We discuss below two policy simulations in the context of the delayed route. In the first one, we investigate the impact of changing the cost sharing characteristics of insurance plans that encourage the use of secondary preventive care. In the second one, we investigate the impact of providing more accurate information to the consumer via secondary preventive care.

6.1 Changing Cost Sharing Characteristics to Encourage Secondary Preventive Care

Recently, many health service providers have advocated the reduction in the cost incurred by the consumer for secondary preventative care. Recall from Section 5.2 that copayment has the strongest impact on the use of secondary preventive care. Thus, in this section, we examine whether and to what extent would the overall healthcare costs (where the overall health care costs are the sum of the healthcare costs incurred by the consumers and healthcare costs incurred by the insurance company) decrease in the long run if the insurance firm were to reduce the copayment for secondary preventive care. Decreasing the copayment for secondary preventive care can have two opposing effects on overall health care costs. On one hand, it can encourage consumers to consume more secondary preventive care, which will result in an increase in health care costs; on the other hand, it can decrease the inefficiencies in the future since it will encourage consumers in segment C to consume more secondary preventive care which will decrease their uncertainty, and thereby make them choose the medium plans in future periods.

To see how reducing the copayment impacts the overall costs, we run a counterfactual in which the insurance company was to reduce the copayment for secondary preventive care in 2005 and onwards by 50%. Table 8 presents the results on the change in consumers' insurance plan choices and change in overall healthcare costs of each of the segments in the subsequent year 2006 with respect to the baseline case (in which there is no reduction in the copayment).

Observe that plan choice probabilities are similar between the baseline and counterfactual cases in 2005. This result follows from our discussion in Section 5.2. Since the contemporaneous insurance choice elasticities are small, changing the cost sharing characteristics will not have a strong immediate impact on insurance choices. This shows that the immediate route is not an effective one for decreasing the extent of moral hazard. On the other hand, there is a greater divergence in the plan choice probabilities across the two cases in 2006, which shows that the delayed route is a much more effective tool for decreasing moral hazard. Consequently, we see that the overall healthcare costs in 2006 are lower in the counterfactual case as compared to the baseline case. To understand the source of this decrease, we next look at the changes in healthcare costs on a segment level basis.

Observe in Table 8 that decreasing the copayment for secondary preventive care leads to a small increase in overall healthcare costs for healthy segment A and moderately healthy segment B, a small decrease in overall healthcare costs for the weak segment D, and a large decrease in overall healthcare costs for moderately healthy segment C. The reason for the increase in overall healthcare costs for healthy segment A and moderately healthy segment B is because most of the consumers in these segments as such choose the medium or the basic plans. Since in medium and basic plans, consumers incur high out of pocket expenses while consuming secondary preventive care, a 50% decrease in copayment amounts to greater savings which encourages them to use more secondary preventive care. And the reason for such a large decrease in overall healthcare costs for moderately healthy segment C follows from our earlier discussion. A 50% decrease in copayment of secondary preventive care increases the probability of consumers in segment C to consume more preventive care in 2005, which in turn increases their probability of purchasing medium plans in 2006. This can be seen in Table 8 – the overall probability in 2006 of choosing the comprehensive plan reduces from 43.3% to 42.1% (between the baseline and counterfactual) and the overall probability of choosing the medium plan increases from 34.1% to 35.4%.

6.2 Improving Health Status Learning (Personalized Medicine)

In the second simulation, we investigate the impact of providing more accurate information to the consumer (regarding her health status) via secondary preventive care on the overall healthcare costs. This is in the spirit of personalized medicine in which health care providers use precise molecular profiling technologies that can provide much more accurate information to consumers about their health status compared to standard diagnostic tests.[16]To do so, we run a counterfactual in which we consider an alternative scenario in which the noise in the secondary preventive care signal were to decrease by half in 2005 and onwards. Since we do not have information on the costs of improved secondary preventive care in the counterfactual case, we assume that its costs are the same as the costs of secondary preventive care in the data.

[16] See "Submitting to the Science of Prevention", Wall Street Journal, Nov. 26, 2008 for an example of individualized health care plan. http://online.wsj.com/article/SB122765661371658079.html

Table 9 presents the results on the change in consumers' insurance plan choices and the change in overall healthcare costs of each segment in the counterfactual case vis-à-vis the baseline case. The pattern of results is similar to that in Table 8. The plan choice probabilities are similar between the baseline and counterfactual cases in 2005, and there is a much greater divergence in the plan choice probabilities across the two cases in 2006. We next compare the change in the overall costs in 2006 across the two cases. On a segment level basis, the overall healthcare costs for segments A, B and D are marginally lower in the counterfactual case as compared to the baseline case. On the other hand, the overall healthcare costs for the moderately healthy segment C are substantially lower in the counterfactual case as compared to the baseline case. The reason for why that is so follows from our earlier discussion. Providing more accurate information decreases the consumers' uncertainty vis-à-vis their health status; and this in turn decreases the probability that consumers in the moderately healthy segment C will purchase more comprehensive plans than what they actually need. This can be seen in Table 9 – notice that by offering more accurate preventive care, the probability of choosing the comprehensive plan in 2006 reduces from 43.3% to 38.3% and the probability of choosing the medium plan increases from 34.1% to 39.4%.

Section 7: Conclusion

In this paper, we propose a dynamic structural model that nests health care consumption into health insurance demand, building upon Grossman's (1972) health investment model. Our model extends the literature by modeling the consumption decisions in terms of preventive and curative care, incorporating investment and informative roles of healthcare services, jointly modeling consumption and insurance plan decisions in a dynamic nested framework and accounting for heterogeneity in consumers' price sensitivities, uncertainty, health status and degradation. Doing so is important not just as a methodological advancement, but also in terms of properly identifying moral hazard from selection and examining the impact of external interventions on the extent of moral hazard.

We apply our model to a novel dataset provided by a large insurance company, which includes consumers' insurance purchases and healthcare consumption decisions over time. Our key results are as follows. *First*, the variation in insurance plan decisions across consumers stems more from heterogeneity in consumers' health status, their uncertainty vis-à-vis their health status and the degradation in their health status over time as compared to heterogeneity in price sensitivities. *Second,* the source of moral hazard is the presence of a sizable segment (13%) of consumers with moderate health status, low price sensitivity and a high risk premium due to high uncertainty in their health status. Instead of buying the medium plan that better matches their health status, these consumers purchase the comprehensive plan; and once in the plan, they opt to go for more expensive curative care even when the illness could be managed through primary preventive care. *Third*, changing the cost sharing characteristics of insurance plans to induce these consumers to switch to medium plans in the same policy year may not work. This is because they have low elasticities in insurance plan decisions with respect to cost sharing characteristics. A more feasible route for decreasing moral hazard is the 'delayed route' in which we incentivize consumers to purchase more secondary preventive care, which reduces their uncertainty and thereby makes them choose the medium plan in the future. We conduct two counterfactuals in the context of the delayed route in which we examine the impact of reducing the copayment for secondary preventive care and providing more accurate information through secondary preventive care on reducing the extent of moral hazard.

We view our paper as the first attempt in the marketing literature to empirically model an *extremely complex* decision making process of insurance plan and healthcare consumption decisions. Consequently, we make a set of simplifying assumptions and relaxing these would a fruitful avenue for future research. Some of our key assumptions are as follows.

First is the aggregation of treatments into a single healthcare option. Ideally, we would have only considered consumers suffering from the same and only one chronic disease in which there are a relatively homogenous set of treatments. Doing so would allay concerns related to aggregation of costs across treatments. However we did not follow this route for three reasons: (a) since we aggregate costs across treatments using a weighted average measure, we do not see a large variation in costs of treatments that fall under a given consumption option (i.e., curative, primary preventive and secondary preventive) for a given disease. (b) If we were to restrict our attention to only those consumers who are suffer from the same chronic disease and consume a homogenous set of treatments with similar costs, we will be left with a very small sample size. The only way to increase the sample size in such a case will be to procure a raw data with more than 100,000 consumers. However, that is prohibitively expensive to procure. (c) Such aggregation of prices is fairly common in the marketing literature. For instance, most prior papers that have used scanner data set have aggregated the prices all flavors and sizes of a brand into a composite brand price or have aggregated the prices of all SKUs within a category into a category composite price.

Second, we assume that it is the patients and not the doctors who make the consumption decisions. This is a standard assumption in the literature. Note that the reality may be that the consumption decision is made jointly by the doctor and the consumer or is made solely by the doctor. If either were the case, the welfare implications could be different. However, we did not pursue this since it is not possible to test these alternative assumptions from the data.

Third, we assume the degradation in health status to be deterministic and not stochastic. Consequently, we do not allow for health shocks that persist over time (the only health shocks that we allow are through the IID econometrician's errors that are observed by the consumer and do not persist over time). We do not model these time persistent health shocks for two reasons: (a) doing so will not change the basic nature of our results, and (b) these health shocks can be only be cleanly identified under restrictive set of assumptions. To understand (b), note that while modeling the time persistent health shocks, it is crucial to first know whether and to what extent does the consumer get to observe the draws of these health shocks in absence of secondary preventive care (i.e., diagnostic tests). The most likely answer is that the consumer would get to 'partially' observe the actual draws of these health shocks in absence of diagnostic tests. In other words, she would be able to observe a part of the health shock by herself (i.e., she would not need a diagnostic test to know the value of a part of the health shock, since she can assess its value herself), but would not be able to observe the other part of the health shock by herself (i.e., she would need a diagnostic test to know the value of the other part of the health shock) [17]. However the conundrum here is that if we allow for partial observability, we would need to identify 'the distribution of the health shocks that the consumer can observe by herself' as well as 'the distribution of health shocks that the consumer cannot observe by herself'. Note that it is easier to identify 'the distribution of time persistent health shocks that the consumer cannot observe by herself' compared to 'the distribution of time persistent health shocks that the consumer can observe by herself'. This is because 'the distribution of time persistent health shocks that the consumer can observe by herself' can only be identified from the randomness in the usage of investment treatments over time[18]; however, that is not as clean as we would like it to be since the learning errors and the econometrician's errors also increase the randomness in usage of investment treatments. On the other hand, 'the distribution of time persistent health shocks that the consumer cannot observe by herself' can be identified because these health shocks only serve to increase the consumer's uncertainty vis-à-vis her health status over time. Since the increase in uncertainty over time increases the purchase probability of diagnostic tests over time, the increase in

[17] This is because if a consumer suddenly falls sick, she would know that she is sick, even though she may not know the full extent of her sickness. This implies that she would observe the health shocks herself (since she knows that she is sick), but not to a full extent (since she may not know the full extent of her sickness unless she gets a diagnostic test).

[18] This is because the health shocks that the consumer can observe by herself only serve to increase the randomness in her mean health status (h_t) over time. This randomness in turn manifests itself as the randomness in the usage of investment treatments over time.

uncertainty in each period (and consequently the distribution of these health shocks) can be identified by the slope of the hazard function that captures the timing of consumer's purchases of diagnostic tests (which is an argument similar to the identification argument for the degradation rate δ_d given in section 4.3). This implies that the clean identification is possible if we assume that in absence of diagnostic tests, consumers do not get to observe the actual draws of the health shocks.

References

Arrow, K. (1963), "Uncertainty and the welfare economics of health care," *American Economic Review*, 53(5), 940–73

Aron-Dine, A., L. Einav, A. Finkelstein and M. Cullen (2012), "Moral Hazard in Health Insurance: How important is Forward-Looking Behavior," working paper, Stanford University

Bajari, P., H. Hong, A. Khwaja and C.A Marsh (2012), "Moral Hazard, Adverse Selection and Health Expenditures: A Semi-parametric Analysis," working paper, University of Minnesota

Carlin, C. and R. Town (2008), "Adverse Selection, Welfare and Optimal Pricing of Employer-Sponsored Health Plans," working paper, University of Minnesota

Cameron, A., P. K. Trivedi, F. Milne and J. Piggott (1988), "Micro-econometric Model of Demand for Healthcare and Health Insurance in Australia," *Review of Economic Studies*, 55(1), 85-106

Cardon, J. and I. Hendel (2001), "Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey," *RAND Journal of Economics*, 32(3), 408–427.

Chan, T. Y and B. Hamilton (2006), "Learning, Private Information, and the Economic Evaluation of Randomized Experiments," *Journal of Political Economy* 114(6), 997–1040

Chiappori P. A. and B. Salanié (2000), "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108(1), 56-78.

Ching A. and M. Ishihara (2010), "The Effects of Detailing on Prescribing Decisions under Quality Uncertainty," *Quantitative Marketing and Economics*, 8(2), 123-165

Crawford, G. and M. Shum (2005), "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73(4), 1137-1173.

David Cutler, D., A. Finkelstein and K. McGarry (2008), "Preference heterogeneity in Insurance Markets: Explaining a Puzzle of Insurance," *American Economic Review Papers and Proceedings,* 98(2): 157-162

Einav, L., A. Finkelstein, S. Ryan, P. Schrimpf and M. Cullen (2013), "Selection on Moral Hazard in Health Insurance," *American Economic Review*, 103(1), 178-219.

Erdem, T. and M. P. Keane (1996), "Decision Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 16(1), 1-20.

Erdem, T., M. P. Keane, T.S. Oncü and J. Strebel (2005), "Learning about Computers: Analysis of Information Search and Technology Choice," *Quantitative Marketing and Economics*, 3, 207-46

Fang, H. and A. Gavazza (2011), "Dynamic Inefficiencies in Employment based health Insurance System: Theory and Evidence," *American Economic Review*, 101(7), 3047-3077

Fang H., M. P. Keane and D. Silverman (2008), "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market," *Journal of Political Economy*, 16(2), 303-350.

Fang H., L. Nicholas and D. Silverman (2013), "Cognitive Ability and Retiree Health Care Expenditure," working paper, University of Pennsylvania

Felder, S. and Mayrhofer, T. (2011), "Medical Decision Making: Health Economic Primer," Springer

Finkelstein, A. and K. McGarry (2006), "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *American Economic Review*, 96(4), 938-958.

Grossman, M. (1972), "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2), 223-55

Grossman, M. and E. Rand (1974), "Consumer Incentives for Health Services in Chronic Illness," *Consumer Incentives for Health Care*, New York: Prodist for Milbank Memorial Fund, 114-151.

Handel, B (2013), "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," forthcoming *American Economic Review*

Handel, B. and J. Kolstad (2013), "Health Insurance for Humans: Information Frictions, Plan Choice and Consumer Behavior." working paper, University of California at Berkeley

Heckman, J. and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions," *Econometrica*, 52(2), 271-320.

Hsieh, C. and S. Lin (1997), "Health Information and the Demand for Preventive Care among the Elderly in Taiwan," *Journal of Human Resources*, 32(2), 308-333.

Jain, D. and N. Vilcassim (1991), "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach," *Marketing Science*, 10(1), 1-23.

Kenkel, D. S (2000), "Prevention," *Handbook of Health Economics*, 1675-1720.

Khwaja, A. (2001), "Health Insurance, Habits and Health Outcomes: A Dynamic Stochastic Model of Investment in Health," Ph.D. Dissertation, University of Minnesota

Lin S., J. Zhang and J. Hauser (2014), "Learning from experience, simply," working paper, Sloan School of Management, MIT

Manning, W., J. Newhouse, N. Duan, E. Keeler, A. Leibowitz (1987), "Health Insurance and Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review*, 77(3), 251-277.

Mehta, N., S. Rajiv and K. Srinivasan (2003), "Price uncertainty and consumer search: A Structural Model of Consideration Set Formation," *Marketing Science*, 22(1), 58-84.

Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold

Zurcher," *Econometrica*, 55(5), 999-1033

Sun, B. (2005), "Promotion Effect on Endogenous Consumption," *Marketing Science*, 24 (3), 430-443

Thorpe, K. E. (2008), "U.S. Senate Committee on Health, Education, Labor, and Pensions Hearing Prevention and Public Health: The Key to Transforming our Sick-care System."

Vera-Hernandez, M. (2003), "Structural Estimation of a Principal-Agent Model: Moral Hazard in Medical Insurance," *RAND Journal of Economics*, 34(4), 670-93

Zhang, J. (2010), "The Sound of Silence: Observational Learning in the U.S. Kidney Market," *Marketing Science*, 29 (2), pp. 315-335.

Zweifei, P. and W. Manning (2000), "Moral Hazard and Consumer Incentives in Health Care," *Handbook of Health Economics*, 1(8), Amsterdam: North Holland.

FIGURES AND TABLES

Figure 1: Cost Sharing for Individuals during Consumption Stage

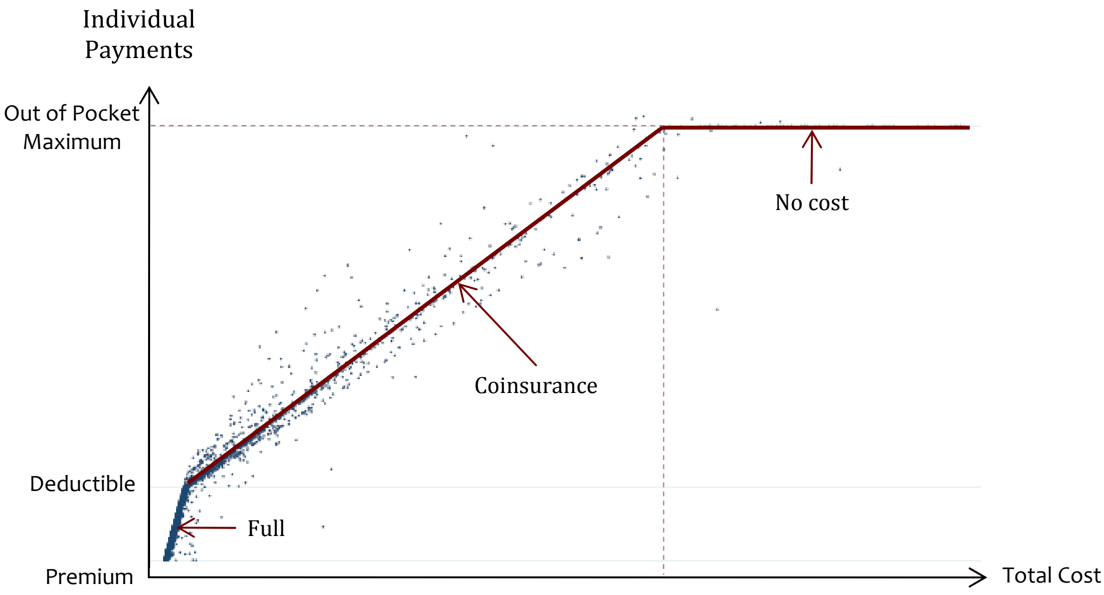

Figure 2: Timeline of Sequence of Health Insurance/Health Care Decision

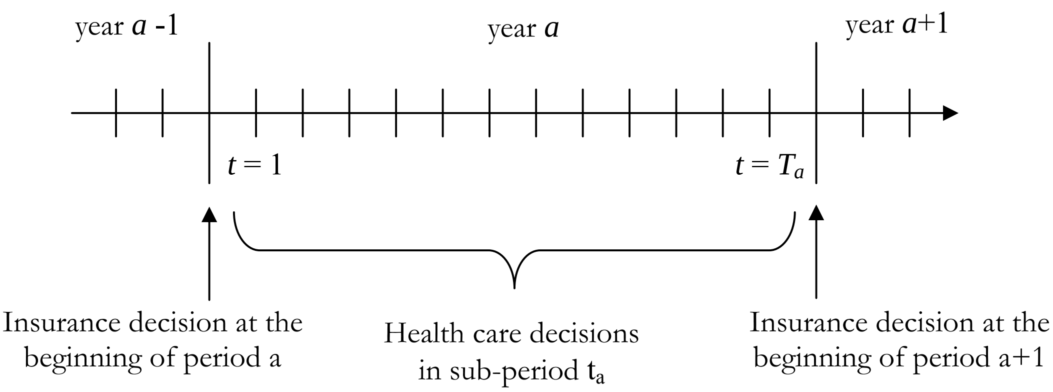

Figure 3: Illustration of Degradation and Investment/Information Effects in period *t*

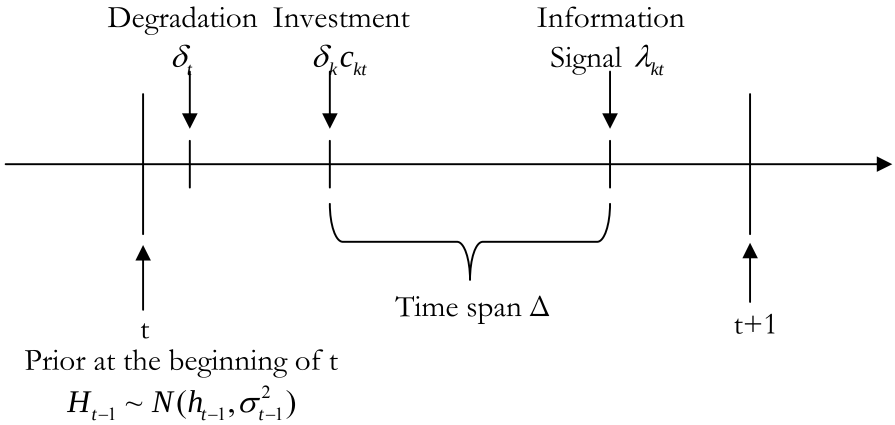

Table 1A: Examples of Chronic Diseases (National Institutes of Health)

Chronic Disease	Morbidity
Heart Attacks and Strokes	1.5 million Americans suffer each year
Breast Cancer	2.8 million
Diabetes	25.8 million (ages 20 or older)
Prostate Cancer	2.6 million men
Hypertension	77.9 million adults

Table 1B: EXAMPLES OF PREVENTIVE AND CURATIVE CARE

Secondary Preventive	Primary Preventive	Curative
Blood pressure test, Cholesterol test, Pap smear, preconception care programs for women with diabetes, mammogram, Prostate-Specific Antigen (PSA) Test, annual checkup, Electrocardiogram (ECG, EKG), Cardiovascular Stress Test, Electrocardiographic Monitoring for 24 hours, Echocardiography, Coronary Angiography, Thallium Stress Test, Cardiac CT angiography, Breast Ultrasound, Breast MRI, Bone Scan	Pneumococcal, Tetanus, HepA, simvastatin, Lipitor, Crestor, Zocor, pravastatin, lovastatin, gemfibrozil, Pravachol, Simcor, Lopid, atorvastatin, Caduet, rosuvastatin, Mevacor, Advicor, Lescol XL, Lescol, amlodipine-atorvastatin, niacin-simvastatin, Altoprev, fluvastatin, niacin-lovastatin, chlorthalidone, methyclothiazide, bromocriptine, Kombiglyze XR, Prandin	Coronary Angioplasty, Coronary Stent Placement, Coronary Artery Bypass, Mastectomy simple complete, Excision of breast lesion, Mastectomy partial (lumpectomy), Mastectomy radical, Mastectomy modified radical, Radiotherapy, Prostatectomy, External beam radiation, Brachytherapy, Orchiectomy, Cryoablation, Colonoscopy with polyp removal, Radical surgical resection, Hemicolectomy, Colectomy.

Table 2A: Summary Statistics of Consumption Options for some Chronic Diseases

		Type 2 Diabetes	Heart Disease	Prostate Cancer
Secondary Preventive Care	Number of different treatments used by consumers in the data	5	6	4
	Weighted Average cost of treatments (Std. Dev. of costs)	159.2 (21.3)	410.5 (29.9)	408.6 (24.9)
Primary Preventive Care	Number of different treatments used by consumers in the data	8	10	6
	Weighted Average cost of treatments (Std. Dev. of costs)	458.9 (22.4)	862.7 (37.8)	752.5 (37.1)
Curative Care	Number of different treatments used by consumers in the data	4	8	4
	Weighted Average cost of treatments (Std. Dev. of costs)	3423.7 (112.9)	5382.9 (153.6)	3682.4 (129.4)

Table 2B: Percentage of Plan Choices

	Basic	Medium	Comprehensive
2005	22.15%	33.92%	43.92%
2006	22.58%	34.06%	43.35%
2007	21.66%	35.72%	42.62%

Table 2C: Switching Patters of Insurance Purchases

2005 \ 2006	Basic	Medium	Comprehensive
Basic	91.49%	3.21%	5.30%
Medium	2.89%	94.31%	2.80%
Comprehensive	3.03%	3.08%	93.89%

2006 \ 2007	Basic	Medium	Comprehensive
Basic	91.09%	3.34%	5.57%
Medium	1.88%	93.26%	4.86%
Comprehensive	1.05%	7.39%	91.56%

Table 2D: Summary Statistics across Insurance Plans

	Basic	Medium	Comprehensive
Premium ($)	873.2	1683.8	2797.5
Copayment ($)	19.2	29.8	15.8
Coinsurance rate (%)	27	14	2
Deductible ($)	1008.8	356.8	112.6
Out-of-pocket maximum ($)	3107.6	1020.5	275.3
Avg. no. of primary preventive care visits/year	8.7	12.7	15.3
Avg. no. of secondary preventive care visits/year	1.4	1.5	2.2
Avg. No. of curative care visits/year	0.9	1.3	4.5
Total Cost paid by consumer and insurer for primary preventive care/year ($)	150.2	340.1	1683.7
Total Cost paid by consumer and insurer for secondary preventive care/year ($)	28.1	28.1	54.9
Total Cost paid by consumer and insurer for curative care/year ($)	1427.5	3928.5	40327.6

Table 3A: Distribution of consumers along two dimensions of usage of secondary preventive care: (a) Number of weeks lapsed from {*t*=1, *a*=1} till the period when the consumer first consumes secondary preventive care, and (b) the number of times secondary preventive care is used after its first usage in the data

% of Consumers	Number of times secondary preventive care is used after its first usage in the data						
		0-1	2-3	4-5	6-7	8-9	10-11
Number of weeks lapsed from {*t*=1, *a*=1} till the period when the consumer first consumes secondary preventive care	1-8	0.10%	1.33%	1.67%	3.43%	0.52%	0.38%
	9-17	0.48%	1.05%	7.44%	10.34%	1.48%	0.81%
	18-26	0.33%	1.62%	6.82%	7.72%	0.52%	0.48%
	27-35	0.62%	0.67%	4.43%	5.29%	1.14%	0.62%
	36-44	0.05%	0.10%	14.97%	10.25%	2.62%	0.57%
	45-53	1.33%	0.00%	4.81%	5.48%	0.29%	0.10%
	54-62	0.00%	0.10%	0.00%	0.00%	0.05%	0.00%

Table 3B: Distribution of consumers along two dimensions of usage of primary preventive care: (a) Number of weeks lapsed from {t=1, a=1} till the period when the consumer first consumes primary preventive care, and (b) the number of times primary preventive care is used after its first usage in the data

% of consumers	Number of times primary preventive care is used after its first usage in the data								
Number of weeks lapsed from {t=1, a=1} till the period when the consumer first consumes primary preventive care		0-7	8-15	16-23	24-31	32-39	40-47	48-54	55-63
	1-8	0.98%	0.33%	0.51%	1.02%	16.39%	5.87%	1.02%	0.09%
	9-17	0.14%	0.09%	0.51%	0.98%	9.36%	11.92%	0.05%	0.14%
	18-26	0.14%	0.51%	0.33%	0.61%	17.00%	13.88%	0.37%	0.33%
	27-35	0.28%	0.14%	1.16%	0.65%	7.27%	4.61%	1.16%	0.79%
	36-44	0.05%	0.00%	0.00%	0.00%	0.19%	0.33%	0.42%	0.05%
	45-53	0.09%	0.09%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%
	54-62	0.00%	0.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 3C: Distribution of consumers along two dimensions of usage of curative care: (a) Number of weeks lapsed from {t=1, a=1} till the period when the consumer first consumes curative care, and (b) the number of times curative care is used after its first usage

% of consumers	Number of times curative care is used after its first usage in the data								
Number of weeks lapsed from {t=1, a=1} till the period when the consumer first consumes curative care		0-1	1-3	4-5	5-6	7-8	9-10	11-12	13-14
	1-8	1.53%	6.41%	8.76%	0.96%	0.26%	0.13%	0.35%	0.09%
	9-17	0.65%	5.27%	7.49%	0.09%	0.04%	0.00%	0.00%	0.04%
	18-26	1.18%	8.24%	7.67%	2.75%	0.96%	0.04%	0.35%	0.04%
	27-35	0.48%	0.87%	8.15%	7.19%	1.92%	1.09%	0.04%	0.00%
	36-44	0.31%	1.61%	6.36%	1.44%	0.00%	0.31%	0.04%	0.00%
	45-53	0.17%	1.05%	1.22%	3.75%	3.01%	0.35%	0.35%	0.09%
	54-62	0.61%	3.88%	0.96%	0.52%	0.17%	0.26%	0.35%	0.17%

Table 4: Model Comparison (N=147512)

	Model I (proposed model)	Model II	Model III
-LL	26,412.6	26842.5	27143.8
AIC	52,814.8	53027.1	54152.7
BIC	52,751.9	53045.2	54178.2

Table 5: Parameter Estimates of the Proposed Model

Parameter	Estimate (Standard Error)			
	Segment A	Segment B	Segment C	Segment D
Segment size	0.225 (0.042)	0.315 (0.035)	0.129 (0.044)	0.331 (0.027)
Mean of Prior Beliefs of Health Status at $\{t=1, a=1\}$, h_0	2.732 (0.512)	1.632 (0.319)	1.612 (0.421)	0.892 (0.201)
Std. Dev. of Prior Beliefs of Health Status at $\{t=1, a=1\}$, σ_0	2.724 (1.008)	2.812 (0.782)	3.986 (0.914)	2.887 (1.105)
Degradation, δ_d	0.0005 (0.0002)	0.0012 (0.0006)	0.0012 (0.0006)	0.0053 (0.0015)
Price Coefficient, α	0.168 (0.042)	0.188 (0.084)	0.131 (0.051)	0.099 (0.038)
Informative Effect: Standard deviation of the consumption Signal of Secondary Preventive care, σ_{sp}	7.143 (1.802)			
Investment Effect of Primary Preventive care, δ_{pp}	0.014 (0.005)			
Investment Effect of curative care, δ_c	0.063 (0.015			

Table 6A: Price Elasticities of Insurance Choice

Seg. A / Seg. C	Seg. B / Seg. D	Insurance Purchase	
Copay of comprehensive plan		0.081	0.102
		0.061	0.054
Deductible of comprehensive plan		0.107	0.145
		0.081	0.073
Coinsurance of comprehensive plan		0.115	0.139
		0.101	0.069
Premium of comprehensive plan		0.054	0.066
		0.042	0.034

Table 6B: Price Elasticities of Insurance Choice

Seg. A / Seg. C	Seg. B / Seg. D	Insurance Purchase	
Copay of medium plan		0.109	0.135
		0.081	0.074
Deductible of medium plan		0.126	0.152
		0.098	0.084
Coinsurance of medium plan		0.134	0.179
		0.115	0.097
Premium of medium plan		0.069	0.088
		0.054	0.049

Table 6C: Price Elasticities of Insurance Choice

Seg. A / Seg. C	Seg. B / Seg. D	Insurance Purchase	
Copay of basic plan		0.125	0.161
		0.086	0.079
Deductible of basic plan		0.152	0.191
		0.104	0.092
Coinsurance of basic plan		0.161	0.191
		0.123	0.099
Premium of basic plan		0.085	0.107
		0.061	0.052

Table 7: Price Elasticities of Consumption

Seg. A	Seg. B	Health Care Consumption					
Seg. C	Seg. D	Secondary Preventive		Primary Preventive		Curative	
Copay		0.529	0.552	0.487	0.515	0.369	0.398
		0.352	0.316	0.374	0.307	0.222	0.216
Deductible		0.451	0.469	0.434	0.471	0.309	0.334
		0.264	0.213	0.291	0.242	0.182	0.169
Coinsurance		0.365	0.402	0.397	0.421	0.389	0.423
		0.205	0.198	0.231	0.221	0.242	0.211

Table 8: Decreasing Copayment of Secondary Preventive Care by 50%

Plan	Plan choice probability in 2005: Baseline	Plan choice probability in 2006: Baseline	Plan choice probability in 2005: counterfactual	Plan choice probability in 2006: counterfactual
Basic	22.2%	22.6%	22.3%	22.5%
Medium	33.9%	34.1%	34.4%	35.4%
Comp.	43.9%	43.3%	43.3%	42.1%

Segment (size)	Change in Healthcare Costs between baseline and counterfactual in 2006
A (22.5%)	+0.7%
B (31.5%)	+0.4%
C (12.9%)	-6.5%
D (33.1%)	-1.7%

Table 9: Personalized Medicine: Increasing Informational Value of Secondary Preventive Care

Plan	Plan choice probabilities in 2005: Baseline	Plan choice probabilities in 2006: Baseline	Plan choice probabilities in 2005: counterfactual	Plan choice probabilities in 2006: counterfactual
Basic	22.2%	22.6%	22.3%	22.3%
Medium	33.9%	34.1%	35.8%	39.4%
Comp.	43.9%	43.3%	41.9%	38.3%

segment (size)	Change in Healthcare Costs between baseline and counterfactual in 2006
A (22.5%)	-0.9%
B (31.5%)	-2.3%
C (12.9%)	-22.7%
D (33.1%)	-3.9%

Appendix I: Aggregation of Costs

We discuss how we aggregate the costs of all treatments (where the cost of a treatment is the cost incurred by the consumer and the cost incurred by the insurance firm) that fall under a given consumption option into a single aggregate cost of that consumption option. The process is similar to what prior papers have done to operationalize the aggregate prices of a category or a brand when using scanner data sets.

First consider a consumer suffering from only one chronic disease. If the consumer chooses a treatment under consumption option *k* (where *k* can be primary preventive, secondary preventive or curative) in period *t*, then the cost of option *k* is taken as the cost of the chosen treatment in period *t* as recorded in the data. For a non-chosen consumption option *k* in period *t*, its cost is taken as the weighted average cost of all treatments that fall under the option *k*, where the weights are the percentage of times the consumer has used each of the treatments that fall under the option *k*, during her entire history. If the consumer has never consumed option *k* during her entire observed history, then the cost of option *k* is taken as the cost of the most popular treatment that falls under option *k* for the disease that the consumer is suffering from.

Next consider a consumer suffering from multiple chronic diseases. Similar to the case before, if the consumer choses a treatment under option *k* in period *t*, then the cost of option *k* is taken as the cost of the chosen treatment in period *t* as recorded in the data. For a non-chosen consumption option *k* in period *t*, the cost is calculated as follows. If the consumer in that period *t* does consume some other consumption option *k*', then the cost of option *k* is taken as the weighted average cost of treatments that fall under the option *k* for the same disease as the one that option *k*' treats (where the weights are the percentage of times the consumer has used each of the treatments during her entire history). If the consumer in that period *t* chooses the no consumption option, then the cost of option *k* is taken as the weighted average cost of all treatments that fall under the option *k* for all chronic diseases the consumer is suffering from.

Appendix II: Approximation and Interpolation of $E\max_t$ Functions

We adapt the Keane and Wolpin (1994) approximation method (originally proposed for continuous choice setting by Bellman et al., 1963) to our model. It is applicable when the state space is large either because there are a large number of discrete state variables or due to the presence of continuous state variables (or both). In this approach the $E\max_t$ functions are computed at a subset of the state points and some method of interpolation is used to evaluate $E\max_t$ at other values of the state space.[19] We summarize our estimation by heavily borrowing in the notation and explanation of Keane et al (2011).

We begin to describe the choice set of our problem. The choice set during each insurance period (annually) a till the last week T_a (except the first week) consists of four consumption alternatives. For each year, the insurance choices include three different plans. There are thus the

[19] Keane and Wolpin (1994) provide Monte Carlo study evidence on the effect of the approximation error on the bias of the estimated model parameters under alternative interpolating functions and number of state points, though there is no formal proof of the proposition. Intuitively, as the subset of the state points that are chosen is enlarged and the dimension of the approximating function is increased, the approximation will converge to the true solution. Marketing and economics literature has extensive application of this method including Erdem and Keane (1996) and Crawford and Shum (2005).

following twelve mutually exclusive choices, given by $D_{it}^{kj} = \{D_{it}^{01}, D_{it}^{11}, D_{it}^{21}, D_{it}^{31}, D_{it}^{02}, D_{it}^{12}, D_{it}^{22}, D_{it}^{32}, D_{it}^{03}, D_{it}^{13}, D_{it}^{23}, D_{it}^{33}\}$, where the first superscript refers to the insurance choice ($j = \{1,2,3\}$) and the second to the consumption choice ($k = \{0,1,2,3\}$). Note, insurance choice j only needs to make at the end of each policy year (the beginning of the next policy year), which mean the choice set only includes the four consumption choices. This helps simplifies the computation of the value function of consumption.

A consumer's objective function is to choose the mutually exclusive alternative at each t_a of year a that maximizes the expected discounted value of the lifetime utility. Define U_{it}^{kj} to be the contemporaneous utility flow for the consumption (conditional on the insurance choice at year a), the alternative-specific value functions for the consumption problem are

$$V_t^{kj}(S_{it}) = U_{it}^{kj}(S_{it}) + \beta E[V_{t+1}(S_{it+1}) \mid S_{it}, D_{it}^{kj}] \qquad \text{(A-1)},$$

where the expectation is taken over the joint distribution of the preference shocks, $f(\varepsilon_t^0, \varepsilon_t^1, \varepsilon_t^2, \varepsilon_t^3)$, which are mutually serially independent. Let $\widetilde{V_t^{kj}}$ be the vector of alternative specific value functions at period t of year a,

$$V_t\ (S_{it}) = \max(\widetilde{V_t^{kj}}(S_{it})) \qquad \text{(A-2)}.$$

The model is solved by backward recursion. The solution requires that the $E\max_t$ function be calculated at different state points and for all t. Consider first the calculation of the $E\max_t$ for any given state space element. At period τ of year A the consumer no longer consumes insurance, so we need to calculate

$$E\max_{\tau_A} = E_{\tau-1_A} \max_{k|j_A}(U_{\tau_A}^{kj_A}(\tilde{\varepsilon})) \qquad \text{(A-3)},$$

where $\tilde{\varepsilon}$ is the vector of shocks.

$E\max_{\tau_A}$ can be calculated using Monte Carlo integration. Let $\widetilde{\varepsilon_m}$ be the m^{th} random draw, $m = 1,\ldots,M$, from the joint distribution, $f(\varepsilon_t^0, \varepsilon_t^1, \varepsilon_t^2, \varepsilon_t^3)$, an estimate of $E\max_{\tau_A}$ at say the n^{th} value of the state space in S_{τ_A}, $S_{\tau_A}^n$, is

$$\widehat{E\max_{\tau_A}^n} = \frac{1}{M}\sum_{m=1}^{M} \max_j [U_{\tau_A}^{0j}(\widetilde{\varepsilon_m}; S_{\tau_A}^n), U_{\tau_A}^{1j}(\widetilde{\varepsilon_m}; S_{\tau_A}^n), U_{\tau_A}^{2j}(\widetilde{\varepsilon_m}; S_{\tau_A}^n), U_{\tau_A}^{3j}(\widetilde{\varepsilon_m}; S_{\tau_A}^n)] \qquad \text{(A-4)}.$$

Because it is infeasible to calculate $\widehat{E\max_{\tau_A}}$ at all points in the state space, we randomly draw N state points (without replacement) and calculate the $\widehat{E\max_{\tau_A}}$ function for these N state space elements as shown above. We then treat these N values of $\widehat{E\max_{\tau_A}}$ as a vector of dependent variables in an interpolating regression

$$\widehat{E\max_{\tau_A}} = g_{\tau_A}(S_{\tau_A}^n; \gamma_{\tau_A}) + \varsigma_{\tau_A}^n \qquad \text{(A-5)},$$

where γ_{τ_A} is a time τ vector of regression coefficients and $g_{\tau_A}(\cdot;\cdot)$ is a flexible function of state variables. With this interpolation function, estimates of the $E\max_{\tau_A}$ function can be obtained at any state point in the set S_{τ_A}.

Given $\widehat{E\max}_{\tau_A}$, we can similarly calculate $V^{kj}_{\tau-1_A}$ at a subset of the state points in $S_{\tau-1_A}$. Continuing this procedure, we can obtain the interpolating functions for all of the $\widehat{E\max}_t$ functions for all $t<T$ where consumers only need to make the consumption choices (conditional on the insurance choices).

At $t=T$ of each policy year a, the choice set now includes the insurance choice. Similarly, define VI^j_{ia} to be the contemporaneous utility flow for the insurance choices when $t=T_a$, the alternative-specific value functions for the insurance problem are

$$W^j_a(S_{ia}) = VI^j_{ia}(S_{ia}) + \beta E[W_{a+1}(S_{ia+1}) \mid S_{ia}, D^j_{ia}] \qquad \text{(A-6)},$$

where, letting $\widetilde{W^j_a}$ be the vector of alternative specific value functions relevant at year a,

$$W_a(S_{ia}) = \max(\widetilde{W^j_a}(S_{ia})) \qquad \text{(A-7)}.$$

At year A the consumer no longer buys insurance, so we need to calculate $E\max_A = E_{A-1}\max_j(VI^j_A(\tilde{\varphi}))$, where $\tilde{\varphi}$ is the vector of insurance shocks. All of the $E\max_a$ functions at $a=1,\ldots,A$ requires numerical integrations over the 12 mutually exclusive choices based on the joint error distribution of consumption shocks $\tilde{\varepsilon}$ and insurance shocks $\tilde{\varphi}$. Following the same logic above, we could simulate and interpolate all of the $\widehat{E\max}_a$ functions.

Here the set of values of the ε and φ vectors determine optimal choices and serving as limits of integration in the probabilities associated with the insurance and consumption alternatives that comprise the likelihood function have no analytical form and the likelihood function requires a multivariate integration. We follow the literature to use the simulation methods for the maximum likelihood estimation. Although these frequency simulations converge to the true probabilities as draws increase, the likelihood is not smooth in the parameters, which precludes the use of the derivative methods (e.g., BHHH). This lack of smoothness forces the use of the non-derivative methods, which converge more slowly, we follow McFadden (1989) to smooth the simulator.

Reference:

1. Bellman, R., R. Kalaba, B. Kootkin. 1963. Polynomial approximation: a new computational technique in dynamic programming allocation processes. *Mathematics of Computation* 1, 155–161.
2. Keane, M., K. Wolpin. 1994. The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *Review of Economics and Statistics* 76(4) 648-672.
3. Keane, M. P., P. E. Todd, K. Wolpin. 2011. The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications. *Handbook of Labor Economics*, Chapter 4. 331-461.
4. McFadden, D., 1989. A method of simulated moments for estimation of discrete response models with numerical integration. *Econometrica* 57(5) 995–1028.