Estimating Switching Costs for Medicare Advantage Plans*

Kathleen Nosal

JOB MARKET PAPER

This draft: October 2011

First draft: April 2009

Abstract

Medicare eligibles have the option of choosing from a menu of privately administered managed care plans, known as Medicare Advantage (MA) plans, in lieu of conventional fee-for-service Medicare coverage ("original Medicare"). These plans often provide extra benefits to enrollees, but may impose large switching costs as a result of restrictive provider networks, differences in coverage across plans, and learning and search costs. I propose a structural dynamic discrete choice model of how consumers who are persistently heterogeneous make the choice among MA plans and original Medicare based on the characteristics of the available MA plans. The model explicitly incorporates a switching cost and changes over time in choice sets and plan characteristics. I estimate the parameters of the model, including the switching cost, using the methods developed by Gowrisankaran and Rysman (2011). The estimates indicate that the switching cost is statistically and economically significant. Through a series of counterfactual analyses, I find that the share of consumers choosing MA plans in place of original Medicare would more than triple in the absence of switching costs, and nearly double if plan exit and quality changes were eliminated. I also find that when switching costs are accounted for the Medicare Advantage program is not very valuable to consumers and even reduces consumer welfare in some years.

*I am grateful to my main advisor, Gautam Gowrisankaran, who has been an excellent source of ideas, suggestions, and encouragement. The input of Kei Hirano, Mo Xiao, and Keith Joiner has also been invaluable. Paris Cleanthous and other participants at the CEPR IO School made helpful comments, as did participants in the University of Arizona Empirical Workshop, including Mauricio Varela, Tiemen Woutersen, and Mario Samano. Josh Lustig kindly shared data for this paper.

1 Introduction

Medicare beneficiaries have the option to choose from a menu of privately administered Medicare Advantage (MA) plans available as an alternative to traditional fee-for-service Medicare coverage ("original Medicare"). In general, the MA plans provide more extensive coverage than original Medicare, but their combined market share has remained low. One explanation for the low market share of the MA plans is the high switching costs in this market. A consumer switching into an MA plan from original Medicare or another MA plan may have to switch doctors because of the plan's provider network, may have to switch treatments because of different coverage under the new plan, and will likely experience some learning and hassle costs. Furthermore, plans can exit the market or reduce the benefits offered in subsequent years, so consumers choosing MA plans may have to switch multiple times to stay in a desirable plan. Since original Medicare is the only plan that is guaranteed to be offered in every period and with essentially unchanging coverage, it is a safe haven for consumers seeking to avoid the switching cost. Therefore, consumers may be choosing not to enroll in otherwise very desirable MA plans because of a combination of high switching costs and knowledge of the consequences of enrolling in a plan that may exit the market or decrease in quality in the future.

In this paper, I develop and estimate a structural model of discrete choice demand for the MA plans.[1] Unlike previous models of demand for MA plans, mine is dynamic and allows for switching costs when consumers change plans.[2] Since switching costs and consumers' rational beliefs about the future of the plans are both crucial determinants of choice in this market, these features of the model contribute to a more complete picture of this market than could be established with a static model. The model also allows for persistent consumer heterogeneity, and incorporates detailed information about the benefits of each MA plan offered.

The importance of the switching cost is borne out in the estimation results. I find a median switching cost of about $4000. While large enough to imply that the switching cost has a significant influence on consumer choice, this estimate is not incomparable with other estimates of health plan switching costs. Using the estimates of the structural parameters of the model, I calculate market shares and consumer welfare in several counterfactual scenarios. I find that either eliminating the switching cost or eliminating exit and quality change of plans dramatically increases the market shares of MA plans at the expense of the original Medicare share and increases consumer welfare.

An important question about the Medicare Advantage program since its inception has been whether it is worth its considerable cost. Medicare as a whole is a very expensive program, costing $452 billion or 13% of the Federal budget in 2010 (Center on Budget and Policy Preparedness, 2011). Furthermore, MA plans are a more expensive way to provide coverage than original Medicare. The capitation payments to plans are on average about 10% higher than the cost of insuring the average fee-for-service enrollee (Miller, 2007), despite the fact that MA plans are thought to attract healthier enrollees. I

[1] The model is based on the dynamic discrete choice demand models of Shcherbakov (2008) and Gowrisankaran and Rysman (2011). I describe later the modifications that I made to these models to fit this market.

[2] Previous papers estimating demand for MA plans include Town and Liu (2003), Brand (2005), Maruyama (2006), Hall (2007) and Lustig (2008). While none of these address switching costs, several of them model other important features of this market such as adverse selection.

address the question of the value of the MA coverage through two counterfactuals, one in which all of the MA plans exit, leaving only original Medicare, and one in which consumers pay a surcharge to compensate Medicare for the additional cost of their coverage. Results from both suggest that consumers place a surprisingly low value on the MA program. In fact, in some years the consumers in MA plans would be better off on average in original Medicare if it weren't for the switching costs they would incur by making the switch to original Medicare.

There is a current political debate about the future of Medicare and Medicare Advantage. As part of the Patient Protection and Affordable Care Act, passed in 2010, capitation payments to MA plans are slated to decrease starting in 2014, which may result in many plans exiting the market. Others have proposed further privatizing Medicare, such as through a voucher system. Knowledge of the magnitude of switching costs in this market could help inform some aspects of this debate. When considering any complete overhaul of the Medicare system that would move many consumers to substantially different plans, the switching costs would be properly included in an assessment of the costs and benefits. The large impact of switching costs in this market on consumer welfare also suggests some comparatively simple policy changes that could make consumers better off. For example, it may be possible to decrease switching costs by further standardizing the benefits that the plans offer or limiting the restrictiveness of the plans' provider networks.

The remainder of the paper is organized as follows. First, I discuss the source and importance of the switching costs and dynamics in this market. Next, I explain how this paper fits in the literature and provide background information about the Medicare Advantage program. After that, I lay out the details of the model, followed by a discussion of the data, identification and estimation. Finally, I discuss the results of the estimation and several counterfactual exercises, and conclude.

2 Switching Costs and Dynamics

2.1 The Source and Importance of Switching Costs

For the purposes of this paper, a switching cost is defined as any one-time cost a consumer incurs as a result of choosing a Medicare plan–either original Medicare or an MA plan–different than the plan chosen in the previous period.

There are numerous reasons to believe that changing MA plans imposes a large switching cost. Since most MA plans restrict patients to physicians in a network, switching plans may entail switching providers. Provider switching can disrupt continuity of care, and it may take a considerable amount of time to build a good physician-patient relationship with the new provider (Emanuel and Dubler, 1995.) Furthermore, staying with a provider instead of switching may allow the provider to build up more knowledge about the patient's health history. Some studies have shown that consistently seeing the same provider can have a positive impact on health outcomes. For example, Gill et al. (2000) find that Medicaid patients who use the same provider multiple times have a lower rate of emergency room visits than those who use different providers each visit. These effects may be amplified for Medicare eligibles because of their age. Strombom et al. (2001) find that older or sicker enrollees in employer health plans change plans less often than their younger, healthier counterparts. Older people may have higher switching costs because they have longer health histories and visit the doctor more often, establishing a stronger physician-patient relationship.

Aside from the provider switching effect, there are other ways that care can be disrupted by switching plans. Since the set of treatments covered vary from plan to plan, a patient may be forced to change treatments as a result of switching plans. For example, a patient may be taking a drug that is covered under his old plan but not his new plan. Under the new plan, the patient may be switched to a different drug that is covered. Both drugs may be indicated for treatment of the patient's condition, but the patient may need to learn to cope with a new set of side effects, experiment with dosage, or take the drug on a different schedule. This is distinctly an issue of switching costs rather than of differing plan quality, which will be handled separately in the model, because even if one treatment isn't a better option for the patient from the outset, there can be direct disutility from the change itself.

Another type of switching cost is a learning cost that MA enrollees incur when they need to determine what is covered under their new plan, what co-pays they are responsible for, and what providers are included in the plan's network. Plans can be complicated, making this process non-trivial. Furthermore, a study (Gazmararian, et al., 2006) suggests that about a third of enrollees in Medicare managed care plans have limitations to their literacy skills or basic health knowledge, which might make it even more difficult or costly to learn about a new plan. Finally, there may be a hassle cost associated with filling out the paperwork to opt into a new plan rather than accepting the default of staying in the same plan, and with updating insurance information with one's providers.

Some readers may prefer to interpret the switching cost as not a "real" cost but a "psychological" cost or a default bias (the latter in the manner of Choi, Laibson and Madrian, 2011). My approach does not rule out this interpretation as long as consumers consider switching costs to be "real" whenever they make a plan choice. In the results tables that include welfare effects, I always separate out the part of welfare that is coming directly from switching costs paid. Adherents to the behavioral interpretation can ignore this component and focus instead of the effects of the switching cost on consumer choice, which in some cases is the larger effect.

If switching plans is so inconvenient and costly, why would a consumer ever switch? One reason is that the consumer's choice set changes from year to year. Since MA plans frequently enter and exit markets, a new plan might enter that a consumer finds more appealing than his existing plan, or the consumer's existing plan might exit, forcing a switch. The coverage offered by a plan can also change over time, and a consumer might want to leave a plan that has diminished in quality. Finally, the consumer's own health might change in a way that changes which plan is best suited to his needs.

2.2 Switching Costs and Dynamics

Switching costs induce state dependence that is best handled in a dynamic model. If there are non-negligible switching costs, a consumer's choice in the current period is influenced by his choice in the previous period, because the previous period choice determines which plans entail switching costs. Furthermore, an optimizing consumer must be forward looking, because the plan chosen in the current period affects switching costs in the future. A consumer facing high switching costs is unlikely to choose a plan that he thinks will drastically diminish in quality, exit the market, or become less suited to his health needs in the next period. In the absence of switching costs, a consumer would merely choose the best plan in each period, switching as often as necessary to do so. Such a decision could be modeled statically. When there are switching costs, a consumer may stick with a plan that would look suboptimal from a myopic viewpoint in order to circumvent a switching cost, or may avoid choosing a plan that he knows he will not keep for many periods. This type of optimization is clearly dynamic.

In addition to the switching costs, other factors contribute to the dynamics in this market. There is a high level of entry and exit of plans, making the choice set potentially different in every period. Also, the coverage offered in a particular plan can change from year to year. In conjunction with the switching cost, the changes in the choice set and plan quality induce an optimal timing decision. If plans are entering that have increasingly higher quality compared to the consumer's current plan, the consumer must decide whether to switch right away or to wait for a better plan.

A dynamic model that includes switching costs allows for a better understanding of the consumer choice problem and related welfare analysis. Suppose several plans add a new feature in a particular year, and we want to estimate how much consumers value that feature. An assumption of zero switching costs would lead to undervaluation of the feature, because we would see fewer consumers switching to the plans with the new feature than we would if there truly were zero switching costs. Consider a policy in which a new plan is added with more extensive coverage. The estimated welfare gain would likely be too large, because we would conclude that more consumers would switch to the new plan than actually would. Clearly, ignoring switching costs can bias policy-relevant welfare analysis.

In addition to the effect of switching costs on consumers, switching costs also have implications on the firm side. High switching costs alter the entry game for firms offering Medicare Advantage plans. Switching costs give incumbent firms an advantage, and make it more difficult for new firms to enter or for firms to offer new plans. Switching costs may also affect the way that firms design benefit packages and premiums. A forward thinking firm may want to initially offer a more attractive plan in order to "lock in" consumers through the switching cost. While the firm side is not explicitly modeled in this paper, consumer awareness of these firm-side dynamics could be contributing to consumer-side dynamics.

3 Relationship to Literature

In terms of the model and methodology, this paper is closely related to papers by Gowrisankaran and Rysman (2011) and Shcherbakov (2008). Gowrisankaran and Rysman propose a model of dynamic consumer demand for durable goods and develop an estimation routine that allows them to estimate the model for the digital camcorder industry. Shcherbakov adapts the model to incorporate switching costs as the source of the dynamics in place of durability, and uses it to estimate switching costs in the cable television industry. He exploits the fact that the same three options for television service are available in every period to simplify the model by letting the product identity be part of the state space. Since I deal with a market where there are many distinct products and large and changing choice sets, Shcherbakov's approach is not tractable in this setting. Therefore, I combine the state-space-reducing features of Gowrisankaran and Rysman's model with the switching-cost-based dynamics of Shcherbakov's model. While my model is very similar in structure to Gowrisankaran and Rysman's, there is an important difference in the way that the decision problem is set up. In my case the key decision is switching versus not switching, while in theirs it is buying versus not buying.

Switching costs in health plan markets have been addressed in other empirical works.[3] Strombom, Buchmueller and Feldstein (2001) empirically study the effect of switching costs in employer sponsored health plan choice in a static setting using a reduced form model. Handel(2010) estimates a structural model of health plan choice that incorporates both switching costs and adverse selection. Both papers use detailed, consumer-level data from a single employer. While there are advantages to using this type of data, the data available about MA plans is on the market share level. Therefore, an important distinction is that I use a model that can be estimated with market share data. Also, the scope of these papers is limited by the fact that they are using data from a single employer, while my data covers a substantial fraction of all of the elderly people in the U.S. In addition, both of these papers look at a working age population. I focus on the elderly, who might have different switching costs. Ericson (2010) considers firms' response to switching costs for health plans. Using data from the market for Medicare Part D prescription drug plans, he finds evidence supporting his theoretical prediction that firms raise prices on existing plans to exploit consumers who are "stuck" because of switching costs and introduce new, cheap plans to attract unattached consumers.

4 Background about Medicare Advantage

Medicare Advantage, formerly known as Medicare + Choice, was created with the goal of introducing private competition into the provision of Medicare coverage and offering more options to Medicare beneficiaries. While private managed care organizations have had some role in Medicare since the 1970's, the current Medicare Advantage system is mostly a result of the Balanced Budget Act of 1997 and the Medicare Modernization Act of 2003.

Medicare consists of four parts, denoted Medicare Part A, B, C and D. Parts A and B comprise original Medicare coverage. Part A is hospital insurance and Part B is medical insurance. Part C consists of MA plans, which Medicare beneficiaries have the option of choosing in lieu of part A and B coverage. MA plans are required to cover the same services as part A and B, but may cover them differently, such as through different cost sharing. Most MA plans also cover additional services not covered by original Medicare. Part D is optional prescription drug coverage for part A and B enrollees. Part D coverage was not available until 2006, after the period studied in this paper.

The firms that offer MA plans enter into one year contracts with the Center for Medicare and Medicaid Services through a non-competitive bidding process in which all bids meeting certain criteria are accepted. Contracts are for a single year and can cover one or more counties. When a firm has

[3] There is also a rich theoretical literature on switching costs, surveyed by Klemperer(1995) and Farrell and Klemperer (2007).

a contract to offer a plan in a particular set of counties, it agrees to offer the same coverage terms to any Medicare eligible person residing in one of those counties who opts into the plan. The amount the firms are paid per person covered is based on a county-specific base capitation rate which is adjusted for the actual demographic mix in the plan.

Medicare beneficiaries choose their coverage choice during an open enrollment period each year. The options to choose from are original Medicare or any MA plan offered that year in the beneficiary's county. Coverage changes take effect the following year. The default is to continue in the same plan, whether an MA plan or original Medicare. If a beneficiary's MA plan is discontinued and no new selection is made, by default the beneficiary is enrolled in original Medicare.

There are several types of MA plans, including Preferred Provider Organizations (PPO), Health Maintenance Organizations (HMO), Private Fee For Service Plans (PFFS), Medical Savings Accounts, and Special Needs Plans. The majority of plans are HMOs or PPOs, which are both forms of managed care.

MA enrollees pay the same monthly amount to Medicare as Part B enrollees, plus an additional premium to the firm offering the plan. Some firms choose to charge a premium of $0, and starting in 2003, firms were permitted to charge a negative premium, essentially paying consumers to be in the plan. The zero and negative premiums are potentially profitable for firms because of the capitation payments they receive from Medicare.

The most recent changes to the Medicare Advantage program went into effect as a result of the Medicare Modernization Act of 2003. The Medicare Modernization Act introduced the Part D drug plans, expanded preventative care coverage under original Medicare, and increased payments to plans. The data used in this paper cover the years 2001-2005, which means that years both before and after the passage of this legislation are covered. The changes were rolled out over a period of several years, and not all of them had taken effect by 2005. In particular, Medicare eligibles were not able to enroll in the Part D prescription drug plans until 2006. Thus, in terms of the types of coverage offered, some of the most important changes (which would have been difficult for the model to handle) did not occur during the period covered by the data. On the other hand, the increase in payments took effect in 2005, and the resulting influx of plans can be observed in the data. This increase in plan offering creates some extra variation in the choice set that helps with estimation of consumer preferences. The downside is that some other unobserved changes resulting from the Medicare Modernization Act could be going on at the same time, and may be a source of bias. An important robustness check to try will be estimating the model without the data from 2005.

Under the Patient Protection and Affordable Care Act, passed by the House in 2009 and the Senate in 2010, additional changes to the Medicare Advantage program will be rolled out in 2014. According to a website maintained by the U.S. Department of Health and Human Services (Healthcare.gov, 2011), the legislation will "reduce excessive payments to private insurance companies in Medicare Advantage while protecting...guaranteed Medicare benefits."[4] Presumably the lower payments will decrease the

[4] Close reading of the actual legislation, available online at http://democrats.senate.gov/reform/patient-protection-affordable-care-act-as-passed.pdf, reveals that the law does not directly reduce payments to the plans, but rather introduces a competitive bidding system. Under the new bidding system, plans are paid based on a weighted average of all plans'

profitability of offering MA plans, and some firms will leave the market resulting in fewer plans available.

5 The Model

The goal of the model is to represent consumer choice of Medicare Advantage plans in a way that captures the effects of current and future switching costs, consumer valuation of plan characteristics, and rational expectations over future plan characteristics and choice sets. The model must be parsimonious enough for estimation to be tractable, which requires compromises on how the state space is defined, and parameters must be identifiable from market share level data. This section begins with a model of consumer preferences over characteristics of the plans and ends with a formula for predicted market shares that can be taken to the data.

5.1 Model

Each consumer, indexed by i, lives in a county, indexed by m. A consumer's county never changes, and the set of consumers is the same every year. The consumer's choice set depends on the year, indexed by t, as well as county. The choice set for consumers in county m in year t is denoted J_{mt}, with elements indexed by j. The element $j = 0$ is the outside good, original Medicare. All other $j \in J_{mt}$ are Medicare Advantage plans.

Every year, the consumer must pick a plan j from the choice set J_{mt}. Consumer i's chosen plan in year t is denoted j_{it}. At the time of the decision, consumers know the plans available, plan premiums (prices), observable and unobservable (to the econometrician) characteristics of the plans, and their own previous period choices. A consumer's choice in year t may be the same as his choice in the previous year, provided his year $t-1$ plan is in the choice set for year t, or it may be different. If the chosen plan is different, the consumer incurs a switching cost.

Consumer $i's$ state at the beginning of period t (just before choosing j_{it}) is:

$$(j_{i,t-1}, J_{mt}, X_{mt}, \xi_{mt}, P_{mt}, E_{it}, \Omega_{mt}) \tag{1}$$

where $j_{i,t-1}$ is the consumer's plan choice in the previous year, X_{mt} is the matrix of observed characteristics of the plans in the consumer's choice set, ξ_{mt} is a vector of the unobserved characteristics of these plans and E_{it} is a vector of type 1 extreme value error terms that are independently and identically distributed across consumers, plans and time. The matrix Ω_{mt} contains all the information that consumers might use to form expectations over future values of the other variables. It is discussed in more detail later.

A consumer's infinite horizon expected utility from choosing a plan in year t consists of three components. The first is a one-period flow utility. The flow utility, representing the single year net benefit from being enrolled in the plan, is a function of plan characteristics, the plan premium, consumer

bids. Since there is little theoretical literature on average-bid auctions, it is not immediately obvious what effect the policy will have on payments. However, claims that the legislation reduces the payment rates are widespread.

preferences, and a random shock. The second component of utility is the switching cost. The switching cost represents any costs incurred due to choosing a plan in year t that is different from the plan chosen in year $t-1$. It is zero if the consumer does not switch plans, and a constant, η, if the consumer does switch plans. The third component is the continuation value, the expected discounted infinite horizon utility for the consumer for year $t+1$ onward, given the state variables at the beginning of period t and the consumer's plan choice. The one-period flow utility and some useful functions of it are defined in the equations that follow. The continuation value is defined recursively by the Bellman equation.

The one-period flow utility to consumer i in period t for plan j in county m is:

$$f_{imjt} = \begin{cases} \alpha_0^i + \alpha_1' x_{jt} + \alpha_2^i p_{jt} + \xi_{mjt} + \varepsilon_{ijt} \text{ if } j \neq 0 \\ \varepsilon_{i0t} \text{ if } j = 0 \end{cases} \tag{2}$$

The variables x_{jt} and p_{jt} are the observed characteristics vector and premium for plan j, respectively. The α coefficients are parameters to estimate. The unobserved characteristic, ξ_{mjt}, represents county, plan, and year level unobserved quality, which can consist of unobserved dimensions of plan benefits or county specific aspects of the plan's quality, such as the quality of the doctors in the plan's network for that county. The error term, ε_{ijt}, is a type 1 extreme value random variable and is independent across consumers, plans and years. Normalization of one flow utility is necessary to establish scaling. This normalization is accomplished by setting the mean flow utility of the outside good to zero. The flow utility of the outside good therefore just equals the error term.

Two of the α coefficients, the constant term and the coefficient on price, are modeled as random coefficients that have a normal distribution across consumers. The other coefficients, contained in the vector α_1, are modeled as constant across consumers. Allowing all of the coefficients to have a non-trivial distribution would allow for more flexible consumer heterogeneity, but restricting the heterogeneity to two coefficients helps make the estimation tractable. The distribution of the random coefficients is defined as follows:

$$\alpha_0^i \sim N(\bar{\alpha}_0, \sigma_0) \tag{3}$$
$$\alpha_2^i \sim N(\bar{\alpha}_2, \sigma_2)$$

where $\bar{\alpha}_0$ and $\bar{\alpha}_2$ are mean (across consumers) coefficients to be estimated and σ_0 and σ_2 are standard deviation parameters to be estimated. I assume that α_0^i and α_2^i are uncorrelated. A consumer's random coefficient draw remains the same every year, which is what creates persistence in consumer tastes. For example, because of the random coefficient on the premium, a consumer who chooses a high premium plan this period is likely to choose another high premium plan next period. Each county is assumed to have a continuum of consumers whose preferences are distributed in this way.[5]

Two more notations related to the flow utility will show up in later equations. The mean flow utility (across values of the i.i.d. error term ε_{ijt}) of a consumer of type i in county m for plan j at time

[5]In the estimation, consumers will be of discrete "types" independently drawn from this distribution.

t is:

$$\overline{f_{imjt}} \equiv \begin{cases} \alpha_0^i + \alpha_1' x_{jt} + \alpha_2^i p_{jt} + \xi_{mjt} \text{ if } j \neq 0 \\ 0 \text{ if } j = 0 \end{cases} \tag{4}$$

Notice that the error term ε_{ijt} doesn't appear because it has been integrated out. The mean flow utility across consumers in county m is:

$$\overline{f_{mjt}} \equiv \begin{cases} \overline{\alpha}_0 + \alpha_1' x_{jt} + \overline{\alpha}_2 p_{jt} + \xi_{mjt} \text{ if } j \neq 0 \\ 0 \text{ if } j = 0 \end{cases} \tag{5}$$

Here, the random coefficient draws have also been integrated out, leaving only the mean coefficients.

Let Ω_{mt} denote a matrix of the current observed and unobserved characteristics and premiums for all plans in county m in period t, and anything else that might affect consumer expectations about the future choice set, plan characteristics and premiums. It might, for example, include the complete past history of plan offerings and characteristics. Assume that Ω_{mt} evolves by some Markov process, $q(\Omega_{mt} \mid \Omega_{m,t-1})$.

Since plans can exit, a consumer's chosen plan from period $t-1$ may not be in the period t choice set. If this happens, the consumer is forced to choose a different plan. Whether the consumer's previous plan is still an option is captured by this indicator function:

$$I(j_{i,t-1}) = \begin{cases} 1 \text{ if } j_{i,t-1} \in Jmt \\ 0 \text{ if } j_{i,t-1} \notin Jmt \end{cases} \tag{6}$$

Assuming an infinite horizon and annual discount factor β, we can now write the Bellman equation. In this form, the Bellman is not useful for the estimation, but writing it this way sets up the choice problem and illustrates why the state space must be reduced. First, consider conditioning on the case where the consumer's incumbent plan is still available in period t (that is, conditioning on $I(j_{i,t-1}) = 1$):

$$\begin{aligned} &V(j_{i,t-1}, \varepsilon_{ijt}, \Omega_{mt} | I(j_{i,t-1}) = 1) = \\ &\quad max\{f_{imj_{it-1}}t + \beta E[V(j_{i,t-1}, \Omega_{mt,+1} | \Omega_{mt})], \max_{j \in J_{mt}, j \neq j_{it-1}} \{-\eta + f_{imjt} + \beta E[V(j, \Omega_{m,t+1} | \Omega_{mt})]\}\} \end{aligned} \tag{7}$$

where the expectation is over future error draws and the future evolution of Ω and the notation $f_{imj_{it-1}}t$ means the period t flow utility for the plan chosen in period $t-1$. The inner maximization on the right side of the Bellman represents the choice of the best plan of all the plans in the current period choice set excluding the consumer's incumbent plan. The outer maximization represents the choice of switching (choosing the plan that is the argmax of the inner maximization) or not switching (choosing the incumbent plan). This two step maximization process is equivalent to choosing the best of all plans in the choice set, but puts the Bellman into a form that will be useful for later simplification.

Now, consider instead conditioning on the consumer's incumbent plan *not* being available in period t (that is, conditioning on $I(j_{i,t-1}) = 0$). In this case, the Bellman consists of only the inner maximization because the consumer does not have the option of staying with the incumbent plan. The consumer must choose the best of the currently available plans even if she would prefer to stay in her (now defunct) incumbent plan:

$$V(j_{i,t-1},\varepsilon_{ijt},\Omega_{mt}|I(j_{i,t-1})=0)=\underset{j\in Jmt}{max}\{-\eta+f_{imjt}+\beta E[V(j,\Omega_{m,t+1}|\Omega_{mt})]\} \tag{8}$$

Further simplification is necessary before the Bellman is tractable to work with. At this point, Ω_{mt} can have an arbitrary number of dimensions and can affect consumer expectations in an entirely unrestricted way. To make the estimation tractable, it is necessary to further specify what information goes into consumer expectations and how these expectations are formed. These restrictions are then used to reduce the dimensionality of the state space.

A consumer in this market forms expectations about two things: the future of the plan currently held, including whether it will exit, and the future of the other plans in the market, including changes to this set from future entry and exit. Expectations about the other plans in the market will be formulated in terms of the *logit inclusive value,* which is defined as:

$$\delta^{imt}(j_{i,t-1},\Omega_{mt})\equiv\ln(\sum_{j\in J_{mt},j\neq j_{it-1}}\exp(\delta_j^{imt}(j_{i,t-1},\Omega_{mt}))) \tag{9}$$

where:

$$\delta_j^{imt}(j_{i,t-1},\Omega_{mt})\equiv-\eta+\overline{f_{imjt}}+\beta E[V(j,\Omega_{m,t+1})|\Omega_{mt}] \tag{10}$$

The logit inclusive value is the expected value of the consumer's *best* plan choice among all available plans excluding the consumer's incumbent plan. The expectation is over the extreme value error term ε_{ijt}, and it takes the closed form above based on properties of the extreme value distribution. The logit inclusive value can be thought of as the value of switching, or as a summary of the quality and selection of other plans in the market, taking into account switching costs and the infinite horizon future value. It is realistic to think that consumers have some such summary of the available plans in mind when they make their choices. With many plans available, each with a complex coverage structure, consumers probably don't have in mind the details of every plan when choosing whether to stay or switch, but rather have a more broad idea of the quality of plans available. Therefore, it is not unreasonable to model consumer expectations about the future states of the market as being based on the logit inclusive value, δ^{imt}. In addition to the changing quality of plans over time, the evolution of the logit inclusive value also captures information about entry and exit. All else equal, the logit inclusive value decreases with exit and increases with entry. Of course, some information is lost, as upon observing a decrease of a particular size, there is no way to distinguish between plan exit, some combination of quality changes, or both.

For the consumer's own incumbent plan, expectations are about the future mean flow utility of the plan, and about the plan's probability of exiting. These expectations are handled separately from expectations about the other plans in the market to allow consumers to have more specific knowledge about their own plan. Focusing on the mean flow utility is also a simplification, though, because in principle consumers could have detailed expectations about the future of each characteristic of the plan.

The following assumptions formalize these ideas and define the processes governing the consumers' expectations.

Assumption 1: Sufficiency of a reduced set of state variables:

$$Pr(\delta^{im,t+1}(j)|\Omega_{mt}) = Pr(\delta^{im,t+1}(j)|\Omega'_{mt}) \text{ if } \delta^{imt}(j,\Omega_t) = \delta^{imt}(j,\Omega'_t) \quad (11)$$

$$Pr(\overline{f_{imj,t+1}}|\Omega_{mt}) = Pr(\overline{f_{imj,t+1}}|\Omega'_{mt}) \text{ if } \overline{f_{imjt}}(\Omega_t) = \overline{f_{imjt}}(\Omega'_t) \quad (12)$$

This assumption states that the distributions of the future logit inclusive value and mean flow utilities depend on Ω_t only through their respective previous value. Given these previous values, the other information in Ω_t doesn't add anything.

Assumption 2: Consumers expect that δ^{imt} evolves according to the following autoregressive processes:

$$\delta^{im,t+1}(j) = \gamma_0 + \gamma_1 \delta^{imt}(j) + u_{jt} \text{ for } j \neq 0 \quad (13)$$

$$\delta^{im,t+1}(0) = \gamma_0^0 + \gamma_1^0 \delta_t(0) + u_{0t} \quad (14)$$

and if j is a consumer's incumbent plan, and j is still available in period $t+1$, then the consumer expects that $\overline{f_{imjt}}$ evolves according to the following autoregressive process:

$$\overline{f_{imjt+1}} = \tau_0 + \tau_1 \overline{f_{imjt}} + \nu_{jt} \text{ for } j \neq 0 \quad (15)$$

$$\overline{f_{im0t}} = 0 \ \forall t \quad (16)$$

where u_{jt}, u_{0t} and ν_{jt} are each an identically and independently distributed normal error term with mean zero, and the γ's and τ's are parameters to estimate. The consumers are assumed to have rational expectations in the sense that the γ's and τ's are such that the consumers are correct on average given the observed data. In practice this means that these parameters will be found via a regression in the estimation. Notice the special treatment when the consumer's incumbent plan is the outside good, original Medicare. Consumers correctly expect that the mean flow utility of the outside good is always zero. For $\delta^{imt}(j)$, the coefficients are allowed to be different in the case where the incumbent plan is an MA plan (and original Medicare is therefore included in δ^{imt}) than in the case where the incumbent plan is original Medicare (and original Medicare is therefore *not* included in δ^{imt}).

The process specified in (15) for the mean flow utility is conditional on the consumer's plan still being in the market. The next step is modeling the consumer's beliefs about the probability of the plan exiting.

Let p_{mjt} be an indicator for plan j exiting market m at time t. Then, the average plan exit rate for time periods one through T across all M markets is:

$$p \equiv \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{j \in J_{mt-1},\ j \neq 0} \frac{1}{|J_{m,t-1}|} p_{mjt} \quad (17)$$

An empirical plan exit rate for the years covered in the data, $\widehat{p}$, can be calculated using this formula. Assume that consumers expect that the probability of any particular plan exiting in the next time period is equal this aggregate exit rate, unless the plan is original Medicare, which has probability 0 of exiting. Then, consumer expectations about plan exit can be expressed as follows:

$$\begin{aligned} I(j_{it-1}) &= \begin{cases} 1 \text{ with probability } 1-\widehat{p} \\ 0 \text{ with probability } p \end{cases} \text{for } j_{it-1} \neq 0, \\ I(0) &= 1 \text{ with probability } 1 \end{aligned} \tag{18}$$

Since several objects defined above depend on whether the plan in question is the outside good, one of the state variables in the reduced state space will be an indicator that takes the value one when the consumer's previous plan choice is the outside good and zero otherwise. Denote this indicator $i(j_{i,t-1})$.

Taking the expectation of (7) and (8) and applying Assumptions 1 and 2 and the definition in equation (9), the Bellman equation can now be written in expected value form, with many fewer dimensions. Since only the expectation of the value function appears in the transition probabilities and market shares, solving this form of the Bellman will be sufficient to construct these objects. Conditioning on whether the consumer's plan has exited,

$$\begin{aligned} &EV(\overline{f_{imj_{i,t-1}}t}, \delta^{imt}(j_{i,t-1}), i(j_{i,t-1})|I(j_{i,t-1})=1) \\ &\quad = \ln(\exp(\overline{f_{imj_{i,t-1}}t} + \beta E[V(\overline{f_{imj_{i,t-1}}t+1}, \delta^{imt+1}(j_{i,t-1}), i(j_{i,t-1}))|\overline{f_{imj_{i,t-1}}t}, \delta^{imt}(j_{i,t-1})]) + \exp(\delta^{imt}(j_{i,t-1}))) \\ &EV(\overline{f_{imj_{i,t-1}}t}, \delta^{imt}(j_{i,t-1}), i(j_{i,t-1})|I(j_{i,t-1})=0) = \delta^{imt}(j_{i,t-1}) \end{aligned} \tag{19}$$

with the expectations over future error draws, the evolution of $\overline{f_{imj_{it-1}}t}$ and $\delta^{imt}(j_{it-1})$, and the probability of plan drop-out. The functional form comes from standard results about expected maxima of independent type 1 extreme value random draws. Using the law of iterated expectations and the probability, p, that a plan drops out, the expected Bellman simplifies to:

$$\begin{aligned} &EV(\overline{f_{imj_{i,t-1}}t}, \delta^{imt}(j_{i,t-1}), i(j_{i,t-1})) = \\ &(1-p) * \ln(\exp(\overline{f_{imj_{i,t-1}}t} + \beta E[V(\overline{f_{imj_{i,t-1}}t+1}, \delta^{imt+1}(j_{i,t-1}), i(j_{i,t-1}))|\overline{f_{imj_{i,t-1}}t}, \delta^{imt}(j_{i,t-1})]) + \exp(\delta^{imt}(j_{i,t-1}))) \\ &+p * \delta^{imt}(j_{i,t-1}) \end{aligned} \tag{20}$$

Conveniently, the expected value is now a function of only three variables: $\overline{f_{imj_{i,t-1}}t}$, $\delta^{imt}(j_{i,t-1})$ and $i(j_{i,t-1})$. This is a sufficient reduction in the dimensionality of the problem to make the estimation tractable.

With the preceding setup established, expressions can be derived for the probability that a consumer of a given type chooses a particular plan. These probabilities are then used to construct a transition matrix for the market shares. The specific functional forms of the probabilities are a consequence of the extreme value error term assumption.

If consumer i chose plan j' in the previous period, his probability of switching is:

$$\Pr^i_{switch}(j') \equiv \Pr(j_{it} \neq j' | j_{it-1} = j') = \frac{\exp(\delta^{imt}(j'))}{\exp(\delta^{imt}(j')) + \exp(\delta^{imt}_{j'}(j'))} \tag{21}$$

Likewise, if consumer i chose plan j' in the previous period, his probability of *not* switching is

$$\Pr^i_{noswitch}(j') \equiv \Pr(j_{it} = j' | j_{it-1} = j') = \frac{\exp(\delta^{imt}_{j'}(j'))}{\exp(\delta^{imt}(j')) + \exp(\delta^{imt}_{j'}(j'))} \tag{22}$$

Conditional on switching and given that j' was the plan chosen in the previous period, the probability of choosing plan j is:

$$\Pr^i_{j|switch}(j') \equiv \Pr(j_{it} = j | j_{it-1} = j', j_{it} \neq j') = \frac{\exp(\delta^{imt}_{j}(j'))}{\exp(\delta^{imt}(j'))} \tag{23}$$

Finally, the total probability of choosing plan j in period t having chosen plan j' in period $t-1$ is:

$$\Pr^i_j(j') \equiv \Pr(j_{it} = j | j_{it-1} = j') = 1_{\{j=j'\}} \Pr^i_{noswitch}(j') + 1_{\{j \neq j'\}} \Pr^i_{j|switch}(j') \Pr^i_{switch}(j')$$

The transition probabilities can be used to express the expected market share in the current period as a function of the previous period's market share for a given consumer type. Let $s_{imj,t-1}$ be the period $t-1$ market share for plan j in county m for consumer type i. Then, the expected county m market share of plan j in year t for consumers of type i can be expressed:

$$\widehat{s_{imjt}} = \sum_{j' \in J_{m,t-1}} s_{mij',t-1} \Pr^i_j(j')$$

Essentially, each $\Pr^i_j(j')$ is an element in a transition matrix relating period $t-1$ market shares to period t market shares for type i. Integrating the market shares over consumer types yields predicted county level market shares:

$$\widehat{s_{mjt}} = \int \widehat{s_{imjt}} dF_i \tag{24}$$

where F_i is the distribution of consumer types, which is defined by the distributions of the random coefficients described in equation (3). As promised, the model makes predictions about county-plan-year-level market shares based on the previous year's market shares, but is based on a model of consumer preferences over plan characteristics, the parameters of which are estimable.

5.2 Discussion of Assumptions and Possible Extensions of Model

The model captures some aspects of the market very well, and is open to improvement in other areas. For example, it easily handles changes in the choice set, with δ^{imt}_j always summarizing the state of the market and the possibility of plan drop-out captured in the process governing expectations over future flow utilities. However, changes in the set of *consumers* are completely ignored, even though in reality a new group becomes eligible for Medicare every year and some existing Medicare enrollees die each year. It is actually to my detriment to leave this out of the model, since it would be possible to exploit the different choices made by new consumers, who have a switching cost for every plan, and existing consumers, who do not have a switching cost for one plan, to help identify the switching cost. This identification strategy would be similar to the one used in Handel (2010).

The scope of consumer heterogeneity is also somewhat limited. While consumers are allowed to have persistently heterogenous taste over two of the characteristics (and could have heterogeneous taste over all of the characteristics in exchange for an increase in the computational burden of estimation), consumers' tastes are assumed to have the same distribution in every county. An extension would be to allow the distribution to depend on county level demographics, such as the age distribution, which may be a more realistic type of heterogeneity. Another limitation is that the switching cost is assumed to be the same in every situation, instead of depending on what type of switch the consumer is making. For example, switching into a fee-for-service plan might be less costly than switching into an HMO. These are all ripe areas for extension of the basic model. While all of these enhancements would increase computation time for the estimation, none in isolation would do so to a degree that makes it intractable.

One modeling choice that requires further discussion is the reduced form AR1 processes that govern consumer expectations. A more sophisticated alternative would be to endow consumers with an understanding of the supply side dynamics that determine the quality and availability of plans over time. Modeling the supply side of this market in detail would be a huge undertaking that would entail a separate research project complementary to this one. Even if such a model were readily available, it is not clear that the best assumption would be that consumers have a deep understanding of what is happening on the supply side in terms of the dynamic games that firms play which determine patterns of entry, exit, and plan design. The way expectations are modeled here, consumers have expectations in three dimensions: they have expectations about the future quality of their own plan, the probability that their plan will exit in a given period, and the quality of other plans in the market. While these are relatively simple concepts that would not be difficult for consumers to think about, it allows for some interesting patterns that might reflect the underlying dynamics. For example, in this framework consumers can have the expectation that a given plan will lessen in quality over time, while the overall quality of plans in a market increases over time. The underlying reason for this pattern might be that new plans enter with high quality to attract consumers, then lower in quality over time, maintaining market share because of the switching costs, while meanwhile additional high quality plans enter in hopes of poaching consumers. Something that is missing is correlation between the three dimensions of expectations[6] (for example, that future plan quality and probability of exit are negatively correlated, or that the future quality of a specific plan is correlated with overall quality in the market). The use of these reduced form expectation processes follows Gowrisankaran and Rysman (2011). They experiment with other simple ways to represent consumer preferences, such as perfect foresight, and settle on the

[6]This is ruled out because the error terms in the two AR1 processes are uncorrelated, and the probability of plan drop-out is assumed to be independent of everything else in the model. Relaxing these assumptions would introduce additional parameters, which would enlarge the computational burden.

AR1 process as the preferred specification.

6 Data

Data comes from the Center for Medicare and Medicaid Services.[7] Three types of data are used: data on market shares, data on characteristics of the plans, and data on county-specific base capitation payment rates. The data covers the years 2001 to 2005. However, 2001 serves as the initial conditions year and only the market share data is used from that year while all three data sets are used for the other years.

A shortcoming of the data is that market shares are reported only at the county-contract level, not the county-plan level. A contract is an agreement between a firm and CMS to offer a particular group of plans in one or more counties. A contract may contain one, several or many plans, but contains the same plans in every county in which it is offered. By matching the plan-level characteristics data to the contract-level share data by contract, the plan-level choice set in each county can be determined. However, there is no way to know how the market share belonging to a contract-county is distributed among the plans in the contract. This limitation of the data is problematic because the plans within a contract can have different benefits, and consumers actually make choices on the plan level, not the contract level. Ideally, a choice model would be on the level of plans, but the parameters of a plan-level model are not identified with only contract-level data available.

There are two ways that the lack of plan-level market shares is dealt with in the literature. The first approach, used by Lustig(2008), is to find another data source that has consumer-level data on exact plan choice. Such data is expensive and difficult to obtain, and only covers a sample of Medicare eligibles, while the contract-level data covers every Medicare eligible. The second approach, used by Hall (2007), and now in this paper, is to select a representative plan from each contract and treat that plan as the only plan in the contract. The contract is then considered to be a single plan that has the characteristics of the selected plan. Hall selects the lowest numbered plan in each contract, arguing that the lowest numbered plan is the base plan, which tends to be the most commonly chosen by consumers. I adopt Hall's selection rule, but I also tried another selection rule as a robustness check.[8]

Throughout the remainder of this paper, the word "plan" refers to these contracts that are being treated as individual plans. In particular, any reference to switching plans actual means switching contracts. If a consumer switches to a different plan within a contract, they are still in the same plan under the operative definition, and they do not incur a switching cost. This assumption is reasonable because generally plans under the same contract will have the same network and similar benefits.

The original characteristics data set consists of the text data underlying the plan comparison tool

[7] Characteristics data is available on the Medicare.gov website only for the current year. Special thanks to Josh Lustig for the 2001-2005 characteristics data, which he obtained through correspondence with a CMS employee.

[8] The other selection rule that I tried was taking the median of each characteristic across plans within the contract to construct a "typical" plan. Many of the parameter estimates were similar to the parameter estimates from the data using the other selection rule, with a notable exception. The coefficient on price was much smaller under this rule, which led to an unrealistically high marginal utility of income. Since the marginal utility of income plays a crucial role in the counterfactuals, I stuck with the selection rule that gave a more reasonable value for it.

provided for Medicare beneficiaries to obtain information about the plans available in their county in each year. This data is extremely detailed but requires extensive cleaning in order to be made into usable variables. Each plan may have one or more text comment in each of about forty fields. The fields are categories of benefits, such as "Vision Services" or "Doctor Office Visits." There is also a field for the premium. The text comments appear to have been selected from a predetermined list, sometimes with a dollar amount or percentage filled in. However, quite varied types of information can appear in the same field. While it is straightforward to use text parsing methods to extract the numbers from the text, it is less obvious how to combine disparate information about, for example, fixed co-pays, coverage limits, and percent of cost covered into a single, meaningful, numerical variable. Of course, this is only partially a data issue. The root of the problem is trying to compare plans that might have a fundamentally different structure of coverage.

From the characteristics data set, I constructed eighteen variables for potential use in the estimation. Definitions of all of these variables appear in the Appendix. The choice of what variables to construct had two motivations. First, I wanted the variables that would be the most empirically relevant. Therefore, I tried to focus on the fields relating to benefits that either most elderly people would use in a given year, like "Doctor Office Visits," or that would represent a large expenditure, like "Emergency Services." Fields corresponding to more obscure benefits, such as "Podiatry" I ignored. The second motivation was one of practicality. Some fields simply had too many distinct comments for it to be possible to distill the information into a numerical variable. Others lent themselves quite nicely to one or two fairly straightforward variables. Of the variables constructed, fourteen were used in the final version of the estimation, in addition to year indicators.

The market share data comes from a data set from CMS called "GeoAreas." For each year and county, the data lists each contract offered in the county, the number of Medicare eligibles residing in that county, and the total number of county residents enrolled in plans included in the contract. Dividing the contract enrollees by the county eligibles yields the market share. Not included in the data are enrollees in contracts that are not currently offered in the county in which they officially reside (for example, because the enrollee has moved since choosing a plan). While other data sets available from CMS do include such enrollees, it is difficult to determine the choice set for each county from that type of data because spurious contract-county combinations show up in the data. I focus on the "GeoAreas" data because it offers a clean match of contracts to the counties in which they are offered, and contains only minor inaccuracies in enrollment.

Some observations are unusable. Due to privacy restrictions, exact market shares are not given for contracts with fewer than 11 enrollees. For counties with a large population, the market share for such contracts is effectively zero, so filling in zero or a small number for shares omitted for this reason is relatively innocuous.[9] For very small counties, however, ten enrollees might be a significant share of the Medicare-eligible population. For this reason, I drop all counties with fewer than 2000 Medicare

[9] In practice, I filled in 5, because it is halfway between the 0 and 10, the upper and lower bounds for the number of enrollees when the number is omitted. My first impulse, filling in zero, proved disastrous when it came time to take the log of the shares! Zeros in the data are also troublesome because the model will never predict a market share of exactly zero, though it can predict an arbitrarily small market share.

eligibles. I also drop observations for any county that does not have at least one MA contract in every year in the sample because of difficulties with the estimation when some counties have zero MA plans in some years.[10] In addition, I drop certain types of contracts that are either very similar to original Medicare, are not part of the choice set of every Medicare eligible in the county, or are systematically missing characteristics data. For example, I drop HCPPs (Health Care Pre-Payment Plans) because they are normally available only to union members and employees of particular companies, and I drop Cost plans because they are very similar to original Medicare. Dropping contracts within a county while keeping others is equivalent to lumping the market share of the dropped contracts with the outside good, original Medicare. This type of omission is distinct from dropping entire counties, which is a reduction in the number of markets sampled. After dropping the unusable observations and joining the share data to the characteristics data, I am left with data on 872 counties (out of 3083 counties in the US and about 2000 counties that show up somewhere in the data) and 300 contracts (out of about 500). Because many counties and contracts are dropped, selection bias is a concern. However, the most common reasons that observations were dropped were that the county had too few eligibles or did not have an MA plan available in every year. Even if the dropped counties are systematically different than those that are left in, the counties that remain in the data cover about two-thirds of people with access to MA plans in this time period. Therefore, this set of counties might be the most relevant from a policy perspective.

Tables 1, 2 and 4 contain summary statistics about the data. Table 4 provides the mean, minimum and maximum for each of the characteristics variables used in the estimation. Many of the mean characteristics move in the direction of increasing coverage over the four years. Some of the coverage diminishes over time, though, such as the vision coverage. There are two reasons for the means to change over time, which can act in the same or opposite directions. The plans that stay in the market can change their coverage, or entry and exit of plans can change the composition of plans in the market.

Tables 1 and 2 provide information about the total number of MA plans and entry and exit. The reported plan numbers do not include original Medicare, which is always an option in every county. Entry and exit is reported at the plan-county level, so that if a plan that is present in ten counties exits, ten exits are counted. The number of plans is increasing over time, with a big jump in 2005 due to a large amount of entry that year. The maximum number of plans available in one county is nineteen, but the median number of plans per county is much lower, ranging from two to four plans over the four year time span. There is a non-negligible amount of entry and exit in every year, though entry and exit rates do fluctuate from year to year.

I did some simple preliminary analyses of the data to determine whether it exhibits patterns consistent with high switching costs. Two features of the data that would be expected under switching costs would be low market shares for plans that have just entered, and persistence in shares from period to period. To test the first prediction, I calculated the mean share of newly entered plans and existing plans (excluding original Medicare) in each period. The results of a difference in means test

[10] In particular, it is not clear how to construct the logit inclusive value, which represents the expected value of choosing the best plan that is a switch, when the notion of switching is ill-defined.

are reported in Table 3. Clearly, the shares are much lower for the newly entered plans. To test the second prediction, I took the subset of plans that are in the data for two periods in a row, and regressed the current period share on the previous period share, controlling for current period characteristics. Results from the regression are in table 5. The coefficient on the previous period share was 0.98 and significant at the 1% level, indicating a high degree of correlation between previous and current period shares. While there are many other explanations that could account for these results, and more sophisticated reduced form tests for switching costs that could be devised, the results are at least suggestive of switching costs.

On the firm side, firms seeking to exploit consumers "locked in" by switching costs may increases prices and decrease quality of existing plans, while introducing new plans with low prices designed to attract unattached consumers. Ericson (2010) finds evidence of this type of firm behavior in the market for Medicare Part D Drug Plans, and it is also described in the theoretical literature about switching costs. To determine whether my data exhibits the same pattern, I divided the plans that were offered in 2005 by their year of entry into the market, and averaged the 2005 values of the plan characteristics within entry years (see Table 6). Compared to the plans that entered in 2003 or earlier, the newest plans have the lowest premiums and are the most likely to offer drug coverage and a category of drugs with no coverage limit. While not all plan characteristics follow this pattern, the evidence is consistent with firms offering new plans with better coverage compared to those plans that already have "locked in" consumers.

7 Instruments and Identification

7.1 Instruments

An instrumental variables approach is necessary because the plan premiums are endogenous. The error term ξ_{mjt} represents plan-county-year level unobserved quality. Unobserved quality can consist of extra dimensions of plan benefits that are not included in the observed characteristics data, or of factors that are county specific like the quality of the network's physicians in that county. Because unobserved quality is likely taken into account when premiums are determined, the premiums cannot be considered exogenous. Instrumental variables are needed that are correlated with the premium but not with the error term.

To find appropriate instruments, it is necessary to understand how premiums are set. When a firm offers an MA plan in multiple counties, it must choose one premium and set of benefits that applies to the plan in all of the counties. In choosing the premium, it therefore should think about market conditions in *all* of the counties in which the plan will be offered. As a simple example, consider a plan, called "plan 1," that is offered in two counties, "county A" and "county B". Suppose we are looking for an instrument for premium for use in county A– that is, we are looking for something that is correlated with the plan 1 premium in county A (which is the same as the premium in county B) but not the unobserved quality of plan 1 in county A (which is different from the unobserved quality in county B). An obvious place to turn is county B. In particular, consider the base capitation payment and plan premiums of the other plans, aside of plan 1, in county B. Since the other plans' premiums and the capitation rates in county B enter into the firm's profit function, they affect the optimal premium charged for the plan, which will be the premium in county A as well as county B. At the same time, the county B premiums and capitation rate should not directly affect unobserved county-specific quality in county A. This is the basic idea of the instruments.

While the role of the other plan premiums in the profit functions is straight forward, the role of the capitation rate is more complicated. It is not obvious whether a higher capitation payment means a market is more profitable or less profitable, all else equal, since the higher capitation payment is also a signal that Medicare believes that costs are higher in that county. The capitation rate would not work as an instrument if it *perfectly* compensated firms for the differing risk profiles of counties, because in that case a county with a higher capitation payment would be neither more nor less profitable. Brown et. al. (2011) find that even as Medicare devises increasingly sophisticated ways to risk-adjust the capitation payments, firms are very effective at engaging in adverse selection in a way that enhances profit. Therefore, it is reasonable to believe that firms are not perfectly compensated for differing risk across counties, and capitation payments do affect profits.

Stepping out of the two county example above and into the more complicated environment of many counties and many plans with overlapping county coverage, I construct the instruments as follows. For a given plan-county combination, I find the set of all *other* counties that contain the plan. I then take the mean, minimum, and maximum of the base capitation rate across these counties. Next, I take the set of all *other* plans (i.e., excluding the original plan) in those other counties, and take the mean, minimum and maximum premium.[11] Notice that the instruments vary on the level of plan-county combinations, because two plans that are together in one county do not necessarily share all the same counties.

The idea of using prices in other markets as instruments is reminiscent of Hausman, Leonard and Zona (1994). The traditional instruments to use in this type of setting are functions of the characteristics of other products within a market, as in Berry (1994) and Berry, Levinsohn and Pakes(1995). Such instruments are less appropriate here because characteristics of health plans are not fixed in the same sense as are the characteristics of cars, the product studied by Berry and Berry Levinsohn and Pakes. In his paper on switching costs, Shcherbakov (2008) suggests also instrumenting with lagged values of the regular instruments to help further identify switching costs through state dependence. I have not yet implemented this approach, but it would be a worthwhile specification to try in the future.

7.2 Identification

Three sets of parameters are estimated: the mean coefficients on the characteristics variables, the standard deviations for the random coefficient distributions, and the switching cost. Identification results in Berry, Levinsohn and Pakes (1995) and Berry(1994) imply identification of the mean coefficient parameters. The switching cost and random coefficient parameters are trickier. Shcherbakov (2008) makes an argument that the switching cost is identified in this type of model, but he relies on assumptions

[11] The minimum and maximum premium instruments are not used in the specification for which the results are reported.

that are specific to the cable television industry and the result cannot be directly applied here. I was able to formulate an identification result for the switching cost in this market, but it requires stronger assumptions than those made in this paper.[12] In absence of a sufficiently general formal identification result, some informal arguments about the identification of the parameters are given below.

The key to identification of the switching cost is the entry of new plans. A first intuition about how switching costs affect the observable market share data is that higher switching costs means higher persistence of market shares. Under high switching costs, a plan that has a large market share in one year will tend to have a large market share in the next, even if plans enter that seem to have more appealing characteristics, because it is costly to change plans. Measuring the degree of persistence of market shares is not quite enough to back out the switching cost parameter, however. For one thing, there is another explanation for share persistence: the consumers' persistent preference for the same plans, which should be picked up by the random coefficients and not the switching cost. Also, the predicted effect of an increase in switching costs on the market share of a particular plan in some year is ambiguous. The share might increase, because fewer consumers switch out of the plan, or it might decrease, because fewer consumers switch into it from other plans. One situation where this relationship is unambiguously monotonic is when a plan has newly entered a market. Then, any consumer who chooses the plan must incur a switching cost, because no consumer in that market chose the plan in the previous period. This creates a strictly decreasing relationship between the switching cost and the market share of such a plan, since an increase in the switching cost can only make the plan less appealing. In theory, this strictly monotonic relationship can be inverted and the switching cost is identified.

The variance of the random coefficients is identified by substitution patterns between plans when a plan exits. The degree of variance of the random coefficient distribution determines whether consumers tend to choose similar plans each time they choose a new plan. For concreteness, consider the random coefficient on the constant term. The constant term represents preference for an MA plan as opposed to original Medicare, since the normalization of the flow utility of original Medicare to zero means that original Medicare is the one plan for which the constant term is turned off. The random coefficient on the constant term allows for some consumers who consistently prefer original Medicare over an MA plan, and some who consistently prefer MA plans. Consider what happens when an MA plan exits. The higher the probability that someone whose MA plan exited chooses another MA plan, the stronger the persistence of preferences in this dimension, and the larger the random coefficient variance should be. This monotonic relationship allows for the identification of the variance parameter. Notice that in the case of plan exit, the switching cost has no effect on the consumer's decision, because he will have to pay the switching cost no matter what. Therefore, the random coefficient variance parameter and the switching cost can indeed be identified separately.

Identification of the mean coefficients on the characteristics variables is straightforward. They are identified by the differences in market shares of plans with different characteristics profiles. For example, if plans with drug coverage systematically have higher market shares than plans without,

[12] This result, or some improvement upon it, will appear in a separate dissertation chapter.

then the coefficient on drug coverage is positive. If the plans with drug coverage have *much* higher market shares, the coefficient should be large and positive.

Three sources of variation in the data help with identification of the parameters: variation in the characteristics of the plans, variation in the choice set for a county over time, and variation in the choice set across counties. These different types of variation work together to allow for simultaneous identification of all of the parameters. Consider a pair of counties that differ in only one way with respect to number of plans, plan characteristics, and entry-exit history. If the difference is in a plan characteristic, market shares from that county pair help pin down one of the mean coefficients. If the difference is in entry-exit history, market shares from that county help pin down the switching cost or the variance on one of the random coefficients. Since making a very large number of such comparisons would eventually determine all of the parameters, the model is (informally) identified.

8 The Estimator and the Estimation Procedure

8.1 The Estimator

The main parameters to estimate are the switching cost,η, the variances on the random coefficients,σ, and the mean coefficients on the characteristics, α. There are also nuisance parameters, such as the γ's and τ's governing the process by which the logit inclusive values and mean flow utilities evolve.

The estimator is based on the GMM estimators in Gowrisankaran and Rysman (2011) and Berry, Levinsohn and Pakes (1995). It is defined as follows:

$$\min_{\alpha,\eta,\sigma} \xi(\alpha,\eta,\sigma)' Z W Z' \xi(\alpha,\eta,\sigma)$$

$$s.t.\ \widehat{s}(\eta,\sigma) = s$$

where $\xi(\alpha,\eta,\sigma)$ is the vector of structural errors at the given parameter values, Z is the matrix of instruments, W is a weighting matrix, $\widehat{s}(\eta,\sigma)$ is the vector of predicted market shares when the dynamic programming problem is solved at the given parameters, and s is the true vector of market shares.

8.2 The Estimation Procedure

The estimation method is a variation on the three-level nested fixed point estimation routine developed by Gowrisankaran and Rysman (2011). The basic idea of Gowrisankaran and Rysman's algorithm is to nest solving a dynamic programming problem inside the market share inversion of Berry, Levinsohn and Pakes (1995).

To simulate the distribution of the random coefficients, 30 fixed draws from a two-dimensional normal distribution are taken at the beginning of the estimation.[13] Each of the 30 draws can then

[13] To reduce variance, the draws can be taken using importance sampling. The details of how to do importance sampling in this setting are described in Berry, Levinsohn and Pakes (1995). Under importance sampling, consumers who are more likely to choose an inside good (in this case, an MA plan) are oversampled, and the draws are reweighted accordingly when the integral over consumer types is taken in the inner loop.

be considered a discrete consumer "type." The steps of the inner loop are repeated for each of the 30 types, and the mean of the resulting shares is taken over types to get the overall county/plan market share.

8.2.1 Inner Loop

The inner loop maps a vector of parameters, (η, σ) and a vector of mean flow utilities, $\overline{f_{mjt}}$, to a vector of predicated market shares $\widehat{s_{mjt}}$ by solving the dynamic programming problem defined by the Bellman equation for a given consumer type and plugging the resulting value function into the formulas for the shares.

The inner loop simultaneously finds fixed points of several equations. It finds the value function that is the fixed point of the Bellman equation. It finds the vectors of $\delta_j^{imt}(j_{i,t-1})$ and $\delta^{imt}(j_{i,t-1})$ that satisfy the recursive definitions of these two objects. Finally, it finds estimated autoregression coefficients, $\widehat{\gamma}$ and $\widehat{\tau}$ that are stable from iteration to iteration.

To make the estimation feasible, the continuous state space must be discretized. The state space dimensions for the variables $\delta^{imt}(j_{i,t-1})$ and $\overline{f_{imj_{i,t-1}t}}$ are each divided into 50 grid points. The minimum and maximum values for the grid are based on guesses about a reasonable range for the variables plus some added leeway. The value function $V(f_{imj_{i,t-1}t}, \delta^{imt}(j_{i,t-1}), i(j_{i,t-1}))$ is defined discretely on each point on the grid. The value of the function when the arguments fall between the grid points is approximated by linear interpolation.

In order to start the inner loop, some initializations are necessary. Initial values of the value function at the grid points, plus initial values of the $\delta^{imt}(j_{i,t-1})$ and $\widehat{\gamma}$ vectors are needed for use in the first iteration. Mean flow utilities, $\overline{f_{mjt}}$, are passed in from the middle loop, as well as values for the parameters η and σ. The discount factor is set to 0.9 on the annual level.[14]

First, $\delta_j^{imt}(j_{i,t-1})$ is calculated for the plans. The expectation of the value function is part of the expression for $\delta_j^{imt}(j_{i,t-1})$. To find this expectation, a simulated two-dimensional integral must be taken over the error terms in the expectation processes. Once $\delta_j^{imt}(j_{i,t-1})$ has been calculated for each plan, the $\delta^{imt}(j_{i,t-1})$ can be updated by taking the log of the sums of the exponentials of $\delta_j^{imt}(j_{i,t-1})$ for all the plans in a county except j_{it-1}, the previous plan choice. The $\delta^{imt}(j_{i,t-1})$ are then regressed on the $\delta^{im,t-1}(j_{i,t-1})$, and the $\overline{f_{imj_{i,t-1}t}}$ on the $\overline{f_{imj_{i,t-1}t-1}}$ to obtain a new $\widehat{\gamma}$ and $\widehat{\tau}$.

There are two options for the next step in the algorithm. Since $\delta_j^{imt}(j_{i,t-1})$, $\delta^{imt}(j_{i,t-1})$, and $\widehat{\gamma}$ all eventually need to converge, it may help to iterate on the preceding steps several times before moving on to the value function. However, because more instances of the simulated integral have to be calculated for the $\delta_j^{imt}(j_{i,t-1})$ than for the value function, I have found in practice that overall convergence of the inner loop tends to be faster if the $\delta_j^{imt}(j_{i,t-1})$ are calculated only once for every time the value function is updated using the Bellman equation.

Updating the value function consists of evaluating the right hand side of the Bellman equation

[14] The discount factor is known to be difficult to estimate in this type of setting, so I do not attempt to estimate it. The value that I set it to, 0.9, is lower than what is typically used, to reflect that the elderly population in Medicare might have a shorter time horizon than a typical population of consumers.

for every point on the grid in order to get a new left hand side. The data enters only through the expectation of the value function, which in turn depends on $\widehat{\gamma}$, which depends on the data through $\delta^{imt}(j_{i,t-1})$. Once the value function has been updated, $\delta^{imt}(j_{i,t-1})$, $\widehat{\gamma}$, and the value function are all checked for convergence. If they have not all converged, another iteration begins, starting with recalculation of the $\delta_j^{imt}(j_{i,t-1})$ based on the new values of the other variables.

After convergence has been achieved, transition probabilities are calculated using the newly computed value function. These transition probabilities are arranged into a transition matrix for each year. The transition matrix, plus shares for an initial conditions year, are used to predict market shares for every plan and county in each year of the data. Once this process has been completed for each random coefficient draw, the shares are averaged over the random coefficient draws.

8.2.2 Middle Loop

The middle loop is the Berry, Levinsohn, Pakes (1995) inversion. This inversion is based on the insight that there is a one-to-one mapping between the mean flow utilities and market shares. It gives an iterative procedure to update the mean flow utilities until the predicted market shares match the observed market shares. While the BLP inversion is a contraction mapping in the static case, it is not guaranteed to be a contraction mapping in the dynamic case.

The mean flow utilities,$\overline{f_{mjt}}$, are updated according to this mapping:

$$\overline{f_{mjt}^{new}} = \overline{f_{mjt}^{old}} + \Psi(\ln(s_{mjt}) - ln(\hat{s}_{mjt}(\overline{f_{mjt}^{old}}, \eta, \sigma))) \tag{25}$$

where Ψ is a tuning parameter, s_{mjt} is the county and plan level market share observed in the data, and $\hat{s}_{mjt}$ is the corresponding estimated market share, which is a function of a mean flow utility and η and σ, the candidate parameter values passed in from the outer loop. The mapping is iterated on until $\overline{f_{mjt}^{new}}$ and $\overline{f_{mjt}^{old}}$ match, up to some tolerance. Calculating $\hat{s}_{mjt}(\overline{f_{mjt}^{old}},\eta, \sigma)$ entails invoking the inner loop, which solves the dynamic programming problem and calculates shares based on the arguments of $\hat{s}_{mjt}(\overline{f_{mjt}^{old}},\eta, \sigma)$. Convergence of the middle loop, therefore, is actually joint convergence of the middle and inner loops.

Once the mean flow utility has converged, the mean coefficient vector α can be found by doing an instrumental variables regression of the flow utilities for each plan and county combination on the plan characteristics. The residuals from this regression form the vector of errors, ξ. Notice that ξ can be thought of as a function of the parameters η and σ, because the values of ξ that come out of the middle loop will depend on the η and σ fed into the inner loop.

8.2.3 Outer Loop

The outer loop is a Generalized Method of Moments procedure, minimizing a nonlinear criterion function over the parameter vector (η, σ). The identifying assumption used here is that the instrument matrix Z is orthogonal to the error vector ξ.

Define the following function:

$$G(\eta, \sigma) = Z'\xi(\eta, \sigma)$$

Then, the minimization problem is:

$$\min_{\eta,\sigma}\{G(\eta, \sigma)'WG(\eta, \sigma)\}$$

where W is a weighting matrix. Initially, W is set to $(Z'Z)^{-1}$. In the second stage, it is updated to the optimal weighting matrix according to standard results.[15]

The algorithm terminates when the outer loop has found a minimum. During the optimization, $G(\eta, \sigma)$ will be evaluated at many different parameter vectors. Each time it is evaluated, the middle loop is invoked to find $\xi(\eta, \sigma)$ The middle loop, in turn, invokes the inner loop many times for each evaluation of $\hat{s}_{mjt}(\overline{f^{old}_{mjt}}, \eta, \sigma)$. Because of this nesting, at termination all three loops will have jointly converged.

8.3 Standard Errors

Because the estimation procedure is a form of the Generalized Method of Moments, the usual formulas for GMM standard errors apply here. A caveat is that using simulation draws for the random coefficients introduces an extra source of variation that will not be accounted for in these standard errors. An alternative would be to bootstrap the standard errors, resampling counties and repeating the estimation many times, but this is not feasible due to the length of time each estimation would take. The regular GMM standard errors estimates are a lower bound for the correctly estimated standard errors, and may not be very inaccurate as long as a sufficient number of simulation draws are taken to reduce simulation error.

9 Results and Counterfactuals

9.1 Coefficients on Plan Characteristics

Parameter estimates and standard errors are in Table 7. The coefficients on the plan characteristics have a mix of expected and unexpected signs.

The coefficient on premium has a negative sign, as expected, and is significant at the 5% level. The magnitude of this coefficient is of particular importance because it is used (along with the type-specific random coefficient contributions) to find the marginal utility of income, as suggested by Small and Rosen (1981). Accurately calculating the dollar value for the switching cost and consumer welfare measures depends on this coefficient being of the correct magnitude. Since the reported dollar values are all plausible, there is no reason to suspect a problem with this coefficient.

The coefficients on the glasses coverage indicator, the sum of the drug limits, the routine eye coverage indicator, and network size all have the expected sign. The coefficients on several plan characteristics

[15] The results reported in this paper are from the first stage only. First stage results are consistent but not efficient,

do not have the expected sign, yet are statistically significant. These are the coefficients on the dental coverage indicator, the cost of an emergency room visit, the drug coverage indicator and the "no limit" drug coverage indicator. The negative coefficient on the prescription drug indicator was particularly unexpected, because anecdotally drug coverage is important to elderly consumers. Interestingly, the coefficient on the Drug Discount Card indicator is positive, large and significant. One possibility is that the regular drug coverage that most plans have is not very good compared to the Drug Discount Card and the consumers' preference for drug coverage is showing up only in the coefficient on the Discount Card indicators. Another possibility is that there is a high degree of collinearity between the drug indicators and an excluded characteristic which consumers do not like. Different specifications that include more detailed drug coverage variables or variables about additional aspects of coverage might remedy the sign problem.

9.2 The Switching Cost

The switching cost parameter has the correct sign and is statistically significant at the 5% level. In dollars, the switching cost for the median consumer is $4162.90. At first glance this number is large, but it is not inconsistent with other estimates of health plan switching costs in the literature. For example, Handel (2011) finds health plan switching costs around $2000, and he studies working age people at a large firm, who are likely to be more amenable to change than the elderly population in my data. Furthermore, the average health expenditure per person for people over 65 is about $11,000, indicating that a switching cost of the estimated magnitude would not completely swamp the value of the health services these consumers are receiving. Hence, the estimated switching cost is large enough to suggest that switching costs are indeed a very important determinant of consumer choices in this market, but not so large as to be implausible.

Another way to look at the switching cost is to consider elasticities with respect to the switching cost. Such elasticities are reported in Table 8. The first set of elasticities are of switching rates with respect to switching costs. The switching rates are the fraction of consumers in a given starting state who make a particular type of switch. The rate of switching between MA plans and from original Medicare to an MA plan both have a large, negative elasticity with respect to switching cost, indicating that a small increase in the switching cost causes a proportionally large decrease in this type of switching. On the other hand, the rate of switching from an MA plan to original Medicare actually increases when the switching cost goes up, though not by much as the elasticity is in the inelastic range. This positive elasticity is surprising, since one might expect all switching rates to decrease in response to an increase in the switching cost. What may be happening is that consumers are reacting to the cost of future switching as well as current switching, and consumers in original Medicare are less likely to switch in the future, as their plan can never decrease in quality or exit. The total share in MA has close to negative unit elasticity with respect to the switching costs, which also supports this idea that consumers tend to flee MA plans in favor of original Medicare as the switching cost increases.

9.3 Random Coefficient Distribution

The estimated standard deviation parameters for the two random coefficient distributions are both close to zero and insignificant, indicating very little persistent consumer heterogeneity. In other words, there is no evidence that consumers tend to choose plans similar to their previous choices when changing plans. This result is surprising. Since consumer choice in this market likely depends on health status, and health status tends to be both variable across consumers and enduring for a given consumer, it seems that a high degree of persistent consumer heterogeneity would be present. There are several explanations for this result other than a true lack of consumer heterogeneity. First, it could be that the random coefficients are on the wrong variables. A random coefficient distribution is estimated only for the price and constant term, but it is possible that the important heterogeneity is in a different dimension, such as a different plan characteristics. Second, this result could indicate an identification problem. It is admittedly difficult to sort out persistence in market shares due to consumer tastes and persistence in market shares due to the switching cost. It is possible that the high estimate for the switching cost and the low estimates for the parameters of the random coefficient distributions indicate that some of the persistence that should have been attributed to consumer tastes was instead attributed to the switching cost. Finally, this result could indicate a problem with the way that initial conditions are handled. By assumption, the random coefficient distribution is the same within the consumers in each plan in the initial conditions year. In reality, a consumer is in a plan in that year because he chose it at some point in the past, so consumers in a given plan will tend to be those consumers who like that type of plan. Not allowing the initial configuration of market shares to inform the random coefficient distribution could be leading to these implausible estimates. All of these issues can be further explored. The misspecification of which variables have random coefficients can be addressed by trying alternative specifications. The identification problem could be mitigated with instruments that better identify the different sources of persistence. Finally, the initial conditions could be handled in a more sophisticated way, either by extending the estimation backwards to the beginning of Medicare HMOs, or by devising a better scheme for assigning random coefficient distributions within shares in the current initial conditions year.

9.4 Counterfactuals

I compute four counterfactuals: a counterfactual in which switching costs are zero, a counterfactual in which consumers pay an annual surcharge when enrolled in an MA plan, a counterfactual in which plans do not enter, exit or change in quality after a particular year, and a counterfactual in which all MA plans exit the market. In this section, I discuss the general procedure used for all the counterfactuals and how to interpret the results of the counterfactuals in a dynamic setting. Details about each counterfactual are in the subsections that follow.

In each counterfactual, consumers are allowed to re-optimize in reaction to the change. In addition to changes in the consumers' choices in a given period, consumers also form different expectations in the counterfactuals. Therefore, the solution to the Bellman equation can be different in the counterfactuals, since the value function depends on expectations, and the inner loop dynamic programming problem must be solved again for each counterfactual. Firms, unlike consumers, are assumed not to re-optimize in any way. Since the model does not have a supply side, it is impossible to simulate how firms would react to these changes. While in reality there are many ways firms could react in the counterfactual scenarios, such as entering or exiting markets or changing plan characteristics or premiums, everything on the firm side is held constant in the counterfactuals except where explicitly noted. Despite this limitation, there is much to be learned from the consumer-focused counterfactuals. Studying the consumer side in isolation allows for the exploration of consumer substitution patterns and what is driving demand in this market.

For each counterfactual, I calculate a welfare change that is essentially a compensating variation– the amount of money that would induce an equivalent change in expected utility, taking into account the consumers' ability to re-optimize. Small and Rosen (1981) derive the formula for compensating variation in a Logit setting, but the random coefficients and the dynamics add extra complications here. The random coefficients are dealt with by calculating compensating variation separately by type, then integrating over types by multiplying each quantity by the number of consumers of that type and summing. The dynamics are taken care of by including the expected discounted infinite horizon value function in the utility of each product. Of course, this means that the interpretation of the welfare change is different because it includes the change in infinite horizon future utility. Since such numbers can be difficult to interpret, I decompose them into a component consisting of the change in expected utility experienced in the current period and a component consisting of the change in the discounted expected utility in all future periods. The key to correctly calculating these numbers is to always include both the current and future components when determining consumers' optimal choices. The separate components can then be calculated by finding the current period utility conditional on these choices and subtracting it from the total. To further ease interpretation, the expected welfare changes are always reported as per-person averages across all Medicare eligibles rather than totals.

The current period change in expected welfare is further decomposed into the component that results from changes in switching costs paid, and the component that results from consumers making different plan choices in the counterfactual. This decomposition serves two purposes. First, it emphasizes the different channels through which consumer welfare is affected in the counterfactuals, and shows their relative importance. Second, it allows for an alternative interpretation of the switching cost. Some think of switching costs as a psychological impediment to decision making or a friction rather than a "real" cost. Under that interpretation, the component of the welfare change due directly to the switching cost should be ignored, as only its effect on consumer decision making, which is accounted for in the other component, is relevant.

In addition to the change in welfare, the actual and counterfactual share in MA are reported. The share in MA is calculated as the number of consumers choosing a MA plan out of all Medicare eligible consumers in the data. While changes in this share under the counterfactuals illustrate broad substitution patterns, changes in shares at the county or plan level, which might go in different directions, are obscured when only observing the share at this level. For example, substitution across MA plans cannot be detected just by looking at the overall MA share.

In each case, the welfare change and counterfactual market shares are reported for the years 2002-2005. These are the years for which the consumers' choice sets and market shares can be observed in the data. While the total welfare change figure for 2002 includes infinite horizon expected future value, it does not include all the information in the 2003-2005 figures because these can account for the realization of the choice sets and plan quality in these years. Since the data stops with 2005, these choice sets are not known past that year, which is why 2005 is the last year for which individual year welfare is reported. However, the expectation for all future years is included in the total 2005 number, given what consumers know at that point.

9.4.1 Counterfactual: No Switching Cost

In this counterfactual, the switching cost is permanently set to zero starting in 2002. The change is unexpected in the sense that consumers are assumed to make all the same decisions before 2002, but consumer expectations are allowed to change from 2002 on. While there is no actual policy that would reduce switching costs all the way to zero, there are some ways that switching costs could be reduced substantially, such as limiting the restrictiveness of provider networks, or further standardizing the benefits that plans offer. This counterfactual is meant to measure the complete effect of the switching cost on consumers, and provide an upper bound for how much these policies aimed at reducing switching costs could help consumers.

The actual and counterfactual market shares are reported in Table 9. In each year, the counterfactual share in MA plans is more than triple the actual share. This finding suggests that the switching cost is one of the main factors preventing more consumers from choosing MA plans over original Medicare.

The expected changes in welfare are reported in Table 10. In every year the expected current period welfare gain is around \$1000 per person. The increase in utility in the absence of switching costs can come from two sources: consumers who switch in either case no longer paying the switching cost, and consumers who only switch in the absence of the switching cost ending up in better plans. Interestingly, most of the welfare gain is through the plan choice and not directly from the switching cost. In terms of determining the actual effect of the switching cost on consumers utility, this \$1000 figure is perhaps a better number to focus on than the \$4162.90 switching cost. It indicates how much the switching cost affects the utility of the average consumer, given that some consumers optimally choose in a way that avoids the switching cost.

9.4.2 Counterfactual: Medicare Advantage Surcharge

In this counterfactual, each consumer who chooses an MA plan pays Medicare a surcharge equal to 10% of the base capitation payment for plans in their county. Consumers choosing original Medicare do not pay the surcharge. As with the previous counterfactual, the policy change occurs in 2002 and is unexpected but permanent. The reason that this is an interesting policy experiment is that it is thought that the plans are overpaid by about 10%, in that it would cost about 10% less to offer the same consumers coverage under original Medicare. In the counterfactual, consumers are taking on the burden of the extra cost instead of Medicare. Table 11 reports the average surcharges under this policy.

The counterfactual market shares are reported in Table 12. Under the counterfactual, the market share of MA in 2002 is about half of the actual market share, and it decreases in subsequent years down to 2.46% in 2005. The clear pattern is that consumers are switching out of MA plans over time once the surcharge goes into effect. The gradual rather than immediate change is attributable to the switching cost. Some consumers need a high ε draw for original Medicare in order to make the switch worthwhile, which can be interpreted as a positive one-time shock to the value to the consumer of original Medicare.

The expected changes in welfare, reported in Table 13, show a pattern consistent with the decline in the market shares. In 2002, the current period welfare difference is -$277.88 per person. At this point, nearly 10% of consumers are still in the MA plans and paying the surcharge. By 2005, few consumers are left in the MA plans, and the welfare difference is only -$89.23. In 2005, much of the utility loss is coming from the consumers who choose original Medicare in the counterfactual, but preferred MA plans without the surcharge. Very few consumers are paying the surcharge in 2005.

In addition to the effects on consumers, this policy also affects Medicare in two ways. First, Medicare receives the revenue generated from the surcharge. Second, Medicare saves money on the consumers who switch to original Medicare who would have been in the more expensive MA plans. Table 14 reports the revenue and approximate savings to Medicare. The approximate savings are found by taking 10% of the county-specific capitation payment for each person who is in original Medicare in the counterfactual but not in the status quo.[16] In every year except 2002, the combined revenue and savings per person exceed the absolute value of the average consumer's current period change in welfare. The savings and revenue would therefore be enough to compensate consumers for their lost welfare under this policy.

Overall, the evidence from this counterfactual suggests that most consumers in MA plans would not be willing to internalize the extra cost of their coverage, and that essentially overpaying the MA plans is not efficient in the long run.

9.4.3 Counterfactual: No Exit, Entry, or Quality Change

In this counterfactual, the set of plans offered and their flow utilities are fixed from 2002 onward. In other words, no plans enter, exit or change their coverage after 2002. This scenario is interesting to consider because the frequent changes in the choice set and plan quality may be driving some of the effects of the switching cost. If the market were more stable in this respect, there might be less benefit to switching, and thus the switching cost would have less bite. While the counterfactual policy begins in 2002, the plans, choice sets, and flow utilities used are those from 2003. The plans from 2003 have the lowest mean flow utility, so using these plans eliminates confusion between the effect of having the

[16] The actual savings would depend on the services used by the consumers because of the fee-for-service structure of the coverage of original Medicare. The estimated savings will be inaccurate if the consumers choosing original Medicare because of this policy systematically consume more or fewer services than the typical fee-for-service enrollee. Also, note that the sum of the savings and revenue is fixed in a sense, because each consumer who would be in MA without the surcharge either must pay the surcharge, or choose original Medicare and save Medicare an amount equal to the surcharge. The differences in this sum from year to year are caused by changes in the total number of Medicare eligibles.

same plans available and the effect of having *better* plans available. Since the plans from 2003 are the worst, an overall increase in utility cannot be coming from an increase in the quality of the plans available in some year.

The counterfactual market shares are reported in Table 15. Under the counterfactual, the market share in MA almost doubles from its actual value by 2005. The higher market share indicates that consumers are more willing to be in MA plans when they are less subject to exit and quality change.

The expected changes in welfare are reported in Table 16. The single period welfare change is negative in three of the four years, indicating that consumers are worse off because of the policy during those years. However, the negativity is attributable only to the switching cost component– the plan choice component is positive in every year, implying that consumers choose plans with higher flow utility in the counterfactul. Furthermore, the total welfare change, which includes future periods, is positive. Consumers are willing to incur the switching cost, which greatly reduces current period utility, in order to choose plans that will make them better off in the future. This trade-off becomes worthwhile in the counterfactual because consumers know that the plans will still be in the market and still have the same flow utility in future periods.

9.4.4 Counterfactual: All Medicare Advantage Plans Exit

In the final counterfactual, all MA plans unexpectedly exit. This counterfactual is calculated separately for exit in each year from 2002 to 2005. The motivation for this counterfactual is twofold: first, to evaluate the impact if the Medicare Advantage program were eliminated, and second to determine the value of the Medicare Advantage program to consumers.

If the Medicare Advantage program were eliminated, everyone who was in an MA immediately prior to the elimination would be automatically switched to original Medicare and incur a switching cost. The program's exit would therefore have a large welfare impact through switching costs in addition to the welfare lost by those who prefer the MA plans to original Medicare. The columns in Table 17 labeled "All S.C." report the expected change in welfare including these forced switching costs. Consumers lose about $700 in current period welfare when the program exits, but most of it acts through the switching cost. Only arount $100-$200 is resulting from the lower quality of original Medicare compared to MA plans.

When considering the value of the MA program, the switching costs resulting from its exit are not relevant. The appropriate counterfactual compares welfare from the program to welfare if the program didn't exist at all. The columns in Table 17 labeled "No S.C. in Counter" calculate the change in welfare without including the switching costs that would result directly from the program's exit. (Switching costs for the case where MA plans are still available are left in.) These numbers represent the dollar value of the program (per Medicare eligible), or the change in welfare in subsequent years after the program's exit when the switching cost has already been paid. Interestingly, when calculated this way the total change in welfare is positive in some years. In these years, the average consumer in an MA plan would be better off if he could costlessly switch from his MA plan to original Medicare. In a sense, consumers actually place a negative value on the program in these years. What are nearly

20% of Medicare eligibles doing in MA plans if on average consumers would prefer to be in original Medicare in absence of switching cost? The MA plans must be valuable to consumers in some years, and consumers subsequently remain in them mostly because of switching costs. This behavior is not myopic as long as consumers are considering expected net present value when making their choices, which might entail trading off some future utility for a better plan today. The welfare numbers for 2005 provide an example where both current and future utility are improved on net by the MA plans. For the 2002 to 2004 numbers to make sense, there must have been some other years with this property prior to 2002. In any case, the overall impression left by this counterfactual is that consumers do not value Medicare Advantage all that highly, and eliminating it would not have a very large impact on utility.

9.5 Conclusion

In this paper, I develop and estimate a structural dynamic model of consumer choice in the market for Medicare Advantage plans. A main focus throughout the paper is on switching costs, which arise in this market because of restrictive provider networks, substantial differences in plan benefits, and learning and search costs associated with choosing a different plan. I find that switching costs are indeed large, and that many more consumers would enroll in MA plans if switching costs were lower. I also find that given the high switching costs and other undesirable features of this market, consumers do not value MA plans very highly, as evidenced by the fact that only a small percentage of consumers would be willing to pay a surcharge to remain in an MA plan. These results offer an important lesson in the design of health insurance exchanges: consumers will not benefit much from having extra coverage options if switching costs are high enough to interfere with efficient matching of consumers to plans.

10 Tables

Table 1: Number of Medicare Advantage Plans

	2002	**2003**	**2004**	**2005**
Distinct Plans	170	198	190	234
Plan-County Comb.	1733	2003	2193	3150
Min Plans/County	1	1	1	1
Max Plans/County	11	13	13	19
Med. Plans/County	2	2	2	4

Note: Original Medicare, which is available in every county in every period, is not included here.

Table 2: Entry and Exit

Year	**Total Plan-Counties**	**Exiters**	**Entrants**
2002	1733	275	286
2003	2003	80	350
2004	2193	268	358
2005	3150	74	1391

Note: Entry and exit are reported on the plan-county level, in that a single plan that exits or enters multiple counties is counted as a separate exit or entry for each county.

Table 3: Average Market Shares of New and Existing Medicare Advantage Plans

	Average Market Share		**Dif. in Means**
Year	**New Plan**	**Existing Plan**	**T-statistic**
2002	0.0238	0.0479	-12.46
2003	0.0045	0.0449	-25.39
2004	0.0038	0.0411	-24.88
2005	0.0017	0.0390	-28.10

Notes: A "new plan" is a plan that entered in the given year. An "existing plan" is a plan that entered in any previous year. Original Medicare is not counted in either category.

Table 4: Summary Statistics for Plan Characteristics

		2002	2003	2004	2005
Premium	Mean	50.48	59.97	51.33	27.83
	Min	0.00	-20.00	-20.00	-78.20
	Max	180.00	196.00	215.00	210.00
Dental	Mean	0.13	0.17	0.16	0.24
(Indicator)	Min	0	0	0	0
	Max	1	1	1	1
Cost of Emergency Visit	Mean	45.95	46.63	44.43	44.68
	Min	0.00	0.00	0.00	0.00
	Max	161.80	166.00	50.00	50.00
Glasses Coverage	Mean	0.32	0.29	0.26	0.18
(Indicator)	Min	0	0	0	0
	Max	1	1	1	1
Drug coverage	Mean	0.43	0.47	0.53	0.64
(Indicator)	Min	0	0	0	0
	Max	1	1	1	1
No limit for some drug category	Mean	0.17	0.22	0.29	0.48
(Indicator)	Min	0	0	0	0
	Max	1	1	1	1
Sum of Drug Limits	Mean	263.34	310.95	258.29	293.43
	Min	0	0	0	0
	Max	3800.00	3800.00	3000.00	7500.00
Cost of Primary Care Visit	Mean	13.33	12.78	12.32	11.25
	Min	0.00	0.00	0.00	0.00
	Max	25.00	27.50	27.50	30.00
Routine Eye coverage	Mean	0.50	0.45	0.40	0.29
(Indicator)	Min	0	0	0	0
	Max	1	1	1	1
Drug Discount Card	Mean	0.00	0.00	0.00	0.42
(Indicator)	Min	0	0	0	0
	Max	0	0	0	1
Demo Plan	Mean	0.06	0.18	0.19	0.18
(Indicator)	Min	0	0	0	0
	Max	1	1	1	1
Private Fee For Service Plan	Mean	0.29	0.28	0.31	0.39
(Indicator)	Min	0	0	0	0
	Max	1	1	1	1
Network Size	Mean	0.00	0.00	0.06	0.05
	Min	0.00	0.00	0.00	0.05
	Max	0.00	0.00	1.82	2.21
No Network Information	Mean	1.00	1.00	0.67	0.75
(Indicator)	Min	1	1	0	0
	Max	1	1	1	1
Number of Observations		1733	2003	2193	3150

Note: Statistics are all computed using plan-county level data.

Table 5: Results from OLS Regression of Current Share on Lagged Share, Controlling for Plan Characteristics

	Estimate	**Standard Error**
Constant Term	0.0080224	0.0011369
Lagged Share	0.9805048**	0.0049084
Premium	-0.0000122**	0.0000047
Dental Coverage	-0.0010996**	0.0004763
Cost of emergency visit	-0.0000220	0.0000170
Glasses Coverage	-0.0019902**	0.0006138
Drug Coverage	-0.0016283**	0.0006554
No limit for some drug category	-0.0005801	0.0006094
Sum of drug limits	0.0000003	0.0000005
Cost of primary care visit	-0.0001616**	0.0000374
Routine Eye Coverage	0.0035117**	0.0006250
Drug Discount Card	0.0022515**	0.0005784
Demo Plan	0.0003359	0.0007106
Private Fee For Service	0.0006321	0.0006205
Network Size	0.0029191*	0.0016685
No network information	-0.0023896**	0.0005185
Year 2002	-0.0006243	0.0005585
Year 2003	-0.0024031**	0.0004171
Year 2004	-0.0017469**	0.0003363

Notes: Plan characteristics and the left-hand-side share are from the same year; the lagged share is from the previous year. A single star indicates statistical significance at the 10% level and a double star indicates statistical significance at the 5% level.

Table 6: Means of Selected Characteristics of Medicare Advantage Plans Available in 2005 by Year of Entry

	Year of Entry				
	2001 or earlier	**2002**	**2003**	**2004**	**2005**
Premium	36.31	35.29	55.14	23.06	14.29
Dental	0.23	0.19	0.61	0.19	0.19
Cost of Emergency Visit	47.93	46.81	48.38	46.54	40.00
Glasses	0.35	0.38	0.07	0.10	0.04
Drug Coverage	0.50	0.58	0.47	0.75	0.79
No Limit for some Drug Category	0.33	0.28	0.35	0.58	0.65
Sum of Drug Limits	309.16	252.94	477.61	325.84	229.22
Cost of Primary Care Visit	11.94	11.91	12.17	12.24	10.06
Routine Eye Coverage	0.53	0.47	0.16	0.12	0.10
Drug Discount Card	0.30	0.20	0.21	0.53	0.56
Number of observations	1464	204	297	358	1391

Notes: Means are calculated using plan-county level data. Characteristics are all from the 2005 data, and the columns divide the plans-counties by the year the plan entered the county.

Table 7: Parameter Estimates from Structural Dynamic Model

	Parameter	Estimate	Standard Error
Non-linear Parameters	**Switching Cost**	4.64800**	2.09795
	σ_0 **(SD for constant term RC)**	0.00046	0.55587
	σ_1 **(SD for premium RC)**	0.00001	0.05562
Coefficients	**Constant term**	-0.34592**	0.07189
	Premium	-0.01340**	0.00223
	Dental Coverage	-0.08659**	0.03406
	Cost of emergency room visit	0.00382**	0.00117
	Glasses Coverage	0.00639	0.03858
	Drug coverage	-0.17108**	0.04121
	No limit for some drug category	-0.26467**	0.05062
	Sum of drug limits	0.00005*	0.00003
	Cost of primary care visit	0.01024	0.00643
	Routine eye coverage	0.22276**	0.05559
	Drug Discount Card	0.63108**	0.05639
	Demo plan	-0.00145	0.05273
	Private fee for service plan	-0.30913**	0.04961
	Network size	0.38020**	0.09557
	No network info	-0.55206**	0.04028
	2002 indicator	0.87197**	0.07863
	2003 indicator	0.80015**	0.09649
	2004 indicator	1.15000**	0.11634

Note: A single star indicates statistical significance at the 10% level. A double star indicates statistical significance at the 5% level.

Table 8: Elasticities with Respect to Switching Cost

		Elasticity
Switching Rates	Between MA Plans	-2.46
	From Orig. Med to MA	-8.29
	From MA to Orig Med	0.61
Share	Total share in MA	-0.96

Notes: Switching rates are calculated as the percentage of consumers making a particular type of switch given a particular starting state. For example, the "Between MA Plans" switching rate is the percentage of consumers who switch to a different MA plan out of the consumers who start in an MA plan. All switching rates are aggregated across years. The total share in MA is the percentage of all Medicare eligibles who are enrolled in any MA plan, aggregated across years.

Table 9: Market Share of Medicare Advantage Plans for Counterfactual with Zero Switching Cost

Year	Factual Share in MA	Counterfactual Share in MA
2002	18.10%	65.56%
2003	16.95%	64.80%
2004	16.76%	69.16%
2005	17.53%	79.96%

Notes: In the counterfactual, the switching cost is permanently set to zero starting in 2002. The "Share in MA" is the fraction of Medicare eligibles who are enrolled in any MA plan in the given year.

Table 10: Expected Change in Welfare for Counterfactual with Zero Switching Cost

	Expected Change in Welfare				
	Current Period Only			Future periods	Total (Sum of
Year	Total	From Switching Cost	From Plan Choice	(discounted)	Current & Future)
2002	$1069	$272	$797	$13,270	$14,339
2003	$941	$194	$747	$13,159	$14,100
2004	$929	$93	$836	$13,497	$14,426
2005	$1150	$104	$1046	$14,240	$15,390

Notes: In the counterfactual, the switching cost is permanently set to zero starting in 2002. All numbers are an average per person change in welfare across all Medicare eligibles. It is assumed that consumers reoptimize in the counterfactual. The "From Switching Cost" column gives the change in current period welfare due to switching costs paid in the status quo but not the counterfactual. The "From Plan Choice" column gives the change in current period welfare due to consumers' different plan choices under the counterfactual. These sum to the current period total. The "Future Period" column gives the discounted expected change in welfare summed across all periods following the current period.

Table 11: Average Surcharge in Counterfactual with Medicare Advantage Surcharge

	Average Surcharge	
Year	**Monthly**	**Annually**
2002	$59.27	$711.23
2003	$60.44	$725.26
2004	$62.48	$749.80
2005	$72.15	$865.75

Notes: The surcharge is set to 10% of the base capitation payment in the consumer's county. The averages are across counties, weighted by the number of Medicare eligibles in the county.

Table 12: Market Share of Medicare Advantage Plans for Counterfactual with Medicare Advantage Surcharge

Year	**Factual Share in MA**	**Counterfactual Share in MA**
2002	18.10%	9.71%
2003	16.95%	5.47%
2004	16.76%	3.51%
2005	17.53%	2.46%

Notes: In the counterfactual, any consumer choosing an MA plan pays a surcharge of 10% of the base capitation payment in his county in addition to the regular plan premium. The "Share in MA" is the fraction of Medicare eligibles who are enrolled in any MA plan in the given year.

Table 13: Expected Change in Welfare for Counterfactual with Medicare Advantage Surcharge

Year	Expected Change in Welfare					
	Current Period Only				Future periods (discounted)	Total (Sum of Current & Future)
	Total	From Switching Cost	From Surcharge	From Plan Choice		
2002	-$277.88	-$195.38	-$71.27	-$11.23	-$200.37	-$478.25
2003	-$113.50	-$55.28	-$40.71	-$17.51	-$59.64	-$173.14
2004	-$76.46	$4.41	-$26.73	-$54.14	-$32.25	-$108.71
2005	-$89.23	$55.93	-$21.42	-$123.74	-$57.70	-$146.93

Notes: In the counterfactual, any consumer choosing an MA plan pays a surcharge of 10% of the base capitation payment in his county in addition to the regular plan premium. All numbers are an average per person change in welfare across all Medicare eligibles. It is assumed that consumers reoptimize in the counterfactual. The "From Switching Cost" column gives the change in current period welfare due to the difference in switching costs paid in the status quo and the counterfactual. The "From Surcharge" column gives the change in current period welfare due to the surcharge payments made in the counterfactual. The "From Plan Choice" column gives the change in welfare due to consumers' different plan choices under the counterfactual. These three columns sum to the current period total. The "Future Period" column gives the discounted expected change in welfare summed across all periods following the current period.

Table 14: Revenue and Savings to Medicare in Counterfactual with Medicare Advantage Surcharge

	Revenue		Savings		Sum of Revenue & Savings	
Year	Total	Per Person	Total	Per Person	Total	Per Person
2002	$1,980,200,000	$71.27	$1,724,300,000	$62.07	$3,704,500,000	$133.34
2003	$1,145,200,000	$40.71	$2,442,700,000	$86.83	$3,587,900,000	$127.54
2004	$761,870,000	$26.73	$2,944,600,000	$103.31	$3,706,470,000	$130.04
2005	$619,220,000	$21.42	$3,903,700,000	$135.05	$4,522,920,000	$156.47

Notes: Total revenue is the sum of all surcharges paid by consumers who choose MA plans. Total savings is the approximate savings from the consumers who choose original Medicare in the counterfactual but an MA plan in the status quo. Because of the way that the surcharge is chosen, the sum of revenue and savings is fixed. The small differences between years results from the different number of Medicare eligibles.

Table 15: Market Share of Medicare Advantage Plans for Counterfactual with no Entry, Exit, or Quality Change

Year	Factual Share in MA	Counterfactual Share in MA
2002	18.1%	24.33%
2003	16.95%	28.22%
2004	16.76%	31.12%
2005	17.53%	33.38%

Notes: In the counterfactual, the set of plans available in 2003 in the status quo becomes the set of plans available in every year from 2002 on, with no entry, exit or changes in plan flow utility. The "Share in MA" is the fraction of Medicare eligibles who are enrolled in any MA plan in the given year.

Table 16: Expected Change in Welfare for Counterfactual with no Entry, Exit, or Quality Change

Year	Expected Change in Welfare				
	Current Period Only			Future periods (discounted)	Total (Sum of Current & Future)
	Total	From Switching Cost	From Plan Choice/ Plans Available		
2002	-$161.28	-$293.16	$131.88	$1114.28	$953.00
2003	-$8.55	-$152.68	$144.13	$1284.35	$1275.80
2004	$2.07	-$114.87	$116.94	$1346.83	$1348.90
2005	-$11.67	-$59.83	$48.16	$1356.57	$1344.90

Notes: In the counterfactual, the set of plans available in 2003 in the status quo becomes the set of plans available in every year from 2002 on, with no entry, exit or changes in plan flow utility. All numbers are an average per person change in welfare across all Medicare eligibles. It is assumed that consumers reoptimize in the counterfactual. The "From Switching Cost" column gives the change in current period welfare due to the difference in switching costs paid in the status quo and the counterfactual. The "From Plan Choice/Plans Available" column gives the change in current period welfare in the current period due to consumers choosing different plans, which here can be due to the different choice set. These two columns sum to the current period total. The "Future Period" column gives the discounted expected change in welfare summed across all periods following the current period.

Table 17: Expected Change in Welfare in Counterfactual where all Medicare Advantage Plans Exit

Year	Expected Change in Welfare							
	Current Period Only					Future periods (discounted)	Total (Sum of Current & Future)	
	Total		From Switching Cost		From Plan			
	All S.C.	*No S.C. in Counter*	*All S.C.*	*No S.C. in Counter*			*All S.C.*	*No S.C. in Counter*
2002	-$706.24	$129.01	-$563.04	$272.21	-$143.20	$75.79	-$630.45	$204.80
2003	-$705.77	$38.20	-$612.98	$130.99	-$92.79	$96.67	-$609.10	$134.87
2004	-$709.88	-$13.57	-$602.51	$93.80	-$107.37	$64.40	-$645.48	$50.83
2005	-$750.99	-$63.04	-$584.23	$103.72	-$166.76	$6.33	-$744.66	-$56.71

Notes: Each line is a seperate counterfactual in which all MA plans exit in the given year, leaving only original Medicare. All numbers are an average per person change in welfare across all Medicare eligibles. It is assumed that consumers reoptimize in the counterfactual. The "From Switching Cost" column gives the change in welfare due to the difference in switching costs paid in the status quo and the counterfactual. The "From Plan" column gives the change in welfare due to the difference in current period flow utility of original Medicare and the plan chosen in the status quo. These two columns sum to the current period total. The "Future Period" column gives the discounted expected change in welfare summed across all periods following the current period. Columns labeled "All S.C." are calculated taking all switching costs into account, including those induced by consumers who involuntarily switch to original Medicare in the counterfactual. Columns labeled "No S.C. in Counter" are calculated assuming that consumers do not pay the switching cost when original Medicare exits in the counterfactual.

11 Appendix: Definitions of Plan Characteristics

General Characteristics Variables

Premium: The amount that the enrollee must pay the firm offering the plan in addition to the amount paid for regular Medicare Part B coverage. In the initial two years that the data covers, the premium can only be zero or positive, but starting in 2002 the premium can be negative. A negative premium means that the Medicare Advantage Organization refunds the enrollee some or all of the amount paid to the CMS for Part B coverage.

Primary care: The amount that the patient would pay for a visit to a primary care physician under the plan. In some cases an exact co-pay is given in the original data. If instead a range is given, I take the midpoint of the range. In a few cases a percentage that the enrollee is responsible for is given, in which case I multiply the percentage by $100, which I take as the average cost of a primary care visit in the absence of insurance.

Dental: An indicator for whether any dental services are covered under the plan.

Emergency: The amount a patient would pay on average for an emergency room visit under the plan. If a range is given, I take the upper end of the range. If a percentage is given, I multiply the percentage times the average cost of an emergency room visit in that year ($751-$829).

Hearing and Vision Variables

Discount Hearing Aid: An indicator for whether the plan offers any coverage for hearing aids. The coverage can take the form of a fixed amount that the patient pays, a non-zero coverage limit for hearing aids, or hearing aids provided free of charge.

Routine eye coverage: Indicator for whether the plan covers routine eye exams (as opposed to eye exams intended to treat diseases of the eye). I include cases where there is an annual coverage limit or the patient pays a co-pay under the coverage case.

Glasses coverage: Indicator for whether the plan offers any coverage for glasses, frames, or lenses. There may be a coverage limit or a fixed amount or percentage that the patient pays.

Prescription Drug Variables

Drug: An indicator for whether the plan offers any drug coverage at all. It is set to zero if the drug field contains "You pay 100% for most prescription drugs" or "You pay 100% for non-Medicare prescription drugs" and one otherwise.

Sum of Drug Limits: The sum of coverage limits across all categories of drugs (for example, brand and generic drugs, or different tiers of a formulary). A zero means that no drug coverage is offered unless the coverage limit variable is zero and the 'no limit' variable is one.

No Limit: An indicator for whether there is unlimited coverage for at least one category of drugs.

DDC: An indicator for whether the enrollees have the option of buying a Drug Discount Card to supplement the plan. Only plans in 2005 can have this option.

Max cost 30: Maximum across categories of drugs of the out of pocket cost to patients for a 30 day supply. For most plans, this variable will capture the cost of brand name drugs or the highest tier of drugs on the formulary.

Min cost 30: Minimum across categories of drugs of the out of pocket cost to patients for a 30 day supply. For most plans, this variable will capture the cost of generic drugs or the lowest tier of drugs on the formulary.

Min percentage, Max percentage, Mean percentage: Similar to "Max cost 30" and "Min cost 30" but for plans where the cost to the patient of the drug is expressed as a percentage of the total cost instead of as an absolute amount.

Network Variable

Netsize: The number of providers included in the plan's network, divided by the number of Medicare enrollees in the county. This variable is both plan and county specific because the plan has a different network in each county, and each county has a different number of Medicare enrollees. The variable is reported as a range in the original data, in increments of 500 or 1000. I took the midpoint of each range. In addition, the data is censored for values greater than 9001, which are reported as 9001. For fee for service plans, which have no network, this variable has the value zero (even though having no network is similar to having a very large network). There is no data on network size for the years 2002 and 2003, so zeros are reported there, too. The indicator variables for years 2002 and 2003 and the fee for service indicator should absorb the average effect of omitting the network size in these cases.

Plan Type indicators

Managed care: Indicator for any type of managed care plan: Health Maintenance Organization, Preferred Provider Organization, or Provider Sponsored Organization.

Fee-for-service: Indicator for fee-for-service plans.

Demo: Indicator for Demonstration plans. These are experimental plans designed to test out new forms of coverage or new benefits. These plans can be either managed care or fee for service. Since characteristics are generally missing for these types of plans, this indicator does more heavy lifting than the others.

Note: any plan that doesn't fit into one of these three categories was removed from the data and its market share was added to the outside good. To preserve linear independence, the managed care indicator is omitted in the estimation.

Year indicators

Indicators for the years 2002, 2003, 2004 and 2005. To preserve linear independence, 2005 is omitted in the estimation.

References

[1] Berry, S. (1994). "Estimating Demand in Discrete Choice Models of Product Differentiation." *Rand Journal of Economics*, 25(2), pp. 242-262.

[2] Berry, S., J. Levinsohn, and A. Pakes. (1995). "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4),pp. 841-890.

[3] Brand, Keith. (2005). "A Structural Model of Health Plan Choice and Health Care Demand in the Medicare Managed Care Program." Working paper, University of Virginia.

[4] Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston.(2011) "How Does Risk-Selection Respond to Risk-Adjustment? Evidence from the Medicare Advantage Program". NBER working paper no. 16977.

[5] Center on Budget and Policy Priorities. "Policy Basics: Where Do Our Federal Tax Dollars Go?" Online at http://www.cbpp.org/cms/index.cfm?fa=view&id=1258, retrieved September 2011.

[6] Choi, James, David Laibson, Brigitte Madrian. (2011) "$100 Bills on the Sidewalk: Suboptimal Investment in 401(k) Plans." *Review of Economics and Statistics*, 93(3):748-763.

[7] Emanuel, Ezekiel J. and Dubler, Nancy N. (1995) "Preserving the Physician-Patient Relationship in the Era of Managed Care." *Journal of the American Medical Association*, 273(4), pp.323-329.

[8] Ericson, Keith M. Marzilli. (2010) "Market Design when Firms Interact with Inertial Consumers: Evidence from Medicare Part D." Working Paper, Harvard University.

[9] Farrell, Joseph and Paul Klemperer. (2007) "Coordination and Lock-in: Competition with Switching Costs and Network Effects." *The Handbook of Industrial Organization, Volume 3*. Ed. M Armstrong and R. Porter. pp.1970-2056.

[10] Gazmararian, Julie, David Butler, Mark Williams, Ruth Parker, Tracy Scott, S. Nicole Fehrenbach, Junling Ren and Jeffrey Kaplan.(1999) "Health Literacy Among Medicare Enrollees in a Managed Care Organization." *Journal of the American Medical Association*, 281(6), pp. 545-551.

[11] Gill, James M. Arch G. Mainous and Musa Nsereko (2000). "The Effects of Continuity of Care on Emergency Department Use." *Archives of Family Medicine*, 9, pp. 333-338.

[12] Gowrisankaran, Gautam and Mark Rysman. (2011). "Dynamics of Consumer Demand for New Durable Goods." Working paper, University of Arizona.

[13] Hall, Anne E.(2007) "The Value of Medicare Managed Care Plans and Their Prescription Drug Benefits." *Finance and Economics Discussion Series*, Federal Reserve Board of Governers.

[14] Handel, Benjamin (2010). "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts." Working Paper, Northwestern University.

[15] Hausman, Jerry, Gregory Leonard and J. Douglas Zona.(1994) "Competitive Analysis with Differentiated Products." *Annales D'Economie Et De Statistique* ,34, pp.159-180.

[16] Klemperer, Paul. (1995) "Competition when Consumers have Switching Cost: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade." *The Review of Economic Studies*, 62(4), pp. 515-539.

[17] Lustig, Josh. (2008). "The Welfare Effect of Adverse Selection in Privatized Medicare." Working paper, Boston University.

[18] Maruyama, Shiko. (2006). "Welfare Analysis Incorporating a Structural Entry-Exit Model: A Case Study of Medicare HMOs." Discussion Paper, Hitotsubashi University Institute of Economic Research.

[19] Miller, Mark E. (2007). "The Medicare Advantage Program and MedPAC Recommendation." Testimony before the Committee on the Budget, U.S. House of Representatives. Online at http://www.medpac.gov/documents/062807_Housebudget_MedPAC_testimony_MA.pdf, accessed September 2011.

[20] Nevo, Aviv. (2000) "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Economics & Management Strategy*, 9(4), 513-548.

[21] Shcherbakov, Oleksandr. (2008). "Measuring Consumer Switching Costs in the Television Industry." Working paper, University of Arizona.

[22] Small, Kenneth A & Rosen, Harvey S,.(1981) "Applied Welfare Economics with Discrete Choice Models," *Econometrica,* 49(1), pp. 105-130.

[23] Strombom, Bruce, Thomas Buchmueller, and Paul Feldstein. (2002) "Switching Costs, Price Sensitivity, and Health Plan Choice." *Journal of Health Economics* 21, 89-116.

[24] Town, Robert and Su Liu. (2003). "The Welfare Impact of Medicare HMOs." *The RAND Journal of Economics*, 34(4), 719-736.