Estimating Dynamic Discrete Choice Models of Product Differentiation: An Application to Medicare Part D with Switching Costs

Daniel P. Miller*

and

Jungwon Yeo†

November 11, 2012

Abstract

This paper proposes an algorithm to estimate dynamic discrete choice models using aggregate market share data. The algorithm achieves a computational advantage by decomposing the complicated mapping between market shares and utility flows into two simpler ones. The first maps observed market shares to mean choice specific values, and the second then maps to mean utility flows. In the application, we estimate switching costs in the Medicare Part D market. Our results indicate a large switching cost of around $1,700, which implies an average welfare loss of $480 as enrollees choose to remain in sub-optimal plans to avoid switching costs.

*Department of Economics, Clemson University; dmille7@g.clemson.edu

†Corresponding author. School of Economics, Singapore Management University; jwyeo@smu.edu.sg

1 Introduction

In many settings, consumers' purchase decisions exhibit dynamics; that is, their choices today affect choices made in the future. Applications with dynamic demand include durable goods (consumer electronics, cars) where the product is expected to last into the future; goods where consumers face switching frictions to switch from one choice to another (TV programs, 401K plans), experience goods (medicines, beauty products) where learning about the product's value impacts repeat purchases; and addiction (tobacco, video games) or resistance goods (antibiotics) where consumers grow addicted or develop resistance to the good.

In this paper, we propose an algorithm to estimate dynamic discrete choice models of product differentiation using aggregate market share data. The method applies to a broad class of models and is relatively straightforward to implement as compared to other methods. Being able to use aggregate data offers practical advantages for applications where individual-level survey or scanner data are not readily available, or available ones are prone to sampling problems such as selection bias and small sample sizes. The method accommodates forward-looking consumer behavior; that is, consumers are assumed to incorporate expectations about the future when making decisions today. Alternative specifications that either omit dynamics or assume consumers act myopically run the risk of biasing estimates of demand parameters, propagating bias to other parts of the model such as supply-side estimates of firm mark-ups derived from demand elasticities, and misguiding policy decisions.

The application in this paper considers switching costs in health insurance. Evidence from individual-level enrollment data documents strong persistence in the demand for health insurance (Carlin and Town, 2009; Handel, 2010), and there are reasons to believe that switching frictions may drive the persistence. For example, switching consumers must first overcome their inattentiveness from having a default option (Madrian and Shea, 2001) to expend time researching alternative choices. After switching plans, they may need to switch health care providers and medications and bear any costs of making mistakes while growing accustomed to a different plan's rules and restrictions. We consider switching frictions from various sources (such as those listed above) as economic costs that consumers must incur to switch plans.

In our application to Medicare Part D prescription drug insurance plans, we model demand as a dynamic discrete choice problem. Each year, enrollees either choose to remain in their existing, default plans or pay a switching cost to select another plan. Consumers are assumed to be rational forward-looking decision makers who think not only about the contemporaneous utility they get from a plan, but also the utility they expect that plan (and other plans) to deliver in the future. Expressed in recursive form, the value of a plan (termed *choice specific* *value* in Hotz and Miller (1993)) is the plan's current flow utility plus the expected discounted stream of utility flows and paid switching costs in the future. We assume consumers choose plans which solve their dynamic programming (DP) problem. We model plans as differentiated products. Consumers derive utility from observable product attributes such as monthly premiums, deductibles, drug copays, and the comprehensiveness of drug formularies (list of covered drugs) and product qualities known to consumers, but unobserved by the econometrician. We let plan characteristics, both observed and unobserved, evolve exogenously. Year to year changes in plan characteristics and consumer preferences are the primary factors driving consumers' decisions to switch plans.

Our estimation procedure is a three-level nested fixed point algorithm. In the outer loop we minimize GMM objective functions based on moment conditions that put restrictions on the serial correlation of unobserved product qualities. We also include moment conditions that require observed product attributes and price instruments be orthogonal to unobserved product qualities. In the middle loop we solve for the fixed point of mean flow utilities that equates model predicted and observed market shares. We break this procedure into two steps. First, we use numerical inversion to solve for the choice specific values that equate model predicted and observed market shares. As shown in Hotz and Miller (1993), this is the dynamic counterpart to the inversion step in Berry (1994) that solves for mean utility values. Second, we employ an iterated fixed point algorithm to recover the mean flow utilities that are consistent with the choice specific values found in step one. The fixed point operator is defined using the functional relationship between mean flow utilities and choice specific values. The second step calls to the inner loop which solves the forward looking consumer's DP problem using value function iteration. While solving the DP, we reduce the dimensionality of the state space—which is large due to the large number of available choices, over 40—following Gowrisankaran and Rysman (2009) (hereafter GR)'s method.

The estimated switching cost ranges from \$1,500 to \$1,700 in our preferred specifications. To put the magnitude in perspective, the estimate is 54% to 60% of the average enrollee's annual prescription drug expenditure (\$2,790) and more than the average out-of-pocket spending on premiums, deductibles, and drug copays (\$893).[1] This large switching cost estimate is consistent with other studies that estimate switching costs for health care plans using individual-level data. The implied average switching probability is 8.8%, which is quite similar to the actual switching probability.

Methodologically, we contribute to the literature on the estimation of dynamic demand

[1] MEDPAC, "A Data Book: Medicare Part D program," March 2010. Dollar figures are for the year 2007, deflated to 2006 dollars.

models using aggregate market share data. Gowrisankaran and Rysman (2009) is the most closely related. Our algorithm shares the same outer and inner loop and uses a similar method to reduce the dimensionality of the state space. The key difference lies in the middle loop. The middle loop in GR solves for the mean utility flows that equate model predicted and observed market shares using a fixed point algorithm that literally mirrors the BLP contraction mapping. It is a slowly converging, cumbersome procedure because it does not operate on any functional relationships generated by the model. The middle loop in our methodology decomposes this fixed point problem into two steps. The first step exploits the model-derived relationship between market shares and choice specific values, and the second step exploits the functional relationship between choice specific values and mean flow utilities. By decomposing the complicated mapping from aggregate market shares to mean flow utilities into two less complex ones, our algorithm achieves significantly faster convergence with more numerical stability. The trade-off to our method's computational simplicity is that it cannot be applied to random coefficient models. The GR algorithm accommodates models with persistent preference heterogeneity through random coefficients. Put simply, GR is the dynamic counterpart to BLP, and our method is the counterpart to Berry (1994). Given the trade-offs, our method is best suited for dynamic discrete choice models where the convergence of the dynamic BLP algorithm may not be easily obtained. Convergence can be hampered in cases where some products are dynamic complements rather than substitutes. Our method is also useful when the identification of random coefficients is somewhat dubious because of limited information on the initial state of the consumer type distribution and when random coefficients are not of first order importance.

Also related is Shcherbakov (2008) who applies GR's nested fixed point algorithm to estimate switching costs in the cable and satellite television industry using aggregate market share data. He models a forward-looking consumer's choice between three alternatives: cable, satellite and antenna TV. His methodological innovation exploits this simple choice set structure to make less restrictive assumptions on the state space of the DP problem than in GR and our method.

Besides the methodological contribution, our application contributes to the health insurance literature. The presence of choice inertia or switching costs in health insurance markets has been well-documented and studied in many papers. Strombom et al. (2002) use enrollee level data on plan choices to explore heterogeneity in price sensitivities. They find that older and sicker demographics and consumers holding a default insurance option with an employer sponsored plan are less price sensitive. They attribute the lower price sensitivity to higher switching costs. Handel (2010) estimates switching costs to study how they alleviate adverse selection in an employer-sponsored insurance setting. He leverages a natural experiment and individual-level panel data on choice and medical claims to cleanly identify switching costs separately from persistent preference heterogeneity. He finds large switching costs on the same order of magnitude as our estimates. A limitation of his method is that he treats consumers as myopic. Ericson (2011) provides descriptive evidence on the presence of switching costs or choice inertia and proposes an optimal default mechanism in the Medicare Part D market. Keating (2007) examines the role of switching costs in Medicare Part D using data over the first two years of the program. He estimates static demand models separately for 2006 and 2007 while allowing simulated consumers to receive extra utility in 2007 from sticking to their default plans. He then infers the switching cost using a similar method to that proposed in Shum (2004). His switching cost estimate is $744. We contribute to this literature by explicitly estimating a dynamic model with forward-looking consumers.

Our paper addresses important policy concerns surrounding Medicare Part D and health insurance reforms. A literature has emerged on Medicare Part D which suggests enrollees may not be making rational plan choices (Abaluck and Gruber, 2011; Heiss et al., 2006, 2010; Lucarelli et al., 2008). These papers propose policy recommendations that would limit the number of choices. We explore whether switching costs and forward looking behavior rationalize some of the apparent non-optimal choices when consumers are assumed to be only myopic. To this end, we use our estimates to simulate enrollees' choices for the years 2006 through 2009. We find the welfare difference between the chosen plan in the dynamic model and the static optimal plan averages just $5.1 in the inception year, 2006. But in subsequent years, after enrollees are "locked-in" to their initial choices, the welfare differences increase to a range of $460 to $480. These larger figures are similar in magnitude to those in Abaluck and Gruber (2011) who attribute the welfare losses to choice inconsistencies. However, their findings are based on data from the inception year, 2006, when we simulate a much smaller welfare loss. Thus, forward-looking behavior due to switching costs does not explain their findings. However, if enrollees tend to make sub-optimal choices when they enroll in Medicare Part D plans for the first time and can resolve the problem by switching plans, then policies aimed at reducing switching costs will be as important as those that help them make the right choices in the first place. Indeed, one set of Medicare Part D reforms from the 2010 Patient Protection and Affordable Care Act has already been enacted to minimize switching costs for low income households.[2]

The paper is organized as follows. In section 2, we briefly describe the institutional features of Medicare Part D and present the model. We describe the estimation algorithm in section 3 and describe the data used for estimation in section 4. In section 5 we report estimates of

[2]HR 3590 Sec 3302 Improvement in Determination of Medicare Part D Low-Income Benchmark Premium, 3303 Voluntary De Minimis Policy for Subsidy Eligible Individuals under Prescription Drug Plans and MA-PDs, 3305 Improved Information for Subsidy Eligible Individuals Reassigned to Prescription Drug Plans and MA-PD Plans.

switching costs and own and cross price elasticities. We also report the average welfare loss due to switching costs. Section 6 concludes.

2 Model

In this section, we describe the model. It builds upon the discrete choice framework of Berry (1994) and Berry et al. (1995), in which a consumer's indirect utility from a product depends on observed product characteristics and product qualities unobserved by the econometrician. Before describing the model, we provide a brief background on Medicare Part D.

2.1 Medicare Part D Background

Medicare Part D is a recently enacted program, authorized under the Medicare Modernization Act of 2003, that provides a prescription drug benefit to the Medicare population. The act requires all Medicare beneficiaries to obtain drug coverage that meets or exceeds a minimum coverage standard; otherwise they face a penalty in the form of higher premiums paid in later years. Instead of the government offering its own plan, the act established a regulated and subsidized insurance exchange where beneficiaries can purchase individual insurance policies sold by competing private insurers. Coverage can also be obtained through a group policy offered by an employer/union under the Retiree Drug Subsidy (RDS) program, an MA+Part D policy that bundles Medicare Advantage (hospital/doctor coverage) with Part D prescription drug coverage, or another drug program such as Veterans Affairs insurance.

The rules permit insurers to freely enter and exit the market. The regulations require each insurer to offer at least one "basic" plan with coverage characteristics that meet the minimum standard. They may also offer enhanced plans with more generous coverage. Coverage characteristics such as annual deductibles, drug copay/coinsurance rates, and drug formularies differentiate plans. Insurers choose the monthly premium and must charge the same amount to all enrollees regardless of age, demographics, or prior experience. The government subsidizes a portion of the premium for all enrollees and provides additional subsidies for low income households under the Low Income Subsidy (LIS) program.

Medicare beneficiaries may choose any plan offered in their home market. There are many choices; the typical market has over 40 plans offered by about 20 insurers. Markets are geographically separated in 39 regions drawn around state boundaries. Newly eligible Medicare beneficiaries (those reaching 65 years old) must actively choose and enroll in a Part D plan during an open enrollment period that runs from November 15th to December 31st. Current enrollees who take no action in the open enrollment period are automatically reenrolled in their current, default plans for the upcoming year. They may switch to a different plan for the upcoming year by calling the Center for Medicare and Medicaid Services (CMS) or the insurers during the open enrollment period. It should be noted that insurers may adjust their premiums and coverage characteristics from year to year as market conditions evolve and regulations adjust. An insurer may also consolidate its plans and carry over default enrollees from the pre-consolidated plans.

The enrollment procedures are different for a subset of LIS eligible households deemed "dual eligibles" that have both Medicare and Medicaid coverage. They compose about 25% of the Medicare population. If a dual eligible does not actively select a plan during the open enrollment period, Medicare will automatically and randomly assign him to an LIS eligible plan. Only basic plans with premiums set below a threshold are eligible to receive automatically enrolled dual eligibles.[3] For the upcoming year a randomly assigned dual eligible is reassigned to his current plan if the plan retains its LIS eligibility. Otherwise, the dual eligible is assigned to another LIS eligible plan.[4]

2.2 Model Set-up

Let $\mathbf{J}_{mt} = \{0, 1, ..., J_{mt}\}$ denote the set of choices available to consumers in market m in year t. Per regulation, all consumers have the same choice set, and we assume they know all plans available in their regions.[5] Choice 0 is the outside option. Let $d_{it} \in \{1, ..., J_{mt}\}$ denote consumer i's default plan that she would be automatically enrolled in for year t. $d_{it} = 0$ if the consumer does not have a default plan. When a plan exits the market, the default for its enrollees becomes the outside option.[6]

Consumer i's indirect utility from plan j is given as a linear function of plan j's current premium (p_{jt}) and other characteristics ($\mathbf{X}'_{jt}$):

[3] The threshold is determined by a weighted average of the premiums set by plans, where weights are calculated using past enrollments.

[4] There is sufficient time for a randomly assigned dual eligible to reject his assignment and actively enroll in another plan. However, if a dual eligible has ever actively enrolled in a plan, he will never be eligible in the future for automatic enrollment.

[5] An alternative modeling approach with explicit search frictions would limit the choice set to the plans that a consumer has discovered through a search process. We do not adopt this interpretation of search frictions because the Medicare Part D insurance exchange is designed to be highly transparent. Medicare disseminates information about all plan offerings through many channels including direct mailings, the Medicare.gov website, and pharmacies. No plan would be hidden from consumers with such information so readily available at their fingertips.

[6] From an enrollee's perspective, exits are treated as a random event of her default becoming the outside option. We do not differentiate consumers whose default is the outside option due to exits of their previously chosen plans from those whose default is the outside option by active choice. Exits in this market rarely occur, and the market shares of the exiting plans are much smaller than the average market shares. See Table 12.

$$
\begin{aligned}
u_{ijt} &= -\alpha p_{jt} + \mathbf{X}'_{jt}\beta_x + \xi_{jt} + \epsilon_{ijt} - sc \cdot 1(j \neq d_{it}) \text{ if } d_{it} \neq 0 \\
&= -\alpha p_{jt} + \mathbf{X}'_{jt}\beta_x + \xi_{jt} + \epsilon_{ijt} \;\; \text{if } d_{it} = 0.
\end{aligned}
$$

ξ_{jt} captures a product attribute or quality that enters a consumer's utility but is not observed by the econometrician. ϵ_{ijt} is assumed to be drawn from the type 1 extreme value distribution independently across consumers, plans, and time periods. We normalize the mean utility flow from the outside option to be zero so that $u_{i0t} = \epsilon_{i0t}$. Persistency in preferences is captured by time-invariant coefficients α and β_x, and heterogeneity in preferences is captured by the Logit error terms, ϵ_{ijt}, as in Berry (1994). Thus heterogeneity in consumer preferences is not persistent. The parameter sc captures the switching cost associated with an enrollee choosing a plan other than her default plan.

Note that we assume consumers moving from the outside option to an inside option do not incur a switching cost. Only consumers moving amongst inside options and those moving from an inside option to the outside option face switching costs. The primary reason for this modeling choice is that our data starts with the inception of the Part D program in 2006. By definition no consumer had a default plan, and there was a massive influx of enrollment in that initial year. We do not want this large wave of enrollment to be rationalized by a low switching cost estimate. By imposing asymmetry on the switching cost, we can better focus on our goal of estimating switching costs that rationalize the choice behavior of consumers who have a default Part D plan.[7] There is also a technical issue we encountered in our estimation algorithm when we allow for switching costs from the outside option. We describe the issue in section 3, and present illustrative estimates of a modified model in section 5.

2.3 Discussion: Sources of Switching Costs

In economic models, switching costs are parameters that rationalize consumers making repeat purchases of the same product. The switching costs in our model impose disutility on consumers who choose plans differing from their default plans. As a result, switching costs induce inertia in consumers' plan choices.

It's worthwhile discussing various sources of choice inertia in our application to health insurance plans. We should note that there are no explicit penalties for switching Part D plans like one would face from breaking a contract to switch cell phone carriers. It is also notable that marketing activities targeting switching behavior, such as new customer sign up discounts and

[7] Handel (2010) also adopts this view.

customer loyalty programs, are not sources of switching costs because the Part D regulations strictly prohibit such practices. Instead there are several indirect switching cost channels. We classify these costs into broad categories: risk aversion toward products with unknown quality, search, learning, transaction, and psychological.

First, choice inertia or repeat purchases can arise when products are experience goods and risk-averse consumers believe that products' values are persistent (Klemperer, 1995; Dubé et al., 2010). In Part D, an enrollee that has developed familiarity with copay prices, restrictions, and drug coverage under her default plan may be reluctant to switch to another plan with unfamiliar coverage. Moreover, if a different plan's formulary restrictions require an enrollee to change medications, there is uncertainty about untested medicines and concerns over possible adverse physiological consequences. As evidence that Part D enrollees base decisions on prior experience, Ketcham et al. (2011) find that an enrollee is more likely to switch plans if she learns, ex-post, that she could have saved on out-of-pocket costs by choosing a different plan in the previous year. Their findings imply that Part D plans are likely experience goods and enrollees expect information they acquired in the previous year, whether about their current plans or their health conditions or the match qualities, to persist.

Second, when products are search goods, their values can be acquired by consumers through search or comparative research, but there are time and effort costs to do so (Moshkin and Shachar, 2002; Dubé et al., 2010). For the case of Part D plans, enrollees have to forecast their demand for a variety of medications and then compare formularies, copays, and coinsurance rates to calculate expected out-of-pocket expenses. Moreover, they need to acquaint themselves with non-monetary characteristics including pharmacy networks and restrictions, such as requirements to obtain prior authorization, to evaluate each plan. Kling et al. (2012) find that comparison friction—the wedge between availability of information and consumers' use of it—has a significant impact on switching behavior. In their experiment, they find that 28% of the group of Medicare beneficiaries receiving a letter with personalized cost information about plans switch for the upcoming year, while only 17% of the comparison group, those receiving a letter with the web address to look up such information, switch. They argue that the intervention reduces the search costs of acquiring information and makes the benefits of switching more apparent, so that more enrollees find it worthwhile to pay the switching cost to changes plans. In our modeling framework, we treat search costs as a form of switching costs and do not distinguish the two.

Note that these two sources regard information asymmetry that can be resolved by experience and search, respectively. They both lead to choice inertia provided that an enrollee knows more about his default plan than alternative plans.

The third source involves learning costs incurred after consumers switch products. Switching enrollees face costs to learn a new set of plan rules and procedures and may make costly mistakes during the learning process. For example, on the first visit to the pharmacy, a new enrollee might forget to obtain a prior authorization and be denied coverage. These learning costs are analogous to those described in Klemperer (1995) where the example is that of a PC user switching to an Apple computer and needing to learn the Apple operating system.

The fourth source is transaction costs. For switching enrollees who need to switch drug regimens and pharmacies, there are transactions costs involved with doctor's visits to obtain new prescriptions and costs to transfer prescriptions to a new pharmacy.[8]

Finally, switching frictions can arise from inattention, procrastination or status-quo biases (Madrian and Shea, 2001; Samuelson and Zeckhauser, 1988) and psychological costs (Klemperer, 1995; Dubé et al., 2010). We feel that these can be quite relevant to the Part D market because enrollees are automatically enrolled in their default plans as the result of them doing nothing. It is also possible that some enrollees simply forget to call CMS to elect a different plan during the open enrollment period.

All of these sources of switching costs will contribute to our estimate of the switching cost, sc. Although policy makers may want to disentangle the sources, doing so is beyond the scope of this paper, especially given that we work with aggregate market share data.

2.4 Forward-looking Consumer's Problem

We assume that the Medicare beneficiary is a forward-looking consumer who maximizes the expected stream of utility flows with an effort to minimize costs from possible switches. Let Ω_t denote an information set that includes current product attributes $\{p_{jt}, \mathbf{X}'_{jt}\}_{j\in\mathbf{J}_{mt}}$ and other factors that influence future product attributes. We assume Ω_t follows a Markov process. Defining utility flow $\delta^f_j(\Omega_t) \equiv -\alpha p_{jt} + \mathbf{X}'_{jt}\beta_x + \xi_{jt}$, the Bellman equation of the forward-looking consumer is given as

$$V(d_{it}, \Omega_t, \boldsymbol{\varepsilon}_{it}) = \max_{d_{it};\ j\in\mathbf{J}_{mt}\setminus\{d_{it}\}} \left\{ \begin{array}{c} \delta^f_{d_{it}}(\Omega_t) + \epsilon_{id_{it}} + \beta E\left[EV(d_{it}, \Omega_{t+1})|\, \Omega_t\right], \\ \delta^f_j(\Omega_t) + \epsilon_{ijt} - sc + \beta E\left[EV(j, \Omega_{t+1})|\, \Omega_t\right] \end{array} \right\}. \tag{1}$$

where $\boldsymbol{\varepsilon}_{it} = [\epsilon_{ijt}]_{j\in\mathbf{J}_{mt}}$ and $EV(j, \Omega_t) = \int V(j, \Omega_t, \boldsymbol{\varepsilon}_{it}) dP(\boldsymbol{\varepsilon}_{it})$.

Let $cv_j(\Omega_t) \equiv \delta^f_j(\Omega_t) + \beta E\left[EV(j, \Omega_{t+1})|\, \Omega_t\right]$ denote the choice specific value function of plan j. Then, as in Rust (1987), the expected value function can be written as

[8] The doctors office for one of this paper's authors charges \$15 per prescription faxed over to a new pharmacy.

$$EV(d_{it}, \Omega_t) = \ln \left[\phi \sum_{j \in \mathbf{J}_{mt} \backslash \{d_{it}\}} \exp\left(cv_j(\Omega_t)\right) + \exp(cv_{d_{it}}(\Omega_t)) \right] + \gamma, \tag{2}$$

where $\phi = \exp(-sc)$ and γ denotes Euler's constant, which we will omit for notational simplicity.[9]

Let $\delta_{-d}(\Omega_t) \equiv \ln(\sum_{j \neq d} \exp\left(cv_j(\Omega_t)\right))$ so that $\exp(\delta_{-d}(\Omega_t)) = \sum_{j \neq d} \exp\left(cv_j(\Omega_t)\right)$. This value, which is similar to the logit inclusive value in Gowrisankaran and Rysman (2009), is monotonic in the sum of the choice specific values of the non-default plans. We call this value the *non-default option value*. Now Equation (2) can be rewritten as

$$EV(d_{it}, \Omega_t) = \ln \left[\phi \exp\left(\delta_{-d_{it}}(\Omega_t)\right) + \exp(cv_{d_{it}}(\Omega_t)) \right]. \tag{3}$$

The possible large dimensionality of the state variables Ω_t could make the DP problem almost impossible to solve. To reduce the dimension of the state space, we make the following two assumptions about the transition probabilities, $P(\delta_d^f(\Omega_{t+1}), \delta_{-d}(\Omega_{t+1}) | \Omega_t)$.

Assumption 1 *If* $\delta_d^f(\Omega_t) = \delta_d^f(\Omega_t')$ *and* $\delta_{-d}(\Omega_t) = \delta_{-d}(\Omega_t')$, *then* $P(\delta_d^f(\Omega_{t+1}), \delta_{-d}(\Omega_{t+1}) | d, \Omega_t) = P(\delta_d^f(\Omega_{t+1}'), \delta_{-d}(\Omega_{t+1}') | d, \Omega_t')$ *for all* t, Ω_t, Ω_t' *and for all* $d \in \mathbf{J}_{mt}$.

In words, if two different information sets Ω_t and Ω_t' lead to the same mean utility flow and *non-default option value* for plan d, then the joint distribution of the plan's future mean utility flow and non-default option value is the same. This implies that δ_d^f and δ_{-d} have all the information relevant to the evolution of the state variables in Ω_t, as in GR. Under this assumption, we can reduce the state variables Ω_t into a pair $(\delta_d^f,\ \delta_{-d})$ and rewrite $EV(d, \Omega_t)$ as $EV(d, \delta_{dt}^f, \delta_{-dt})$. As noted by GR, in principle, this assumption is restrictive. For example, when δ_d^f and δ_{-d} are high, it implies consumers expect δ_d^f and δ_{-d} to evolve in the same fashion regardless of whether δ_{-d} is high because there are many non-default products with low choice specific values or because there are just a few non-default products with high choice specific values. However, in our application, the number of plans and the distribution of their market shares, which could be relevant omitted state variables, are quite stable over the sample period.

Assumption 2 *If* $\delta_{dt}^f = \delta_{d't}^f$ *and* $\delta_{-dt} = \delta_{-d't}$, *then* $P(\delta_{dt+1}^f, \delta_{-dt+1} | d, \delta_{dt}^f, \delta_{-dt}) = P(\delta_{d't+1}^f, \delta_{-d't+1} | d', \delta_{d't}^f, \delta_{-d't})$ *for all* t *and* $d, d' \in \mathbf{J}_{mt}$.

In words, if plan d and plan d' have the same levels of mean utility flow and non-default option value, then consumers expect the distributions of future values of the two variables to

[9] See Rust (1987) for proof.

be identical. This assumption further reduces the dimension of the state space from three to two by assuring that the expected value function depends on the default plan d only through the values of the two state variables, δ_d^f and δ_{-d}. This assumption will not be appropriate if plans exhibit systematic differences in their dynamic behavior and consumers are aware of these differences. For example, there is concern if some plans tend to keep the mean utility levels constant over time whereas the others steadily decrease them. This assumption is neither necessary nor appropriate for many studies. In our application, it is useful because we need to estimate transition probabilities using data with a large number of choices in each market over a short sample period (4 years). Also, the Medicare Part D market is new; we believe no insurer has established a reputation for consumers to expect plans to behave differently from one another.

Under the two assumptions, we can rewrite $EV(d, \Omega_t)$ as $EV(\delta_{dt}^f, \delta_{-dt})$ and $cv_d(\Omega_t)$ as $cv(\delta_{dt}^f, \delta_{-dt})$. Now the expected value function is given as

$$
\begin{aligned}
EV(\delta_{dt}^f, \delta_{-dt}) &= \ln\left[\phi \exp\left(\delta_{-dt}\right) + \exp\left(cv(\delta_{dt}^f, \delta_{-dt})\right)\right], && (4)\\
\delta_{-dt} &\equiv \ln\left(\sum\nolimits_{j \neq d} \exp\left(cv(\delta_{jt}^f, \delta_{-jt})\right)\right), && (5)\\
cv(\delta_{jt}^f, \delta_{-jt}) &= \delta_{jt}^f + \beta \cdot EEV(\delta_{jt}^f, \delta_{-jt}) \quad \forall\, j \in \mathbf{J}_{mt} && (6)
\end{aligned}
$$

where $EEV(\delta_{jt}^f, \delta_{-jt}) = E\left[EV(\delta_{jt+1}^f, \delta_{-jt+1}) \middle| \delta_{jt}^f, \delta_{-jt}\right]$. Let $\boldsymbol{\delta}_t^f = \left[\delta_{jt}^f\right]_{j \in \mathbf{J}_{mt} \setminus \{0\}}$ denote the vector of mean utility flows of all products except the outside option. Note that the choice specific value function $cv(\delta_{jt}^f, \delta_{-jt})$ is ultimately a function of $\boldsymbol{\delta}_t^f$, given the Markov transition probability distribution, because the non-default option value δ_{-jt} is an implicit function of $\boldsymbol{\delta}_t^f$. For notational simplicity, we let cv_{jt} denote the choice specific value for plan j in time t.

Let $s_{jt}(d)$ denote the probability that a consumer whose default plan is d chooses to purchase plan j in time t. When $d \neq 0$, the probability differs depending on whether plan j is the default plan or not:

$$
s_{jt}(d) = \frac{\exp(-sc \cdot 1(j \neq d)) \exp(cv_{jt})}{\phi \exp\left(\delta_{-dt}\right) + \exp(cv_{dt})} \tag{7}
$$

The probability that a consumer whose default is the outside option purchases plan j is given as

$$
s_{jt}(0) = \frac{\exp(cv_{jt})}{\exp\left(\delta_{-0t}\right) + \exp(cv_{i0})}.
$$

With a positive switching cost, equation (7) shows the probability of choosing the default plan is higher than the probability of choosing a non-default plan that has an identical choice specific value. We want to emphasize that choice probabilities in a dynamic Logit model are functions of choice specific values whereas they are functions of mean utility flows in a static Logit model. This implies that choice probabilities, as well as switching probabilities, can be computed using only knowledge of products' choice specific values $\{cv_{jt}\}_{j \in \mathbf{J}_{mt}}$. It is not necessary to know how the choice specific values are decomposed into current mean utility flows and continuation values.

To derive the aggregate market share expression, we need to account for the annual influx of newly eligible Medicare beneficiaries entering the market and outflux exiting the market. Senior citizens become eligible at age 65 and remain eligible for the rest of their lives. We assume they exit the market upon death. Let N_{ot} denote the number of people who were eligible in $t-1$, and N_{nt}, the number of people who became newly eligible in t. ρ is one minus the death rate for the population over 65. Define $n_{ot} = \dfrac{\rho N_{ot}}{\rho N_{ot} + N_{nt}}$ as the fraction of the Medicare population eligible in $t-1$ who remain in the Medicare Part D market in year t.[10]

The market share of plan j in time t is

$$M_{jt} = n_{ot} \sum_{d=0}^{J_{mt}} s_{jt}(d) M_{dt-1} + (1 - n_{ot}) s_{jt}(0) \tag{8}$$

The first term is the probability that consumers who were eligible in period $t-1$ choose plan j and the second term is the probability that newly eligible consumers in period t choose plan j. Equation (8) shows that the current market share of plan j is determined by the distribution of lagged market shares of all plans $\{M_{dt-1}\}_{d \in \mathbf{J}_{mt}}$ and all pairwise switching probabilities $\{s_{jt}(d)\}_{d \in \mathbf{J}_{mt}}$. Note that $M_{j0} = s_{j0}(0)$ where we let $t = 0$ denote the inception year of the program, 2006.

3 Estimation Algorithm and Identification

We estimate the switching cost and flow utility parameters using a three-level nested fixed point algorithm. In the outer loop we minimize GMM objective functions based on moment conditions that put restrictions on the serial correlation of unobserved product qualities. We also include moment conditions that require observed product attributes and price instruments be orthogonal to unobserved product qualities. The middle loop recovers unobserved product qualities by inverting the non-linear system of market share equations into mean utility flows. The inner loop solves the dynamic programming problem described by equation (4) through (6).

[10] We do not account for under 65 individuals who are eligible for Medicare through disability. This omission will not introduce bias if their exit rate from Medicare is similar to that of the over 65 Medicare eligibles.

Our estimation algorithm is most similar to Gowrisankaran and Rysman (2009) in that they share the same outer and inner loop. The key difference lies in the middle loop that recovers mean utility flows. Our middle loop proceeds in two steps. The first step recovers choice specific values by numerically inverting the non-linear system of market shares equations in (8) after substituting observed market shares for model-predicted ones. This step is the dynamic counterpart of the inversion in Berry (1994). Because the market share equations are solely functions of choice specific values, the first step inversion can be performed without knowledge of mean utility flows. The second step searches for the vector of mean utility flows that generate the recovered choice specific values. Recall that the choice specific value function $cv(\delta_{jt}^f, \delta_{-jt})$ is an explicit function of the mean utility flow δ_{jt}^f and the non-default option value δ_{-jt}. As the latter is itself an implicit function of the vector of mean utility flows of all other alternatives, the choice specific value function is a function of the vector of mean utility flows $\boldsymbol{\delta}_t^f$. We apply a fixed point algorithm to solve for $\boldsymbol{\delta}_t^f$ in this step.

The method in Gowrisankaran and Rysman (2009) skips the first step of recovering (mean) choice specific values. Instead, their method directly inverts out mean utility flows from the market share equations by literally applying the BLP fixed point algorithm to mean utility flows (with a small modification to slow down the convergence process). For a non-random coefficient Logit model, we have found that our method performs significantly better than theirs in terms of convergence speed.[11]

3.1 Estimation Algorithm

First, we discretize the state space $(\delta_d^f,\ \delta_{-d})$ to solve the DP problem in the inner loop.[12] We assume that the discount factor, β and the survivor rate, $(1-\ \rho)$ are known.[13]

3.1.1 Outer Loop

The outer loop searches over the parameter space to satisfy the orthogonality condition $E(Z'\xi) = 0$ where Z is a matrix of instrumental variables. The parameter estimate $\widehat{\boldsymbol{\theta}}$ is chosen to minimize

[11]For comparison, we ran the GR algorithm using one consumer type for various trial values of the switching cost. Only 4 to 11 markets achieved convergence after around 800 to 900 iterations and most markets failed to converge within the specified maximum of 1,500 middle loop iterations. Under our algorithm, most markets achieved convergence within around 400 iterations and just a few required over 700 iterations. Only one market, at one particular value of the switching cost, failed to converge within 1500 iterations.

[12]We descretize δ_d^f into 40 evenly-spaced grid points and δ_{-d} into 60 evenly-space grid points.

[13]We do not explicitely model how the non-zero death rate enters the senior citizen's dynamic programming problem.

the corresponding GMM sample criterion function. That is,

$$\widehat{\boldsymbol{\theta}}=\arg\min_{\boldsymbol{\theta}\in\Theta} G(\boldsymbol{\theta};\widehat{\boldsymbol{\delta}_t^f};\widetilde{X},Z)'WG(\boldsymbol{\theta};\widehat{\boldsymbol{\delta}_t^f};\widetilde{X},Z) \tag{9}$$

where $\widetilde{X}=\{p_j,X_j\}_{j\in\mathbf{J}_{mt}\setminus\{0\}}$, $\boldsymbol{\theta}=\{sc,\alpha,\beta_x\}$, $G(\boldsymbol{\theta};\widehat{\boldsymbol{\delta}_t^f};\widetilde{X},Z)=\frac{1}{N}Z'\boldsymbol{\xi}(\boldsymbol{\theta};\widehat{\boldsymbol{\delta}_t^f};\widetilde{X},Z)$ is the sample analog of the moment conditions, and W is a weight matrix.

We can concentrate out the linear parameters, and perform our numerical minimization of the GMM criterion function over the switching cost parameter. For each trial value of the switching cost, we obtain a vector of mean utility flows $\widehat{\boldsymbol{\delta}_t^f}$. As in Berry (1994) and Berry et al. (1995), we obtain a vector of unobserved product qualities $\left[\xi_{jt}\right]_{j\in\mathbf{J}_{mt}\setminus\{0\}}$ as the residual of IV regression of $\widehat{\delta_{jt}^f}$ on $(p_{jt},\mathbf{X}_{jt})$.

To construct moment conditions, we consider two different cases: when the unobserved product quality ξ_{jt} is serially independent, and when it is serially correlated. We instrument for the premium as it is likely set in response to the unobserved demand shock ξ_{jt}. Our instruments for the premium include lagged prices of the same plan in other regions. We maintain the assumption that other observed product attributes are exogenous. We elaborate on instrumental variables and identification of the switching cost and linear preference parameters in section 3.2.

Our algorithm begins with an initial guess of the switching cost. Let sc_k denote the k-th trial value of the switching cost.

3.1.2 Middle Loop

First Step: For the given switching cost sc_k passed from the outer loop, the first step of the middle loop recovers the *relative choice specific value* for each product. Let $\Delta cv_{jt}\equiv cv_{jt}-cv_{0t}$ denote what we call product j's relative choice specific value. At this step, we can only solve for the J_{mt} relative choice specific values, $\Delta cv_{1t},\ldots,\Delta cv_{J_{mt}t}$, not the $J_{mt}+1$ choice specific values because equation (8) only gives J_{mt} linearly independent market share equations. This follows from the fact that $s_{0t}(d)=1-\sum_{j\neq 0}s_{jt}(d)$, and can be seen by dividing the numerator and denominator of the choice probability in the market share equation by cv_{0t} :

$$\begin{aligned} s_{jt}(d) &= \frac{\exp(-sc\cdot 1(j\neq d))\exp(cv_{jt}-cv_{0t})}{\exp(-sc)\sum_{k\neq d}\exp\left(cv_{kt}-cv_{0t}\right)+\exp(cv_{dt}-cv_{0t})} \\ &\equiv \frac{\exp(-sc\cdot 1(j\neq d))\exp(\Delta cv_{jt})}{\exp(-sc)(1+\sum_{k\neq d,0}\exp\left(\Delta cv_{kt}\right))+\exp(\Delta cv_{dt})}. \end{aligned}$$

Given a trial value of the switching cost sc_k and the shares of newly entering consumers $\{(1-n_{ot})\}$, we substitute model predicted market shares in equation (8) with actual market shares, and then solve for the vector of unknown relative choice specific values, $\mathbf{\Delta cv}_t = [\Delta cv_{jt}]_{j \in \mathbf{J}_{mt} \setminus \{0\}}$. Note that $\mathbf{\Delta cv}_t$ is only determined by the value of the switching cost sc and observed market shares $\mathbf{M_t} \equiv [M_{jt},\ M_{jt-1}]_{j \in \mathbf{J}_{mt} \setminus \{0\}}$. Therefore, the value of $\mathbf{\Delta cv}_t$ is fixed until a new trial value of the switching cost is passed by the outer loop.

This step is the dynamic counterpart of the inversion in Berry (1994). Unlike the static model in Berry (1994) that has an analytical solution, the inversion must be performed numerically.[14] However, it is not computationally expensive and can be implemented with standard solver packages such as fsolve in Matlab.

Second Step: In the second step we recover J_{mt} values of mean utility flows $\boldsymbol{\delta}_t^f$ from the J_{mt} functional values of $\mathbf{\Delta cv}_t$ recovered in the first step. Because the choice specific value function $cv(\delta_{jt}^f, \delta_{-jt})$ is an implicit function of $\boldsymbol{\delta}_t^f$ and δ_{0t}^f (normalized to zero), this requires another inversion of a function. We turn this into a fixed point problem.

To illustrate the core idea of this step, consider a simplified model where we assume that the function $EEV(\cdot,\cdot)$ is known. This is equivalent to assuming that the Markov transition probability distributions of δ_d^f and δ_{-d} are known. Note that cv_{0t} is a function of $\mathbf{\Delta cv}_t$ as shown in $cv_{0t} = \beta \cdot EEV(0, \ln(\sum_{j \neq 0} \exp(\Delta cv_{jt} + cv_{0t})))$. Let's suppose, with $EEV(\cdot,\cdot)$ known, we could invert the above expression to solve for cv_{0t} and hence could compute $\delta_{-jt} = \ln(\sum_{k \neq j} \exp(\Delta cv_{kt} + cv_{0t}))$. Recall that $\mathbf{\Delta cv}_t$ is solely determined by sc and observed market shares $\mathbf{M_t}$. Thus, cv_{0t} and δ_{-jt} are also functions of sc and $\mathbf{M_t}$. To emphasize this, we rewrite them as $cv_{0t}(sc, \mathbf{M_t})$ and $\delta_{-jt}(sc, \mathbf{M_t})$. Now we can solve for δ_{jt}^f based on the definition of the choice specific value function:

$$\Delta cv_{jt}(sc, \mathbf{M_t}) + cv_{0t}(sc, \mathbf{M_t}) = \delta_{jt}^f + \beta \cdot EEV\left(\delta_{jt}^f, \delta_{-jt}(sc, \mathbf{M_t})\right). \tag{10}$$

Because δ_{jt}^f in equation (10) cannot be solved for analytically, we turn to a fixed point algorithm. Define the operator $\widetilde{T}(sc, M_\mathbf{t}) : R^{\ J_{mt}} \to R^{\ J_{mt}}$ pointwise by

[14] The log-linearity exploited in the analytical Berry inversion does not hold with a positive switching cost because a plan's market share is derived from the sum of consumers switching into the plan from all other plans. If the switching cost is zero, $s_{jt}(d) = s_{jt}$ regardless of default d, and therefore equation (8) implies $M_{jt} = s_{jt}$. Also, $\beta E\left[EV(d_{it}, \Omega_{t+1}) | \Omega_t\right] = \beta E\left[EV(\Omega_{t+1}) | \Omega_t\right]$ as the default does not affect future decisions. These together lead to an analytical inversion

$$\frac{M_{jt}}{M_{0t}} = \frac{s_{jt}}{s_{0t}} = \exp(cv_{jt} - cv_{0t}) = \exp(\delta_{jt}^f + \beta E\left[EV(\Omega_{t+1}) | \Omega_t\right] - (0 + \beta E\left[EV(\Omega_{t+1}) | \Omega_t\right])) = \exp(\delta_{jt}^f)$$

just as in Berry (1994).

$$\widetilde{T}^j(sc, \mathbf{M_t})\left[\delta^f_{jt}\right] = \Delta cv_{jt}(sc, \mathbf{M_t}) + cv_{0t}(sc, \mathbf{M_t}) - \beta \cdot EEV\left(\delta^f_{jt}, \delta_{-jt}(sc, \mathbf{M_t})\right) \tag{11}$$

for $j = 1, ..., J_{mt}$. A fixed point $\boldsymbol{\delta}^f_t$ of the operator $\widetilde{T}$ solves equation (10). The operator $\widetilde{T}$ is a contraction mapping under some mild condition (see Appendix B for proof), and hence there is a unique fixed point.

In the above hypothetical scenario, we assume $EEV(\cdot,\cdot)$ is known and we can analytically invert $EEV(\cdot,\cdot)$ to obtain cv_{0t}. In reality, $EEV(\cdot,\cdot)$ is unknown because the parameters of the transition probability distributions are unknown. We estimate them using $\boldsymbol{\delta}^f_t$. As such, the second step nests the value function iteration. Thus, $EEV(\cdot,\cdot)$ is re-estimated every time $\boldsymbol{\delta}^f_t$ takes on different values. We also need to modify the fixed point operator in equation (11) because cv_{0t} cannot be solved for analytically from $EEV(\cdot,\cdot)$.

Our second step begins by passing the initial guesses of $cv_{0t(0)}$ and $\boldsymbol{\delta}^f_{t(0)}$ to the inner loop. Let $EEV^{(n)}(\cdot,\cdot)$ denote the function that the inner loop returns after the n-th iteration of the second step. Given $EEV^{(n)}(\cdot,\cdot)$, the middle loop's second step updates values for $cv_{0t(n+1)}$ and $\delta^f_{jt(n+1)}$ using the operator $T(sc, \mathbf{M_t})\left[cv_{0t(n)}, \boldsymbol{\delta}^f_{t(n)}\right] : R^{\ J_{mt}+1} \rightarrow R^{\ J_{mt}+1}$ defined pointwise as

$$\begin{aligned} T^0(sc, \mathbf{M_t})\left[cv_{0t(n)}\right] &= \beta \cdot EEV^{(n)}(0, \ln(\textstyle\sum_{j\neq 0} \exp(\Delta cv_{jt} + cv_{0t(n)}))), \\ T^j(sc, \mathbf{M_t})\left[\delta^f_{jt(n)}\right] &= \Delta cv_{jt} + T^0(sc, \mathbf{M_t})\left[cv_{0t(n)}\right] - \beta \cdot EEV^{(n)}(\delta^f_{jt(n)}, \delta_{-jt(n)}) \end{aligned}$$

for $j = 1, ..., J_{mt.}$ Because the inner loop works on a discretized state space, we need to compute the continuation value $EEV^{(n)}(\delta^f_{jt(n)}, \delta_{-jt(n)})$ for each $j \in \mathbf{J}_{mt}$ using an interpolation technique such as cubic splines. This operator T differs from the one in equation (11) because it includes the additional pointwise operator T^0 and because the function $EEV^{(n)}$ is not fixed at each iteration. In summary, updating $cv_{0t(n+1)}$, $\delta^f_{jt(n+1)}$ and $EEV^{(n+1)}(\cdot,\cdot)$ proceeds in three sequential steps. First, we apply the pointwise operator T^0 to update the value for $cv_{0t(n+1)}$. Second, we substitute $cv_{0t(n+1)}$ into the pointwise operators T^j to update the values for $\delta^f_{jt(n+1)}$.[15] Third, we pass $cv_{0t(n+1)}$ and $\delta^f_{jt(n+1)}$ to the inner loop to update $EEV^{(n+1)}(\cdot,\cdot)$. The loop continues until $err_n = \underset{j\in\mathbf{J}_{mt}\setminus\{0\}}{max}\left\{\left|cv_{0t(n+1)} - cv_{0t(n)}\right|, \left|\delta^f_{jt(n+1)} - \delta^f_{jt(n)}\right|\right\} < \epsilon$ where ϵ is a pre-specified criterion for convergence.[16]

[15] In early iterations (up to 200 iterations), we set $cv_{0t(n+1)} = T^0\left[cv_{0t(n)}, \boldsymbol{\delta}^f_{t(n)}\right]$ and $\delta^f_{jt(n+1)} = T^j\left[cv_{0t(n)}, \boldsymbol{\delta}^f_{t(n)}\right]$. For later iterations, we set $cv_{0t(n+1)} = .2\ cv_{0t(n)} + .8\ T^0\left[cv_{0t(n)}, \boldsymbol{\delta}^f_{t(n)}\right]$ and $\delta^f_{jt(n+1)} = .2\ \delta^f_{jt(n)} + .8 T^j\left[cv_{0t(n)}, \boldsymbol{\delta}^f_{t(n)}\right]$ to stabilize the algorithm. For quite a few markets and many switching cost values, the middle loop converged with less than 200 iterations.

[16] We tested with $\epsilon = 1e-09$ and $\epsilon = 1e-06$. The difference in the resulting vectors is within the precision.

Unlike the operator $\widetilde{T}$ in equation (11), we cannot prove that $T(sc, \mathbf{M_t})\left[cv_{0t(n)}, \boldsymbol{\delta}^f_{t(n)}\right]$ is contractive and hence cannot prove there is a unique fixed point. We use many starting values of $cv_{0t(0)}$ and $\delta^f_{t(0)}$, and find that uniqueness is not a problem in practice. When the switching cost is greater than 4, the operator converges to the same vector of $\boldsymbol{\delta}^f_t$ for all markets and for all starting values. When the switching cost is less than 4, uniqueness fails in at most one or two markets. Uniqueness does not hold for the alternative model where consumers whose default option is the outside option also pay the switching cost to enroll in a plan. The lack of uniqueness is the technical reason for not using this alternative model.

3.1.3 Inner Loop

The inner loop takes values of $cv_{0t(n)}$ and $\boldsymbol{\delta}^f_{t(n)}$ passed from the middle loop and computes $\delta_{-jt(n)} = \ln\left(\sum_{k\neq j} \exp(\widehat{\Delta cv_{jt}} + cv_{0t(n)})\right)$ for each $j \in \mathbf{J}_{mt}$. Then it solves the dynamic programming problem to find a fixed point of the expected value function in equation (4) and returns those functional values, $EEV^{(n)}(\cdot,\cdot)$, back to the middle loop.

To solve the DP problem, it necessary to estimate the Markov transition probability distributions for the two state variables δ^f_{jt} and δ_{-jt}. We make the following simplifying assumption on the transition probabilities:

$$P(\delta^f_{jt+1}, \delta_{-jt+1} \mid \delta^f_{jt}, \delta_{-jt}) = \Pr(\delta^f_{jt+1} \mid \delta^f_{jt}) \cdot \Pr(\delta_{-jt+1} \mid \delta_{-jt}). \tag{12}$$

This assumption states that the two state variables evolve independently. It implies the consumer believes that the flow utility the plan is expected to deliver in the next period does not depend on the value of other plans as embodied in δ_{-jt}. This is a restrictive assumption in general, but is innocuous if δ^f_{jt} is set by the insurer in response to other plan's values embodied in δ_{-jt}. In such case, conditioning on δ^f_{jt-1} will not bring in extra information about the evolution of δ^f_{jt}.

We further assume that each state variable follows an AR(1) process as follows:

$$\begin{aligned} \delta^f_{jt} &= \gamma_{11m} + \gamma_{12m}\delta^f_{jt-1} + \nu_{jmt},\ \nu_{jmt} \sim \Phi(0, \sigma^2_{1m}), \\ \delta_{-jt} &= \gamma_{21m} + \gamma_{22m}\delta_{-jt-1} + \widetilde{\nu}_{jmt},\ \widetilde{\nu}_{jmt} \sim \Phi(0, \sigma^2_{2m}). \end{aligned} \tag{13}$$

For the outside option, we assume $Prob\ (\delta^f_{0t+1} = 0 \,\middle|\, \delta^f_{0t} = 0) = 1$. The above specification implies consumers on average correctly forecast the future. As mentioned earlier, we do not model the exit of a plan as a discrete event because there are very few exits observed in the data.

Estimates of the probability of exit based on the frequency will not be reliable. Instead, we include exiting plans as observations of which mean utility flows fall to zero when we estimate the above AR(1) equation, which allows for $Prob\ (\delta^f_{jt+1} = 0 \,\Big|\, \delta^f_{jt} \neq 0) \neq 0$.

Let $P^{(n)}(x', y' \mid x, y)$ denotes the transition probability distribution we estimate using $\boldsymbol{\delta}^f_{t(n)}$ and $\{\delta_{-jt(n)}\}$. Given $P^{(n)}$, we solve for $EV^{(n)}(\cdot,\cdot)$ in equation (14) through contraction mapping to obtain $EEV^{(n)}(\cdot,\cdot) = \int EV^{(n)}(x', y') \cdot dP^{(n)}(x', y' \mid x, y)$.

$$EV^{(n)}(x, y) = \ln\left[\phi \exp(y) + \exp\left(x + \beta \int EV^{(n)}(x', y') \cdot dP^{(n)}(x', y' \mid x, y)\right)\right]. \tag{14}$$

3.2 Identification and Instruments

Identification of switching costs using only aggregate market share data is a subtle matter. In this section, we describe how our moment conditions identify switching costs jointly with other demand parameters. We consider two cases. The first case assumes the unobserved product qualities ξ_{jt} are serially independent under the true data generating process (DGP). The second case allows ξ_{jt} to be serially correlated. For each case, we use distinct moment conditions.

Conceptually, the switching cost is identified through persistency in the distribution of market shares. However, market share persistency can also be generated by persistency in observed product characteristics and unobserved product qualities. To illustrate this point, first consider an example of non-identification. Suppose there are three alternatives available in period 0 and period 1. The first alternative's market share is 50% while each of the remaining alternatives has a market share of 25% in period 0. Assume that all three alternatives' market shares do not change over time and their observed product characteristics and premiums do not change. Without variation in product characteristics and market shares across time, switching costs are not identified. One cannot tell whether the first alternative remains popular in period 1 because of a high switching cost that prevents people from switching away or because its observed product characteristics and/or unobserved product quality remain desirable. Now let's modify the example so that there is variation in observed product characteristics across time. For example, suppose the first and second plan's observed product characteristics and premiums flip with each other in period 1. In this case, the persistent high market shares of product 1 cannot be explained by persistency in its observed product characteristics and premium. Instead it can be attributed to either a high switching cost or serial correlation of unobserved quality ξ_{10} and ξ_{11}. If the unobservable is known to be serially independent, we may exclude the latter explanation.

For the case of serially independent ξ_{jt}, we identify the switching cost off the fact that remaining across-time variation in flow utilities, after netting out the effect of across-time variation in product characteristics and premiums, is serially independent. Consider the mechanics of the estimation algorithm. For any given trial value of the switching cost parameter, the middle loop returns a vector of flow utilities $\delta^f_{j0}, \delta^f_{j1}, ...$ for each product that is consistent with the observed persistency in market shares. The vector of flow utilities will differ at different values of the switching cost parameter. Therefore, across-time variation in observed product characteristics and premiums—$(X_{j0}, X_{j1}, ...)$ and $(p_{j0}, p_{j1}, ...)$—will generate a different level of persistency in the unobserved product quality remainder terms, $\xi_{j0}, \xi_{j1}, ...$, at each switching cost value. At the true switching cost and linear preference parameters, the vector of recovered flow utilities must preserve the serial independence of ξ_{jt}.

Thus, we base our moment condition on the serial independence condition, $E(\xi_{jt-1}\xi_{jt}) = 0$. Because ξ_{jt-1} is unobserved, we use lagged premium, $p_{jt-1,}$ as an instrument for ξ_{jt-1}: $E(p_{jt-1}\xi_{jt}) = 0$. Note that lagged premium is orthogonal to ξ_{jt} because ξ_{jt} is not in the insurer's information set when making pricing decisions in period $t-1$, but it is correlated with ξ_{jt-1} because insurers set prices in response to contemporaneous demand shocks. We also impose orthogonality between lagged values of product characteristics X_{jt-1} and ξ_{jt}, so that across-time variation in product characteristics can help identify the switching cost parameter.

As in static models, we use contemporaneous values of the exogenous product characteristics X_{jt} as instruments for themselves to identify the β_x coefficients. For the endogenous premium coefficient, α, we use a lagged Hausmann instrument H_{jmt-1}: average prices of the same plan in other markets which are intended to reflect common cost shocks across markets. However, if the Hausmann instrument also picks up common demand shocks, contemporaneous H_{jmt} is not a valid instrument. Even so, the lagged Hausmann instrument can be valid. This is because yesterday's demand shock in market m', reflected in H_{jmt-1}, is uncorrelated with today's shock in market m by the serial independence assumption. As cost shocks have long lasting effects through labor contracts and other business commitments such as contracts with pharmaceutical manufacturers in our case, the lagged value of a variable that captures common cost shocks is expected to be correlated with the current price.

In summary we use the following moment conditions for the case of serially independent unobserved product qualities.

$$E(Z'_{jt}\xi_{jt}) = 0 \quad \text{where} \quad Z_{jt} = [p_{jt-1}, \mathbf{X}_{jt-1}, H_{jt-1}, \mathbf{X}_{jt}] \tag{15}$$

We consider another case where the unobserved qualities ξ_{jt} are serially correlated under the true DGP. In this case, across-time variation or persistency in recovered flow utilities can be explained by across-time variation in observable product attributes and serial correlation in the unobserved product qualities. The above moment conditions are no longer valid. Because $E(\xi_{jt-1}\xi_{jt}) \neq 0$ in this environment, insurers will set prices today in expectation of future demand shocks, thus $E(p_{jt-1}\xi_{jt}) = 0$ does not hold.

To sort out the effect of serial correlation in the unobservable on (across-time variation or persistency in) flow utilities, one needs information about the way ξ_{jt} is serially correlated. We assume that ξ_{jt} follows an AR1 process with $corr(\xi_{jmt-1}, \xi_{jmt}) = corr(\delta^f_{jmt-1}, \delta^f_{jmt}) = \gamma_{12m}$. This assumption leads to the condition that the resulting white noise ν_{jt+1} should be uncorrelated with ξ_{jt} at the true parameter values. Thus, we identify the switching cost off the assumption that the remaining across-time variation in flow utilities, after netting out the effect of across-time variation in product attributes and serial correlation of ξ_{jt}, is due to purely temporary demand shocks that are orthogonal to lagged unobserved product qualities.

As in the serially independent case, we use contemporaneous values of the exogenous product characteristics X_{jt} as instruments for themselves to identify the linear preference parameters β_x. For the endogenous premium coefficient, we use the lagged Hausmann instrument.

In summary we use the following moment conditions for the case of serially correlated unobserved product qualities.

$$E(\widetilde{Z}'_{jt}\xi_{jt}) = 0 \quad \text{where} \quad \widetilde{Z}_{jt} = [\nu_{jt+1}, H_{jt-1}, \mathbf{X}_{jt}] \tag{16}$$

Comparing the two cases, there are trade-offs in the identification strategy. The first case imposes the strong restriction that unobservable plan qualities are serially independent. But it has the advantage of using robust instruments. Lagged premium is a robust instrument because it is based on a weak exogeneity assumption that future demand shocks are not in the insurers' current information set. Also note that in some specifications, we include insurer and market fixed effects in X_{jt} which makes it more plausible that the residual error ξ_{jt} represents serially independent demand shocks. The second case imposes a weaker restriction that allows unobservable plan qualities to be serially correlated. But it has the disadvantage of using less robust instruments based on the assumption that $corr(\delta^f_{jmt-1}, \delta^f_{jmt}) = corr(\xi_{jmt-1}, \xi_{jmt}) = \gamma_{12m}$. The temporary demand shocks used to identify the switching cost come from the inner loop AR1 regression, which may be misspecified due to its simple parametric form. Misspecification is a particular concern when the assumption about the independence of the two state variables (as stated in equation (12)) is thought to be restrictive.

When data on switching frequencies are available, one can instead impose a micro moment condition on the switching probability predicted by the model. For example, the CMS press release on January 31, 2008 revealed that 12% of enrollees switched their plans between 2007 and 2008. To compute the average switching probability, one only needs to know choice specific values relative to the value of the outside option $\mathbf{\Delta cv}_t(sc)$. This implies we can pin down the value of the switching cost, $\widehat{sc}$, by setting the predicted average switching probability in 2008 equal to 12%. Then we would search for the fixed point $\boldsymbol{\delta}_t^f$ that is consistent with the given $\mathbf{\Delta cv}_t(\widehat{sc})$ and obtain estimates of linear coefficients on product characteristics that minimize a GMM objective function. We do not estimate our model using this idea. Instead we use this extra information to see whether the switching cost parameter estimated using the above moment conditions can produce switching probabilities similar to the actual one.

4 Data

We collected publicly available data from CMS. The data include the monthly premium, enrollment, and benefit type (basic or enhanced) for all plans offered since the inception of Part D in 2006 through 2009. They also publish statistics on the number of people eligible for Medicare. Across all years there is an average of 42 million eligible Medicare beneficiaries.

We also purchased detailed data on plan characteristics from CMS. There are four files. The formulary file lists all drugs on a plan's formulary, the beneficiary cost file describes copay/coinsurance cost sharing rates, the pharmacy network file lists all preferred and non-preferred pharmacies, and the pricing file reports average drug transaction prices for every drug and plan. The pricing file was first published in 2009, the other files are available in all years. We've converted all dollar values into real terms, expressed in 2006 dollars.

4.1 Market Shares and Premiums

Across the four years, 102 insurers offer stand-alone Part D plans. The average insurer offers plans in 9 of the 39 regions in the US and its territories. We use only the 34 regions in the continental US for estimation. The size distribution of insurers is quite skewed: each year an average of 10 insurers offer plans in all 34 regions, while about a half offer plans in just one region. Consumers have many plans to choose from. The average region includes 48 plans offered by 21 insurers. Insurers typically offer 1, 2, or 3 plans in each region. At least one must be a basic plan. Table 8 shows the average number of insurers and plan offerings per market over the sample period. The number of plans tended to increase in the first three years until it fell in 2009. Basic plans outnumbered enhanced plans in a typical market until 2007.

Table 9 presents the average monthly premium in 2006 dollars for each year during the sample period. The premium averages $38.2 across regions and years. These premium figures are net of subsidies. The average monthly premium fell in the second year, but experienced a moderate increase in 2008 and a 15% increase in 2009. The large variance in the monthly premium is largely due to the presence of two segments in the benefit structure; the average monthly premium for basic plans is $30.0 with a standard deviation of $9.6, and it is $46.6 with a standard deviation of $18.7 for enhanced plans. The second panel in Table 9 reports the market share weighted average premium across the 34 regions for each year. The mean increased gradually, from $26.88 to $32.48, over the sample period. The fact that the weighted average premium is lower than the simple average premium implies that more popular plans charge lower monthly premiums than less popular plans. It is interesting that the market share weighted premium rose in 2007, yet the non-weighted average premium fell. One possible interpretation is that many enrollees remained in their default plans even though they were faced with premium increases.

The first panel in Table 10 reports statistics on market-level enrollment rates in stand-alone Part D plans as a percentage of all Medicare eligible beneficiaries. The average enrollment rate is 34.4% in all years. Some markets have a low enrollment rate around 18% and the maximum ranges from 55% to 59%. The second panel reports national-level enrollment rates. National enrollment in stand-alone Part D plans remained stagnant across years. CMS reports about 90% of Medicare beneficiaries have drug coverage from a stand-alone Part D plan or another source. Along with non-enrollment, these other sources compose the outside option in our model and include bundled MA+Part D plans, RDS plans, and plans from other government programs. The final column in the table indicates enrollment in MA+Part D has increased over the years.

Table 11 reports summary statistics on plan-level market shares. An average plan's market share is around 0.77%. This small figure is consistent with a market level enrollment rate of 38% and a large number of plans in an average market, 48. Markets shares are quite skewed. The market share of the most popular plan in a region averages 7.6% across markets and years. Some garner particularly high market shares of around 16%. At the insurer level, United Healthcare is the market leader, capturing around 18.26% of enrollees. It offers a plan named AARP MedicareRx Preferred, which is the market leader in 24 markets on average over the four year period. Humana Insurance Company is a close runner-up with an 18.24% share of national enrollment, although its plans are market leaders in only 4 markets on average.

Some plans leave the market overtime. We supplement the enrollment data with information in the "Crosswalk" data file to identify whether these plans were consolidated with another plan or were truly terminated. When a plan consolidates into another plan, enrollees are automatically transferred to the other plan. This requires a slight modification to the computation of the past markets shares of consolidating plans. For example, if plan k is consolidated into plan j in year t, $\widetilde{M}_{jt-1} = M_{jt-1} + M_{kt-1}$ and we use $\widetilde{M}_{jt-1}$ as the endowed past market share that determines the number of consumers whose default option is plan j.[17]

Some plans, though not many, truly exit the market. If a plan exits the market in year t, its enrollees from year $t-1$ are not automatically enrolled in another plan. To obtain coverage, those enrollees would have to actively choose and purchase another plan. Therefore, we designate their default option to be the outside option. This requires a modification to the lagged market share of the outside option: $\widetilde{M}_{0t-1} = 1 - \sum_{j \in \mathbf{J}m_t \setminus \{0\}} \widetilde{M}_{jt-1}$.[18] Table 12 presents the number of terminated plans and their average market shares for each year. As shown, exits do not occur frequently, and exiting plans usually have market shares much smaller than the market average.

To estimate the survivor rate $(1-\rho)$, we use the number of deaths and the death rates among the population over age 65 in 2006 published by the Center for Disease Control (CDC).[19] The estimated survivor rate $(1-\rho)$ is 95.3%.

4.2 Plan Characteristics

We follow the cost sharing structure laid out in the Part D regulations to construct variables that measure the generosity of plans' coverage. The regulations set a minimum coverage standard. Figure 1 depicts this basic benefit structure. It describes the out-of-pocket price an enrollee pays to fill a prescription as a function of the enrollee's annual drug expenditures. The price schedule can be thought of as a five part tariff. The first region is the premium; it is paid regardless of drug expenditures. Next, an enrollee pays 100% of drug expenditures out of pocket until he reaches the deductible threshold ($250 in 2006). For the next $2,000 of expenditures, coverage begins, and the enroll pays 25% out of pocket. This is called the initial coverage region. After that, coverage ends, and the enrollee pays 100% of expenditures. This is the so-called donut hole region. Finally, for very high expenditures above $5,100, coverage resumes with an out-of-pocket price of just 5%. This is called catastrophic coverage. Medicare categorizes plans that meet (or are actuarially equivalent to) the minimum standard as basic plans. Enhanced plans typically offer some combination of a lower deductible, lower out-of-pocket costs in the initial coverage zone, or added coverage in the donut hole.

The regulations give plans a lot of scope in selecting what drugs to include on their drug

[17] We also incorporate this crosswalk information in estimation of the transition probabilities of the two state variables. Continuing with the example, the consolidation event will give two separate observations, $(\delta^f_{jt-1}, \delta^f_{jt})$ and $(\delta^f_{kt-1}, \delta^f_{jt})$, to be included in the estimation of the transition probability distribution of the utility flow.

[18] When plan l exits the market, it generates observations of $(\delta^f_{lt-1}, 0)$ and $(\delta_{-lt-1}, \delta_{-0t})$ to be included in the estimation of the transition probability distributions of the two state variables.

[19] CDC statistics retrieved from http://www.disastercenter.com/cdc/Table_3_2006.html.

formularies and in setting copay/coinsurance rates and usage restrictions on a drug-by-drug basis. This flexibility is intended to give plans leverage in negotiating discounts and rebates with drug manufacturers and wholesalers. The regulations require plans to include some drugs from all of the major therapeutic classes, but plans have the discretion to choose exactly which drugs to include. Indeed, some plans offer almost the entire universe of Part D drugs, and some restrict the set of covered drugs.[20] For each drug on the formulary, plans can set usage restrictions in the form of quantity limits, prior authorization, and step therapies. Plans set copay/coinsurance rates by placing drugs on pricing tiers such as "preferred," "non-preferred," and "specialty." Higher tiered drugs have higher copay/coinsurance rates. To meet the basic benefit structure, Medicare must approve the cost sharing rules as being "actuarially equivalent." Despite this requirement, idiosyncratic differences in cost sharing rules and negotiated drug prices create a lot of variation at the aggregate level, particularly amongst the most popular drugs.

We construct three plan-level variables to measure the generosity of coverage using the detailed drug-level information purchased from CMS on formularies, copay/coinsurance rates, and drug prices. The first variable is the deductible. The regulations set the maximum annual deductible to \$250 in 2006 and it rose to \$295 (nominal) in 2009. Table 15 reports summary statistics.

The second and third coverage characteristics are intended to measure coverage generosity in the initial coverage zone and donut hole. For each zone, we construct price indices of out-of-pocket prices paid by enrollees for the top 100 most popular drugs ranked by prescriptions filled. The first price index measures the out-of-pocket cost for an enrollee to fill a 30 day supply for a basket of the 100 drugs when he is in the initial coverage zone. The second index measures out-of-pocket costs in the donut hole. The basket of drugs evenly weight each drug (1/100th).

Table 13 reports summary statistics on the number of top 100 drugs included on formularies. The average plan covers about 90 of the 100 drugs. In later years, plans have included slightly fewer drugs on average, and the standard deviation has increased. This may indicate plans are becoming more differentiated with respect to formulary coverage. The table also reports the number of unique formularies. There are fewer formularies than plans because insurers often share formularies across their plan offerings.

Constructing out-of-pocket costs is straightforward for drugs covered by copay; the copay is directly available in the beneficiary cost file. For drugs covered by coinsurance, it is necessary

[20] Medicare's formulary review board determines the set of drugs that are considered Part D drugs and classifies them into therapeutic categories. Some drugs, such as prescription sleep aids, are not considered Part D drugs. Plans may include non-Part D drugs, but then the plan is classified as enhanced and the cost of offering additional drugs is not subsidized.

to know the negotiated price of the drug. We use the 2009 pricing file. For drugs excluded from the formulary, enrollees do not receive coverage; therefore, the out-of-pocket cost is the full retail price.[21] We set the retail price to the average price in the region.[22] In summary, there are three sources of variation underlying the price indices: copay/coinsurance rates, negotiated drug prices, and formulary comprehensiveness. Table 14 reports summary statistics on the out-of-pocket price index variables. Average prices are higher in the donut hole because many basic plans offer zero coverage. There is a lot of variation across plans, particularly in the initial coverage zone. Much of the variation stems from variation in formulary comprehensiveness.

Pharmacy networks introduce another dimension of horizontal product differentiation. Our final plan characteristic (pharm per eligible) measures the number of network pharmacies per eligible Medicare beneficiary in a plan's region. Table 15 reports summary statistics.

To capture any additional coverage benefits for enhanced plans that are not captured by the above variables, we include a dummy variable (BenefitEA dummy) that takes on the value 1 if the plan is an enhanced plan. An example of an additional benefit could be the inclusion of a non-Part D drug. Table 16 presents summary statistics on this variable in the first panel. This variable is also important because enhanced plans are not eligible for auto-enrolled LIS beneficiaries. While the share of enhanced plans was less than 50% in the first year of the program, it increases over the sample period and becomes the dominant type of plan in all 34 markets in 2009.

We include another dummy variable (LIS dummy) that takes on the value 1 if the plan is eligible to receive auto-enrolled LIS beneficiaries. This variable is possibly endogenous because only basic plans with a price set below a threshold (unknown a priori) become eligible. According to table 16, 29 percent of plans were LIS-eligible in 2006. This figure increases to 36% in 2007, but then drops such that only 18% of plans are LIS-eligible in 2009. Since our market shares include automatically assigned enrollees, we expect the coefficient on this dummy to be economically and statistically significant.

Table 17 presents summary statistics on year-to-year changes in these covariates. The variation in these changes is large. While the monthly premium on average increases by $3 every year, the standard deviation of the changes is $11. On the extremes, there is a plan that experienced a premium decrease of $50 and another plan that experienced an increase of $75. The out-of-pocket drug price indices decline on average, but many plans experienced a sharp

[21] To fill a prescription for an off-formulary drug, an enrollee would have to purchase it on his own or in some cases the enrollee may be able to go through an appeals process with the insurance company.

[22] We use the average national price in rare cases where a region price does not exists. For the years 2006 to 2008 we construct the price indices in the same manner using 2009 prices. For plans that did not exists in 2009, we use average regional prices. Drug prices, coinsurance rates, and copays differ across preferred, non-preferred, and mail order pharmacies. All of our calculations are based on preferred pharmacies.

increase. For most variables, the standard deviation of year-to-year changes is larger than the average. This across-time variation is important for identifying the switching cost parameter in our model because changes in plan characteristics are a driving force in an enrollee's decision to switch plans.

Table 18 presents estimation results of AR(1) processes for each product characteristic. Based on the AR(1) coefficient estimates, the process is stationary for all variables. The goodness of fit for the AR(1) models, measure by R^2, is reasonably high, yielding R^2 of 0.6 to 0.8 for most variables.

4.3 Preliminary Analysis

Before estimating the dynamic model, we present reduced form evidence that documents persistence in the market share data. Table 1 reports estimates from various OLS regressions of log market shares on product attributes. In 2006, all enrollees were new to the Part D program and were not locked-in to a default plan. All consumers who decided to enroll would choose a plan that offers the highest expected value of utility flows. However in subsequent years, once consumers became locked-in, switching costs make it disadvantageous to enroll in any plan besides the default plan. Consumers would only switch to another plan if the expected value of utility flows is high enough to justify the switching cost. Therefore, if premiums and product characteristics change across years, static demand estimates should appear as if people were less sensitive to premiums and plan characteristics in 2007-2009 than the initial year 2006. Comparing the coefficients on premium in the first two columns of table 1 indicates that consumers appeared more price sensitive in 2006 than the later years. Although this is not affirmative evidence that there are switching frictions in the market, it shows that the environment consumers faced in 2006 differs from that in later years.

We then examine how much a plan's current market share is affected by the past market share by regressing a plan's market share on its lagged market share and demand controls. If there exists a cost associated with switching plans, the data will exhibit strong dependence of present market shares on past market shares. We report the results in the last three columns. Our results show a large positive impact of lagged market shares on current market shares. The effect diminishes, but is not eliminated, when we add insurer and plan fixed effects. Because the fixed effects, particularly plan fixed effects, help control for serially correlated unobserved product qualities, these results suggest that the state dependence in demand is likely structural (caused by switching frictions) rather than spurious (an artifact of serially correlated unobservables).

Table 1: Preliminary OLS Regression Results

sample period	2006	2007-9	2007-9	2007-9	2007-9
lagged log share			.7251*	.6069*	.3307*
			(.0055)	(.0066)	(.0141)
Premium	-.0606*	-.0256*	-.0219*	-.0216*	-.0164*
	(.0041)	(.0021)	(.0009)	(.0010)	(.0012)
Deductible/12	-.0143*	-.0461*	-.0056*	-.0055*	.0037
	(.0050)	(.0033)	(.0015)	(.0015)	(.0036)
BenefitEA dummy	.5428*	-.2206*	.2693*	.2488*	-.0299*
	(.1008)	(.0752)	(.0317)	(.0307)	(.0437)
Initial coverage PI	-.0112*	-.0436*	-.0068*	-.0229*	-.0199*
	(.0047)	(.0024)	(.0010)	(.0017)	(.0016)
Donut Hole PI	-.0899*	.0036	-.0339*	-.0226*	-.0394*
	(.0078)	(.0066)	(.0029)	(.0033)	(.0052)
pharm per eligible ($\times$1000)	.7771*	1.4064*	.1589*	.2542*	-1.0432*
	(.1743)	(.1218)	(.0536)	(.0510)	(.2758)
LIS dummy	2.3470*	1.8981*	1.0587*	1.2867*	.9144*
	(.1194)	(.0721)	(.0315)	(.0303)	(.0321)
Insurer FE	N	N	N	Y	N
Plan FE	N	N	N	N	Y
N obs	1415	5139	4270[a)]	4270	2889[b)]
Adjusted R^2	0.5101	0.2769	0.8700	0.8982	0.9450

[a)]: Only plans with lagged market shares are included. [b)]: Only plans that existed for the whole sample period (2006-2009) are included.

5 Estimation Results

In this section, we present the estimation results. Table 2 reports the estimates of the switching cost and coefficients on product attributes based on the moment conditions for the case of serially independent unobserved product qualities. We consider 6 specifications that differ with respect to the inclusion of insurer and region fixed effects as well as instrumenting strategies for the LIS dummy variable. Specifications group I includes the BenefitEA dummy as a demand control whereas group II uses it as an instrument for the LIS dummy.[23]

Across the 6 dynamic model specifications, the switching cost estimate ranges from 6.43 to 7.17. These switching cost levels predict an average switching probability that ranges from 6.2% to 7.9%. The switching cost expressed in dollar terms (converted using the premium coefficient) ranges from $2,366 to $2,697.

[23]The LIS dummy is potentially endogenous because a plan's LIS status is determined by the plan's premium relative to a weighted average of other plans' premiums. The BenefitEA dummy can serve as an instrument for the LIS dummy because enhanced plans are not LIS eligible regardless of the premium set by insurers. The BenefitEA dummy variable may be excluded from the demand equation because other demand controls (deductibles, donut hole prices, etc) already control for the difference between enhanced and basic plans.

Table 2: Case A: Serially Independent Unobserved Qualities

Specification	I-1	I-2	I-3	II-1[a)]	II-2	II-3	Static II-3	Alt II-3
Switching Cost (SC)	7.00*	6.43*	6.41*	7.17*	6.54*	6.51*		3.75*
	(0.1389)	(0.1324)	(0.1326)	(0.1673)	(0.1476)	(0.1472)		(0.2015)
Dollar Value of SC	$2366	$2651	$2473	$2538	$2697	$2464		$1485
Ave. Switching Prob.	6.23%	7.83%	7.90%	5.86%	7.47%	7.57%		40.35%
Premium	-0.0355*	-0.0291*	-0.0311*	-0.0339*	-0.0291*	-0.0317*	-0.0476*	-0.0303*
	(0.0019)	(0.0019)	(0.0019)	(0.0020)	(0.0019)	(0.0020)	(0.0027)	(0.0023)
Deductible/12	-0.0246*	-0.0161*	-0.0164*	-0.0171*	-0.0119*	-0.0116*	-0.0354*	-0.0194*
	(0.0020)	(0.0020)	(0.0019)	(0.0025)	(0.0023)	(0.0023)	(0.0032)	(0.0027)
BenefitEA dummy	0.3984*	0.2422*	0.2232*					
	(0.0473)	(0.0417)	(0.0409)					
Initial coverage PI	-0.0160*	-0.0120*	-0.0122*	-0.0107*	-0.0093*	-0.0097*	-0.0426*	-0.0021
	(0.0014)	(0.0018)	(0.0018)	(0.0017)	(0.0019)	(0.0019)	(0.0023)	(0.0022)
Donut Hole PI	-0.0480*	-0.0375*	-0.0401*	-0.0556*	-0.0393*	-0.0428*	-0.0487*	-0.0336*
	(0.0049)	(0.0043)	(0.0042)	(0.0049)	(0.0044)	(0.0044)	(0.0055)	(0.0053)
pharm per eligible (×1000)	1.0212*	0.8887*	0.6982*	1.1603*	0.9476*	0.8499*	0.5571*	1.2280*
	(0.0656)	(0.0643)	(0.1474)	(0.0816)	(0.0662)	(0.1619)	(0.1913)	(0.1837)
LIS dummy	1.6202*	1.6830*	1.6226*	0.8210*	1.1438*	1.0688*	2.3562*	1.5280*
	(0.0476)	(0.0434)	(0.0445)	(0.0832)	(0.0717)	(0.0742)	(0.1004)	(0.1023)
Insurer FE	N	Y	Y	N	Y	Y	Y	Y
Region FE	N	N	N	N	N	Y	Y	Y

a): Specification 2 uses the BenefitEA dummy as an instrument for LIS.

If plan j in period t did not exist in $t-1$ or if $t = 2006$, p_{jt-1}, x_{jt-1}, and H_{jt-1} are set to be zero.

Table 3: Case B: Serially Correlated Unobserved Qualities

Specification	I-1	I-2	I-3	II-1[a)]	II-2	II-3
Switching Cost (SC)	6.95*	6.15*	6.12*	7.01*	6.18*	6.15*
	(0.2464)	(0.3120)	(0.3154)	(0.2672)	(0.3265)	(0.3303)
Dollar Value of SC	$1283	$1488	$1684	$1135	$1498	$1689
Ave. Switching Prob.	6.34%	8.89%	9.02%	6.21%	8.77%	8.89%
Premium	-0.0650*	-0.0496*	-0.0436*	-0.0741*	-0.0495*	-0.0437*
	(0.0083)	(0.0238)	(0.0231)	(0.0092)	(0.0240)	(0.0231)
Deductible/12	-0.0246*	-0.0282*	-0.0256*	-0.0134*	-0.0235*	-0.0223*
	(0.0027)	(0.0069)	(0.0064)	(0.0040)	(0.0062)	(0.0059)
BenefitEA dummy	0.3744*	0.1327*	0.0972			
	(0.0605)	(0.0620)	(0.0564)			
Initial coverage PI	-0.0101*	-0.0058*	-0.0052	-0.0132*	-0.0055	-0.0050
	(0.0020)	(0.0029)	(0.0028)	(0.0022)	(0.0029)	(0.0028)
Donut Hole PI	-0.1009*	-0.0766	-0.0673	-0.1141*	-0.0769*	-0.0679*
	(0.0137)	(0.0366)	(0.0356)	(0.0150)	(0.0370)	(0.0358)
pharm per eligible ($\times$1000)	1.1803*	1.0797*	1.0487*	1.3431*	1.1206*	1.1231*
	(0.0995)	(0.1656)	(0.2717)	(0.1114)	(0.1792)	(0.2983)
LIS dummy	1.2691*	1.5617*	1.5327*	0.2474	1.2190*	1.2782*
	(0.0998)	(0.1417)	(0.1549)	(0.1864)	(0.2571)	(0.2509)
Insurer FE	N	Y	Y	N	N	Y
Region FE	N	N	Y	N	Y	Y

a): Specification II uses the BenefitEA dummy as an instrument for LIS.
If plan j in period t did not exist in $t-1$, H_{jt-1} is set to be zero. If plan j in period t exited the market in $t+1$, ν_{t+1} is set to be zero. All 2009 plans are excluded.

Table 3 reports the parameter estimates based on the moment conditions for the case of serially correlated unobserved qualities. The switching cost estimate ranges from 6.12 to 7.01, slightly lower than the range found for the serially independent case. The implied switching probability varies between 6.2% and 9.0%, which is higher than the serially independent case. The switching cost expressed in dollar terms (converted using the premium coefficient) ranges between $1,135 and $1,689, which is much lower than the previous case. The dollar values are lower primarily because the premium coefficient is larger in magnitude, ranging between -0.044 and -0.074, as opposed to a range of -0.029 to -0.036 for the serially independent case. The intuition behind the different switching cost estimates found for the two cases follows from the identifying assumptions. In the serially independent case, the model attributes all of the persistency in market shares that remains after accounting for persistency in observables to switching costs, whereas the serially correlated case attributes some of the persistency to the serial correlation of unobserved plan qualities, which tends to lower the switching cost estimate.

The switching cost estimates are large and economically significant regardless of which case represents the true data generating process. It is useful to put the dollar-valued switching cost into perspective. According to a 2007 Medicare Payment Advisory Commission (MEDPAC) report, the average annual prescription drug spending per enrollee was \$2,790 and the average out-of-pocket drug spending was \$467.[24] The average annual premium was \$426. Based on the estimation results from the serially correlated case, the dollar-valued switching cost estimates are around 40% to 60% of the annual drug spending per enrollee and around 130% to 190% of the average total Part D spending (out-of-pocket drug spending plus premiums).

Our results are more or less consistent with previous studies on health care plans. Handel (2010) estimates switching costs amongst employer-provided health plans using individual-level data. The average switching cost is estimated around \$2,000, which is around 50% of the average total employee health spending. He admits, given that he estimates switching costs among PPO plans that are differentiated only by financial characteristics, his estimate can be viewed as a lower bound on switching costs in a more general setting, like ours, where switching costs can stem from having to adapt to different non-financial characteristics, such as pharmacy networks and drug formularies. Also, his estimate is based on choices made by people of working age. Several studies have predicted or indirectly shown older people are likely to have higher switching costs.[25]

Most of the coefficients on the product characteristics have expected signs and are statistically significant at the 5% level. The results are similar across the two cases and 6 specifications. The only large difference across specifications comes on the LIS dummy which depends on whether or not BenefitEA is used as an instrument. Consumers prefer more generous coverage (lower deductibles and out-of-pocket prices in the initial coverage zone and donut hole) and more extended pharmacy networks. Medicare beneficiaries appear to be more sensitive to a \$1 increase in the premium than in the deductible, which is sensible given that there is uncertainty associated with meeting the deductible. This is also consistent with the Abaluck and Gruber (2011)'s estimation results based on individual data on plan choices and drug utilization. The coefficients on the price indices are both negative, more so for the donut hole price index. The LIS dummy shows that LIS eligible plans experience a large shift in demand from receiving randomly assigned individuals.

[24] MEDPAC, "A Data Book: Medicare Part D program," March 2010. Dollar value figures deflated to 2006 dollars.

[25] Strombom et al. (2002) relates the negative relationship between age and price sensitivity they found to a possibly positive relationship between age and switching costs. Handel (2010) finds switchers are slightly younger than stayers.

5.1 Comparison with A Static Demand Model

We present the coefficient estimates for a static demand model in the second to the last column of table 2. As static models are misspecified in this environment, it is useful to compare static and dynamic results to understand the direction and size of the bias. This is particularly important in applications where supply-side mark-ups are based on estimates of price elasticities. The static model is estimated by the method described in Berry (1994) using the same moment conditions as the dynamic model with serially independent plan qualities.

Compared to the dynamic model results, consumers appear to be more price sensitive (higher in magnitude price coefficient) in the static model. This may seem counter-intuitive at first glance. Suppose a plan experiences a small drop in its market share when its premium increases. Static models that ignore state dependence would recover the present year's mean utility level only slightly lower than the past year's and would hence deduce a low price sensitivity. Explaining this small drop in the market share in a dynamic setting with high switching costs may require the present year's utility level to be much lower than the past year's. As such, one may expect a high price sensitivity with a dynamic model that incorporates state dependence.

The above intuition, however, only considers the fact that switching costs give rise to state dependence in consumers' choices, and does not account for the forward-looking nature of consumer behavior. It is important to recognize that it is the choice specific value that the dynamic model recovers. How this value affects today's flow utility level depends on flow utility transition probabilities. In our application, the estimated transition probability distribution implies that the current and future utility flows are positively related. Consequently, a drop in today's flow utility gets amplified through the negative effect on the continuation value. That is, when a plan experiences a small drop in its market share when its price increases, the dynamic model would recover a large drop in the choice specific value. But, that drop could separate out into a relatively small drop in the current year's flow utility and a relatively large drop in the continuation value.

A properly specified dynamic model disentangles the effect of today's price change on today's flow utility from its effect on the expected utility flows. Because a static demand model does not subtract out the latter effect, it may over- or under-estimate the marginal effect of a price change on flow utility. In our case, price sensitivity is biased up under the static model specification. In the opposite case where the current level of flow utility is negatively related to expectation about future utility flows, the static model specification is likely to under-estimate the price sensitivity. Note that in a world with switching costs and myopic consumers, there is no need to account for the continuation value, and price sensitivity estimates from a static specification without switching costs would be unambiguously biased downwards.

5.2 Robustness

To evaluate whether the parameter estimates of the switching cost are reasonable, we compare predicted switching probabilities to actual ones. Recall that the average switching probability was 12% between 2007 and 2008. This 12% switching figure includes low income subsidy (LIS) beneficiaries who were randomly reassigned after their previous plans became ineligible for the full low income subsidy. The average switching probability among non-LIS beneficiaries is only 6% implying significant choice inertia.[26]

We do not model choices made by LIS beneficiaries subject to auto-enrollment differently from those made by the non-LIS population because they are also allowed to freely choose plans and because we do not have data on the population of LIS beneficiaries for each market.[27] Whereas the aggregate market share data used for estimation reflect the 12% switching frequency and changes in market shares induced by reassignment, our model assumes all switches are voluntary and subject to the same level of switching costs. This discrepancy implies that the estimated model may not be able to generate an average switching probability of 12% or of 6%. Indeed, the implied switching probabilities from our estimates for the serially correlated case lie in the middle, around 8%. Because the market shares used for estimation are less persistent than they would otherwise be with a sample that includes only voluntary switches amongst non-auto enrollees, we expect our model's predicted switching frequency to be higher than 6%. On the other hand, the fact that our model requires everyone to pay the switching cost—including those automatically reassigned—tends to suppress the predicted switching probability because it would associate the observed changes in market shares with more stays and fewer switches than a model that incorporates random reassignment for the subset of auto-enrollees. Hence, the predicted market share is likely to be lower than 12%. As such, the model's implied switching frequency of 8% compares favorably to the benchmark figures from the 2008 micro data. This suggests that our model and estimation algorithm based on aggregate market share data performs fairly well with respect to replicating switching frequencies. It also implies that the switching cost estimate reflects the average value of switching costs that auto- and non-auto enrollees are facing in this market.[28]

[26] This is based on the figure given in the Center for Medicare and Medicaid Services Press Release, January 31, 2008: "Medicare Prescription Drug Benefit's Projected Costs Continue to Drop." (http://www.cms.gov/apps/media/press/release.asp?Counter=2868&intN)

[27] The switching cost for a randomly reassigned beneficiary would be zero if all that mattered were the financial and psychological costs of finding an alternative plan. But it could be positive for other reasons, such as the costs associated with adapting to the new plan's drug formulary. As a dual eligible must take an action if she wants to stay with the current plan, the switching cost may be negative for the auto-enrollment eligibles.

[28] We account for the effect of random assignment on market shares by including a dummy variable for plans that are eligible to receive randomly assigned LIS beneficiaries. The LIS dummy (across-time changes in the values) makes sure that some switches are because of changes in observable characteristics rather than changes

In the last column in table 2, we present tentative estimates of the dynamic model where we assume switchers from the outside option also pay the switching cost. The presented estimates are only illustrative because the middle loop converges to different values of the mean utility flow vector for different initial guesses. For each value of the switching cost, we solve the middle loop for 12 different initial guesses, and then evaluate the GMM function for each resulting mean utility flow vector. Although there was some trend, for example, the GMM values tend to be smaller around a switching cost value of 3.7 than around, say, 4.0, some initial guesses result in higher GMM values than others. We present the estimates that obtained the smallest GMM value of all tested initial guesses and switching cost values. Although illustrative only, the switching cost estimate under this model specification is much smaller both in the parameter value (3.75) and in dollars ($1,485). This result is expected because the model requires the level of switching costs to be small enough to rationalize the massive influx from the outside option in the inception year 2006.

5.3 Dynamic Price Elasticities

In this subsection, we present various price elasticities calculated based on specification II-3 of the serially correlated case. An increase in the premium affects the market share of the plan through three channels: 1) it lowers its own choice specific value by decreasing the current utility flow; 2) it further lowers the continuation value through the estimated transition probabilities; 3) it reduces other plans' choice specific values by lowering their continuation values through decreases in the non-default option values.

Note that the effect of the second and the third channel highly depends on the estimated transition probability distributions. Table 4 presents the AR(1) process estimates of the two state variables, utility flow (δ^f_{jt}) and non-default option value (δ_{-jt}) for specification II-3 of the serially correlated case. Both of the processes are stationary and exhibit positive autocorrelation. The positive autocorrelation of the utility flow variable confirms our previous argument on how consumers could appear to be more price sensitive in misspecified static models. The fact that the autocorrelation of δ^f_{jt} is positive for all markets implies the expected value function $EV(\delta^f_{jt}, \cdot)$ is an increasing function in δ^f_{jt} for all markets. Similarly, the positive autocorrelation of δ_{-jt} causes the continuation value function $EEV(\cdot, \delta_{-jt})$ to be increasing in δ_{-jt}.

Table 5 presents own price elasticities. In the table, Dyn-1 gives the price elasticity when only the first channel is accounted for, which is equivalent to a case where consumers believe the price shock is purely temporary. Dyn-2 accounts for the first two channels, and Dyn-3

in the unobservable or small switching costs. Therefore, without the LIS dummy variable, the given changes in the market shares are likely to be rationalized by a smaller switching cost estimate.

Table 4: Estimation Result: AR(1) Process of the State Variables

	δ^f_{jt}				δ_{-jt}			
	mean	s.d.	min	max	mean	s.d.	min	max
Number of observations	157.26	16.07	106	198	160.26	16.08	109	201
Constant	-.338	.123	-.663	-.120	7.969	3.848	.190	17.620
AR(1) coefficient	.446	.052	.360	.618	.571	.149	.258	.950
σ	.864	.138	.678	1.220	.563	.072	.429	.725
R^2	.349	.073	.208	.534	.114	.032	.043	.183

for all three channels. While the static model that assumes zero switching cost yields similar price elasticities for both 2006 and later years, the dynamic switching cost model yields much higher price elasticities for 2006 than the later years. The reason why Dyn-1 yields elasticities lower than the static model is that the premium coefficient estimate is slightly smaller for our dynamic model (0.0437) than for the static model (0.0476). The dynamic model tends to yield price elasticities much larger than the static model when the dynamic channels are accounted for. The table shows the dynamic channels, particularly the second channel, are important in determining the price elasticity. The third channel brings down price elasticities compared to the case where consumers do not adjust their prospects of the other plans (Dyn-2), but the effect is minimal. As we argued above, static models do not differentiate the first and second channel, and hence cannot analyze how a policy that changes expectations about the market environment, such as regulations that limit radical changes in monthly premium and other plan characteristics, will affect market shares over time.

Table 6 presents lagged own price elasticities. In our model, an increase in a plan's premium

Table 5: Own Price Elasticity

		obs	median	mean	s.d.	min	max
$t = 2006$	Static	1415	1.7062	1.7617	0.6080	0.0816	4.6011
	Dyn-1		1.5664	1.6173	0.5581	0.0749	4.2241
	Dyn-2		2.3954	2.4917	0.8772	0.1885	10.1945
	Dyn-3		2.3742	2.4723	0.8709	0.1791	10.014
$t \geq 2007$	Static	5139	1.6016	1.8198	0.8580	0.4218	6.2785
	Dyn-1		1.2343	1.4047	0.8053	0.1032	5.7637
	Dyn-2		1.7640	1.9565	0.9512	0.2412	8.3334
	Dyn-3		1.7591	1.9504	0.9525	0.2361	8.3120

Dyn-1 considers a case where a change in the premium of plan k affects neither expectation about the future values of the mean utility flow of plan k nor the non-default option values of the other plans while Dyn-3 is where it affects both of them and also expectation about the future values of the non-default option values. Dyn-2 is when it affects only affects the expectation about the plan's stream of utility flows but not the other plans' non-default option values.

Table 6: Lagged Own Price Elasticity

		$t+1$		$t+2$		$t+3$	
		obs	mean	obs	mean	obs	mean
$t = 2006$	Dyn-1	1198	0.6128	1103	0.3643	963	0.2088
	Dyn-3		1.1612		0.7379		0.4348
$t \geq 2007$	Dyn-1	3072	0.3881	1361	$0.1641^{\dagger}$		
	Dyn-3		0.6255		$0.2974^{\dagger}$		

† There we found three incidents with positive lagged own price elasticity.

in year t will affect the plan's market share not only in that year but also all the subsequent years. A premium increase in 2006 has considerable effects on the market share of the same plan in the following years, but the effects gradually decrease over time. These effects are of course greater when the increase alters consumers' expectation about the future premium (Dyn-3) compared to the case where the expectation is fixed (Dyn-1). The effects of a 1% premium increase in 2007 on the future market shares are much smaller than, almost a half of, the effects from a 1% premium increase in 2006. The lagged own price elasticity does not always have to be negative because an increase in a plan's premium is likely to increase other plans' market shares in that period, which in turn could increase the pool of potential switchers. In fact, in three out of 1,361 cases, a premium increase leads to an increase in the market share two years later, although the magnitude is on the order of 1/10000.

Table 7 presents contemporary and lagged cross price elasticities. The mean cross price elasticity is slightly higher for the dynamic model than the static model in 2006. As expected, the dynamic model yields contemporary cross price elasticities much lower in 2007 compared to 2006. While contemporary cross price elasticities are all positive in 2006, some are negative in 2007, albeit only 224 out of over 256,000 cases. This occurs through the third channel: if a plan's premium increase lowers the other plans' choice specific values by decreasing their non-default option values, some plan that is most negatively affected can experience a drop in its market share. This shows how competing products can be complements in a dynamic setting.

The mean lagged cross price elasticities decrease as the time lag increases. While the majority of lagged cross price elasticities are still positive, there are many pairs of plans that exhibit negative values of lagged cross price elasticity. This occurs when plan A's premium increase results in a decrease in plan B's contemporary market share (through the third channel) or a decrease in the pool of potential switchers for plan B. The share of pairs of plans with negative lagged cross price elasticities increases as the time lag increases.

Table 7: Cross Price Elasticity

		t		$t+1$		$t+2$		$t+3$	
		mean	neg (%)	mean	neg (%)	mean	neg (%)	mean	neg (%)
$t = 2006$	Static	0.0116	0						
	Dyn-1	0.0106	0	0.0070	9.3	0.0046	20.4	0.0031	29.7
	Dyn-3	0.0196	0	0.0138	21.4	0.0091	32.9	0.0061	40.0
$t \geq 2007$	Static	0.0101	0						
	Dyn-1	0.0047	0	0.0030	3.0	0.0018	13.1		
	Dyn-3	0.0081	0.1	0.0056	9.7	0.0035	21.7		

neg(%) gives the share of pairs of plans with negative cross price elasticities.

5.4 Simulation: Welfare Loss

Abaluck and Gruber (2011) present a puzzle for this market that documents enrollees making inconsistent choices. Using enrollee level data on Part D plan choices in 2006, they find that the average enrollee chooses a plan such that he overpays by 30.9 percent of the average total Part D spending on premiums, deductibles and copays versus the lowest-cost plan. In dollar terms, the average enrollee overpays by about \$300 based on the average total Part D spending of around \$900. We attempt to estimate how much of this can be explained by enrollees' efforts to avoid switching costs in future years.

We simulate the choices of Part D eligibles based on model II-3 of the serially correlated case. We compute the differences in utility levels between the plans chosen by forward-looking consumers, who face a switching cost of \$1,689, and plans they would choose if there were no switching costs. In the initial year, 2006, the average welfare loss is \$5.1, which increases to \$465 for 2007, and around \$480 for the years 2008 and 2009. The loss in the initial year is small because the plan with the highest static flow utility is the same as the plan with the highest choice specific value for most simulated consumers. The losses become significantly larger in later years as many enrollees choose to stick to their default plans as long as the present value of the momentarily best plan is not sufficiently higher than that of the default plan to justify paying the switching cost.

The welfare loss caused by switching costs in the years after 2006 is on the same order of magnitude as the results in Abaluck and Gruber (2011). However, it should be noted that the result in Abaluck and Gruber (2011) is based on 2006 choices before enrollees became locked-in, and our simulated welfare loss due to forward-looking behavior in that year is very small. Thus, forward-looking behavior does not resolve the choice inconsistency puzzle, and the behavioral and informational explanations examined in their study remain valid. However, given our simulations, we would expect results from their analysis performed in years after 2006 to yield larger welfare losses as enrollees got locked-in to their sub-optimal choices.

The issue of choice inconsistency is an important policy matter. Based on behavioral and informational explanations, one of the proposed policy prescriptions is to limit the number of plan offerings. However, if enrollees tend to make sub-optimal choices when they enroll in Medicare Part D plans for the first time and can resolve their sub-optimal choices by switching plans, then policies aimed at reducing switching costs—such as those enacted in the 2010 "Patient and Protection and Affordable Care Act" for low income households— will be as important as those that help them make the right choices in the first place.

5.5 Switching Costs and Persistent Consumer Heterogeneity

Our model incorporates persistency in consumer preferences via time-invariant coefficients on observed product characteristics. However, heterogeneity in consumer preferences is only momentary. This may lead our switching cost estimates to be biased upward. Without individual-level data, there are two ways of modeling persistent heterogeneity in consumer preferences: persistent random coefficients and serially correlated idiosyncratic preference shocks. The latter could potentially absorb a lot of choice inertia and hence lower switching cost estimates. However, such a model is infeasible because the degree of serial correlation cannot be identified separately from the switching cost.

The former alternative does not suffer from non-identification because variation in the distribution of plan characteristics across markets in the program's initial year 2006 would identify the random coefficients, as in Berry et al. (1995).[29] However, in contrast to the autocorrelated error term alternative, a random coefficient model is not guaranteed to yield a lower switching cost estimate. Generally speaking, under the environment where product attributes of default options also evolve, time-invariant random coefficients may or may not generate more persistent matches between plans and enrollees at the aggregate level than time-invariant non-random coefficients. It ultimately depends on the interaction between distributions of random coefficients and the evolution of product characteristics. We believe in our case that omitting random coefficients does not necessarily cause a systematic bias of the switching cost parameter.

The main issue with a dynamic random coefficient model is the computational burden.

[29] When one lacks the period 0 data, identification of consumer heterogeneity (levels of standard deviations of random coefficients), mean preference, and the switching cost is very tricky. In the static model, disproportionate substitution patterns among products with similar product attributes in the same period is associated with a high standard deviation. However, in the dynamic model, products with similar attributes in period t may have very different distributions of consumer types while products with different attributes may have a very similar distribution of consumer types to start with. Therefore, knowledge of or assumptions on the initial condition of the distribution of consumer types attached to each product is critical in identifying degrees of consumer heterogeneity.

With random coefficients, we cannot use our method to recover values of the (mean) choice specific value function from observed markets shares by simply solving a system of equations, and we also cannot use the functional relationship between the mean utility flow and the (mean) choice specific value function.[30] Instead, we would need to simulate forward-looking consumers, solve a dynamic programming problem for each simulated consumer, and compute model-predicted market shares at each trial value of the vector of mean utility flows. The method proposed by Gowrisankaran and Rysman (2009) take exactly this approach to incorporate random coefficients.

Under such a method, having a reasonable number of simulation draws to obtain precise estimates can be extremely time-consuming.[31] The computational burden is further increased when a fixed-point algorithm is not monotonically convergent. For example, the dynamic analog of the BLP fixed point algorithm in Gowrisankaran and Rysman (2009) is not likely a contraction mapping, even when it is applied to a non-random coefficients model. A sufficient condition for the BLP algorithm to be contractive is that all products are (weakly) substitutes.[32] In a dynamic discrete choice model, competing products can be complements. For example, in our switching cost model, suppose plan 1's predicted market share is smaller than the observed one whereas plan 2's is larger. Then the BLP algorithm will increase plan 1's mean utility while decreasing plan 2's for the next iteration. This leads to a lower non-default option value for plan 1 and a higher non-default option value for plan 2 than in the previous iteration. Because the choice specific value function is increasing in the non-default option value, plan 1's predicted market share may further decrease and plan 2's market share further increases after the iteration. In this example, plan 1 and plan 2 are complements.[33] As a result, the convergence process of a dynamic analog of the BLP fixed point algorithm can be stalled, and it is not easy to derive a fixed point operator with a more efficient updating algorithm.[34]

[30] Suppose each consumer type's utility flow can be written as: $\delta^f_{ij} = \overline{\delta^f_j} + \mu_{ij}$ and that we parameterize the distribution of μ_{ij}. Let $\overline{csv}_j$ denote the mean choice specific value function, and suppose we write each consumer type's choice specific value function as $csv_{ij} = \overline{csv}_j + \widehat{\mu}_{ij}$. Note that the distribution of $\widehat{\mu}_{ij}$ is different from the distribution of μ_{ij}. Also, the mean choice specific value function $\overline{csv}_j$ is different from the choice specific value function of the consumer type whose $\delta^f_{ij} = \overline{\delta^f_j}$. Hence, modifying our method to incorporate random coefficients is not tractable in practice.

[31] Nosal (2011) rejects persistent preference heterogeneity in the Medicare Advantage market in her estimation with 20 simulated Medicare beneficiaries.

[32] See Berry (1994) and Berry et al. (1995) for the proof.

[33] This occurred occasionally through the course of convergence when we tried to estimate a dynamic random coefficient model using the method by Gowrisankaran and Rysman (2009).

[34] Also, the numerical instability of the logit-class of models that occurs for simulation draws on the tail of the extreme value distribution (see Skrainka and Judd (2011)) can hinder rapid convergence, especially near the boundaries of a sparsely discretized state space. A densely discretized state space with a sufficiently small lower and large upper bound will improve the numerical instability, but it will increase the computational burden.

6 Conclusion

In this paper, we develop an algorithm to estimate dynamic discrete choice models of product differentiation using aggregate market share data and apply the method to estimate switching costs in the Medicare Part D market.

We model consumers in the Medicare Part D market as forward-looking decision makers who choose plans that solve their dynamic programming problems. If enrollees choose plans other than their default plans, they will have to pay switching costs. They base their decisions on the plans' current utility flows, derived from product characteristics such as prices and the generosity of drug coverage, as well as expectations about future product characteristics and the switching costs they face now and will in the future.

We estimate the switching cost parameter and coefficients on plan characteristics by recovering the mean flow utility for each plan and then forming moment conditions that put restrictions on unobserved plan qualities. To this end, we develop an estimation algorithm that maps aggregate market shares to mean flow utilities in two steps. We first back out values of the choice specific function for all plans at each switching cost value. We do this by solving a system of non-linear equations that relate the distribution of the current market shares to conditional choice probabilities and the distribution of past market shares. We then search for the vector of mean utility flows that is consistent with the recovered choice specific values using a fixed point algorithm, defined using the functional relationship between the two variables.

The advantage of breaking down the complicated mapping between aggregate market shares and mean utility flows into two simple mappings is that the algorithm can achieve a faster convergence rate than the dynamic analogue of the BLP algorithm proposed by Gowrisankaran and Rysman (2009). The disadvantage is that it does not allow for random coefficients. Given the trade-offs, our method is best suited for dynamic discrete choice models where the convergence of the dynamic BLP algorithm may not be easily obtained. Convergence can be hampered in cases where some products are dynamic complements rather than substitutes. Our method is also useful when the identification of random coefficients is somewhat dubious because of limited information on the initial state of the consumer type distribution and when random coefficients are not of first order importance.

The estimation results indicate a high switching cost around \$1,500 to \$1,700. The estimated dollar values of the switching cost are around 50% to 60% of the annual drug spending per enrollee and more than the sum of the annual premium payment and out-of-pocket drug expenditure. The values are consistent with previous studies on health care plans. We then perform a counterfactual simulation to calculate the welfare loss due to switching costs. The average welfare loss ranges from \$450 to \$480 against the hypothetical zero switching friction world.

In our application to the Medicare Part D, there are many sources for switching frictions. Although it may important to disentangle the sources to enact effective policies, doing so is beyond the scope of this paper, especially given that we work with aggregate market share data. One may be able to separately identify psychological switching costs from the others by looking at enrollment patterns of auto-enrolled LIS households whose plans lose LIS eligibility. This subset of households has to pay costs from breaking down procrastination or inattention to actively enroll in their current plans to avoid the switching costs related to learning and transaction costs. This extension would require data on the population eligible for auto-enrollment (full Medicaid coverage) in each market, which is not yet publicly available.

More interesting work can be added by modeling the supply side in a dynamic setting to study how insurers interact with inertial consumers under imperfect competition. Given the large switching cost estimates, supply-side incentives are likely to be dictated by a trade-off between "investing" with low prices to attract new consumers and "harvesting" with high prices to take profits from locked in consumers (Klemperer, 1995; Farrell and Klemperer, 2007; Dubé et al., 2009). There is anecdotal evidence that Part D plans have engaged in this behavior.[35] The current Medicare Part D regulations may further encourage "invest-then-harvest" incentives through the way in which premium subsidies are tied to lagged market shares. As shown in Miller and Yeo (2012), even in the static world, insurers with large lagged market shares face more inelastic residual demand and hence have a stronger incentive to raise prices, which will lead to a rise in the government subsidy amount. One can examine how subsidy rules interacts with "invest-then-harvest" incentives in insurers' pricing decisions in search for welfare-improving subsidy rules that better reflect the market's dynamic environment. We leave this to future work.

[35] In the first year of the program, 2006, Humana priced its plans at an average of \$24 lower than the national average and captured the number 2 market share. In subsequent years it raised prices, up to \$6 above the nation average in 2009, and was able to retain its number 2 market share.

References

Abaluck, J. and J. Gruber (2011). Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *American Economic Review 101*(4), 1180–1210.

Berry, S. (1994). Estimating Discrete-choice Models of Product Differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Carlin, C. and R. Town (2009). Adverse Selection, Welfare and the Optimal Pricing of Employer-Sponsored Health Plans. *University of Minnesota working paper*.

Dubé, J., G. Hitsch, and P. Rossi (2009). Do Switching Costs Make Markets Less Competitive? *Journal of Marketing Research 46*(4), 435–445.

Dubé, J., G. Hitsch, and P. Rossi (2010). State Dependence and Alternative Explanations For Consumer Inertia. *The RAND Journal of Economics 41*(3), 417–445.

Ericson, K. (2011). Market Design When Firms Interact with Inertial Consumers: Evidence from Medicare Part D. *Working Paper, Boston University*.

Farrell, J. and P. Klemperer (2007). Coordination and Lock-in: Competition with Switching Costs and Network Effects. *Handbook of Industrial Organization 3*, 1967–2072.

Gowrisankaran, G. and M. Rysman (2009). Dynamics of Consumer Demand for New Durable Goods. *NBER Working Papers #14737*.

Handel, B. (2010). Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts. *Working Paper, University of California, Berkeley*.

Heiss, F., D. McFadden, and J. Winter (2006). Who Failed to Enroll in Medicare Part D, and Why? Early Results. *Health Affairs 25*(5), w344.

Heiss, F., D. McFadden, and J. Winter (2010). Mind the gap! Consumer Perceptions and Choices of Medicare Part D Prescription Drug Plans. *Research Findings in the Economics of Aging*, 413–481.

Hotz, J. and R. Miller (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies 60*(3), 497–529.

Keating, B. (2007). Switching Costs in an Evolving Industry: Assessing Beneficiary and Firm Behavior in the Medicare Prescription Drug Program. *Ph.D. Dissertation, Stanford University*.

Ketcham, J., C. Lucarelli, E. Miravete, and M. Roebuck (2011). Sinking, swimming, or learning to swim in medicare part d. *Working Paper*.

Klemperer, P. (1995). Competition When Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade. *The Review of Economic Studies 62*(4), 515–539.

Kling, J., S. Mullainathan, E. Shafir, L. Vermeulen, and M. Wrobel (2012). Comparison Friction: Experimental Evidence From Medicare Drug Plans. *Quarterly Journal of Economics forthcoming*.

Lucarelli, C., J. Prince, and K. Simon (2008). Measuring Welfare and the Effects of Regulation in a Government-Created Market: The Case of Medicare Part D Plans. *NBER Working Paper #14296*.

Madrian, B. and D. Shea (2001). The Power of Suggestion: Inertia in 401 (k) Participation and Savings Behavior. *Quarterly Journal of Economics 116*(4), 1149–1187.

Miller, D. and J. Yeo (2012). The Consequences of a Public Health Insurance Option: Evidence From Medicare Part D Prescription Drug Markets. *Working Paper, Clemson University*.

Moshkin, N. and R. Shachar (2002). The Asymmetric Information Model of State Dependence. *Marketing Science*, 435–454.

Nosal, K. (2011). Estimating Switching Costs for Medicare Advantage Plans. *Working Paper, University of Arizona*.

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

Samuelson, W. and R. Zeckhauser (1988). Status Quo Bias in Decision Making. *Journal of risk and uncertainty 1*(1), 7–59.

Shcherbakov, O. (2008). Measuring Consumer Switching Costs in the Television Industry. *Unpublished manuscript, University of Arizona*.

Shum, M. (2004). Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market. *Journal of Economics & Management Strategy 13*(2), 241–272.

Skrainka, B. and K. Judd (2011). High performance quadrature rules: How numerical integration affects a popular model of product differentiation. *CeMMAP working paper CWP03/11*.

Strombom, B., T. Buchmueller, and P. Feldstein (2002). Switching Costs, Price Sensitivity and Health Plan Choice. *Journal of Health economics 21*(1), 89–116.

A Figures and Tables

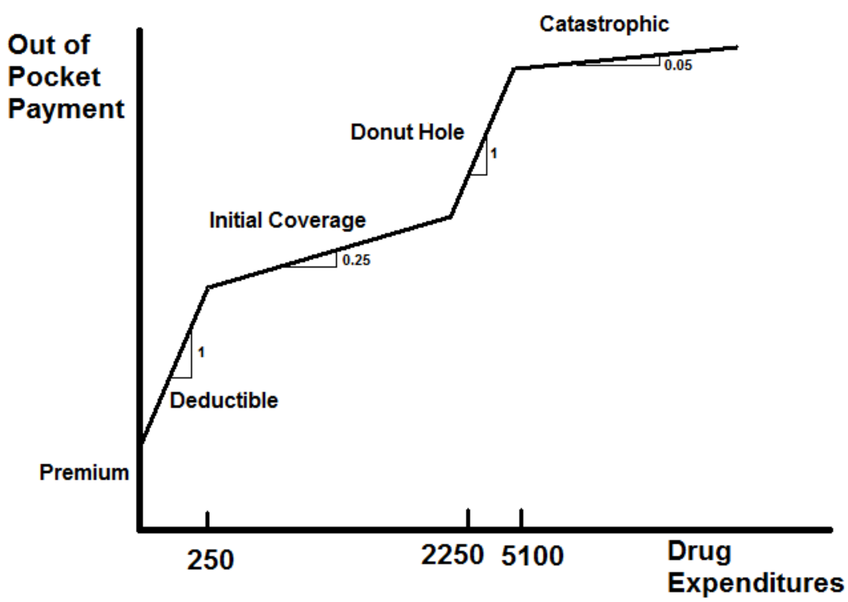

Figure 1: Part D Basic Benefit Structure

Table 8: Number of Insurers and Plan Offerings

	Insurers	All Plans	Basic	Enhanced
2006	17.8	41.6	23.9	17.7
2007	22.4	52.1	26.4	25.7
2008	22.0	52.7	25.9	26.8
2009	20.9	46.4	21.0	25.4

Number of insurer and plan offerings averaged across 34 markets.

Table 9: Average Monthly Premium

	Monthly Premium					Market share-weighted Premium ($)				
	obs	mean	s.d.	min	max	obs	mean	s.d.	min	max
2006	1415	37.26	12.67	1.87	99.9	34	26.88	2.96	21.24	35.59
2007	1770	35.51	14.89	9.24	131.94	34	26.95	2.83	19.66	32.66
2008	1791	37.24	18.62	9.18	100.66	34	28.23	3.29	20.32	33.79
2009	1578	43.11	19.43	9.68	128.55	34	32.48	3.31	23.70	38.90

All figures are in 2006 dollars.

Table 10: Rate of Enrollment in Stand-alone Part D Plans

	Market-Level Enrollment Rate (%)					National-Level Enrollment Rate (%)	
year	obs	mean	s.d.	min	max	Stand-alone Part D	MA + Part D
2006	34	37.8	9.2	18.5	58.5	36.9	14.7
2007	34	38.1	8.9	18.7	59.1	37.4	16.1
2008	34	37.8	8.7	18.4	57.9	37.2	19.6
2009	34	34.7	8.6	14.6	54.8	36.4	21.5

Table 11: Market Shares of Stand-alone Part D Plans

	Market Shares (%)			Share of Market Leader (%)				
year	obs	mean	s.d.	obs	mean	s.d.	min	max
2006	1415	0.91	1.69	34	8.2	2.6	3.8	16.5
2007	1770	0.73	1.57	34	8.2	2.8	4.0	18.1
2008	1791	0.72	1.42	34	7.2	2.8	3.9	18.5
2009	1578	0.75	1.40	34	6.7	2.0	3.5	12.0

Those who choose not to enroll in any stand-alone part D plan include enrollees in MA plans without any PDP coverage, enrollees in MA+PDP plans, enrollees in RDS(Retiree Drug Subsidy) sponsored plans, and non-enrollees. b: in 2006 dollars

Table 12: Exited Plans

Mean	Number of plans		Sum of market shares (%)		Average market share (%)	
	All plans	Exited plans	All	Exited	All	Exited*
2006	41.618	0.059	37.778	7.65e-04	0.922	0.0131
2007	52.059	1.176	38.141	0.0285	0.738	0.0242
2008	52.676	2.382	37.816	0.0804	0.723	0.0338
2009	46.412	0.588	34.347	0.1097	0.752	0.1864

The figures are the mean of each variable across 34 markets except the average market share of exited plans, which takes the average across exited plans.

Table 13: Number of Top 100 Drugs Included in Formularies

year	# of plans	# of formularies	# of top 100 drugs included	
			mean	s.d.
2006	1415	93	90.54	5.96
2007	1770	101	92.02	5.98
2008	1791	88	89.12	7.61
2009	1578	89	86.72	9.12

Table 14: Out-of-Pocket Drug Price Indices

		Initial Coverage PI				Donut Hole PI			
	obs	mean	s.d.	min	max	mean	s.d.	min	max
2006	1415	63.78	9.19	38.53	85.23	95.06	5.58	51.57	101.08
2007	1770	59.05	9.16	41.48	77.78	93.69	5.72	49.77	106.32
2008	1791	55.12	10.34	27.29	73.93	93.96	4.89	74.15	106.32
2009	1578	52.29	10.35	22.56	68.59	93.83	5.43	77.06	106.32

All figures are in 2006 dollars. For off-formulary drugs, the consumer pays full retail price. As we do not have data on retail drug prices, we assume the consumers pay the regional or national average negotiated drug price. All 100 drugs are evenly weighted in the index.

Table 15: Deductible and Network Pharmacies

		Deductible (monthly) (in 2006 dollars)				# Network Pharmacies per Eligible Beneficiary[a]			
	obs	mean	s.d.	min	max	mean	s.d.	min	max
2006	1415	7.66	9.67	0	20.83	0.142	0.023	0.063	0.376
2007	1770	7.24	9.75	0	21.47	0.138	0.020	0.056	0.193
2008	1791	8.10	10.04	0	21.46	0.135	0.020	0.061	0.190
2009	1578	8.69	10.70	0	23.10	0.136	0.018	0.073	0.186

[a]: in hundredth.

Table 16: Low Income Subsidy (LIS) and Enhanced Plan (BenefitEA) Dummy

		BenefitEA dummy				LIS dummy			
	obs	mean	s.d.	min*	max*	mean	s.d.	min*	max*
2006	1415	.425	.494	.365	.500	.289	.453	.143	.390
2007	1770	.493	.500	.455	.531	.356	.479	.164	.491
2008	1791	.509	.500	.472	.558	.276	.447	.096	.384
2009	1578	.547	.498	.527	.583	.176	.381	.023	.304

* min and max are the minimum and maximum over 34 markets' mean values.

Table 17: Year-to-Year Changes in Product Characteristics

	$x_{t+1} - x_t$				
x	obs	mean	s.d.	min	max
Monthly Premium	4270	2.960	10.767	-50.136	74.536
Monthly Deductible	4270	-0.241	5.359	-21.472	21.472
BenefitEA dummy	4270	0.043	0.227	-1	1
Initial Coverage PI	4270	-4.295	7.865	-42.339	19.783
Donut Hole PI	4270	-0.374	3.242	-12.882	37.035
pharm per eligible (×1000)	4270	-.009	0.099	-2.232	0.854
LIS dummy	4270	-0.073	0.379	-1	1

Table 18: AR(1) Process of Product Characteristics

X	Constant	AR(1) coefficient	R^2
Monthly Premium	4.80*	0.95*	0.65
Monthly Deductible	0.77*	0.87*	0.73
BenefitEA dummy	0.09*	0.90*	0.81
Initial Coverage PI	12.69*	0.71*	0.49
Donut Hole PI	19.79*	0.79*	0.64
pharm per eligible (×1000)	0.29*	0.78*	0.78
LIS dummy	0.06*	0.60*	0.42

The regression equation is $X_t = \gamma_0 + \gamma_1 X_{t-1}$. The results are based on 4270 observations across years and regions. *: estimates are significant at 1%

B Appendix

We maintain the assumptions on the transition probability distribution given in 3.1.3:
$P(\delta^f_{jt+1}, \delta_{-jt+1} \mid \delta^f_{jt}, \delta_{-jt}) = \Pr(\delta^f_{jt+1} \mid \delta^f_{jt}) \cdot \Pr(\delta_{-jt+1} \mid \delta_{-jt})$ and the conditional distribution of δ^f_{jt+1} is normal with mean $\gamma_{11}\delta^f_{jt}+\gamma_{12}$ and standard deviation σ_1 and the conditional distribution of δ_{-jt+1} is also normal with $\gamma_{21}\delta_{-jt}+\gamma_{22}$ and σ_2. Let $\mu_i(\cdot, x)$ denotes the normal distribution function with mean $\gamma_{i1}x+\gamma_{i2}$ and standard deviation σ_i for each $i = 1,\ 2$. Let Z denote the standard normal random variable and $\Phi(z)$ denote the distribution function.

Proof Assume $\beta\frac{|\gamma_{11}|}{\sigma_1} < 1$. Consider $x_1,\ x_2 \in R$. Without loss of generality, suppose $x_2 - x_1 > 0$. We want to show that $\widetilde{T}(sc, \mathbf{M_t})(x) = \Delta cv_{jt}(sc, \mathbf{M_t}) + \beta \cdot EEV\left(0, \delta_{-0t}(sc, \mathbf{M_t})\right) - \beta \cdot EEV(x, y)$ is contractive for any value of y. That is, $\left|\widetilde{T}(sc, \mathbf{M_t})(x_2) - \widetilde{T}(sc, \mathbf{M_t})(x_1)\right| = \beta\,|EEV(x_2, y) - EEV(x_1, y))| \leq \lambda\,|x_2 - x_1|$ for some $0 < \lambda < 1$.

Note that $EEV(x, y) = \int\int EV(x', y')d\mu_2(y', y)d\mu_1(x', x) = \frac{1}{\sigma_1}\int\int EV(\sigma_1 z + \gamma_{11}x + \gamma_{12},\ y')d\mu_2(y', y)d\Phi(z)$ with the change of variables $z = \frac{x' - (\gamma_{11}x + \gamma_{12})}{\sigma_1}$. Without loss of generality, suppose $\gamma_{11} > 0$ for illustration purpose.

$$
\begin{aligned}
& |EEV(x_2, y) - EEV(x_1, y)| \\
= \ & \frac{1}{\sigma_1}\left|\int\int EV(\sigma_1 z + \gamma_{11}x_2 + \gamma_{12}, y')d\mu_2(y', y)d\Phi(z) - \int\int EV(\sigma_1 z + \gamma_{11}x_1 + \gamma_{12}, y')d\mu_2(y', y)d\Phi(z)\right| \\
= \ & \frac{1}{\sigma_1}\left|\int\int \left[EV(\sigma_1 z + \gamma_{11}x_2 + \gamma_{12}, y') - EV(\sigma_1 z + \gamma_{11}x_1 + \gamma_{12}, y')\right] d\mu_2(y', y)d\Phi(z)\right| \\
= \ & \frac{1}{\sigma_1}\left|\int\int \left[\ln\left(\phi\exp(y') + \exp(\sigma_1 z + \gamma_{11}x_2 + \gamma_{12})\right) - \ln\left(\phi\exp(y') + \exp(\sigma_1 z + \gamma_{11}x_1 + \gamma_{12})\right)\right] d\mu_2(y', y)d\Phi(z)\right| \\
= \ & \frac{1}{\sigma_1}\left|\int\int \ln\left[\frac{\phi\exp(y') + \exp(\sigma_1 z + \gamma_{11}x_2 + \gamma_{12})}{\phi\exp(y') + \exp(\sigma_1 z + \gamma_{11}x_1 + \gamma_{12})}\right] d\mu_2(y', y)d\Phi(z)\right| \\
\leq \ & \frac{1}{\sigma_1}\left|\int\int \ln\left[\frac{\exp(\sigma_1 z + \gamma_{11}x_2 + \gamma_{12})}{\exp(\sigma_1 z + \gamma_{11}x_1 + \gamma_{12})}\right] d\mu_2(y', y)d\Phi(z)\right| \\
= \ & \frac{1}{\sigma_1}\left|\int\int \ln\left[\exp(\gamma_{11}x_2 - \gamma_{11}x_1)\right] d\mu_2(y', y)d\Phi(z)\right| = \frac{|\gamma_{11}|}{\sigma_1}\,|x_2 - x_1|
\end{aligned}
$$

The inequality follows from the mathematical fact that $\frac{t_2 + c}{t_1 + c} \leq \frac{t_2}{t_1}$ if $c \geq 0$ and $t_2 \geq t_1$. Hence, $|T(sc, \mathbf{M_t})(x_2) - T(sc, \mathbf{M_t})(x_1)| = \beta\ |EEV(x_2, y) - EEV(x_1, y)| \leq \beta\frac{|\gamma_{11}|}{\sigma_1}\,|x_2 - x_1|$. Assuming $\beta\frac{|\gamma_{11}|}{\sigma_1} < 1$, the operator T is contractive. ■