CEPR Centre for Economic Policy Research [redacted] The Pinhas Sapir Center for Development Tel Aviv University [redacted] TEL AVIV UNIVERSITY The Gordon Faculty of Social Sciences Department of Public Policy

EIGHTH CEPR/JIE SCHOOL ON APPLIED INDUSTRIAL ORGANIZATION

Hosted by

The Eitan Berglas School of Economics, Recanati School of Business, Department of Public Policy at Tel Aviv University

Supported by

The Pinhas Sapir Center for Development, Tel Aviv University
Recanati School of Business, Tel Aviv University
Department of Public Policy, Tel Aviv University
Journal of Industrial Economics (JIE)
CEPR

Tel Aviv; 24 May 2011

Estimating Switching Costs for Medicare Advantage Plans

Kathleen Nosal (University of Arizona)

The views expressed in this paper are those of the author(s) and not those of the funding organisation(s), which take no institutional policy positions.

Estimating Switching Costs for Medicare Advantage Plans*

Kathleen Nosal

This draft: May 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 traditional fee-for-service Medicare coverage. I propose a structural dynamic discrete choice model of how persistently heterogeneous consumers make this choice based on the characteristics of the MA plans and traditional Medicare. The model explicitly incorporates switching costs, which are an important determinant of consumer choice in this context and the source of the dynamics. The model generalizes Shcherbakov's (2007) model of dynamic discrete choice with switching costs by allowing the choice set to be different across markets and time. I describe how to estimate the parameters of the model, including the switching costs, using the three-level nested fixed point estimation routine of Gowrisankaran and Rysman (2007). Preliminary estimates are reported for a simplified version of the model. The estimated switching cost is large compared to estimates of other parameters in the model, indicating that switching costs are, in fact, exerting a substantial influence on consumer choice in this market.

*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. Josh Lustig kindly shared data for this paper. Mario Samano provided some useful programming tips and reading recommendations. All mistakes are mine.

1 Introduction

Medicare beneficiaries have the option to choose from a menu of privately administered Medicare Advantage plans available as an alternative to traditional Medicare coverage. These plans, contracted by the Center for Medicare and Medicaid Services, can vary greatly in benefits provided and premiums charged. In addition, the set of plans available to beneficiaries is different across counties and changes every year. I estimate a model of discrete choice demand for these plans. Unlike previous models of demand in this market, mine is dynamic and allows for switching costs when beneficiaries switch among plans. This paper extends the previous literature on Medicare Advantage plan choice in three ways. First, it incorporates a switching cost into a discrete choice model of plan choice. Consumer switching costs are an important feature of this market that have not been addressed in previous literature. Second, it introduces dynamics and allow consumers to be forward-looking about their choice of plan. Third, it uses a richer data set on plan characteristics.

Estimating a demand model of this market that incorporates switching costs helps address several important policy issues. The goal of many papers in this literature is to quantify the welfare impacts of offering Medicare Advantage plans in addition to original Medicare. The total payments to these plans beyond what it would cost to provide original Medicare benefits is estimated to be in the billions of dollars per year (Biles et al., 2009). Of course, this spending might be justified if the value of the extra benefits provided under the plans exceed the extra costs. Clearly, some way of determining how Medicare eligibles value these benefits is necessary in order to make this type of welfare comparison. As I will argue in more detail in section 3, these calculations will be inaccurate if switching costs are simply ignored as they are in most demand models.

Quantifying the switching costs in this market is also interesting in its own right. Many of the plans offered under the Medicare Advantage program are forms of managed care, where enrollees are restricted to providers within a network. A major source of the switching costs may be that

enrollees are forced to change providers when they switch to a new plan with a different network. One interpretation of switching costs in this market is that they reflect a cost of these restrictive networks. Thus, the magnitude of these switching costs has implications in the debate over whether a managed care model or a fee for service model is a better way to provide health care.

A big open question about Medicare Advantage is what it tells us about how competition affects the provision of health care. Does competing against other firms encourage Medicare Advantage Organizations to strive to provide more benefits at lower cost, or are firms simply making a grab for the ample payments without much incentive to provide coverage in the most efficient way? This paper cannot answer this question directly, since it models only the demand side of the market. However, a good model of the demand side is a first step towards being able to model both sides of the market and beginning to tackle this larger question. A full supply and demand model would also allow for predicting both firm and consumer reactions to changes in policy.

2 Related Literature

Several authors have estimated discrete choice models of demand for Medicare Advantage plans. Town and Liu (2003) propose a static nested logit model, and estimate consumer surplus and firm profits resulting from private plan choice in Medicare. The set of plan characteristics that they use is limited, and their data is only through the year 2000, after which several interesting policy changes have occurred. Brand (2005) estimates a model that allows for adverse selection and incorporates individual consumer characteristics which influence plan choice. He presents a more nuanced welfare analysis, in which different consumer types may be affected differently. Maruyama (2006) combines a discrete choice demand model with an entry game on the firm side. He conducts counterfactuals which measure the welfare effect on consumers of subsidizing entry for firms. Other papers along similar lines include Lustig(2008) and Hall(2007). Clearly, there is no shortage of

research investigating demand for Medicare Advantage plans. The innovation of this paper is that it incorporates switching costs and dynamics.

In terms of the model and methodologically, this paper is closely related to papers by Gowrisankaran and Rysman (2007) and Shcherbakov (2007). Gowrisankaran and Rysman propose a model of dynamic consumer demand for durable goods, and develop an estimator and a three level nested fixed point estimation routine that allows them to estimate the model for the digital camcorder industry. Shcherbakov shows that the model can be adapted to incorporate switching costs instead of durability. His insight is that switching costs create dependence between current and future choices in the same way that durability does. He uses a version of Gowrisankaran and Rysman's estimation routine to estimate switching costs in the cable television industry. Shcherbakov's model and estimation are simplified by the fact that consumers chose among the same three products in every period. Because there are so few products, he can index the value function by the individual products without making the model intractable. I generalize Shcherbakov's model to allow for an arbitrary number of choices and a changing choice set, which entails borrowing some techniques for reducing the size of the state space from Gowrisankaran and Rysman.

There is a rich theoretical literature on switching costs, which is surveyed in Klemperer(1995) and Farrell and Klemperer (2007). In addition, the topic of switching costs in the context of health plans has been addressed in other empirical works. 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 cost and adverse selection. A difference between Handel's approach and mine is that he uses detailed consumer-level data, while my model can be estimated using only aggregate market share data. In addition, these other papers focus on the working-age population, while mine focuses on the elderly, who might have higher switching costs.

3 The Importance of Switching Costs and Dynamics

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 Advantage plan different than the plan chosen in the previous period, which can be either another Medicare Advantage plan or original Medicare.

There are numerous reasons to believe that switching Medicare Advantage plans imposes a large switching cost. Since most Medicare Advantage plans restrict patients to physicians in a network, switching plans may entail switching providers in many cases. 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 of 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 patients 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, because even if one treatment isn't objectively a better option for the patient than the other, there can be direct disutility from the change itself.

Another type of switching cost is a learning cost that enrollees incur when they have 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 suggests that about a third of enrollees in Medicare managed care plans have limitations to their literacy skills or basic health knowledge (Gazmararian, et al., 2006), 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 the default, and with updating insurance information with one's providers.

Switching costs induce state dependence that must be 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, the consumer has an incentive to be forward looking, because the plan chosen in the current period affects switching costs in the future. A consumer who is cognizant of 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 can 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. Suppose instead that we want to do a counterfactual analysis involving the addition of a new plan with more extensive coverage. Overestimation of the welfare gain might result, because we would conclude that more consumers would switch to the new plan than actually would. Clearly, ignoring switching costs can bias welfare analysis.

Switching costs also have implications on the firm side. High switching costs alter the entry and exit 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. For these reasons, it is important to incorporate information about switching costs into the demand model if it is to be usefully combined with a supply model.

4 Institutional Background

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. Parts A and B comprise original Medicare coverage. Part A is hospital insurance and Part B is medical insurance. Part C consists of Medicare Advantage plans, which Medicare beneficiaries have the option of choosing in lieu of part A and B coverage. Medicare Advantage 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 Medicare Advantage plans also cover additional services not covered by traditional Medicare. Part D is optional prescription drug coverage for part A and B enrollees. Medicare Advantage enrollees obtain drug coverage through their Medicare Advantage plan, if available, rather than through Part D.

There are several types of Medicare Advantage 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. As of 2007, 73% of Medicare Advantage enrollees were in plans of one of these two types (Gold, 2008).

The monthly premium for a Medicare Advantage plan may be different from the regular premium for part B coverage. Thus, Medicare Advantage enrollees may pay an additional monthly premium, may receive a refund of some portion of their part B premium, or may pay no premium beyond the standard part B premium. During the five year period studied in this paper, the rules about premiums changed, with the negative premiums allowed only in the last two years.

Medicare beneficiaries may change their coverage choice during an open enrollment period each year. The options to choose from are traditional Medicare or any Medicare Advantage plan offered in the beneficiary's county. Coverage changes take effect the following year. The default is to continue in the same plan as the previous year, 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.

The private firms who offer Medicare Advantage plans are known as Medicare Advantage Organizations. Medicare Advantage Organizations enter into contracts with the Center for Medicare and Medicaid Services through a non-competitive bidding process. Contracts are approved annually and can cover one or more counties. When a Medicare Advantage Organization has a contract to offer a plan in a particular county, it agrees to offer the same coverage terms to any Medicare eligible person residing in that county who opts into the plan. The plans available vary by county and year. Modeling competition between the firms providing the plans is a topic for future research, but for now I assume that the selection of plans available in each county in each year is exogenous.

As of 2008, 19% of Medicare eligibles were enrolled in a Medicare Advantage plan, and a majority of Medicare eligibles lived in counties where at least three Medicare Advantage plans were offered (Gold 2008).

Under the Patient Protection and Affordable Care Act, passed by the House in 2009 and the Senate in 2010, 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 changes will "reduce excessive payments to private insurance companies in Medicare Advantage while protecting...guaranteed Medicare benefits."[1] Presumably the lower payments will

[1] 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

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

5 The Model

The goal of this section is to model 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. The model must be parsimonious enough for estimation to be tractable, which requires some 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.

Each consumer, indexed by , lives in a county, indexed by . The consumer's county never changes, and the overall set of consumers is the same every year. The consumer's choice set depends on year, indexed by , as well as county. The choice set for consumers in county in year is denoted $_{mt}$, with elements . The element $= 0$ is the outside good, original Medicare. All other $\in {}_{mt}$ are Medicare Advantage plans.

Every year, the consumer must pick a plan from the choice set $_{mt}$. Consumer 's chosen plan in year is denoted $_{it}$. At the time of the decision, consumers know the plans available, premiums, observable and unobservable (to the econometrician) characteristics of the plans, and their own previous period choices. A consumer's choice in year may be the same as his choice in the previous year, provided his year -1 plan is in the choice set for year , or it may be different.

of all plans' 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.

If the chosen plan is different, the consumer incurs a switching cost. The plan choice can be divided into two parts. First, the consumer decides whether to retain his current plan or switch. Second, if the consumer has decided to switch, the consumer chooses a new plan.

Consumer $'$ state at the beginning of period (just before choosing $_{it}$) is:

$$(\ _{i,t-1} \quad _{mt} \quad _{mt}\ T_{mt} \quad _{mt} \quad _{it}\ \Omega_{mt})$$

where $_{i,t-1}$ is the consumer's plan choice in the previous year, $_{mt}$ is the matrix of observed characteristics of the plans in the consumer's choice set and T_{mt} is a vector summarizing the unobserved characteristics of these plans. The characteristics in $_{mt}$ and T_{mt} are assumed to be exogenous. The vector of the premiums of the available plans, $_{mt}$, is assumed to be endogenous and potentially correlated with T_{mt}. $_{it}$ 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 will be discussed in more detail later.

A consumer's utility from choosing a particular plan consists of three components. The first is a one-period flow utility. The flow utility, representing the net benefit from the plan that the consumer obtains during the year, is a function of plan characteristics, the plan premium, consumer preferences, and a random shock.. The second component of utility is a switching cost. The switching cost represents any costs incurred due to choosing a plan in period that is different from the plan chosen in -1. The third component is a continuation value, the expected infinite horizon utility for the consumer for period $+1$ onward given the value of the state variables in period t and the consumer's choice. The one-period flow utility and the switching cost are defined in the equations that follow, along with some useful functions of them. The continuation value is defined recursively by the Bellman equation.

The one-period flow utility to consumer in period for plan in county is:

$$
{}_{imjt} = \begin{cases} \varepsilon\,{}^{i}_{0} + \varepsilon\,{}'_{1}\ {}_{jt} + \varepsilon\,{}^{i}_{2}\ {}_{jt} + T_{mjt} + j_{ijt} \text{ if } \ \neq 0 \\ j_{i0t} \text{ if } \ = 0 \end{cases}
$$

where ${}_{jt}$ and ${}_{jt}$ are the observed characteristics vector and observed premium for plan , respectively. The unobserved characteristic, T_{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 in that county. The error term, j_{ijt}

of as a discrete consumer type and in others it will be convenient to think of it as a point in the continuum of types.

The random coefficients allow consumers to have heterogeneous preferences that persist from year to year over the characteristics. For example, the random coefficient on premium means that a consumer who chooses a low premium plan in one year will be likely to choose another low premium plan in the next year.

Establishing more notation that will be used later on, the mean (across values of the i.i.d. error term) utility of a consumer of type for plan at time is:

$$\overline{\quad}_{imjt} \equiv \begin{cases} \varepsilon_0^i + \varepsilon_1' \;_{jt} + \varepsilon_2^i \;_{jt} + T_{mjt} \text{ if } \neq 0 \\ 0 \text{ if } = 0 \end{cases}$$

and the mean flow utility across consumers in county is:

$$\overline{\quad}_{mjt} \equiv \begin{cases} \overline{\varepsilon}_0 + \varepsilon_1' \;_{jt} + \overline{\varepsilon}_2 \;_{jt} + T_{mjt} \text{ if } \neq 0 \\ 0 \text{ if } = 0 \end{cases}$$

The next component of utility is the switching cost. The consumer incurs the switching cost, F, by choosing a plan in year different from the plan chosen in -1 The effective switching cost from choosing plan $_{it}$ when one's previous choice was $_{i,t-1}$ can be represented by the following function:

$$F(\;_{i,t-1} \;_{it}) \equiv \begin{cases} F \text{ if } \;_{i,t-1} \neq \;_{it} \\ 0 \text{ if } \;_{i,t-1} = \;_{it} \end{cases}$$

Let Ω_{mt} denote a matrix of the current plan characteristics, premiums, and error terms for county and anything else that might affect consumer expectations about future choice set, plan characteristics and premiums. Assume that Ω_{mt} evolves by some Markov process, $(\Omega_{mt} \mid \Omega_{m,t-1})$.

Since plans can exit, a consumer's chosen plan from period -1 may not be in the period choice set. If this happens, the consumer is forced to choose a different plan. This possibility can

be captured with an indicator function:

$$(\ _{i,t-1}) = \begin{cases} 1 \text{ if } \ _{i,t-1} \in \\ 0 \text{ if } \ _{i,t-1} \in \end{cases}$$

Assuming an infinite horizon and annual discount factor . , we can now write the Bellman equation. First, consider conditioning on the case where the consumer's incumbent plan is still available in period t (that is, conditioning on $(\ _{i,t-1}) = 1$):

$$\begin{aligned} (\ _{i,t-1}\ \Omega_{mt} |\ (\ _{i,t-1}) &= 1) = \quad \{\ _{\mathsf{imj_{it-1}}t} + .\ \ [\ (\ _{i,t-1}\ \Omega_{mt,+1}|\Omega_{mt})] \\ &\ _{\mathsf{j \in J_{mt}; j \neq j_{it-1}}} \{-F(\ _{it-1}\ ) + \ _{imjt} + .\ \ [\ (\ \ \Omega_{m,t+1}|\Omega_{mt})]\}\} \end{aligned}$$

where the expectation is over future error draws and the future evolution of Ω. The notation $\ _{\mathsf{imj_{it-1}}t}$ means the period flow utility for the plan chosen in period -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. Notice that one of the arguments of the value function is the consumer's incumbent plan, $\ _{it-1}$. The incumbent plan is a crucial state variable because it affects the consumer's utility for any plan choice through the switching cost. This link between the previous plan choice and the the utility of the next plan choice is what makes the problem dynamic.

Now, consider instead conditioning on the consumer's incumbent plan *not* being available in period t (that is, condition on $(\ _{i,t-1}) = 0$). In this case, the Bellman consists of only the inner maximization because the consumer does not have a choice between switching and not switching. The consumer must choose the best of the currently available plans even if she would prefer to stay

in her (now defunct) incumbent plan:

$$(\ {}_{i,t-1} \ \Omega_{mt}| \ (\ {}_{i,t-1}) = 0) = {}_{j \in Jmt}\{-F(\ {}_{it-1} \quad) + \ {}_{imjt} + . \quad [\ (\quad \Omega_{m,t+1}|\Omega_{mt})]\}$$

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. Clearly, it is necessary to further specify what information goes into consumer expectations and how these expectations are formed. Towards this end, the *logit inclusive value* is defined as:

$$/^{imt}(\ {}_{i,t-1} \ \Omega_{mt}) \equiv \quad (\sum_{j \in J_{\mathrm{mt}}, j \neq j_{\mathrm{it-1}}} \quad (/_j^{imt}(\ {}_{i,t-1} \ \Omega_{mt}))) \tag{1}$$

where:

$$/_j^{imt}(\ {}_{i,t-1} \ \Omega_{mt}) \equiv -F(\ {}_{it-1} \quad) + \overline{\ {}_{imjt}} + . \quad [\ (\quad \Omega_{m,t+1})|\Omega_{mt}] \tag{2}$$

The logit inclusive value is the expected value of the consumer's *best* plan choice among all available plans except the consumer's incumbent plan. The expectation is over the extreme value error term ∂_{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 option value of switching, or as a summary statistic about the quality of the 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 know 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 evolution of the logit inclusive value, $/^{imt}$. In addition to capturing a way that consumers might summarize the vast amount of information about the different plans, basing consumer expectations on $/^{imt}$ has the advantage of reducing the dimensionality of the state

space. The following assumptions formalize these ideas.

Assumption: Sufficiency of a reduced set of state variables:

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

The assumption states that consumers can predict the evolution of $/^{imt}$ based only on its previous value, ignoring the other information contained in Ω. More specifically, assume that consumers rationally expect it to follow the following autoregressive process:

$$/^{im,t+1}(\ ) = ,_0 + ,_1/^{imt}(\ ) + \ _{at} \qquad \neq 0 \tag{3}$$

$$/^{im,t+1}(0) = ,^0_0 + ,^0_1/_t(0) + \ _{0t} \tag{4}$$

where $\ _{at}$ and $\ _{0t}$ are each identically and independently distributed normal error term with mean zero, and the gammas are parameters to estimate. Since original Medicare is much less subject to change than Medicare Advantage plans, the process is allowed to be different in the case where the incumbent plan is a Medicare Advantage plan (and original Medicare is therefore included in $/^{imt}$) than it is in the case where the incumbent plan is original Medicare (and original Medicare is therefore *not* included in $/^{imt}$).

In addition to expectations about other plans in the market, captured through the evolution of $/^{imt}$, consumers must have some expectation about their incumbent plan in future periods. Since observed and unobserved characteristics of a particular plan can change from period to period, consumer expectations about the evolution of $\overline{\ _{imjt}}$ could be modeled in a way similar to $/^{imt}$ However, since individual plan quality changes less than the quality of the overall set of plans, which changes largely due to entry and exit of plans, it is also reasonable to think that on average consumers do not expect a specific plan's quality to change over time, conditional on the plan not exiting the market. The next step is modeling the consumer's beliefs about the possibility of his own incumbent plan exiting the market.

Let $_{mjt}$ be an indicator for plan exiting market at time . Then, the average plan exit rate for time periods one through across all markets is:

$$\equiv \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{j \in J_{\mathrm{mt}-1},\ j \neq 0} \frac{1}{|\ _{m,t-1}|}\ _{mjt}$$

An empirical plan drop-out rate for the years covered in the data can easily be calculated using this formula. Assume that consumers expect that the probability of any particular plan exiting in the next time period is simply this aggregate drop-out rate, unless the plan is original Medicare, which has probability 0 of exiting. Then, consumer expectations about plan exit can be expressed as follows:

$$\text{for } _{it-1} \neq 0 \quad (\ _{it-1}) = \begin{cases} 1 \text{ with probability } 1 - \\ 0 \text{ with probability } \end{cases}$$
$$(0) = 1 \text{ with probability } 1$$

Furthermore, consumers must know that they cannot choose a plan anymore once it has exited. One way to represent this is to say that a plan that has exited has utility $-\infty$ Based on the assumption made earlier, consumers expect that if their plan has not dropped out, it will simply have the same flow utility that it had in the previous period.

$$(\overline{\ _{imj_{\mathrm{it}},t+1}} | \overline{\ _{imjt}} \quad (\ _{i,t-1})) = \begin{cases} \overline{\ _{imjt}} \text{ if } (\ _{i,t-1}) = 1 \\ -\infty \text{ if } (\ _{i,t-1}) = 0 \end{cases}$$

Making all of the above assumptions, the relevant state space is reduced to:

$$(\overline{\ _{imj_{\mathrm{it}-1}t}}\ /^{imt}(\ _{i,t-1}) \quad (\ _{i,t-1}) \quad (\ _{i,t-1}) \quad _{it})$$

where $(\ _{i,t-1})$ is an indicator that is equal to zero if the argument is zero (the previous choice was an MA plan) and equal to one otherwise (the previous choice was original Medicare). With this reduced state space, the state consists of the new mean flow utility of the incumbent plan,

the expected utility of choosing a non-incumbent plan, the utility of each non-incumbent plan, indicators for whether the incumbent plan is still available and whether it is the outside good, and the vector of error terms.

The expected Bellman can be written as a function of the variables contained in the reduced state space:

$$(\overline{\quad}_{imj_{\mathrm{i;t-1}}t} \ /^{imt}(\ _{i,t-1}) \ (\ _{i,t-1})| \ (\ _{i,t-1})) =$$

$$\begin{cases} (\quad (\overline{\quad}_{imj_{\mathrm{i;t-1}}t} + \cdot \quad [\ (\overline{\quad}_{imj_{\mathrm{i;t-1}}t+1} \ /^{imt+1}(\ _{i,t-1}) \ (\ _{i,t-1}))| \overline{\quad}_{imj_{\mathrm{i;t-1}}t} \ /^{imt}(\ _{i,t-1})]) \\ \qquad + \quad (/^{imt}(\ _{i,t-1}))) \text{ if } (\ _{i,t-1}) = 1 \\ \qquad /^{imt}(\ _{i,t-1}) \text{ if } (\ _{i,t-1}) = 0 \end{cases}$$

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

$$(\overline{\quad}_{imj_{\mathrm{i;t-1}}t} \ /^{imt}(\ _{i,t-1}) \ (\ _{i,t-1})) =$$

$$(1- \ ) * \ (\quad (\overline{\quad}_{imj_{\mathrm{i;t-1}}t} + \cdot \quad [\ (\overline{\quad}_{imj_{\mathrm{i;t-1}}t+1} \ /^{imt+1}(\ _{i,t-1}) \ (\ _{i,t-1}))| \overline{\quad}_{imj_{\mathrm{i;t-1}}t} \ /^{imt}(\ _{i,t-1})])$$
$$+ \quad (/^{imt}(\ _{i,t-1}))) + \ * /^{imt}(\ _{i,t-1})$$

With the preceding setup established, expressions can be derived for transition probabilities of the market shares. The specific functional forms of these probabilities are a consequence of the extreme value error term.

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

$$\Pr_{switch}(\ ') \equiv \quad (\ _{it} \neq \ '| \ _{it-1} = \ ') =$$

$$\frac{(/^{imt}(\ '))}{(/^{imt}(\ ')) + \quad (/^{imt}_{j'}(\ '))} \tag{5}$$

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

$$\mathrm{Pr}_{noswitch}(\ ') \equiv \quad (\ _{it} = \ '|\ _{it-1} = \ ') =$$

$$\frac{(/_{j'}^{imt}(\ '))}{(/^{imt}(\ ')) + \quad (/_{j'}^{imt}(\ '))} \tag{6}$$

Conditional on switching and on $'$ being the plan chosen in the previous period, the probability of choosing plan is:

$$\mathrm{Pr}_{j|switch}(\ ') \equiv \quad (\ _{it} = \ |\ _{it-1} = \ ' \quad _{it} \neq \ ') =$$

$$\frac{(/_{j}^{imt}(\ '))}{(/^{imt}(\ '))} \tag{7}$$

Finally, the total probability of choosing plan in period having chosen plan $'$ in period -1 is:

$$\mathrm{Pr}_j(\ ') \equiv \mathrm{Pr}(\ _{it} = \ |\ _{it-1} = \ ') = 1_{\{j=j'\}}\,\mathrm{Pr}_{noswitch}(\ ') + 1_{\{j\neq j'\}}\,\mathrm{Pr}_{j|switch}(\ ')\,\mathrm{Pr}_{switch}(\ ')$$

The transition probabilities can be used to express the expected market share in the current period conditional on the previous period's market share. Let $_{imj,t-1}$ be the period -1 market share for plan in county for consumer type Then, the expected county market share of plan in year for consumers of type can be expressed:

$$\widehat{\ _{imjt}} = \sum_{j' \in J_{\mathrm{m;t-1}}} \ _{mij',t-1}\,\mathrm{Pr}_j(\ ')$$

Integration over types yields the predicted county level market share for plan :

$$\widehat{\ _{mjt}} = \int \widehat{\ _{imjt}} \quad _i \tag{8}$$

where $_i$ is the distribution of consumer types, which is defined by the distributions of the random coefficients described earlier in this section. The model therefore makes predictions about county-plan-year level market shares based on the previous year's market shares.

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 $/_j^{imt}$ 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 may be possible to exploit the fact that new consumers have a switching cost for every plan to help identify the switching cost, as in Handel (2010). Also, the scope of consumer heterogeneity is 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, which may be a more realistic type of heterogeneity. Finally, 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. Future incarnations of the model may contain these enhancements.

6 Data

Data comes from the Center for Medicare and Medicaid Services.[2] Three types of data are used: data on market shares, data on characteristics of the plans, and data on county-specific 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.

[2] 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.

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 it is not clear that parameters of a plan-level model are identified with only contract-level data.

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 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 trying other selection rules (for example, selecting the highest numbered plan or a randomly drawn plan) would be a possible robustness check.

Throughout the other sections 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 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. Every 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. For example, in the "Vision Services" field, text comments include:

- You are covered up to a $95 allowance for eye wear every year.
- You are covered up to $150 for eye exams and eye wear every two years.
- You pay 75% of the cost for each routine eye exam.
- You pay $0 for lenses, limited to 1 pair every year.
- You pay $25 to $60 for each routine eye exam.
- You pay 100% for routine eye exams and glasses.

This is just a small sample of the distinct comments that can appear in this 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 fixed copays, 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 use in the estimation. 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.

The complete list of characteristics variables that I constructed and their definitions is included in Appendix A. Since many ad hoc decisions went into constructing the variables, it is important to be clear about exactly how they were are defined.

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 their coverage.) 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 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[3]. 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 eligibles. I also drop observations for any county that does not have at least one MA contract in every year in the sample because the estimation procedure cannot handle counties that have zero MA contracts in some year. 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 873 counties (out of about 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

[3] 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 disastorous when it came time to take the log of the shares! Zeros in the data are also troublesome because the model can never predict a market share of exactly zero, though it can predict an arbitrarily small market share.

or did not have a Medicare Advantage 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 a large majority of people with access to Medicare Advantage plans in this time period. Therefore, this set of counties might be the most relevant from a policy perspective.

Summary statistics about the data appear in Appendix B. Some surprising features of the data show up in these summary statistics. Looking at the mean characteristics data over time in the second table, it does not appear that there is a time trend of increasing quality. In fact, some characteristics move monotonically in the direction of less coverage over the five years, while others stay about the same. The standard deviation on most of the characteristics is quite large, showing that plans do differ from one another quite substantially The first table shows that there is a substantial amount of entry and exit in the five year time period. Clearly, the choice set that consumers face changes from year to year.

7 Instruments

An instrumental variables approach is necessary because the plan premiums are endogenous. The error term T_{mjt} represents plan-county level unobserved characteristics and 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 premiums but not with the structural errors.

To find appropriate instruments, it is necessary to understand how premiums are set. When a firm offers a MA plan in multiple counties, it chooses one premium and set of benefits that applies to the plan in all of the counties. When the firm chooses the premium, it therefore must think

about competing plan quality and 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 of plan 1 in county B). Consider the number of plans of various types that plan 1 competes against in county B, and the characteristics of these plans. These factors affect the premium that the firm optimally charges for plan 1, because they indicate how many and what kind of competitors plan 1 has in county B. However, there is no reason that they would be correlated with unobserved quality in county A, especially since the sets of plans in counties A and B may be disjoint other than plan 1. Therefore, counts and average characteristics of plans in other markets in which the plan competes can be used as instruments. In the tradition of Berry(1994), average characteristics in the county of interest also work as long as they are exogenous.

Constructing the instruments requires first defining the set of competitors for each plan. A plan's competitors are simply the set of all of the plans that are in at least one county with the plan. In general, this set will be distinct for each plan. The instruments are then computed across the plans in this set. The instruments are the counts of managed care, fee-for-service, and demonstration plan in the set, and the mean of each of the exogenous characteristics across the plans in the set.

A second type of instrument that can be used here is a cost or profit shifter. One such shifter is the capitation payment that MA organizations receive from the CMS for each enrollee. A base capitation rate for every county is set by the CMS each year based on county-specific factors that affect cost. The county benchmark capitation payment is then adjusted based on the risk

profile of the enrollees in the plan to determine the actual payment that the MA organizations will receive. While the county benchmark and the risk adjustments are meant to fully compensate for the cost differences associated with insuring different populations of Medicare eligibles, evidence suggests that MA organizations still find ways to distinguish and enroll Medicare eligibles who will be more profitable even given these adjustments (Brown et. al, 2011). Thus, the base capitation payment can be correlated with cost in two ways. First, a higher benchmark capitation payment might signal a higher cost county, and those higher costs may not be fully compensated for in the capitation payment itself. Second, a higher captitation payment directly means that the MA organization will be payed more for insuring people in that county. If the MA organization is able to find a way to insure a population in that county whose costs are less than those that the higher captitation payment was based on, such a county is potentially more profitable than one with a lower benchmark capitation payment. In either case, the capitation payment effects costs or profits, which should have an effect on the optimal price the plan charges. Furthermore, since the CMS bases the benchmark capitation rate on attributes of the counties and not of the plans themselves, there is no reason it should be correlated with the unobserved characteristic term. I average the benchmark capitation rate across counties the plan is offered in to get a single capitation rate for each plan.

Another possible cost related instrument would be a variable that measures the average health of the residents of a county. Obviously, healthier people are less expensive to insure. A problem with using a measure of health, though, is that a healthier population of MA enrollees can be a result of having better MA plans available in the county. This feedback effect creates a possible link between the unobserved plan quality and the instrument. An alternative is to use the county infant mortality rate. Infant mortality rates have been shown to be very highly correlated with broader measures of population health (Reidpath and Allotey, 2002). At the same time, there is no

way for the quality of MA plans to impact infant mortality, because both infants and their parents are much too young to qualify for Medicare. The infant mortality rate is thus a way to capture overall health in the county that is exogenous with respect to the quality of the plans.

Finally, Shcherbakov (2007) suggests an instrument that helps identify the switching costs. Interacting *lagged values* of the instrumental variables (such as the ones described above) with current values of the unobserved quality term creates a new instrument that helps to identify state dependence.

The current estimates use the first two sets of instruments described. The infant mortality rate and the Shcherbakov instruments may be constructed and used in later versions of the estimation.

8 The Estimator

The main parameters to estimate are the non-linear parameters F and X (the switching cost and the variances on the random coefficient) and the linear parameters ε (the mean coefficients on the characteristics.) There are also nuisance parameters, such as the , 's governing the process by which the logit inclusive values evolve.

The estimator is based on the GMM estimators in Gowrisankaran and Rysman (2007) and Berry, Levinsohn and Pakes (1995), who show that it is consistent. It can be defined as follows:

$$\min_{\alpha,\eta,\sigma} T(\varepsilon \quad F \; X)' \qquad {}' T(\varepsilon \quad F \; X)$$

$$\hat{}(F \; X) =$$

where $T(\varepsilon \quad F \; X)$ is the vector of structural errors at the given parameter values, is the matrix of instruments, is a weighting matrix, $\hat{}(F \; X)$ is the vector of predicted market shares when the dynamic programming problem is solved at the given parameters, and is the true vector of market shares.

9 The Estimation

The estimation method is a variation on the three level nested fixed point estimation routine developed by Gowrisankaran and Rysman (2007). 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.[4] Each of the 30 draws can then 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.

Inner Loop

The inner loop maps a vector of parameters, $(F\ X)$ and a vector of mean flow utilities, $\overline{\quad}_{mjt}$, to a vector of predicated market shares $\widehat{\quad}_{mjt}$ by solving the dynamic programming problem defined by the Bellman equation for each consumer type and plugging the resulting value functions 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 $/_j^{imt}(\ _{i,t-1})$ and $/^{imt}(\ _{i,t-1})$ that satisfy the recursive definitions of these two objects. Finally, it finds estimated autoregression coefficients, $\hat{;}$, that are stable from iteration to iteration.

To make the estimation feasible, the continuous state space must be discretized. The state spaces

[4] 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.

for the variables $/^{imt}(\ _{i,t-1})$ and $\overline{\quad_{imj_{i;t-1}t}}$ are each divided into 50 grid points. The minimum and maximum values for the grid are based on the initial range of the variables plus some added leeway. The value function $(\ _{imj_{i;t-1}t}\ /^{imt}(\ _{i,t-1}))$ is defined discretely on each of the 2500 points of the two-dimensional grid of the state space. 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 $/^{imt}(\ _{i,t-1})$ and $\widehat{;}$ vectors are needed for use in the first iteration. Mean flow utilities, $\overline{\quad_{mjt}}$, are passed in from the middle loop, as well as values for the parameters F and X The discount factor is set to 0.9 on the annual level.[5]

First, $/_j^{imt}(\ _{i,t-1})$ is calculated for the plans. The expectation of the value function is part of the expression for $/_j^{imt}(\ _{i,t-1})$ To find this expectation, a simulated integral must be taken over the error term in the $/^{imt}$ autoregressive process. Once $/_j^{imt}(\ _{i,t-1})$ has been calculated for each plan, the $/^{imt}(\ _{i,t-1})$ can be updated by taking logs of sums of exponentials of $/_j^{imt}(\ _{i,t-1})$ for all the plans in a county except $\ _{it-1}$ The $/^{imt}(\ _{i,t-1})$ are regressed on the $/^{im,t-1}(\ _{i,t-1})$ to obtain a new $\widehat{;}$

There are two options for the next step in the algorithm. Since $/_j^{imt}(\ _{i,t-1})$, $/^{imt}(\ _{i,t-1})$, and $\widehat{;}$ 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 $/_j^{imt}(\ _{i,t-1})$ than for the value function, I have found in practice that overall convergence of the inner loop tends to be faster if the $/_j^{imt}(\ _{i,t-1})$ 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

[5] 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 depends on $\widehat{,}$, which depends on the data through $/^{imt}(\ _{i,t-1})$. Once the value function has been updated, $/^{imt}(\ _{i,t-1})$, $\widehat{,}$ and the value function are all checked for convergence. If they have not all converged, another iteration begins, starting with recalculation of the $/_j^{imt}(\ _{i,t-1})$ based on the new values of the other variables.

After convergence has been achieved, transition probabilities are calculated using the newly calculated 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.

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, $\overline{\quad}_{mjt}$, are updated according to this mapping:

$$\overline{\quad}^{new}_{mjt} = \overline{\quad}^{old}_{mjt} + \Psi(\ln(\ _{mjt}) - \quad (\hat{\ }_{mjt}(\overline{\quad}^{old}_{mjt}\ F\ X))) \tag{9}$$

where Ψ is a tuning parameter, $\ _{mjt}$ is the county and plan level market share observed in the data, and $\hat{\ }_{mjt}$ is the corresponding estimated market share, which is a function of a mean flow utility and F and X, the candidate parameter values passed in from the outer loop. The mapping is iterated on until $\overline{\quad}^{new}_{mjt}$ and $\overline{\quad}^{old}_{mjt}$ match, up to some tolerance. Calculating $\hat{\ }_{mjt}(\overline{\quad}^{old}_{mjt}\ F\ X)$ requires invoking the inner loop, which solves the dynamic programming problem and calculates shares

based on the arguments of $\hat{}_{mjt}(\overline{\ {}^{old}_{mjt}}\ F\ X)$. 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 ε parameter vector 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 structural errors, T. Notice that T can be thought of as a function of the parameters F and X, because the values of T that come out of the middle loop will depend on the F and X fed into the inner loop.

Outer Loop

The outer loop is a Generalized Method of Moments procedure, minimizing a nonlinear criterion function over the parameter vector $(F\ X)$. The identifying assumption is that the instrument matrix is orthogonal to the structural error vector T.

Define the following function:

$$(F\ X) = \quad {}'T(F\ X)$$

Then, the minimization problem is:

$${}_{\eta,\sigma}\{\ \ (F\ X)' \qquad (F\ X)\}$$

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

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

10 Standard Errors

Because the estimation procedure is a form of the Generalized Method of Moments, the usual results about 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.

Let . be the parameter vector $(F\, X\, \varepsilon\,)$, and $\widehat{.}$ be the estimator $(\widehat{F}\, \widehat{X}\, \widehat{\varepsilon}\,)$ as defined in the previous section Then, the variance-covariance matrix for $\widehat{.}$ is:

$$_{GMM} = \frac{1}{\ }\left([\frac{}{.}({}'_{mjt} T_{mjt}(.\ ))]' \ [{}'_{mjt} T_{mjt}(.\ ) T'_{mjt}(.\ ) \ {}_{mjt}]^{-1} \ [\frac{}{.}({}'_{mjt} T_{mjt}(.\ ))]\right)^{-1}$$

where is the number of observations, which is the total number of plan-county-year combinations in the data. It can be estimated by its sample analog:

$$\widehat{\ }_{GMM} =$$

$$[(\sum_t \sum_m \sum_{j \in J_{mt}} [\frac{}{.}({}'_{mjt} T_{mjt}(\widehat{.}))])'(\sum_t \sum_m \sum_{j \in J_{mt}} [{}'_{mjt} T_{mjt}(\widehat{.}) T'_{mjt}(\widehat{.}) \ {}_{mjt}])^{-1} \sum_t \sum_m \sum_{j \in J_{mt}} [\frac{}{.}({}'_{mjt} T_{mjt}(\widehat{.}))]]^{-1}$$

with the necessary partial derivatives approximated numerically by perturbing the parameter vector.

11 Identification

Three sets of parameters are estimated: the mean coefficients on the characteristics variables, the standard deviations for the random coefficients, and the switching cost. Identification results in

Berry, Levinsohn and Pakes (1995) and Berry(1994) imply identification of the first two groups of parameters. The switching cost is trickier. Shcherbakov (2007) makes a somewhat formal argument that the switching cost is identified in this type of model, but he relies on assumptions that are specific to the cable television industry and the result cannot be directly applied here. I have made some progress in further formalizing Shcherbakov's argument and adapting it to the model and industry in this paper, and intend to have an identification result in future drafts. In absence of a formal identification result, 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 large switching costs, a plan that has a big market share in one year will tend to have a big 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 are other explanations for share persistence, such as unobserved aspects of consumer preferences that do not change over time. 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. The one case where this relationship is unambiguously monotonic is when a plan has newly entered the 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. 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 a Medicare Advantage plan as opposed to original Medicare, since the flow utility for original Medicare is normalized to zero. The random coefficient on the constant term allows for some consumers who consistently prefer original Medicare over a Medicare Advantage plan, and some who consistently prefer Medicare Advantage plans. An extreme case would be zero variance of the random coefficient. This is equivalent to the case where everyone has the same constant term, and random coefficients are unnecessary. Without random coefficients, independence of irrelevant alternatives would hold as in the standard multinomial logit model. Under independence of irrelevant alternatives, if a plan exited then the consumers who were previously in that plan would not be more likely to choose a Medicare Advantage plan in the next period than would someone starting out in original Medicare. If this assumption is violated, the variance on the random coefficient must be non-zero. The more strongly it is violated, the more dispersion of preferences is implied.

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, then the coefficient on drug coverage must be 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.

12 Simplified Estimation and Results

The results reported in appendix B come from a simplified estimation using a previous version of the model and data. A new estimation that matches what is described in this paper is in progress, but results are not yet available. In this section, I explain what was different about the old data set and the simplified estimation.

The data set used in the old estimation differ from the current data set in two main ways, one of which was an error and the other of which was simply a different aggregation decision. The error was in how the set of plans in each county and year was determined. The "State County Plan" data set available on the CMS website lists every combination of county and contract that has a positive market share, and the market share itself as long as greater than 10 people are enrolled. When constructing the old data set, I made the assumption that every combination of contract and county on the list represented a contract in that county's choice set. However, there are ways that a contract can have a positive market share in a county in which it is not offered. For example, enrollees might move but initially still be enrolled in plans from their former county. This assumption led to many erroneous matches of county and contract. While any contract not actually offered in a county is likely to have only a very small market share, getting the choice sets wrong has the potential to bias the estimation. The second difference was that the data were

not aggregated up to the contract level. If there were multiple plans within a contract, they were considered to be distinct plans.

The different data set affected the computational burden of the estimation. First, between having a multitude of extra county and contract combinations, and having each county and contract combination multiplied by the number of plans in the contract, the number of observations in the data was more than ten times larger than the current data. Second, a modification to the estimation procedure was necessary to accommodate the level of disaggregation. Since shares are available at only the county and contract level, the BLP inversion had to be done at the county and contract level and not at the county and plan level. While this limitation could be dealt with by assuming that the structural errors are the same across plans in a contract, it ruled out the possibility of estimating the ε coefficients with a simple instrumental variables regression between the middle and outer loops. Instead, the ε's had to be treated as additional non-linear parameters, which greatly increased computation time.

The immense computational burden meant that it was impractical to estimate the full model using that data set. I reduced the computational time by making several modifications. No random coefficients were used. In other words, consumers were assumed to all have the same value of the ε parameters. Since the number of times the dynamic programming problem must be solved is multiplied by the number of random coefficient draws, this modification reduced computational time linearly in the number of draws that would have been taken. In addition, only a small subset of the characteristics variables were used in the estimation: the constant term, the premium, the drug coverage indicator, and the average cost of an emergency room visit. An indicator for original Medicare was also included, but in error because the flow utility of original Medicare should have been normalized to zero. Finally, I did the estimation only for 54 counties in California. This decision reduced the number of observations, but left a fairly large number of plans and contracts

in the data since Medicare Advantage has a large presence in California.

Given the shortcomings of the way they were obtained, the results from the simplified estimation are not very informative. However, they do give some indications of how to proceed with the new estimation.

Perhaps the most alarming feature of the results is the positive coefficient on the premium. It is quite unrealistic to think that consumers prefer plans with higher premiums, all else equal. There are two possible explanations for this unexpected sign. The first is that the instrumental variables approach failed to control sufficiently for the correlation between the premium and the unobserved characteristics, and the coefficient is reflecting the fact that the plans with higher premiums have better coverage. The second explanation is that persistent consumer heterogeneity is particularly important with respect to the disutility of paying a higher premium (perhaps because it depends heavily on income, which varies substantially across consumers), and a random coefficient on the premium is necessary in order to properly account for this heterogeneity. Nevo (2000) points out that for most discrete choice demand models the aggregation assumptions necessary to justify non-random coefficients are unrealistic, and often very different results are obtained when random coefficients are added to the model. The latter explanation is the preferable one, because it suggests that the sign problem might be resolved with the inclusion of random coefficients.

The estimated switching cost has the expected sign. Also, the magnitude is quite large– it is three times the size of the estimated coefficient on the indicator for drug coverage. In other words, if a consumer were contemplating switching from a plan without drug coverage to a similar plan with drug coverage, the one-period added utility of the drug coverage would not be nearly enough to overcome the disutility of switching. The large magnitude of the estimate of the switching cost reinforces the idea that switching costs are important in this market, and justifies proceeding with this framework that emphasizes the role of switching costs..

Standard errors have not been calculated for these parameter estimates. For the new estimates, they will be calculated as described in section 9.

13 Conclusion

In its current version, this paper accomplishes three goals. First, it adapts Shcherbakov's switching cost model to a more complicated setting where the choice set varies across markets and time. Second, it describes how to use the estimation procedure of Gowrisankaran and Rysman to estimate the full model, and gives estimation results for a simplified version, demonstrating that the procedure is a feasible way to approach the problem. Third, it details the construction of a data set appropriate for use in the estimation from several disparate and messy sources.

While the preliminary estimation results are not very believable or useful on their own, they provide guidance for the next stage of this research. The relatively large magnitude of the estimated switching cost justifies continued focus on the switching cost as an important factor in consumers' choice of MA plans. The dubiously signed coefficient on the premium variable suggests that random coefficients are necessary in order to get meaningful estimates. Including more characteristics variables or finding better instrumental variables may also help.

With random coefficients and more characteristics variables included, the correct choice sets known for each county, and the convenient aggregation of plans to contracts, the new estimation should be computationally simpler and more empirically enlightening. Once the new estimates have been computed, the next step is to use the estimates for evaluations of the welfare impacts of the Medicare Advantage program and for counterfactuals. The real power of structural estimation is the ability to do this type of analysis.

From a policy perspective, there are many interesting questions to ask about Medicare Advantage. Is the availability of MA plans on net welfare enhancing? Does the presence of competition

between private firms furnishing Medicare coverage create incentives to cut costs? Under alternative policy regimes, would Medicare Advantage Organizations provide higher (or lower) quality coverage, or have lower (or higher) costs? These questions can be addressed by doing counterfactual analysis using a structural model of both the consumer and firm side of the market for MA plans. The present paper develops and preliminarily estimates the consumer side of such a model. Future stages of this research will entail improving upon the consumer side of the model and eventually adding the firm side.

14 Appendix A: Characteristics Variables

General Characteristics Variables

Premium: The amount that the enrollee must pay the Medicare Advantage Organization 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

Coverage Limit: 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.

No coverage: An indicator for the statement: "You pay 100% for most prescription drugs" or "You pay 100% for non-Medicare prescription drugs." These statements imply extremely limited drug coverage. Note: In the estimation, I use one minus this variable to make it easier to interpret.

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, 2002 is omitted in the estimation.

15 Appendix B: Tables

Table 1: Entry, Exit, and number of plans.

	2001	**2002**	**2003**	**2004**	**2005**
Unique plans	172	153	179	171	211
Total Plan/county combinations	1722	1733	2003	2193	3510
Entrants (plan/county combinations)	-	286	350	358	1391
Exiters (plan/county combinations)	-	275	80	168	74
Mean plans per county	1.97	1.99	2.30	2.51	4.03
Median plans per county	1	2	2	2	4
Max plans per county	11	11	13	13	19

Table 2: Summary Statistics for Plan Characteristics. The statistics are calculated across plans in each year, except for Netsize, which is calculated across plan/county combinations. Netsize and Drug Discount are not available in some years.

	All years		**Yearly Means**				
Variable	**Mean**	**St. Dev**	**2001**	**2002**	**2003**	**2004**	**2005**
Premium	29.24	46.76	0	0	51.17	55.77	33.88
Primary Care	9.99	5.91	9.18	10.19	11.07	10.9	9.2
Dental	0.38	0.48	0.31	0.25	0.40	0.38	0.49
Emergency	45.85	16.05	41.98	46.86	47.62	46.36	46.56
Hearing aid	0.52	0.5	0.87	0.70	0.55	0.42	0.27
Routine eye	0.36	0.48	0.58	0.50	0.39	0.29	0.19
Glasses	0.39	0.49	0.37	0.36	0.35	0.36	0.45
Coverage limit	738	2026	2071	338	493	405	460
No limit	0.43	0.49	0.33	0.34	0.34	0.43	0.58
No coverage	0.32	0.47	0.32	0.34	0.35	0.32	0.29
Drug Discount	–	–	–	–	–	–	0.50
Max cost 30	16.5	32.2	17.38	15.35	14.27	15.74	18.26
Min cost 30	4.95	4.68	3.81	4.29	5.35	5.59	5.4
Min percentage	4.57	16.19	1.93	3.09	3.82	6.37	6.29
Max percentage	4.78	16.83	1.94	3.10	3.86	6.4	6.91
Mean percentage	4.71	16.57	1.92	3.10	3.83	6.38	6.73
Netsize (total)	–	–	–	–	–	3013	3011
Netsize (per eligible)	–	–	–	–	–	0.19	0.21

Table 3: Preliminary parameter estimates from the simplified estimation described in section 11.

Parameter	**Estimate**
Constant term	-21.4
Premium	0.010
Drug coverage	1.7
Emergency	-0.011
Original Medicare	5.8
Switching cost	5.1
$, _0$	-41.5
$, _1$	-0.091

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] Biles, Brian, Jonah Pozen and Stuart Guterman. (2009)."The Continuing Cost of Privatization: Extra Payments to Medicare Advantage Plans Jumps to $11.4 Billion in 2009." Commonwealth Fund Issue Brief, online at www.commonwealthfund.org.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] Gold, Marsha. (2008). "Medicare Advantage in 2008." Henry J. Kaiser Family Foundation Issue Brief.

[11] Gowrisankaran, Gautam and Mark Rysman. (2007). "Dynamics of Consumer Demand for New Durable Goods." Working paper, University of Arizona.

[12] 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.

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

[14] 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.

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

[16] 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.

[17] 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.

[18] Reidpath, D.D. and P. Allotey.(2002) "Infant Mortality Rate as an Indicator of Population Health." *Journal of Epidemiology & Community Health*,57,pp344-346.

[19] Shcherbakov, Oleksandr. (2007). "Measuring Consumer Switching Costs in the Television Industry." Working paper, University of Arizona.

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

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