Journal of Public Economics 164 (2018) 106–138

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets[☆]

Jason Abaluck[a, b, *], Jonathan Gruber[b, c], Ashley Swanson[b, d]

[a] Yale, USA
[b] NBER, USA
[c] MIT, USA
[d] University of Pennsylvania, USA

ARTICLE INFO

Article history:
Received 19 October 2016
Received in revised form 3 May 2018
Accepted 8 May 2018
Available online 14 June 2018

Keywords:
Moral hazard
Nonlinear budget sets
Salience
Medicare
Prescription drugs

ABSTRACT

Medicare Part D enrollees face a complicated decision: they dynamically choose prescription drug consumption in each period given difficult-to-find prices and a nonlinear budget set. We use Part D claims data to estimate a flexible model of consumption that accounts for nonlinear prices, dynamic responses, and salience. We use reduced form price responses from a linear regression of consumption on coverage range prices to compare performance under several models of behavior. We find small price elasticities, substantial myopia, and that salient characteristics impact consumption beyond their effect on prices. A hyperbolic discounting model with salience fits the data best.

1. Introduction

In many insurance markets, the structure of benefits is complex and difficult to understand. Coverage generosity often varies nonlinearly as a function of total insured consumption to date, and generosity varies across products in order to steer consumption toward more cost-effective options. From a policy perspective, we would like to know how variations in plan design impact consumption. But the complex nature of insurance contracts creates a dual problem: optimal behavior for rational consumers is hard to calculate, and consumers likely depart from this rational benchmark.

In the context we study, the Medicare Part D prescription drug benefit, plans vary widely in their deductibles, in the copayments and coinsurance for prescriptions above the deductible, and in their coverage of drugs in the infamous Part D "donut hole" where the standard plan offers no coverage. Moreover, prices within a given plan and coverage phase vary across drugs due to some drugs' favorable formulary placement. Enrollees face a complicated nonlinear budget constraint for their consumption decisions, wherein both current and future prices are a function of consumption to date, uncertainty may be realized gradually over time, and prices for each unit of consumption may be difficult to discern. The complexity of this optimization problem may be particularly onerous for an elderly population such as those enrolled in Part D.

In this paper, we present a model of consumption with nonlinear budget sets which accounts for dynamic incentives given uncertainty and myopia as well as variation in the salience of different aspects of the insurance contract. We use reduced form regressions to determine how enrollees respond to price throughout the budget set and relate these to several structural models of consumption behavior. We identify causal responses to prices throughout the budget set using year-to-year changes in coverage phase-specific prices. Notably, our identification of responses to price *changes* will be appropriate for modeling counterfactuals such as filling in the donut hole, in that the vast majority of enrollees at any given time were enrolled in the same plan in the previous year. This framework is very flexible and admits a variety of different price responses – in particular, it allows for the possibility that inertial consumers do not notice year-to-year changes in cost-sharing but are very sensitive to within year changes in prices.

☆ We are grateful to Zack Cooper, Christina Dalton, Liran Einav, Amy Finkelstein, Gautam Gowrisankaran, Jerry Hausman, Kyoungrae Jung, Amanda Kowalski, Fiona Scott Morton, Robert Town, and numerous seminar participants for their helpful comments. Kathleen Easterbrook, Ayesha Mahmud, and Adrienne Sabety provided outstanding research assistance. Financial support was provided by the National Institute of Aging.

* Corresponding author.
E-mail address: jason.abaluck@yale.edu (J. Abaluck).

https://doi.org/10.1016/j.jpubeco.2018.05.005
0047-2727/

The standard approach to estimating demand in the presence of non-linear budget sets is to estimate a nonlinear structural model assuming a particular model of optimization behavior as in Hausman (1985) and Kowalski (2015). However, traditional models of a response to nonlinear budget sets assume that all price responses reflect a single underlying parameter that determines price elasticities. More recent examples relax this assumption to incorporate behavioral responses such as myopia, typically modeled using a discount factor; see Einav et al. (2015) and Dalton et al. (2015) for two excellent recent examples in the Part D context. A central element of these newer approaches is to achieve identification from the response of individuals near "kinks" in the budget constraint, such as the one created by the donut hole in Part D plans.

These approaches face three key limitations. First, as has been discussed in detail in the energy economics literature, lack of information or understanding about prices may lead consumers in nonlinear contracts (such as insurance contracts and electricity contracts) to use rules of thumb beyond discounting in determining their consumption.[1] Second, just as consumers may not be perfectly forward-looking in their consumption behavior, they may also be confused about how visible changes in benefit coverage impact the prices they face. A large recent literature in economics highlights the role of price "salience," which leads consumers in complex or nontransparent environments to be inattentive to some prices (see Chetty et al., 2009 for a review). The structural empirical literature in health care has not typically allowed consumer responses to vary with price salience. Third, most individuals who consume prescription drugs in Part D are not near the budget constraint kink created by the donut hole. Indeed, only 10% of individuals in a given year end the year within $200 of the donut hole, and that 10% is observably sicker than the average enrollee, with higher medical spending and more chronic illnesses.[2] Moreover, many individuals who are impacted by policy counterfactuals such as filling in the donut hole are those who end the year well past the donut threshold – according to our estimates, only 14% of the effect of filling in the donut hole is accounted for by enrollees ending the year within $200 of the donut hole kink. It is therefore unclear whether the results identified by this particular variation extend more broadly to the population of Part D enrollees.

We present a new model of prescription drug demand that addresses all of these concerns. To do so, we draw on the spirit of the nonparametric estimation approach in Blomquist and Newey (2002). Instead of imposing one structure that incorporates all budget segments (e.g., that a particular level of myopia completely explains the relative responses to current and future prices), this approach allows the data to tell us how enrollees' consumption throughout the year responds to different budget set segment prices.[3] We also allow enrollees to respond to variables capturing nominally large changes in benefit coverage that may be more salient. This reduced form model is then embedded in a structural model of drug consumption which allows us to better assess which beneficiaries are impacted by changing prices and how these prices impact the time path of consumption.

We demonstrate that linear regression methods can be used to recover parameters from structural models of consumption for individuals whose marginal prices are in the interiors of budget set segments. This approach allows us to estimate price sensitivity for individuals throughout the Part D spending distribution and not just within a relatively narrow spending range near the budget set kink.[4] One advantage of our approach relative to existing approaches is that we can estimate structural parameters without assuming knowledge of the full distribution of consumer uncertainty. Instead, we assume only that we can forecast with very high probability the likelihood that consumers will end the year in a particular portion of the budget set.[5] We also use our estimates to simulate nonlinear consumption behavior: we can test whether price responses for individuals in the estimation sample can be generalized to the (much smaller set of) individuals near the budget set kink.

The main data source is a 20% sample of Medicare Part D claims provided by the Centers for Medicare and Medicaid Services (CMS). The claims data include information on drugs consumed, as well as the date, quantity, retail price, and amount paid by both the insurer and beneficiary for each claim. Our primary identification strategy utilizes the significant year-to-year variation in the cost-sharing features of Part D plans for existing Part D enrollees. While new enrollees might be expected to pay more attention to prices, 90% of Part D enrollees remain in a given plan across years, so that the response of these consumers is most predictive for changes in the structure of Part D, such as filling in the donut hole. For this counterfactual and others impacting existing (inertial) enrollees in health insurance contracts, this strategy is more representative than others in the literature that consider price responses to the introduction of insurance. However, we separately present results allowing for switching of plans between years and find our conclusions unchanged.[6]

We begin by estimating reduced form regressions of year-to-year consumption changes on changes in the price along each budget segment and changes in salient coverage characteristics. Our results offer mixed support for previous approaches. We find that consumers' responses to different coverage phase prices vary steeply with the proportion of enrollees currently in those coverage phases, even holding marginal coverage phase fixed. This is consistent with myopia, but would not occur if all consumers were responding to marginal prices. On the other hand, we find evidence of substantial price "salience." In particular, we find that nominal changes in "donut hole" coverage impact consumption more than would be expected given their impact on either current or expected marginal price. Our most striking evidence of salience is that even low-spending individuals who are highly unlikely to enter the donut hole coverage phase are nonetheless responsive in their consumption to the presence or absence of donut hole coverage.

Our reduced form results suggest price elasticities of around −0.13 on average, which is of a similar magnitude to the previous literature on prescription drug and health care services demand. The dynamics in the observed marginal price responses imply an estimated (quarterly) β (in a $\beta-\delta$ discounting model) of 0.31, suggesting a very high degree of myopia. But we also estimate sizable salience effects – we find for example that low spending enrollees who are unlikely to hit the donut hole reduce consumption by about 6% when

1 For example, Liebman and Zeckhauser (2004) note that individuals may respond average or local average prices, or to "spot" prices rather than to future marginal prices. Brot-Goldberg et al. (2015) find that individuals in nonlinear health insurance contracts respond only to spot prices, even when there is very little uncertainty regarding end of year price.

2 See Appendix A.2 for further discussion comparing our estimation sample to individuals near the Part D kink.

3 Unlike Blomquist and Newey (2002), which employed rich variation in the locations of budget set kinks as well as in the level of prices in each budget set segment, we only have variation in the latter.

4 Our approach involves explicitly dropping individuals from our estimation sample who are "close" to the kink. It is important to note that, for identification purposes, we are using a greater range of identifying variation than other studies focusing on behavior near the kink for identification.

5 A concern with this approach is that consumers may still be responding to their low probability of bunching or switching. To probe robustness, we estimate a version of our model which explicitly allows for uncertainty and endogenous coverage phase transitions using conventional methods and find that our results are largely unchanged.

6 See, e.g., Duggan et al. (2008) regarding the response of new enrollees when Part D was introduced.

they lose gap coverage. Given our elasticities, this implies that they respond to losing gap coverage (which they are unlikely to ever use) the same way they would respond to an 81% increase in prices.[7] This very large effect is assumed to be zero in the current literature on utilization effects.

The reduced form model cannot account for beneficiaries who switch between coverage ranges or bunch at kink points. To do so, we estimate a model where we identify structural parameters using our linear model for individuals away from the kink point and then use those parameters to forecast the behavior of individuals both at and away from kink points. This contrasts with the usual approach which identifies parameters at kink points and extrapolates behavior away from these kink points. We show that a model of myopia similar to Aron-Dine et al. (2015) supplemented by price salience effects fits the data well. Notably, the preferred structural model captures consumption patterns not only in the estimation sample but also in predicting nonlinear consumption behavior for individuals near the kink (i.e., "bunching"). Our model substantially outperforms reduced form models in which beneficiaries respond to average or marginal price. Among beneficiaries near the kink point (all of whom are out of sample), the structural model misstates their average consumption by −0.34 % while the average price model misstates their consumption by 4.63% and the marginal price model misstates their consumption by 7.84%.

Using the structural parameters implied by our linear estimates, we simulate consumption responses to filling in the Part D donut hole as required under the Patient Protection and Affordable Care Act of 2010. We demonstrate that, given our estimates, it matters not just what prices are changed, but also when they change in the year and how the price changes are presented. Filling in the donut hole will lead consumption to increase by about 6% (or $114 per beneficiary), but such increases will be realized unevenly over the year and will affect even low-spending parts of the Part D population due to price salience. Salience effects account for over 30% of the total consumption response to donut hole coverage.

The rest of the paper proceeds as follows. In Section 2, we describe the background of the Part D program and the related literature on elderly decision-making, moral hazard in health care, and nonlinear prices. In Section 3, we outline a model of how consumption responds to prices in the presence of both myopia and salience in a dynamic setting. Section 4 describes our identification strategy and data, and provides details on price variable construction. In Section 5, we present the reduced form empirical estimates. Section 6 translates our price coefficients into structural parameters and shows the results of our counterfactual simulations. Section 7 concludes.

2. Related literature

Our project builds on several literatures. First, there is a rich literature on the impact of cost-sharing on health care utilization; this literature is reviewed in great detail in Chandra et al. (2010). Of particular note is the RAND Health Insurance Experiment (HIE), which is summarized in Manning et al. (1987). The HIE showed that consumption of medical services was modestly price responsive, with an overall estimated arc-elasticity of medical spending in the range of −0.2. A large subsequent literature has investigated utilization effects specifically in the context of prescription drugs. This literature is reviewed in Goldman et al. (2007), which finds elasticities ranging from −0.2 to −0.6. Previous studies focusing specifically on the elderly have consistently found evidence that drug utilization responds to out-of-pocket prices, but the magnitude of the estimates varies dramatically across studies.[8] Our data include a representative sample of the entire universe of Medicare Part D claims and will thus shed light on the elasticity of demand for the full sample of unsubsidized enrollees.

Another literature on healthcare utilization models health care consumption elasticities in the presence of non-standard pricing. Kowalski (2015) studies the aggregate utilization of medical care in a nonlinear budget set environment with a static consumption decision and finds consumers to have quite low price elasticities, thus concluding that generous coverage leads to modest deadweight losses from moral hazard. Aron-Dine et al. (2015) model dynamic consumption of medical services in the presence of a varying effective deductible and show that individuals respond not only to their expected marginal price but also to the spot price they face before reaching coverage thresholds. Einav et al. (2015) consider Part D enrollees specifically by focusing on dynamic incentives due to enrollees entering into Part D contracts at different points in the year (as they age into Medicare) and estimate an overall price elasticity from the degree of bunching observed among individuals whose total drug expenditures place them near the donut hole threshold at the end of the year. Using variation in dynamic incentives for enrollees aging into the Medicare benefit, they estimate a weekly δ of 0.96, which translates roughly to a quarterly β of 0.5; they find static price elasticities ranging from −0.3 to −0.5. In contrast, Dalton et al. (2015) estimate dynamic preferences based on consumption changes as individuals predictably cross the donut threshold and price elasticities based on cross-drug substitution as individuals cross the threshold; they find small, but significant, price elasticities that decline in drug cost, and full myopia ($\beta = 0$).

Our strategy builds on this literature to estimate elasticities with respect to variation in prices at multiple points in the budget set, for a broad range of the overall spending and age distributions. The method accommodates nonlinear price responses of the sort encountered in the energy economics literature, in which lack of information or understanding about prices leads consumers in nonlinear contracts (such as insurance contracts and electricity contracts) to use rules of thumb beyond discounting in determining their consumption. Liebman and Zeckhauser (2004) note that individuals may respond to nonlinear price schedules by "ironing" (responding to average or local average prices), or by "spotlighting" (responding to immediate "spot" prices rather than to future marginal prices). In the electricity setting, Ito (2014) finds that consumers respond to average price rather than marginal or expected marginal price. While the empirical distribution of Medicare Part D spending shows clear evidence of bunching at kink points, which is not observed in Ito's setting, it remains that there are several similarities between the energy setting and consumption under a nonlinear insurance contract: consumers may find it difficult to track their consumption to date and they have to form expectations regarding future consumption in order to determine their marginal prices. For this reason, we estimate a reduced form model that can accommodate average price, marginal price, and current-future price responses.

Finally, we consider decision-making in a complex setting by an elderly population. Issues considered in behavioral economics, such as myopia and salience, may be particularly acute among the elderly given that the potential for cognitive failures rises at older ages.

[7] This calculation is based on a comparison of the initial coverage range ("ICR") price and stark coverage change ("Stark") coefficients in Table 5: $-0.099 * (ln(p_2) - ln(p_1)) = -0.059$ implies p_2 is 81% higher than p_1.

[8] Lichtenberg and Sun (2007) examine the change in drug expenditures for elderly and non-elderly consumers following the introduction of Part D and estimate that Part D led to a 12.8% increase in prescription drug utilization (from an 18.4% reduction in patient cost sharing, an arc-elasticity of −0.70); Yin et al. (2008) report a 5.9% increase in utilization among Part D enrollees in data from a large pharmacy chain. Using different data sources but a similar methodology, Ketcham and Simon (2008) estimate an arc-elasticity of −0.47. Chandra et al. (2010) analyze another group of retirees, from the California Public Employees Retirement System and find an arc-elasticity of prescription drug consumption of −0.08 to −0.15.

A recent study by Agarwal et al. (2009) shows that in ten different contexts, ranging from credit card interest payments to mortgages to small business loans, the elderly pay higher fees and face higher interest rates than middle-aged consumers.[9] Several studies of these issues apply specifically to the context of Part D. Heiss et al. (2006); Kling et al. (2008); Abaluck and Gruber (2011); and Ketcham et al. (2012) each study plan choice under Medicare Part D and find striking evidence in a variety of settings that elders do not make cost-minimizing choices of Part D plans (there is some disagreement regarding whether choices improve over time).

Price "salience" may be particularly important in this setting. Empirical studies of taxation and insurance have shown that consumers in complex or nontransparent environments can respond suboptimally to prices. Chetty et al. (2009) and Feldman et al. (2015) estimate the effects of tax salience; Abaluck and Gruber (2011) estimate the sensitivity of Part D enrollees' plan choices to more and less salient aspects of plan generosity. Crucially, such behavior has been found to depend on consumers' knowledge and how information is presented. See, e.g., Loewenstein et al. (2013) and Handel and Kolstad (2015). In the Part D setting, we draw on this literature and allow for consumers to respond to more salient price changes over and above how those changes impact out-of-pocket prices faced by individuals. Our results suggest that perhaps the same features that lead elders to make errors in financial choices or in choosing the appropriate Medicare Part D plan lead them also to deviate from rational, forward-looking behavior in responding to cost-sharing features.

3. A dynamic model of prescription drug consumption

In this Section, we lay out a dynamic model of prescription drug consumption given a nonlinear budget set, and present an extension that allows for myopia and salience. We then discuss the challenges of identifying the parameters of this model. Finally, we show how the structural parameters of this model can be recovered by estimating a series of linear equations among beneficiaries who end the year in the interior of budget set segments.

3.1. Dynamic consumption responses

We begin with a simple model, in which individuals choose prescription drug consumption subject to a nonlinear budget set in a single period.[10] Specifically, suppose they have the following out-of-pocket (OOP) expenditure function over total quantity Q purchased:

$$E^{OOP}(Q) = \begin{cases} p_1 * Q & \text{if} \quad Q \leq \frac{\bar{X}}{R} \\ p_1 * \frac{\bar{X}}{R} + p_2 * \left(Q - \frac{\bar{X}}{R}\right) & \text{if} \quad Q > \frac{\bar{X}}{R} \end{cases}$$

where OOP prices p_1 and p_2 are indexed by coverage phase, there is a single coverage threshold at $\bar{X}$, and the drug's retail price (i.e., the total price paid by the plan *plus* enrollee) is R. As in many settings with nonlinear budget sets, the OOP price paid by consumers changes as a function of total spending: for the first $\frac{\bar{X}}{R}$ units, the unit OOP price is p_1; for the remaining $\left(Q - \frac{\bar{X}}{R}\right)$ units, the unit OOP price is p_2. In our application of interest, $\bar{X}$ represents a convex budget set kink with OOP prices $p_1 < p_2$ which occurs at the Medicare Part D donut threshold.[11]

Suppose that consumers have a quasilinear utility function $u(Q) - E^{OOP}(Q)$, where $u(.)$ is a general quasiconcave, continuously differentiable function and utility for non-medical consumption is linear. If there is a single period, then the solution to the consumer's optimization problem is:

$$Q^* = \begin{cases} u'^{-1}(p_1) & \text{if } u'\left(\frac{\bar{X}}{R}\right) \leq p_1 \\ \frac{\bar{X}}{R} & \text{if } p_1 < u'\left(\frac{\bar{X}}{R}\right) \leq p_2 \\ u'^{-1}(p_2) & \text{if } u'\left(\frac{\bar{X}}{R}\right) > p_2 \end{cases}. \tag{1}$$

At the optimum, individuals either consume as they would under a linear price schedule with OOP price p_1 or p_2, or they "bunch" at the coverage threshold $\bar{X}$.[12] This makes intuitive sense: those consumers whose marginal utility of consumption at $\bar{X}/R$ is less than p_1 prefer to consume below the threshold $\bar{X}/R$ at linear price p_1, and would continue to do so for any p_2 with $p_2 \geq p_1$; their consumption is thus unaffected by the nonlinearity of the budget set. Similarly, those consumers whose marginal utility of consumption at $\bar{X}/R$ exceeds p_2 prefer to consume past the threshold under linear price p_2, and would continue to consume beyond the threshold if prices for marginal *or* inframarginal units drop (the additional savings on inframarginal units are akin to a transfer given the assumed quasilinear utility function). Those consumers who would be willing to pay p_1 for an additional unit of consumption at the kink but who are not willing to pay the post-kink price p_2 for that unit will bunch at the donut hole by consuming exactly $\bar{X}/R$ for the year.

The above is the solution for consumption in a single-period model. We next extend the model to allow for consumption decisions to take place in multiple periods within a single nonlinear budget set, and to allow for both myopia and salience. We make the additional assumption that utility over consumption is additively separable across periods. For this case, let V_t denote the value function from period t forward given previous expenditure X_t. Allowing for hyperbolic discounting, we arrive at the following program:

$$V_T(X_T) = W_T(X_T) = max_q u(q) - E^{OOP}(X_T, q)$$
$$V_t(X_t) = max_q u(q) - E^{OOP}(X_t, q) + \beta W_{t+1}(X_t + q * R) \forall t < T$$
$$W_t(X_t) = max_q u(q) - E^{OOP}(X_t, q) + W_{t+1}(X_t + q * R).$$

In each $t < T$, total utility in all future periods W_t is discounted by β: we allow for hyperbolic discounting as in a $\beta - \delta$ discounting model, but impose that $\delta = 1$. This is a reasonable assumption, as the literature on financial decision-making generally finds long-run annualized discount factors to be very close to 1; see Laibson et al. (2015) for a review.[13]

9 See also Salthouse (2004), which shows clear evidence that the performance on a series of memory and analytic tasks declines sharply after age 60; and Fratiglioni et al. (1999) for evidence on the relationship between the onset of dementia and age.

10 Several modeling choices below, such as stipulating that coverage phases end when a certain level of total expenditures are reached, are motivated by prescription drug insurance and specifically the Medicare Part D institutional setting. But given an appropriate specification of the budget set, the results regarding the recovery of structural parameters governing myopia and salience from behavioral responses to different budget set segments should hold in more general settings with a non-linear budget set.

11 In a 2006 Part D plan with standard cost-sharing and no deductible, for a single \$100 drug, we would have $p_1 = \$25$ in the first coverage phase (ICR), $p_2 = R = \$100$ in the second coverage phase (the "donut hole"), and $\bar{X} = \$2,250$. We focus on response to ICR and donut hole prices in this project, as very few individuals end the year in the deductible or catastrophic phase in Medicare Part D. However, the model easily generalizes to accommodate more coverage phases.

12 If we expand the model to allow for kinks at the deductible and catastrophic thresholds, the model optimum is similar, but there is no "bunching" at deductible or catastrophic thresholds. At thresholds where the OOP price decreases, the individual would prefer to "jump" past the threshold – if an individual's marginal utility for an additional unit of consumption exceeds the pre-threshold price p_1, then it also exceeds $p_2 < p_1$.

13 We use hyperbolic discounting to capture "myopia" in the form of present-biased price responses, but we note that such responses could be the result of multiple models of decision-making, including one in which consumers are not attentive to the full schedule of prices that they face (as opposed to being perfectly attentive, but valuing future returns lower than current returns). In the fully-attentive case, our model corresponds to naive hyperbolic discounting.

Consumption in the final period T will be the solution to the static optimization problem as in Eq. (1) for a nonlinear budget set with a kink at $\tilde{X} = \max\left\{\bar{X} - X_T, 0\right\}$. Consumption in $t = 1, \ldots, T-1$ will maximize current-period utility $u(q_t)$ less out-of-pocket expenditure E^{OOP}, plus (potentially discounted) future utility. The model is dynamic in the sense that current consumption impacts future value W_{t+1} via the nonlinear budget set. For all individuals ending both the current period $t < T$ and the year in the interior of a coverage phase, consumption in each period will be given by:

$$q_t^* = u'^{-1}(\beta MP + (1-\beta)CP_t)$$

where the terms MP and CP_t denote the marginal end of year price and the current price in period t, respectively.

If $\beta = 1$, this simplifies such that individuals will only respond to *marginal price* in each period, and consumption in *each period* will mirror the static single-period solution in Eq. (1) above. That is, the functional form of the price response will be the same in each period, though consumption may fluctuate throughout the year due to changes in drug needs and seasonality. If instead consumption is present-biased ($\beta < 1$), then q_t^* will be a function of a weighted average of current price and future price, where the weight on future price equals the discount factor β.[14] This parallels the current-future price specification in Aron-Dine et al. (2015).

We next specify a functional form for utility to allow for the conventional diminishing returns to consumption in each period and for price salience. A number of empirical studies demonstrate that consumers underreact to less salient indicators of price (e.g., insurance plan coverage or taxes) and overreact to more salient indicators of price.[15] In our empirical context (discussed in greater detail below), we observe indicators for price salience, so we specify utility as:

$$u(q) = \tilde{\alpha}_0 + \left(\tilde{\alpha}_1 + \tilde{\theta} * Sal\right) * q + \tilde{\alpha}_2 * q^2.$$

The $\tilde{\theta} * Sal$ term captures the fact that consumers may act like goods with salient characteristics (prices) have higher (lower) marginal utility, and thus consume more (less) of them. The solution becomes:

$$q_t^* = \alpha_1 + \alpha_2\left(\beta MP + (1-\beta)CP_t\right) + \theta Sal_i \tag{2}$$

with $\alpha_1 = -\frac{\tilde{\alpha}_1}{2\tilde{\alpha}_2}$, $\alpha_2 = \frac{1}{2\tilde{\alpha}_2}$ and $\theta = \frac{-\tilde{\theta}}{2\tilde{\alpha}_2}$. We obtain a similar functional form if consumers respond to $p_1 + \tilde{\theta}_1$ and $p_2 + \tilde{\theta}_2$ rather than p_1 and p_2; more detail on the functional form of our "salience" terms is provided in Section 5.3 below, where we discuss why certain prices capture salience in the context of Medicare Part D.

The above analysis takes as given that individuals end both the current period and the year in the interior of a budget set range, and know with certainty that they will do so. Of course, any structural relationship between coverage range prices and consumption will be complicated by endogenous coverage phase transitions and may change in the presence of uncertainty. In the linear demand model, the above result holds even if there is uncertainty about which coverage range is current/marginal as long as price changes do not alter the probabilities of coverage phase transitions and as long as there is no bunching. With other utility functions, this result no longer holds exactly. We contend with these issues in two ways. First, in our baseline analyses, we attempt to minimize the importance of uncertainty, as well as the potential for endogenous coverage phase transitions in response to prices, by focusing our analyses on individuals predicted to end the year well away from kink points. In Appendix A.7.1, we show that our estimates are not sensitive to more conservative sample restrictions. Second, in Appendix A.7.2, we derive and prove a simple theorem characterizing optimal consumption under uncertainty, in which consumption in periods prior to the final period T are optimized given the individual's *expectation* of hitting or crossing the budget set threshold $\bar{X}$ in future periods. That is, consumption is determined as a function of current price and expectations regarding future prices rather than current price and actual marginal price. We demonstrate the limited impact of uncertainty and endogenous phase transitions by re-estimating a richer version of the preferred structural model with endogenous future prices and uncertainty; the results with these modifications are unchanged.

3.2. Derivation of reduced form model

In the model with $\beta = 1$, the extent to which consumers respond to prices in each coverage range of their budget set will depend on the probability that each coverage range is marginal. By breaking the marginal price MP down into its component parts for individual i in plan j in year y, we can write:

$$q_{ity} = \alpha_1 + \alpha_2 MP_{iy} = \alpha_1 + \alpha_2 \mathbb{1}(m_{iy} = Pre) * p_{Pre,jy} + \alpha_2 \mathbb{1}(m_{iy} = Post) * p_{Post,jy}$$

where $\mathbb{1}(m_{iy} = C)$ is an indicator for the event that coverage phase C is the marginal coverage phase for individual i in year y, and $p_{C,jy}$ is the out-of-pocket price in coverage phase C in plan j in year y. In this general setting, $C = Pre$ is the pre-threshold coverage phase; $C = Post$ is the post-threshold coverage phase. Rewriting this as

$$E(q_{ity}) = \alpha_1 + \alpha_{Pre,y} * p_{Pre,jy} + \alpha_{Post,y} * p_{Post,jy}, \tag{3}$$

the coverage phase-specific coefficients will be such that $E(\alpha_{Pre}) = E(\alpha_2 * \mathbb{1}(m_{iy} = Pre)) = \alpha_2 * \Pr(m_{iy} = Pre)$ and similarly for the post-threshold price coefficient.[16]

In the more general current/future model with myopia and salience, the specification as a function of coverage phase prices becomes:

$$\begin{aligned} E(q_{ity}) = \alpha_1 &+ \alpha_2 * \beta * \left(\mathbb{1}(m_{iy} = Pre) * p_{Pre,jy} + \mathbb{1}(m_{iy} = Post) * p_{Post,jy}\right) \\ &+ \alpha_2 * (1-\beta) * \left(\mathbb{1}(c_{ity} = Pre) * p_{Pre,jy} + \mathbb{1}(c_{ity} = Post) * p_{Post,jy}\right) \\ &+ \theta Sal_{ity} = \alpha_1 + \alpha_{Pre,ty} * p_{Pre,jy} + \alpha_{Post,ty} * p_{Post,jy} + \theta Sal_{ity}, \end{aligned} \tag{4}$$

where $\mathbb{1}(c_{it} = C)$ is an indicator for the event that C is the current coverage range for individual i in period t. This implies that the coefficients on the pre- and post-threshold prices respectively will be

[14] See Appendix A.7.2 for further detail regarding the solution to the dynamic consumption problem for a general utility function with bunching (i.e., not imposing that enrollees end the period/year in the interior of a coverage range) and uncertainty.

[15] Salience effects contemplated in this Section may be a function of both limited awareness of plan characteristics and incomplete understanding of those characteristics' implications for prices. Limited awareness has been found to be important in the health insurance setting; see, e.g., Loewenstein et al. (2013), Handel and Kolstad (2015). Incomplete translation of salient characteristics into prices has a substantial impact in the tax context in Feldman et al. (2015).

[16] We are implicitly assuming here that the LATE estimated given our instrument will equal the ATE. This assumption is equivalent to asking whether compliers – consumers for whom the price change of their year t-1 plan impacts prices today – have systematically different marginal prices than non-compliers. If they do, the above probabilities will be the probabilities among compliers rather than among the whole sample.

$E(\alpha_{Pre,ty}) = \alpha_2 * (\beta * \Pr(m_{iy} = Pre) + (1-\beta) * \Pr(c_{ity} = Pre))$ and $E(\alpha_{Post,ty}) = \alpha_2 * (\beta * \Pr(m_{iy} = Post) + (1-\beta) * \Pr(c_{ity} = Post))$.

In other words, our reduced form model of consumption responding to pre- and post-threshold prices subsumes a model of consumption responding to current and future price; however, we do not *require* that the consumption response to coverage phase prices be scaled by current and future probability weights under a specific model of expectations, and can accommodate behavior such as "ironing," where individuals respond to average price rather than a combination of current and marginal prices. In this sense, we can regard our reduced form specifications as a generalization of the "current-future" price regression which has been estimated previously in the literature; this generalization allows for myopia, salience, and any other misperception which causes consumers to fail to respond to coverage range prices in a manner proportionate to the probability that those prices are current or marginal.

The linear functional form specified in Eq. (4) allows for estimation using canned techniques such as two-stage least squares; we use 2SLS in our application, and describe our instruments in Section 4.2 below. To recover the structural parameters α_2 and β from this model, we will need to estimate Eq. (4) separately in samples where the underlying structural parameters are the same, but the probability that given coverage ranges are current or marginal varies. We obtain this variation by looking at consumption decisions for the same individual at different points in time.

Estimating these reduced form models will also require either exogenous variation in prices or as many instruments as we have coverage ranges. In the prescription drug context we consider below, prices are endogenous for at least three reasons. First, prices within plans are not randomized and may vary with consumer demand. Second, consumers may actively choose which plans to enroll in depending on the coverage they offer. Third, marginal prices mechanically depend on consumption – if the price increases after a coverage threshold, we expect to see a mechanical positive relationship between out-of-pocket (OOP) price and consumption, since sicker individuals are more likely to cross the threshold. In our application, we discuss how we can construct instruments to deal with all three problems.

3.3. Recovering structural parameters

Here, we relate the structural parameters to reduced form demand parameter estimates and discuss how one can estimate the structural parameters by GMM. The empirical specification derived above is the following:

$$q_{ity} = \alpha_1 + \alpha_{Pre,ty} * p_{Pre,jy} + \alpha_{Post,ty} * p_{Post,jy} + \theta Sal_{ity} + u_{ity}$$

The structural model to which we wish to link the empirical specification is:

$$q_{ity} = \gamma + \eta\left(\beta MP_{iy} + (1-\beta)CP_{ity}\right) + \kappa Sal_{ity}.$$

Note that we have replaced α_1, α_2, and θ in Eq. (2) with γ, η, and κ to avoid confusing the structural and reduced form parameters. The term γ is predicted spending at zero prices and absent salience effects. The static price response is η– this is the price response we would observe under a linear price contract. The hyperbolic discount factor is β. κ captures consumption shifts in response to shocks to the salience of prices or characteristics, which we model as distinct from the budget set responses captured by η and β.

The relationship between the structural and reduced form parameters is:

$$\gamma = \alpha \tag{5}$$

$$\eta * (\beta * \Pr(MP_y = Pre) + (1-\beta) * \Pr(CP_{ty} = Pre)) = \alpha_{Pre,ty} \tag{6}$$

$$\eta * (\beta * \Pr(MP_y = Post) + (1-\beta) * \Pr(CP_{ty} = Post)) = \alpha_{Post,ty} \tag{7}$$

$$\kappa_{ity} = \theta_{ity} \tag{8}$$

Using the above expressions, we may employ a GMM procedure to infer α, η, and β from the estimated reduced form coverage phase price responses $\alpha_{Pre,ty}$ and $\alpha_{Post,ty}$ and the appropriate coverage phase probabilities $\Pr\{CP_{ty} = Pre\}$, $\Pr\{MP_{ty} = Pre\}$, $\Pr\{CP_{ty} = Post\}$, and $\Pr\{MP_{ty} = Post\}$. The coverage phase probabilities can be based on whatever model of expectations is considered appropriate. In our estimates, we assume "perfect foresight" regarding current and marginal coverage phases; we discuss alternative models of expectations after presenting our main results.

4. Background, identification strategy, and data

4.1. Background

The Part D program passed in 2003, and was implemented in 2006 to provide, for the first time, subsidized prescription drug insurance for the elderly.[17] The most noticeable innovation of Part D is that this new Medicare benefit is not delivered by the government, but rather by private insurers under contract with the government. Beneficiaries can choose from three types of private insurance plans for coverage of their drug expenditures. The first type are stand-alone plans called Medicare Prescription Drug Plans (PDPs) (plans that just offer prescription drug benefits). For example, in 2006, there were 1429 total PDPs offered nationally, with most states offering about forty PDPs. The second are Medicare Advantage (MA) plans, plans that provide all Medicare benefits, including prescription drugs, such as HMO, PPO, or private FFS plans. There were 1314 MA plans nationally in 2006. Finally, beneficiaries could retain their current employer/union plans, as long as coverage is "creditable" or at least as generous (i.e. actuarially equivalent) as the standard Part D plan, for which they would receive a subsidy from the government. We focus on PDP plans so that other health benefits are held constant.

Under Part D, recipients are entitled to basic coverage of prescription drugs by a plan with equal or greater actuarial value to the standard Part D plan. The standard plan for the year 2006 covers: none of the first $250 in drug costs each year (the deductible coverage phase); 75% of costs for the next $2000 of drug spending, up to $2250 total (the initial coverage phase, or ICR); 0% of costs for the next $3600 of drug spending, up to $5100 total (the infamous "donut hole"); and 95% of costs above $5100 of drug spending (the catastrophic phase). Coverage thresholds for the standard plan have increased in each year since first implementation of the program; the standard plan deductible and donut threshold in 2009, the last year of our sample, were $295 and $2700, respectively. The government also placed restrictions on the structure of the formularies that plans could use to determine which prescription medications they would insure. In practice, the vast majority of enrollees have chosen plans with non-standard cost-sharing; over 90% of beneficiaries in 2006 were not enrolled in the standard benefit design, but rather were in plans with low or no deductibles, flat payments for covered drugs following a tiered system, or some form of coverage in the donut hole. The ACA mandates that the donut hole be "filled in" gradually by 2020. For the 2014 benefit year, enrollees in plans without coverage in the donut hole are entitled to a 52.5% discount on

[17] Duggan et al. (2008) provide a detailed overview of the Part D program and many of the economic issues it raises, so we just provide a brief overview here.

branded drugs and a 21% discount on generics while in the donut hole.[18]

Enrollment in Part D plans was voluntary for Medicare eligible citizens. In order to mitigate adverse selection, Medicare recipients not joining during their first eligibility period (and who did not have other creditable prescription drug coverage) were subject to a financial penalty if they eventually joined the program.[19]

4.2. Identification

In this study, we focus on responses to prices in the initial coverage phase and donut hole; these are equivalent to the pre- and post-threshold prices described in the model in Section 3. The ideal variation to identify the impact of budget sets on consumption would include independent variation in each segment of the budget set and random assignment of individuals across plans. Unfortunately for our study, as well as all others using Medicare Part D data, prices are endogenous for several reasons. First, prices result partly from the consumers' decision of which plans to choose in light of their expected drug needs – even in the presence of the potential cognitive failures described above, sicker enrollees may choose more generous coverage. Second, prices chosen by pharmaceutical companies rise and fall in response to changes in consumer demand. Third, the nonlinear budget set means that marginal prices mechanically depend on consumption – if the price increases after the donut hole threshold, we expect to see a mechanical positive relationship between out-of-pocket (OOP) price and consumption, since sicker individuals are more likely to end up in the donut hole.

To deal with the first and second issues, we instrument for prices using variation generated by changes in Part D plan cost sharing rules, following the typical approach of health economics demand studies such as Chandra et al. (2010). This approach will be biased to the extent that individuals switch plans due to cost sharing changes. In our primary identification strategy, we address this by creating the instrument using the change in cost-sharing in the initial plan of the enrollee, regardless of their ultimate plan choice.[20] Under the assumptions (A) that individuals did not choose their initial plans based on anticipated changes in plan cost sharing the next year, and (B) that plans do not alter plan generosity in response to unobservable characteristics of enrollees that predict differential trends in consumption, cost sharing changes in the initial plan are exogenous to plan choice and to enrollee characteristics. We consider these assumptions to be quite reasonable in this setting. Regarding (A), Medicare Part D consumers are very insensitive to out of pocket costs even in the current year (Abaluck and Gruber, 2011, 2016; Ho et al., 2015) and so a fortiori are unlikely to choose plans based on anticipated future changes in coverage. Regarding (B), we control for a rich set of individual demographic and spending characteristics (including data on spending patterns in Medicare Parts A and B, which plans would not be privy to) in order to enable an apples-to-apples comparison of spending trends for individuals in plans that do and do not alter their coverage characteristics. While plans with different sets of enrollees may differentially change the coverage they offer, and while they may even do so in a way that is correlated with changes in consumption, our identifying assumption is only violated if they do so in response to aggregate characteristics of enrollees not included in our rich set of controls for patient claims and expenditures.[21] An additional, related concern would be that we identify price responses based only on the behavior of inertial individuals. Given the low switching rate, we consider this to be the optimal strategy for determining how the average (inertial) Part D enrollee would respond to future generosity changes. However, we also show robustness analyses in which we construct our instruments without holding plan enrollment fixed.

The standard approaches to the third problem – nonlinear budget sets – are either to estimate a nonlinear structural model assuming a particular model of optimization behavior as in Hausman (1985) or, more recently, Kowalski (2015) and Einav et al. (2015); or to estimate a nonparametric model with higher order terms for each segment and threshold of the budget set as in Blomquist and Newey (2002). In our analyses, we employ a simplified version of Blomquist and Newey by considering the linear response to budget set segments and by limiting our sample to individuals who are extremely likely to end the year well into the interior of a budget set segment. Robustness checks using higher order polynomial terms to isolate individual phase price responses in a manner analogous to Blomquist-Newey show similar patterns. We then relate the price responses from the flexible estimation specification to several candidate models of consumption behavior and examine which models perform better in terms of predicting in- and out-of-sample consumption.

The structural analysis serves several purposes. First, it allows us to explore counterfactuals – the impact of completely filling in the donut hole will depend on whether beneficiaries respond to nominal features of plans (i.e. "does this plan offer donut hole coverage?") or the actual price they would face conditional on being in the donut hole. Second, the clinical implications would for example be quite different if the impact of the donut hole arises primarily from beneficiaries stopping consumption the moment they enter the donut hole versus consuming more conservatively throughout the year if they expect to hit the donut hole. Third, understanding the underlying mechanisms provides a test of myopia and salience in a setting of great economic interest.

Our identification approach also has several nice features. First, we have variation in prices in both the initial coverage phase and donut hole. Over 90% of Medicare Part D enrollees end the year in one of these two phases, so this allows us to identify a marginal price response for enrollees over a wide range of total drug expenditures, rather than focusing on behavior around the convex kink in the budget set at the donut hole for price variation. Second, variation in both "current" and "future" price as enrollees spend more over the course of the year allows us to estimate "current" and "future" price elasticities separately in our dynamic analyses and thus to determine whether consumers are primarily forward-looking or primarily myopic. Aron-Dine et al. (2015) separately identify myopia and static

[18] Drug manufacturers offer the branded discount under the Medicare Coverage Gap Discount Program; Medicare covers the 21% generic discount in the donut hole (CMS, 2015).

[19] One group was automatically enrolled: low income elders who had been receiving their prescription drug coverage through state Medicaid programs (the "dual eligibles"). These dual eligibles were enrolled in Part D plans by default if they did not choose one on their own. The Part D plans for dual eligibles could charge copayments of only $1 for generics/$3 for name brand drugs for those below the poverty line, and only $2 for generics/$5 for name brand drugs for those above the poverty line, with free coverage above the out-of-pocket threshold of $3600. In addition, other low income groups were eligible for the Low Income Subsidy (LIS) or for other subsidy programs that lowered their premiums and cost sharing.

[20] The very low rate of plan switching from year to year (roughly 10%) implies that the instrumental variable approach holding initial plan fixed does not have much impact. As discussed in Appendix A.7.3, our results are very similar if we exclude switchers from the analysis. See Ho et al. (2015) for a discussion of switching behavior. A more subtle problem is that the monotonicity constraint might be violated if, for example, "switchers" respond to coverage generosity decreases in their initially chosen plan by switching to a plan even more generous than their initial choice. We cannot test the monotonicity assumption directly, but as noted above we find that our results are not driven primarily by the behavior of active switchers – the coefficient estimates are similar between the full sample and the sample of non-switchers only.

[21] We also examine these assumptions indirectly in Appendix A.3. First, we explore whether plan choice in the current year responds to expected future changes in out of pocket costs. Second, we examine whether the consumption trend between year 1 and year 0 is correlated with the price trend between year 2 and year 1; that is, we test for differential pre-trends among those experiencing price changes in the current year pair. The results of both tests support our assumptions (A) and (B).

price elasticities by comparing their future price elasticity estimates with price elasticity estimates calibrated from the RAND experiment; our variation allows us to make this comparison without relying on any external calibrations. Finally, relying on price changes within contracts for existing enrollees allows us to simulate price responses to policies that are most relevant for Medicare Part D enrollees: those in which existing plans' generosity is altered but consumers are not forced to re-enroll. In contrast, relying on price responses for new enrollees would run the risk of over-estimating consumers' price sensitivity to changes in plan design once they are enrolled.

4.3. Data and variable construction

We analyze data from a 20% sample of Part D enrollees from 2006 through 2009. The claims data contain information on drugs consumed, date of claim, quantity consumed (measured in days' supply purchased on the claim – this is our outcome variable in all specifications), total retail price, and out-of-pocket price for each individual claim. The beneficiary data contain demographic variables and enrollment details. The plan and tier files contain detailed information on drug coverage in each coverage phase as well as nonlinear threshold information. In order to control as richly as possible for heterogeneity across individuals, we merge the Part D data with data on utilization in Medicare Parts A and B as well.

For our main analyses, we exclude individuals under 65, individuals who ever received low-income subsidies (and who thus were not subject to the majority of cost-sharing variation), and individuals who were enrolled in employer-sponsored plans. We focus on enrollees in standalone PDPs only. We analyze data for individuals enrolled in their Part D plan for the full year in each year pair of analysis and who had at least one claim in each year. There are 451,632 sample enrollees in 2006–7; 1,126,682 sample enrollees in 2007–8; and 1,129,200 enrollees in 2008–9. The first year pair 2006–7 included 1372 plans, while 2007–8 and 2008–9 each included over 1700 plans.[22] Summary statistics on sample plans and enrollees are in Table 1. The majority of sample enrollees are white and female, with a mean age 75. Between the first and second year of each year pair, a small proportion (9–11 %) of enrollees switched plans.

The standard plan thresholds moved in each year of the program; the standard deductible increased from $250 to $295 between 2006 and 2009, and the standard donut threshold increased from $2250 to $2700. However, as noted above, the majority of enrollees were not enrolled in standard Part D plans. Between 18 and 24% of enrollees were in plans with the standard deductible, but 70–80 % of enrollees were in plans with no deductible, and a small fringe of enrollees were in plans with positive, but nonstandard, deductibles. Furthermore, some enrollees had coverage in the donut hole; between 2006 and 2009, the proportion of sample enrollees with any donut hole coverage ranged from 13 to 20%.

Sample enrollees purchased 1200 to 1400 days' supply of prescription drugs per year on average, for a total expenditure (out-of-pocket plus plan expenditure) of about $2000 to $2400 per year. Note that, due to the extended enrollment period in the first year of the program, individuals enrolled throughout the entirety of 2006 had higher consumption than the average enrollee in later years, as expected if sicker enrollees signed up earlier in 2006.

Our analyses require a single actual price and price instrument for each enrollee, for each coverage phase, for each year of each sample year pair. We construct actual prices and price instruments in each coverage phase using plan coverage information at the coverage phase-drug (NDC) level, and aggregate those phases using enrollee-specific quantity weights based on days supply of drugs consumed.[23] For year pair (year 1, year 2), the actual price in phase c of year y is the weighted average price the individual would face in phase c given the year y plan's year y cost-sharing rules; weights use the individual's year y consumption (days supply) across all sample drugs observed. That is, the price P_{icy} for individual i enrolled in plan p in coverage phase c, for year y of year pair year 1-year 2, is defined as

$$P_{icy} = \sum_{d \in D_{i,cs}} CS_{dcy,p} * RP_{dy,p} * w_{idy} + \sum_{d \in D_{i,cp}} CO_{dcy,p} * w_{idy}$$

where $CS_{dcy,p}$ and $CO_{dcy,p}$ are coverage phase-specific coinsurance rates and copays, respectively, for drug d in plan p, $RP_{dy,p}$ is retail price for drug d in plan p, and the consumption weight for drug d is calculated as $w_{idy} = q_{id,y} / \sum_{d \in D_{i,y}} q_{id,y}$ using observed quantity consumed $q_{id,y}$ for each individual-drug-year combination. $D_{i,y}$ is the set of all drugs consumed by individual i in year y. Prices are for 30-day supplies of drugs. The retail price for a given plan-drug-year combination is calculated as the average retail price (total expenditure per 30-day supply) across all observations of that plan-drug in the claims data for that year.[24]

The price *instrument* in phase c of year y is the weighted average price the individual would face in phase c given the year 1 plan's year y cost-sharing rules; weights use the individual's year 1 consumption. For coinsurances, we apply the coinsurance rate to the retail price appropriate for the given plan-drug combination in year 1. That is, the IV price P^{IV}_{icy} is defined as

$$P^{IV}_{icy} = \sum_{d \in D_{i,cs}} CS_{dcy,p(\,year1)} * RP_{d,year1,pooled} * w_{id,year1} + \sum_{d \in D_{i,cO}} CO_{dcy,p(\,year1)} * w_{id,year1}.$$

Variation in the instrumental variable prices in the first year of each year pair are shown in the top panel of Table 2. Note that these prices are per 30 days supply; the average beneficiary in our sample purchases 1250–1350 days supply of drugs per year, so the impact of even a small change in prices per 30 days supply is substantial. As in the standard plan, the average price decreases, then increases, then decreases again as one moves from the deductible to the initial coverage range (ICR), from the ICR to the donut hole, and from the donut hole to the catastrophic phase. Average differences between phases are not as large as they would be in the standard plan (which has 100% coinsurance in the deductible and donut, 25% coinsurance in the ICR), because many enrollees have no deductible (in which case the "deductible" price in the Table is effectively the ICR price), and some enrollees have coverage in the donut hole.

Year-to-year changes in individuals' IV prices are shown in the bottom panel of Table 2; price changes are shown broken out by coverage phase and year pair. Note that, because plan choice, consumption weights, and retail prices are held fixed at year 1 values, conditional on those year 1 variables, price differences are a function

[22] The sample in 2006–7 is smaller due to our focus on individuals enrolled for the full year of each year pair. 2006, the first year of our sample and of the Part D program, had an extended enrollment period through May. Our results are not sensitive to the inclusion of individuals enrolling later in 2006. See Appendix A.2 for further description of our sample restrictions.

[23] See Appendix A.1 for a detailed example.

[24] When possible, claims for 30-day supplies only are used to calculate average retail prices. When 30-day supply claims are not observed for particular plan-drug combinations, retail price per 30-day supply is imputed by scaling average prices per one-day supplies observed in claims for all other quantities.

Table 1
Summary statistics.

	2006–7 Sample		2007–8 Sample		2008–9 Sample	
	Mean sample characteristics					
Num. beneficiaries	451,632		1,126,682		1,129,200	
Num. plans	1372		1720		1722	
%White	95.93%		94.94%		94.92%	
%Black	2.35%		3.05%		3.05%	
%Female	62.28%		62.94%		62.53%	
Age	74.52		75.30		75.29	
%Switchers	11.06%		9.65%		9.98%	
	Standard plan features					
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Deductible	250	265	265	275	275	295
Donut threshold	2250	2400	2400	2510	2510	2700
	Sample plan characteristics					
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
%No deductible	73.41%	78.65%	77.60%	79.51%	79.43%	79.06%
%Standard deductible	24.04%	19.30%	20.48%	19.65%	19.58%	17.66%
%Full donut coverage	6.09%	2.18%	1.43%	0.00%	0.00%	0.00%
%Any donut coverage	13.74%	19.82%	16.35%	14.18%	13.83%	13.05%
	Sample consumption					
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Mean Q	1233	1409	1220	1327	1274	1344
(Std. dev.)	827	867	836	863	858	865
Mean exp.	$2108	$2355	$1981	$2078	$1997	$2121
(Std. dev.)	$2187	$2688	$2380	$2729	$2626	$3108

Notes: Table displays enrollee and plan summary statistics for full 20% random sample of unsubsidized, elderly Part D enrollees enrolled continuously in standalone PDPs throughout both years of each year pair (Medicare Advantage and employer sponsored enrollees excluded). In addition, we exclude from our sample enrollees with zero claims in either year of the year pair.

Table 2
Sample price instrument variation.

	2006–7 Sample		2007–8 Sample		2008–9 Sample	
	Mean	(SD)	Mean	(SD)	Mean	(SD)
Year 1 IV prices						
Deductible	26.81	(31.43)	23.04	(34.65)	21.88	(44.22)
Initial coverage range	16.58	(14.73)	15.41	(15.76)	14.50	(28.47)
Donut hole	50.80	(60.12)	49.62	(60.02)	49.19	(98.43)
Catastrophic range	3.99	(2.98)	3.95	(2.89)	3.90	(4.83)
Year 2 IV price - Year 1 IV price						
Initial coverage range	1.01	(7.93)	1.32	(3.49)	3.19	(7.33)
Donut hole	5.64	(36.77)	7.10	(20.04)	1.28	(16.54)

Notes: Prices per 30-day supply. Price instruments generated using plan-drug-coverage phase-specific copays/coinsurances and enrollee-specific consumption weights on drugs. For prices specified as coinsurances, average retail price for each plan-drug combination used as the basis to which plan-drug-phase coinsurances are applied. Enrollee-specific consumption weights based on days supply used to generate a weighted average price for each individual in each coverage phase. For second year of each year pair, the consumption weights, retail prices, and plan choice from first year are imposed to isolate price effect of changes in cost-sharing characteristics holding consumption and enrollment fixed. Donut price change shown only for plans with any gap coverage in either year (17% of sample overall; price instrument difference is mechanically zero for other plans).

only of year 1 plan cost-sharing changes between year 1 and year 2. For the sake of brevity, price differences are shown only for the ICR and donut hole.[25] There is considerable variation in year-to-year ICR price changes across sample individuals; the sample of individuals experiencing price changes in the donut hole is smaller since the IV price change is zero by construction for those individuals with no gap coverage in both years. Price elasticities are identified by comparing consumption trends across similar individuals experiencing different price trends; thus all individuals, including those with and without some gap coverage, are used to identify model parameters. In robustness analyses in which we do not leverage the inertia of plan enrollment, instruments are constructed as in the above description, but without holding plan fixed. That is, price in year 2 is a function of year 2 plan characteristics, year 1 consumption weights, and year 1 drug prices.

[25] Individuals with deductibles in both years of the year pair have no effective IV variation in the deductible because retail prices are held fixed across years. Similarly, catastrophic price variation is based only on retail prices and cost-sharing minimums, the former of which are held fixed between years in the IV and the latter of which vary across years, but not across plans. Thus, the majority of our variation within coverage phase comes from the ICR and donut hole cost-sharing changes. In some analyses below, we also analyze responses to variation coming from changes in the location of deductible thresholds between years.

In looking across the three year pairs of our analysis sample, we note several patterns of interest. First, we see in the bottom panel that the ICR and donut price changes are slightly positive on average, though the standard deviation indicates that there are many substantial positive and negative price changes in the sample. That is, within a given contract, the average enrollee experiences diminishing plan generosity for their fixed bundle of drugs between years 1 and 2. On the other hand, a comparison of prices across years in the top panel of Table 2 indicates that average prices across existing *and* new plans actually decrease between 2006 and 2008. Thus, changes in generosity over time impact existing enrollees less favorably than new enrollees (a similar pattern applies to premiums for new versus existing enrollees, as documented in Ericson (2014) – this phenomenon is termed invest-then-harvest behavior in the context of premium-setting). Second, the variation in the IV donut price change is falling over time, which is consistent with the donut hole data in Table 1 – fewer enrollees have donut hole coverage in 2009 than in 2006, and no sample enrollees have full donut coverage in 2009, as opposed to 6% in 2006. Finally, we note that the magnitude of the price variation generated by our instrument is substantial. Comparing, e.g., the 5th and 95th percentile ICR price change in 2006, a difference of $12.63 per 30 days supply would translate to more than a $500 difference in OOP costs given mean year 1 consumption.

The distributions of year-to-year log price change for the ICR and donut phases are shown in Fig. 1; in each budget set segment, there is a large mass at zero, so the zeroes are omitted from the plots. The log price changes in the ICR are distributed between roughly −0.5 and 0.5; the donut price changes are less symmetric, distributed mostly on the range −0.2 to 0.8.

5. Reduced form price response estimates

In this Section, we present the results of empirical specifications aimed at examining the reduced form impact of prices in different coverage ranges on consumption in the Part D setting. Our approach is similar in spirit to the nonparametric estimation approach in Blomquist and Newey (2002). Instead of imposing one structure that incorporates all budget segments (e.g., imposing that a particular level of myopia completely explains the relative responses to current and future prices), the nonparametric approach in Blomquist and Newey allows enrollees' consumption throughout the year to respond flexibly to different aspects of the budget set. Given sufficient variation, we could follow Blomquist-Newey to estimate fully nonparametric consumption responses to all budget set parameters. In the current empirical setting, we have substantial ICR and donut price variation, but limited variation in locations of budget set kinks. Therefore, we instead estimate consumption responses in each period to ICR and donut prices, and focus our attention on individuals who are predicted ex ante to respond primarily to prices in the interior of coverage ranges. As noted in Section 3, these specifications can be thought of as a generalization of the model with current and future prices sometimes estimated in the literature. We also allow enrollees to respond to variables capturing nominally large changes in benefit coverage that may be more salient. In the next Section, we develop a flexible model that relates these estimates to structural primitives and thus allows us to consider counterfactuals.

Our model includes tests for "myopia" and "salience." We use these terms in a broad sense to refer to patterns in the data that are inconsistent with the usual models rather than a specific psychological phenomenon. Myopia refers to the responsiveness of consumption to the current price conditional on the marginal price: if we observe individuals responding more strongly to marginal price when they enter their marginal coverage phase, then we would consider that differential response as evidence of myopia. Our test for salience examines whether nominally significant categorical price changes enter utility even when they have no real impact on current or expected future prices.

5.1. Sample: individuals away from kink points

A major challenge in modeling consumption with nonlinear budget sets is to deal with bunching and coverage range switching at convex kink points. That is, as we show in greater detail in the next Section, consumption can respond to the nonlinear nature of the budget set, particularly for those ending the year near a budget set kink. To avoid these difficulties, we focus on individuals whose spending can be predicted at the start of the year to be highly likely to lie in the interior of a given coverage range. We also consider several tests of the robustness of this assumption.

Our baseline sample includes individuals for whom the donut is highly likely to be marginal and individuals for whom the ICR is highly likely to be marginal (excluding individuals likely to be near the threshold between the two budget set segments based on previous year spending). Specifically, for the 2006–2007 analysis, we look only at individuals consuming less than or equal to $1500 in 2006 and, separately, at individuals consuming between $3000 and $5000 in 2006. We restrict the samples similarly for 2007 and 2008, adjusting the cutoffs in each year pair to account for secular trends in standard plan thresholds. The first group is a set of individuals who are almost certainly going to have the ICR price as their marginal price; the second group is a set of individuals almost certain to have the donut price as their marginal price.

In practice, 3.7% of individuals in the low-spending group cross the donut hole threshold in year 2 and 14.2% of individuals in the

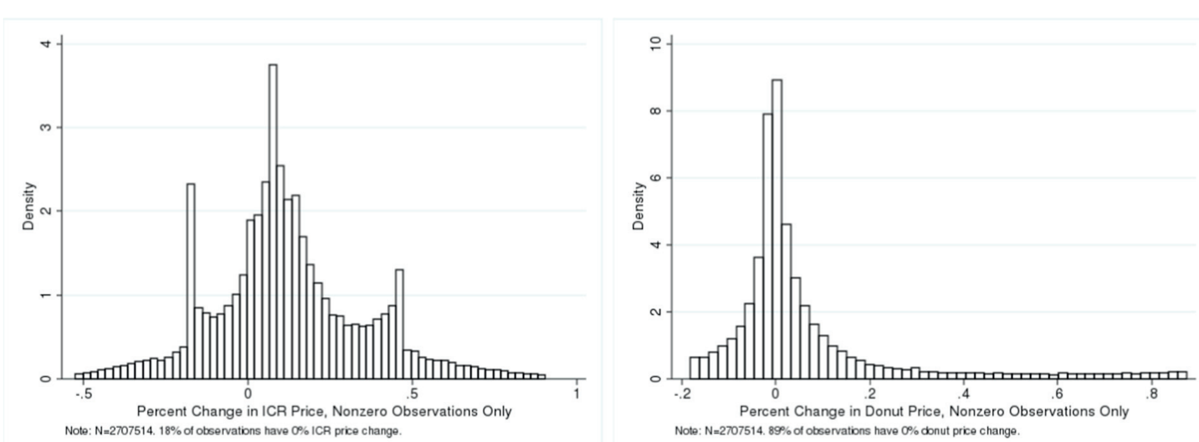

Fig. 1. Sample price instrument variation – Full distributions.

high-spending group do not. Thus, our sample restrictions do not completely eliminate switching behavior and may be impacted by uncertainty. Further, the extent to which coverage phase switching responds endogenously to prices may be an additional source of bias. However, we can use these samples to examine behavior in the presence of far more limited marginal price uncertainty and scope for switching than we would expect in the full sample.

5.2. Quarterly price responses

We empirically estimate a model of prescription drug consumption as a function of the ICR and donut price. We modify Eq. (3) in two ways. First, we rely on the plausible exogeneity of our price change instrument; hence, we take first differences between both price and consumption. Second, the log functional form is employed because the distributions of both price and consumption are skewed positive and there is a heavy upper tail of consumption even within the restricted regression sample. (We omit the salience terms for now and add them to the specification in Section 5.3 below.) For each quarter t and each pair of years $(y, y-1)$, we estimate:

$$\begin{aligned} log(q_{it,y}) - log(q_{it,y-1}) = \alpha_{1t} + \alpha_{ICR,ty}\left(log\left(p_{ICR,j,y}\right) - log\left(p_{ICR,j,y-1}\right)\right) \\ + \alpha_{Donut,ty}\left(log\left(p_{Donut,j,y}\right) - log\left(p_{Donut,j,y-1}\right)\right) \\ + X_{ity}\phi_{ty} + u_{ity}. \end{aligned} \quad (9)$$

Using the identification strategy outlined in Section 4.2 above, we instrument for price changes between $y-1$ and y using changes in cost-sharing between $y-1$ and y in the plan the individual was initially enrolled in:

$$\begin{aligned} & log(p_{ICR,j,y}) - log(p_{ICR,j,y-1}) \\ & = \omega^{ICR}_{1t} + \omega^{ICR}_{ICR,ty}\left(log\left(p^{IV}_{ICR,j,y}\right) - log\left(p^{IV}_{ICR,j,y-1}\right)\right) \\ & + \omega^{ICR}_{Donut,ty}\left(log\left(p^{IV}_{Donut,j,y}\right) - log\left(p^{IV}_{Donut,j,y-1}\right)\right) + X_{ity}\phi^{ICR}_{ty} + v^{ICR}_{ity} \end{aligned} \quad (10)$$

$$\begin{aligned} & log(p_{Donut,j,y}) - log(p_{Donut,j,y-1}) \\ & = \omega^{Donut}_{1t} + \omega^{Donut}_{ICR,ty}\left(log\left(p^{IV}_{ICR,j,y}\right) - log\left(p^{IV}_{ICR,j,y-1}\right)\right) \\ & + \omega^{Donut}_{Donut,ty}\left(log(p^{IV}_{Donut,j,y}) - log\left(p^{IV}_{Donut,j,y-1}\right)\right) + X_{ity}\phi^{Donut}_{ty} + v^{Donut}_{ity}. \end{aligned} \quad (11)$$

As described in Section 4.3, in this and each of the following regressions, we use an instrumental variables strategy based on how beneficiaries respond to changes in copays. For a given pair of years, cost-sharing characteristics used to generate the year 2 price instrument are the year 2 cost-sharing parameters (copays/coinsurances, coverage thresholds, etc.) of the year 1 chosen plan. In all regressions, we include rich controls for individual demographics and individual health care consumption patterns.[26] The rich controls for base year consumption are particularly important because base year consumption weights are used to generate the year 2 price instrument and thus there is potential for mean reversion; see Saez et al. (2012). Finally, we include rich controls for plan characteristics in order to generate the apples-to-apples comparison described in Section 4.2 – e.g. two individuals with identical plans and observables in 2006, one of whom experiences a change in cost-sharing in 2007. We also control for dummies for year 1 deductible and donut hole coverage and thresholds, polynomials of year 1 prices in each budget segment, and 50 quantiles each of year 1 quantity (days supply), expenditure, out-of-pocket spending, and average retail price of drugs consumed for the average person in the year 1 plan. All non-price-change controls are included in X_{ity} and are omitted from the Tables for the sake of brevity. As noted in the introduction, if consumers respond only to within year changes in prices, we should still detect an effect, because we are identified by looking at changes in the donut hole price holding fixed the ICR price – these changes necessarily alter the within year price differential. We report the IV results here and the OLS, reduced form, and IV results in Appendix A.4.

Consider first the low-spending group. Results are in Table 3. The proportion of individuals in their marginal coverage phase is slightly increasing over the course of the year (beginning at 83% in Q1, ending at 97% in Q4) as the small proportion of individuals with deductibles enter the ICR.[27] Considering each year pair individually, the ICR price response is either flat (2006–7) or slightly increasing in magnitude over the course of the year (2007–8 and 2008–9). On balance, the results for all years pooled show that the ICR response is fairly flat across quarters even though the proportion of individuals in the marginal coverage phase increases slightly – given that most individuals (83%) either do not have a deductible or exit the deductible in Q1, these results are limited in their usefulness for detecting myopic behavior. We do not observe a substantial response of low-spending individuals with respect to the donut hole price, as would be expected given even imperfectly forward-looking behavior – in some samples, we observe a small positive sign on the donut hole price. The magnitude of the static price elasticities suggested by these estimates is on the lower end of the elasticities found in the literature (−0.04 to −0.05).

Consider next the high-spending group. Results are in Table 4. Among high-spending individuals, there is a steep increase in the proportion of individuals in the marginal coverage phase (the donut hole) between quarters 1 and 4 (rising from 0.5% to 95% based on year 1 consumption, the latter number reflecting that some high-spending individuals are in plans with no donut hole). Concurrent with this increase, we observe also that the donut hole price response is quite steep over the course of the year in each year pair individually and in the regression that pools all year pairs. High-spending individuals have a large and significant donut price response in quarter 4 (ranging from −0.14 to −0.24 in the individual year pair samples, equalling −0.16 in the pooled analysis) which is significantly larger than the donut response at the beginning of the year (ranging from −0.05 to −0.09 in the individual year pair samples, equalling −0.05 in the pooled analysis). This fact provides striking evidence of myopia, given the low degree of uncertainty that high-spending individuals will be in the donut hole at the end of the year – individuals are on average more than three times as responsive to donut price changes when they are actually *in* the donut hole than they are prior to crossing the donut threshold. Regarding the “spot" price enrollees face prior to the donut hole (the ICR price): in 2006–7 and 2007–8, individuals never significantly respond to the ICR price change, while in 2008–9, the ICR price coefficient is −0.05 to −0.06 at the beginning of the year. The pooled ICR response is significant but small in Q1 and Q2 (−0.02 to −0.03) but shrinks toward

[26] The demographic controls include dummies for age, sex, race, and state. The consumption controls include rich controls for total base year (the first year of each year pair) individual drug consumption and prices (polynomial controls for coverage phase-specific prices, and 100 quantiles each of days supply purchased, total drug expenditure, out-of-pocket (OOP) drug expenditure, and retail price per prescription) and total base year individual medical spending (dummies for nonzero spending overall and in several subcategories – office visits, inpatient emergency, inpatient non-emergency, outpatient, and other – as well as polynomial controls for medical spending overall and level controls for spending in each subcategory). We also include dummies for having any drug spending in each generic therapeutic class (GTC), which is a summary measure of target medical condition (First Data Bank classifies each NDC into one of forty such classes, the most popular of which are “Cardiovascular,” “Autonomic Drugs,” “Cardiac Drugs,” and “Diuretics” among our sample enrollees), and interact the GTC dummies with separate indicators for generic and branded drugs.

[27] The remaining 3% do not exit the deductible in year 1.

Table 3
Results of quarterly ICR and donut price regressions – Low-spending group.

Period	Price	% in ICR	All years pooled		2006–7		2007–8		2008–9	
			Coef	SE	Coef	SE	Coef	SE	Coef	SE
Q1	ICR	83.0%	−0.051	0.006 **	−0.047	0.014 **	−0.022	0.010*	−0.061	0.009 **
Q1	Donut		0.023	0.008 **	−0.008	0.023	0.022	0.011 *	0.018	0.010
Q2	ICR	92.1%	−0.041	0.007 **	−0.046	0.018 *	−0.023	0.010 *	−0.056	0.009 **
Q2	Donut		0.026	0.008 **	0.041	0.026	0.017	0.011	0.029	0.011 *
Q3	ICR	95.4%	−0.043	0.007 **	−0.057	0.017 **	−0.032	0.011 **	−0.044	0.008 **
Q3	Donut		0.008	0.009	0.026	0.027	0.000	0.013	0.003	0.012
Q4	ICR	97.0%	−0.051	0.009 **	−0.039	0.016 *	−0.045	0.015 **	−0.067	0.011 **
Q4	Donut		−0.011	0.011	−0.051	0.027	−0.008	0.016	−0.005	0.017

Notes: Results of quarterly regressions of log consumption change on log change in ICR and donut prices, low-spending individuals with positive consumption in each quarter only. N=919,650 across all years; N=128,412, 388,454, and 402,784 in year pairs 2006–7, 2007–8, and 2008–9, respectively. Proportion of (pooled years) sample for whom initial coverage range (ICR) is marginal in each quarter of first year noted next to estimated coefficients. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

Table 4
Results of quarterly ICR and donut price regressions – High-spending group.

Period	Price	% in Donut	All years pooled		2006–7		2007–8		2008–9	
			Coef	SE	Coef	SE	Coef	SE	Coef	SE
Q1	ICR	0.5%	−0.034	0.010**	−0.018	0.018	−0.012	0.015	−0.057	0.017 **
Q1	Donut		−0.053	0.014 **	−0.054	0.021 **	−0.050	0.024 *	−0.088	0.030 **
Q2	ICR	22.2%	−0.023	0.011 *	−0.003	0.021	−0.013	0.015	−0.053	0.017 **
Q2	Donut		−0.073	0.015**	−0.077	0.025 **	−0.069	0.022**	−0.085	0.030**
Q3	ICR	88.0%	−0.023	0.012	−0.005	0.022	−0.020	0.017	−0.029	0.018
Q3	Donut		−0.086	0.018 **	−0.111	0.029 **	−0.116	0.031 **	−0.069	0.028 *
Q4	ICR	94.9%	−0.008	0.014	−0.021	0.025	−0.017	0.021	0.017	0.022
Q4	Donut		−0.164	0.025 **	−0.167	0.027 **	−0.243	0.047 **	−0.138	0.034 **

Notes: Results of quarterly regressions of log consumption change on log change in ICR and donut prices, high-spending individuals with positive consumption in each quarter only. N=294,898 across all years; N=61,198, 127,052, and 106,648 in year pairs 2006–7, 2007–8, and 2008–9, respectively. Proportion of (pooled years) sample for whom donut hole is marginal in each quarter of first year noted next to estimated coefficients. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

zero in Q3–Q4. Each of these patterns is consistent with myopia. Interestingly, while each of the donut hole coefficients in Table 4 is statistically significantly different from the donut hole coefficients for the low-spending sample, the ICR coefficients are not statistically significantly different until Q4, at which point the two samples' location in the budget set is most different.

The above results provide estimates of a small, significant price elasticity of demand for prescription drugs throughout the spending distribution. We also observe strong evidence of myopic utilization behavior, in that enrollees' marginal price response is much more evident at the end of the year, once they have entered their marginal coverage phase. The evidence from high-spending enrollees in particular allows us to reject that individuals are fully forward-looking and to reject that individuals are fully myopic. We can reject forward-lookingness due to the significant increase in price response as more individuals enter the donut hole; we can reject full myopia due to the negative and significant donut hole price response in Q1 (in which 99.5% of high-spending individuals are still in the ICR). Relating these responses to a parameterized model of myopic behavior requires that we explicitly link each quarter's consumption responses to the distribution of enrollees across coverage phases; we leave that exercise for the next Section. Next, we examine whether individuals respond to "prices" other than the out-of-pocket costs that should be most relevant for them given their observed drug consumption patterns.

5.3. *Salience results*

Some plan characteristics that may specifically be more salient are the presence or absence of a deductible, and the presence or lack of donut hole coverage, each being particularly visible in plan benefit materials and tools such as the Medicare Plan Finder on CMS's website. Regarding the deductible, we expect consumers to respond to the presence or absence of coverage either because they are myopic[28] and/or because the deductible is a particularly salient coverage benefit; in our setting, we cannot distinguish these two mechanisms because there is no within-phase price variation conditional on deductible coverage.

However, we can examine whether individuals respond to categorical donut hole coverage conditional on out-of-pocket price in the donut hole. Some of the donut hole coverage variation encountered by our sample enrollees is of the sort of clear variation used to estimate price sensitivity in previous work, such as the RAND experiment or Chandra et al. (2010); for example, whether the plan includes coverage of branded drugs in the donut hole, or has no donut hole coverage. Other donut hole coverage variation is more subtle and would be difficult for enrollees to translate into prices; for example, whether the plan includes coverage of "many preferred" branded drugs. As discussed in Appendix A.5, we may expect consumers to respond to nominally large changes in coverage (denoted "stark changes" below), over and above how those changes translate into out-of-pocket price in the donut hole.

There is ambiguity about how best to model consumers' responses to more salient characteristics that are not entirely captured in out of pocket prices. In the interest of simplicity, we extend the above empirical specification to include indicators for

[28] The deductible is not the marginal price for all but a small sliver of the Part D population.

whether the plan includes a deductible, as well as whether it includes generous, easy-to-understand ("stark") donut hole coverage. Taking differences, we then obtain a reduced form model with two terms capturing variation in salient coverage characteristics. That is, we use the exact same specification as in Section 5.2, but with two additional variables. The deductible change variable $\Delta\{Ded\}$ equals −1 if the deductible threshold is decreased between years, and equals 1 if the deductible threshold is increased between years by more than the standard deductible change (e.g., \$250 to \$265 between 2006 and 2007); it equals 0 otherwise.[29] The second variable captures changes in categorical coverage in the donut hole. The variable $\Delta\{Stark\}$ equals 1 if coverage for an entire class of drugs (e.g., all generics) is dropped between year 1 and year 2, and equals −1 if coverage for an entire class of drugs is added; it equals 0 otherwise. Thus, the variables $\Delta\{Ded\}$ and $\Delta\{Stark\}$ take the place of the salience term *Sal* in Eq. (4):

$$\begin{aligned} log(q_{it,y}) - log(q_{it,y-1}) = \alpha_{1t} &+ \alpha_{ICR,ty}(log(p_{ICR,j,y}) - log(p_{ICR,j,y-1})) \\ &+ \alpha_{Donut,ty}(log(p_{Donut,j,y}) - log(p_{Donut,j,y-1})) \\ &+ \theta_{Ded,t} * \Delta\{Ded\}_{i,y} + \theta_{Stark,t} * \Delta\{Stark\}_{i,y} \\ &+ X_{ity}\phi_{ty} + u_{ity}. \end{aligned} \quad (12)$$

Using a similar identification strategy to that in Section 4.2 above, we instrument for actual categorical coverage changes between $y-1$ and y using changes in categorical coverage between $y-1$ and y in the plan the individual was initially enrolled in ($\Delta\{Ded\}^{IV}_{i,y}$ and $\Delta\{Stark\}^{IV}_{i,y}$, respectively):

$$\Delta\{Ded\}_{i,y} = \omega^{Ded}_{1t} + \theta^{FS,Ded}_{Ded,t} * \Delta\{Ded\}^{IV}_{i,y} + \theta^{FS,Ded}_{Stark,t} * \Delta\{Stark\}^{IV}_{i,y} \quad (13)$$

$$+\omega^{Ded}_{ICR,ty}\left(log\left(p^{IV}_{ICR,j,y}\right) - log\left(p^{IV}_{ICR,j,y-1}\right)\right) \quad (14)$$

$$+\omega^{Ded}_{Donut,ty}\left(log\left(p^{IV}_{Donut,j,y}\right) - log\left(p^{IV}_{Donut,j,y-1}\right)\right) + X_{ity}\phi^{Ded}_{ty} + v^{Ded}_{ity} \quad (15)$$

$$\Delta\{Stark\}_{i,y} = \omega^{Stark}_{1t} + \theta^{FS,Stark}_{Ded,t} * \Delta\{Ded\}^{IV}_{i,y} + \theta^{FS,Stark}_{Stark,t} * \Delta\{Stark\}^{IV}_{i,y} \quad (16)$$

$$+\omega^{Stark}_{ICR,ty}\left(log\left(p^{IV}_{ICR,j,y}\right) - log\left(p^{IV}_{ICR,j,y-1}\right)\right) \quad (17)$$

$$+\omega^{Stark}_{Donut,ty}\left(log\left(p^{IV}_{Donut,j,y}\right) - log\left(p^{IV}_{Donut,j,y-1}\right)\right) + X_{ity}\phi^{Stark}_{ty} + v^{Stark}_{ity}. \quad (18)$$

The regression results for the full year are shown in Table 5; the first column shows all years pooled, and the second through fourth columns show 2006–7, 2007–8, and 2008–9 separately. In general, the results show that Part D enrollees do respond on average to stark changes in donut hole coverage beyond such changes' effects on expected donut hole prices. Results are shown separately for high- and low-spending enrollees (for brevity, the regression for all years pooled is shown in the Table). In all year pairs but 2008–9, the "stark" donut hole coverage response is negative for both spending groups, i.e. *even among low-spending individuals* who have essentially zero probability of reaching the donut hole. The point estimates indicate that low-spending enrollees, observing that their plan dropped (added) generic or branded coverage to their plan benefit, would decrease (increase) annual consumption by 6% even though they never expect to encounter the donut prices. The "stark" coverage change response is smaller in magnitude for high-spending individuals, indicating a 2.5% decrease in spending among

Table 5
Results of full year ICR and donut price regressions, with stark donut coverage and deductible variables – Low- and high-spending enrollees.

	All years pooled				2006–7				2007–8				2008–9			
	Low-spend enrollees		High-spend enrollees		Low-spend enrollees		High-spend enrollees		Low-spend enrollees		High-spend enrollees		Low-spend enrollees		High-spend enrollees	
Price	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
ICR	−0.099	0.010 **	−0.026	0.008 **	−0.109	0.022 **	−0.014	0.016	−0.076	0.017 **	−0.023	0.012	−0.108	0.012 **	−0.041	0.012 **
Donut	0.043	0.014 **	−0.063	0.013 **	0.018	0.051	−0.081	0.022 **	0.032	0.022	−0.057	0.021 **	0.038	0.017 *	−0.066	0.023 **
Ded. Chg	−0.033	0.009 **	−0.023	0.008 **	0.016	0.023	−0.015	0.018	−0.043	0.013 **	−0.033	0.011 **	−0.046	0.013 **	−0.015	0.013
Stark	−0.059	0.015 **	−0.025	0.010 *	−0.036	0.021	−0.025	0.013	−0.085	0.024 **	−0.008	0.017	0.014	0.031	−0.019	0.019
N	1,326,301		307,004		200,758		64,065		556,878		132,030		568,659		110,909	

Notes: Results of full year regressions of log consumption change on log change in ICR and donut prices, as well as changes in stark gap coverage and deductible coverage. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

[29] For plans with any deductible in both years of the year pair, there is no within-phase price instrument variation because the deductible coinsurance equals 100%.

high-spending enrollees losing gap coverage. The results look somewhat different in 2008–9, as there is little in the way of "stark" gap coverage changes for that year pair; as discussed in detail in Appendix A.5, plans in 2008–9 generally had only slight alterations in their coverage in the donut hole.

We also observe a substantial effect of deductible coverage on consumption, except in 2006–7. The deductible is not marginal for the vast majority of enrollees, but all enrollees with deductibles spend some portion of the year in that phase, so that the deductible response may be due to myopia (overreacting to the individual-specific effective price change earlier in the year) or to salience (reacting to the deductible based on its visibility in benefit presentation rather than based on implied out-of-pocket price change). The deductible change response is larger in magnitude for low-spending than for high-spending enrollees, but not significantly so on average.[30]

Taken together, these results imply that high-level coverage changes (such as changes in deductible or donut generosity) have a large impact on individuals' behavior. The deductible response may be further evidence of myopia. However, the significant response to stark changes in donut hole coverage goes beyond what we would expect given a rational, forward-looking calculation of what those changes imply for marginal prices – our most striking evidence of price salience is that observed for low-spending individuals, who are very unlikely to encounter the donut hole during the year. Our results do not distinguish between a variety of underlying mechanisms that might be responsible for price salience – consumers may be failing to properly forecast future expenditures or they may be using donut hole coverage as a heuristic because the true coverage characteristics of plans are too complicated to calculate. Whatever the reason, our results suggest that beneficiaries respond to prices that are neither current nor marginal. We show in our structural model that this response is economically significant in magnitude.

5.4. Robustness

The previous reduced form analyses embedded several decisions regarding sampling, identification, and modeling. Here, we briefly summarize the results of several analyses we undertook to test robustness of these decisions. The results are described in greater detail in Appendix A.7.

First, we estimate a simple model of consumption as a function primarily of prices in two coverage phases – the initial coverage phase and the donut hole. Given sufficient variation, we could extend this approach as in Blomquist-Newey to estimate flexible nonlinear consumption responses to all budget set parameters. In the current empirical setting, we have substantial ICR and donut price variation, but limited variation in locations of budget set kinks. Therefore, in order to limit the impact of nonlinear price responses on our ICR and donut price coefficient estimates (e.g., individuals bunching at budget set kinks or endogenously responding to prices by switching their marginal coverage phases), we focused our attention on individuals whose spending can be predicted at the start of the year to be highly likely to lie in the interior of a given coverage range. In the first analysis presented in Appendix A.7, we limit the sample further to eliminate bias due to marginal coverage phase uncertainty or switching. We show that the estimates are not sensitive to more conservative sample restrictions that reduce the probability of individuals ending the year in the "wrong" coverage phase; hence, we consider our estimates based on high- and low-spending enrollees to perform well in capturing linear consumption responses within a nonlinear budget setting.

Second, our preferred instrumental variables approach leverages inertia in plan choice among those individuals already enrolled in any Part D plan. We consider this approach to be most appropriate for obtaining estimated price responses for current enrollees (and, in turn, for analyzing counterfactuals that will overwhelmingly impact current enrollees). However, in Appendix A.7.3, we also show robustness analyses in which we construct our instruments without holding plan enrollment fixed, and in which those who switch plans between years are eliminated from the sample. The former analysis is intended to analyze price responses across more and less inertial enrollees; the latter demonstrates the sensitivity of the results to the explicit exclusion of less inertial enrollees. Both sets of results are essentially unchanged, providing reassurance that the estimates are not biased due to the particulars of how our analysis deals with the small sample of enrollees that switch.

Finally, our richest reduced form specification includes two distinct but closely related price responses: the response to the individual-specific price in the donut hole, and the response to categorical changes in donut hole coverage. This decision is important to our analysis of the role of price salience in drug consumption, and one might be concerned that results could be driven in part by collinearity between the two "donut" variables. Reassuringly, as shown in Appendix A.7.4, the consumption response to the "Stark" variable is similar regardless of whether the specification includes a separate control for donut price.

6. Structural estimates

Our reduced form results provide evidence of significant price responses overall, substantial myopia, and price salience effects. In order to construct counterfactual estimates for individuals outside our regression sample (i.e., near the donut hole kink), we require a more general model of consumption behavior. Traditional candidates for this model would include marginal price response and average price response. Dynamic models of consumption along the lines of Aron-Dine et al., 2015 or Dalton et al. (2015) allow beneficiaries to respond either to the current spot price or the end of year marginal price. Motivated by our findings in Section 3.2, we also allow consumers to respond to current or future prices, and we allow price salience to factor into consumption patterns. Our reduced form parameters tell us how enrollees in-sample respond to price changes at different points in the budget set; the exercise in this Section relates the reduced form price coefficients to structural parameters that capture price sensitivity, myopia, and salience. The relationship between the structural parameters and the reduced form coefficients is mediated by the distribution of enrollees across the different budget set segments at different points in time. After estimating our dynamic structural model, we compare our results to marginal and average price models as well as models which allow for a myopic response to current prices or marginal prices but do not allow for price salience.

The dynamic "current-future" model with salience fits the data well both inside and outside the estimation sample and outperforms alternative models – both the rational marginal price response model and the behavioral average price response model (that might arise from "ironing"). The model performs well even though it is

[30] Note that we do not interpret these results as indicating a larger response by low-spending enrollees to changing deductible or gap coverage – consumption enters the regression in logs and thus the coefficients for high and low spenders are not directly comparable. For example, if we scale the pooled "Stark" coefficients by year 1 spending, the implied linear response to the "Stark" variable is −44 among low-spending enrollees versus −51 among high-spending enrollees. In a model with hyperbolic discounting, we would expect the deductible price response to vary inversely with the overall magnitude of spending, but given the standard errors and log specification, we cannot disentangle the effects of myopia and salience using the deductible coefficients. We do observe in Table 6 in the following Section that deductible responses are stronger at the beginning of the year than at the end of the year, which is stronger evidence that the deductible response is driven in part by myopia.

estimated using linear consumption responses; in Appendix A.7.2, we derive and estimate a structural model with bunching and coverage phase switching and show that the performance of the richer model is weakly worse than the simpler specification. We then use the preferred model to simulate the effects of filling in the donut hole.

One could interpret the "behavioral" patterns we document as arising from a model in which demand responds more strongly to donut hole price changes upon entering the donut, or directly to nominal donut hole price changes even conditional on "real" donut hole price changes, because consumers make mistakes in predicting their out of pocket costs or alternatively, because they are more attentive to some features of plans than others for a given forecast. For this reason, we focus on a counterfactual analysis that is likely to include the same behavioral responses we observe in the support of our data. These alternative psychological explanations may yield empirically distinguishable predictions in other contexts, however.

6.1. Parameter estimates

Inferring a set of structural parameters from our reduced form evidence requires comparing the estimated coefficients across individuals, coverage phases, and points in time, as discussed in Section 3.3. However, our specific context requires log-transforming prices and quantities. Thus, rather than using the relations in Eqs. (5)–(8) directly, this exercise requires an additional simple transformation. Specifically, we first differentiate the log-log model in Eq. (12) to determine what it implies for (locally) linear demand.

Letting z_y be the vector of year y prices, we linearize the specification around z_1. Details are given in Appendix A.6. This gives us an expression for each of the below coefficients from our structural model (omitting subscripts for the sake of brevity). These relations show that the linear structural constant γ varies with year 1 consumption and enrollee characteristics, so that this exercise yields a *local* linearization of enrollees' expected year 2 consumption as a function of year 1 observed consumption and predicted year-to-year trends given observables. The ICR and donut price coefficients vary as a function of observables as well as the ratio of year 1 consumption q_1 to each respective year 1 price.

$$\gamma = q_1 * \exp(\alpha_1 + X * \phi + u)(1 - \alpha_{ICR} - \alpha_{Don}) \tag{19}$$

$$\eta * (\beta * \Pr(MP = ICR) + (1-\beta) * \Pr(CP = ICR)) = \left(\frac{q_1}{p_{ICR,1}}\right) * \exp(\alpha + X * \phi + u) * \alpha_{ICR} \tag{20}$$

$$\eta * (\beta * \Pr(MP = Don) + (1-\beta) * \Pr(CP = Don)) = \left(\frac{q_1}{p_{Don,1}}\right) * \exp(\alpha + X * \phi + u) * \alpha_{Don} \tag{21}$$

$$\kappa_{Ded} = q_1 * \exp(\alpha + X * \phi + u) * \theta_{Ded} \tag{22}$$

$$\kappa_{Stark} = q_1 * \exp(\alpha + X * \phi + u) * \theta_{Stark} \tag{23}$$

Table 6 below displays the results of our richest reduced form specification, allowing for both myopia and salience (using the same identification strategy as in all results in Section 5.2): we regress year to year quarterly consumption changes on ICR and donut price changes as well as the deductible change and "Stark" variables, for high and low-spending enrollees only.[31] We allow ICR and donut price responses to vary by spending group, as the proportion of individuals for whom each phase is marginal or current in each period will differ across groups; we hold all other coefficients fixed across groups. These are the full set of reduced form results we use to infer structural parameters.

Table 6
Results of quarterly ICR and donut price regressions, with stark donut coverage and deductible variables – Low- and high-spending enrollees, all years pooled.

		All years pooled (pooled regression)			
		Low-spending enrollee response		High-spending enrollee response	
Period	Price	Coef	SE	Coef	SE
Q1	ICR	−0.053	0.006 **	−0.026	0.008**
Q1	Donut	0.019	0.007 **	−0.048	0.010**
Q1	Ded. Chg	−0.049	0.005 **	−0.049	0.005 **
Q1	Stark	−0.014	0.009	−0.014	0.009
Q2	ICR	−0.045	0.006 **	−0.011	0.009
Q2	Donut	0.026	0.007 **	−0.044	0.013 **
Q2	Ded. Chg	−0.024	0.005 **	−0.024	0.005 **
Q2	Stark	0.002	0.009	0.002	0.009
Q3	ICR	−0.043	0.006 **	−0.003	0.009
Q3	Donut	0.012	0.008	−0.083	0.012 **
Q3	Ded. Chg	−0.016	0.006 **	−0.016	0.006 **
Q3	Stark	−0.007	0.009	−0.007	0.009
Q4	ICR	−0.056	0.008 **	0.024	0.012 *
Q4	Donut	0.003	0.010	−0.168	0.016 **
Q4	Ded. Chg	−0.012	0.007	−0.012	0.007
Q4	Stark	−0.047	0.011 **	−0.047	0.011 **
	N	1,214,548			

Notes: Results of quarterly regressions of log consumption change on log change in ICR and donut prices, as well as changes in stark gap coverage and deductible coverage. Regression allows for separate ICR and donut responses for each spending group; "Stark" gap coverage and deductible change response held fixed across spending groups. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

The results shown in Table 6 are consistent with the patterns described in Section 5.2: low-spending enrollees' marginal price (ICR price) response is flat over the year and high-spending enrollees' marginal price (donut price) response is steeply increasing in magnitude over the year.[32] Enrollees' deductible response is decreasing in magnitude over the course of the year (this is consistent with myopia as enrollees only encounter deductible prices early in the year). Finally, the response to the "Stark" variable is non-monotonic over the year – it is largest in Q1 and Q4.[33]

We use the results of the above regression to infer our structural model parameters. We pool price response estimates across groups of observations (individual-quarters) whose price responses are expected to be similar by classifying sample observations by quintile of $\left(\frac{q_{it,1}}{p_{i,1}}\right) * exp(\hat{\alpha}_t + X_i * \hat{\phi}_t + \hat{u}_{it,2})$, for each price p_i, giving us 25 groups overall – this classification allows us to define groups of individuals based on similarity in their expected price changes, both

[31] The results are shown by year in Appendix Table A14.
[32] See Appendix Table A14 for results by year pair.
[33] This feature of the results is driven by low-spending enrollees' non-monotonic response to stark changes in donut hole coverage – as we see in Appendix Table A12, which compares the results in Table 6 to the same results obtained from separate regressions for the high- and low-spending samples, low-spending enrollees' response to the "Stark" variable is negative and significant in Q1 and Q4 (the Q4 response is larger, but not significantly so), while high-spending enrollees' response to the "Stark" variable is only significant in Q4. These results are consistent with the low-spending individuals' response to stark donut hole coverage being driven by salience effects and the high-spending individuals' response being driven at least in part by myopia.

in and outside the regression sample. We then estimate a single η_g for each group g.[34] We impose a single discount factor β across all sample individuals. Using the expressions including the parameters η and β in the above Taylor expansions (Eqs. (20) and (21) above), we use a GMM procedure to estimate our 25 static linear price response parameters η and our hyperbolic discount parameter β. Eqs. (20) and (21) require $\Pr\{CP = ICR\}$, $\Pr\{MP = ICR\}$, $\Pr\{CP = Donut\}$, and $\Pr\{MP = Donut\}$ as inputs – for the estimates reported, we assumed "perfect foresight" regarding current and marginal coverage phases, in that we used the actual observed probabilities of each phase being current or marginal in year 2 for each individual and each quarter.[35]

It is worth noting here that, even after allowing marginal price responses to vary across 25 groups of individuals, we are overidentified. We are inferring 28 structural parameters from 32 reduced form price responses, pooling information across individuals, coverage phases, and points in time. No simple structural model is expected to fit all coefficients exactly; for example, none of the models under consideration would predict the small positive donut price responses among low-spending individuals. The goal instead is to determine which model of behavior best fits all the observed data and to compare performance across models.

Counterfactual simulation requires that we extrapolate the structural parameters for individuals outside the regression sample (high and low spending enrollees). We extrapolate γ, κ_{Ded}, and κ_{Stark} using Eqs. (19), (22), and (23) above. We obtain η and β from the GMM procedure – individuals outside the regression sample falling in group g are assigned the structural static price response $\hat{\eta}_g$ and discount factor $\hat{\beta}$. In order to ensure that we extrapolate only to individuals who are similar to our regression sample, we exclude outliers, defined as individuals whose $exp(X_{it} * \hat{\phi}_t + \hat{u}_{it,2})$ lie below the $1st$ percentile or above the $99th$ percentile of the same metric in the regression sample.

Table 7 displays the mean values for all model parameter estimates for individuals in and out of the regression sample. At zero prices, our sample is predicted to consume 375 days supply of drugs per quarter. The average linear price response $\eta = -1.7$ implies an average static price elasticity of −0.13 (evaluated at marginal prices). The quarterly hyperbolic discount factor $\beta = 0.31$ suggests substantial myopia – it implies that, prior to entering their marginal coverage phase, individuals are more than twice as responsive to the price they currently face as they are to their marginal price. The 95% confidence interval for β ranges from 0.15 to 0.47, which would reject both fully forward-looking behavior and full myopia. Finally, on average, eliminating the deductible or adding stark donut hole coverage is predicted to increase consumption per quarter slightly, by 9 and 6 days supply, respectively.

The static price elasticity of −0.13 estimated using this procedure is similar in magnitude to the literature (see, e.g., Chandra et al., 2010), but is notably smaller than the static elasticities estimated in Einav et al. (2015) in the Part D context – they estimate elasticities ranging from −0.3 to −0.5. Their identification strategy infers price sensitivity from the magnitude of bunching behavior observed at the Part D kink and is thus focused on a small sample we explicitly exclude from our regression analysis – for this reason, we may not expect to arrive at the same behavioral elasticity.[36]

Our estimate of myopia, based on variation from enrollees across the spending and age distributions and price changes at different points in the budget set, is between those estimated in two recent papers using narrower variation. Einav et al. (2015) estimate a weekly exponential discount factor of 0.96 using variation in timing of enrollees joining Part D plans; if we assumed that this model of discounting were true, it would imply that an individual discounts all future quarters by an average factor of 0.49 across Q1–Q3, implying significantly more forward-looking consumption behavior than we estimate. In contrast, Dalton et al. (2015), relying on identification from behavior changes right at the donut kink, find evidence of full myopia.

In order to investigate the behavioral determinants of myopia and salience in the results in Table 7, we repeated our entire

Table 7
Estimated structural model parameters, all years pooled.

Description	Structural parameter	Central estimate	Standard error
Days supply ($P = 0$)	γ	374.910	9.147
Myopia	β	0.312	0.081
Marginal price effect	η	−1.663	0.147
Deductible effect	κ_{deduct}	−8.915	1.602
Stark gap effect	κ_{stark}	−6.409	2.708
Implied elasticity	ε	−0.126	0.011

Notes: Authors' calculations. N=1,985,127. Data are shown only for individuals within the 1st to 99th percentiles of the distribution of the predicted trend in consumption. All parameters shown are averages except for the hyperbolic discount factor. Standard errors from nonparametric bootstrap of entire procedure, 200 iterations.

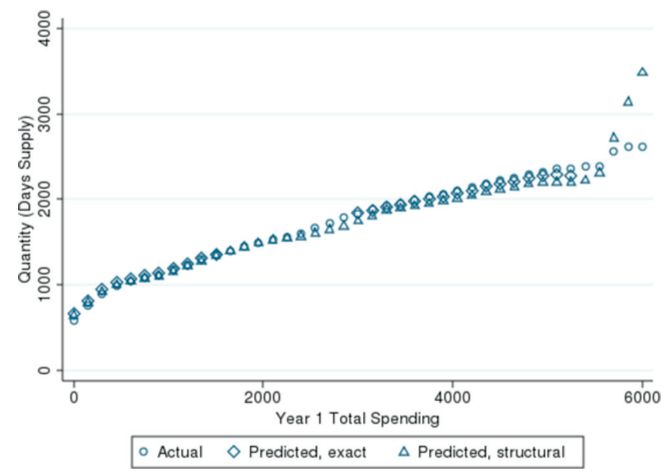

Fig. 2. Actual and predicted Year 2 quantity consumed – Outliers excluded.

[34] The error term $u_{it,2}$ is not directly observed, so we use Duan's smearing technique to scale all transformed coefficients based on the distribution of the regression residuals (Duan, 1983). We allow for heteroscedasticity and let the smearing factor vary in demographic variables.

[35] Of course, enrollees' expectations could instead be that they will consume in year 2 exactly as they did in year 1, or they could expect that they will be a random draw from the observed distribution of consumption among *similar* individuals based on year 1 consumption. Our estimates are not sensitive to this assumption – re-estimating the structural parameters using the individual's actual year 1 phase probabilities or "rational expectations" phase probabilities generates similar estimates, as discussed with regard to model fit below. In the rational expectations case, we classify individuals in each year pair into 100 cells by centiles of year 1 total spending and run the year 2 claims of 200 persons in each cell through the cost parameters for the plan for each individual in the cell and take the means of the resulting phase probabilities.

[36] In their follow-up to the 2015 paper, Einav et al. (2017) observe that different structural models of the behavioral response to the kink point yield very different structural elasticities, and thus different out of sample predictions for the response away from the kink point when prices change within coverage ranges. The model used in Einav et al. (2015) implies large away from kink elasticities while the model used in Saez (2010) implies much smaller elasticities. We estimate away from kink elasticities much closer to those implied by the Saez model.

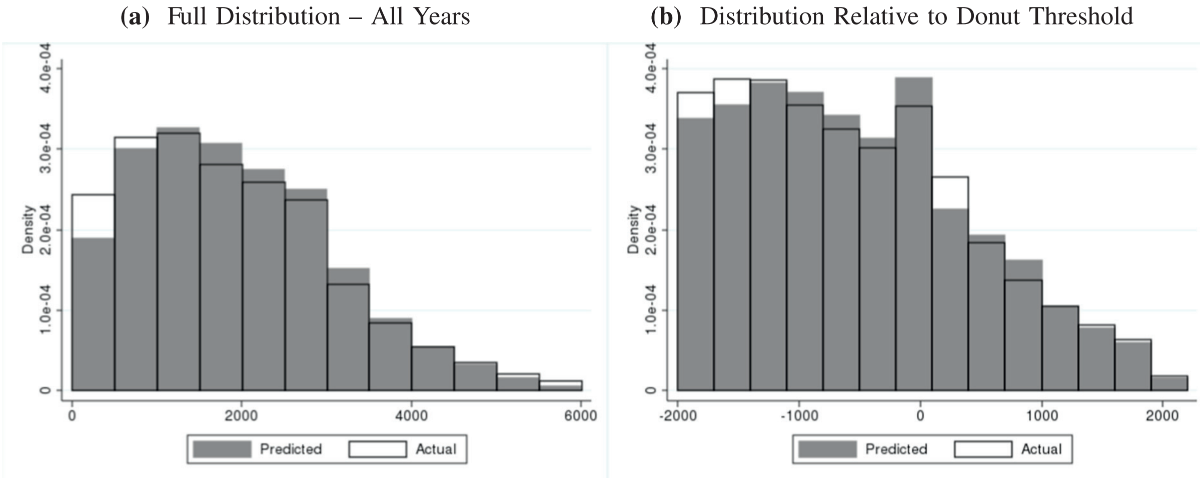

Fig. 3. Distribution of actual and predicted Year 2 spending – Outliers excluded.

estimation procedure separately for individuals with and without chronic conditions, and for groups of individuals defined by specific chronic conditions.[37] This allows us to examine whether individuals taking maintenance drugs on a predictable basis (as opposed to taking drugs for acute conditions) exhibit the same deviations from rationality as our overall sample. The results comparing the chronically ill to the non-chronically ill and the results for the most popular chronic conditions in our sample (hypercholesterolemia, hypertension, and diabetes) are shown in Appendix Table A13. The estimated hyperbolic discount factor β is larger in magnitude for the chronically ill (particularly among hypercholesterolemics and diabetics, who exhibit βs of 0.46 and 0.42, respectively) than for the non-chronically ill, which is consistent with the chronically ill being more forward-looking. However, the most striking feature of the table is the consistency in the parameter estimates across groups – even those with chronic conditions are substantially more sensitive to marginal prices after they encounter those prices than earlier in the year.

Using the model from Section 3 and the estimated structural parameters from Table 7, we solve for the optimal dynamic consumption path in response to the full nonlinear budget set for all sample individuals, including those near budget set kinks. Fig. 2 displays actual and predicted consumption (summed over the full year) for individuals throughout the year 1 spending distribution.[38] Two versions of the prediction are displayed – the exact quantity predicted by the regression model, in sample, and the simulated results using the structural model parameters. Both predictions work quite well for individuals in the regression sample (high and low-spending individuals) – the structural model under predicts actual year 2 spending by 0.34% on average. Notably, the structural model also replicates consumption for enrollees near the donut hole threshold quite well – we overpredict year 2 spending by 0.18% on average for this sample. The prediction performs less well as we approach the very high part of the spending distribution and is poor for the small number (1% of the sample) of non-outliers above the top of the "high" spending group (the $5000 cutoff).[39] In the counterfactual simulations, individuals above the "high-spending" cutoff are excluded.[40]

The absolute and relative distribution of actual and predicted spending are shown in Fig. 3.[41] Predicted spending is close to actual spending on average, as implied by the comparison in Fig. 2. The left panel of Fig. 3 shows that we slightly underpredict low spending and overpredict high spending. The right panel shows the distribution of actual and predicted spending relative to the donut threshold (recall that donut thresholds increase in each year of the sample). Even though we estimated price sensitivity using only individuals away from the donut threshold, we are able to replicate bunching at the donut kink quite well. The slight overprediction of bunching behavior exactly at the kink is due to our assumption of no uncertainty in the simulation model – allowing for uncertainty would yield some dispersion in excess mass around the kink, as we observe in the actual spending distribution just to the right of 0.

Our reduced form and structural analyses showed evidence of imperfectly forward-looking behavior, which rejects a model of consumers responding rationally to marginal price. However, the structural model we estimated above, while intuitive, could miss other behavioral consumption patterns, such as response to an average price as has been found in the empirical literature on electricity consumption. A comparison of the full structural model (with myopia and salience terms) to two alternative models – an average price response model and a marginal price response model – is shown in Table 8. The Table shows that the structural model (0.18% error) performs better in predicting out-of-sample spending than either the average price (6.82% error) or marginal price model

[37] Following Goldman et al. (2004), we identify seven chronic illnesses – hypercholesterolemia, hypertension, diabetes, gastritis, arthritis, asthma, and affective disorders – using diagnosis codes from the individuals' medical claims histories.

[38] Predicted consumption is simulated consumption throughout the year given actual year 2 prices.

[39] The small 1% sample of non-outliers in the above-$5,000 range accounts for 3% of non-outliers' spending. In the full sample, those spending above $5,000 account for 6% of enrollees and 18% of overall spending, as there is a very long tail to the expenditure distribution. Lack of overlap in covariates between the full range of very high-spending enrollees and our regression sample limits our ability to extrapolate price responses to that part of the distribution.

[40] Results are similar whether the perfect foresight approach is used as the model of expectations in the GMM estimation or whether, alternatively, we use the individual's year 1 consumption (0.3% error) or rational expectations (-0.4% error).

[41] Here, because coverage thresholds are defined based on total drug spending rather than quantity consumed, we compare actual and predicted spending rather than consumption as in Fig. 2.

Table 8
Actual and predicted Year 2 spending – Structural model, average price model, and marginal price model.

	Structural model [N=1,960,008]				Average price model [N=1,870,853]				Marginal price model [N=1,818,830]			
	Actual	Predicted			Actual	Predicted			Actual	Predicted		
		Mean	% diff	MSE		Mean	% diff	MSE		Mean	% diff	MSE
In-sample (Low/high spending enrollees)	1,718.239	1,703.561	−0.85%	996,251	1,755.345	1,799.548	2.52%	1,726,621	1,716.097	1,810.151	5.48%	1,705,993
Out-of-sample (Medium spending enrollees)	2,324.124	2,328.278	0.18%	990,186	2,320.353	2,478.690	6.82%	1,816,324	2,316.501	2,555.971	10.34%	1,915,258
All spending groups	1,972.521	1,965.746	−0.34%	993,706	1,993.213	2,085.467	4.63%	1,764,386	1,963.783	2,117.827	7.84%	1,792,322

Notes: Comparison of actual and simulated spending, structural model vs. average price model and marginal price model. In each comparison, data are shown only for individuals within the 1st to 99th percentiles of the distribution of the predicted trend in consumption. Average and marginal prices determined by applying perfect foresight weights to coverage phase-specific individual prices. The same controls as in Section 5 are used in the average and marginal price regressions.

Table 9
Estimated effect of filling in the donut hole, pooled all years.

Estimated impact of filling in the donut hole	\$118.47	5.91%
Based on price response alone	\$80.78	4.03%
Additional effect of price salience	\$37.69	1.88%

Notes: Authors' calculations. N=1,985,127. Data are shown only for individuals within the 1st to 99th percentiles of the distribution of the predicted trend in consumption. All parameters shown are averages except for the hyperbolic discount factor. Standard errors from nonparametric bootstrap of entire procedure, 200 iterations.

(10.34% error). Similarly, the mean-squared error in prediction (MSE) is much smaller for the structural model than either alternative, both in- and outside the regression sample.

6.2. Counterfactual – filling in the donut hole

We next use the estimated structural parameters to simulate the effect of filling in the donut hole on total spending for low and high-spending individuals as well as for individuals near the donut hole kink. We impose that "filling in the donut hole" takes the form of setting the donut hole price in each plan equal to the ICR price (so that there is no donut kink). For all individuals with no donut hole coverage or generic only donut hole coverage in year 1 of the relevant year pair, we also set the "Stark" variable equal to −1, as filling in the donut hole would entail a stark increase in donut hole coverage.[42]

Table 9 shows the effect of filling in the donut hole on mean spending, in levels and percentages. Filling in the donut hole would increase spending by \$118 on average, or 6%. \$81 of this increase is due to the price response (setting the donut price equal to the ICR price), but \$38 (or 32% of the overall effect) comes from what we call the "salience" effect, the coefficient on the stark increase in donut hole coverage.

Fig. 4 shows heterogeneity in the effect of filling in the donut. The left panel shows how the effect of filling in the donut varies with position in the spending distribution. That is, we plot the mean increase in spending when we fill in the donut versus year 1 spending. The price and salience effects are shown separately. The price effect is monotonically increasing in the magnitude of spending – individuals whose marginal price is the donut hole price are impacted more by the price change than individuals who do not hit the donut. Moreover, individuals who consume more of their prescription drugs while in the donut hole (higher spenders within the donut-marginal group) are more affected by the price change due to myopia. On the other hand, the "stark" or what we called the "salience" effect is present throughout the spending distribution, so that even low spending individuals are expected to increase spending in response to the donut hole being filled in. Notably, much of the effect of filling in the donut hole is due to price responses for

[42] An important caveat is necessary here. This counterfactual is performed (1) relative to observed year 2 prices in each sample year pair; and (2) holding all other plan features fixed. Thus, we should interpret the results as capturing *what 2007, 2008, and 2009 consumption would have been if each plan's donut hole were filled in, holding enrollment and other features of plan generosity fixed*. In order to determine the effect of changes to the donut hole generosity between now and 2020, one would need to account for trends in generosity between our sample period and the present, as well as how plans adjust other plan features in response to the imposed change in donut generosity (which may impact, in turn, sorting of patients across plans). Without a model to inform supply side responses to the ACA's donut policy changes, we leave the more sophisticated counterfactual to future research.

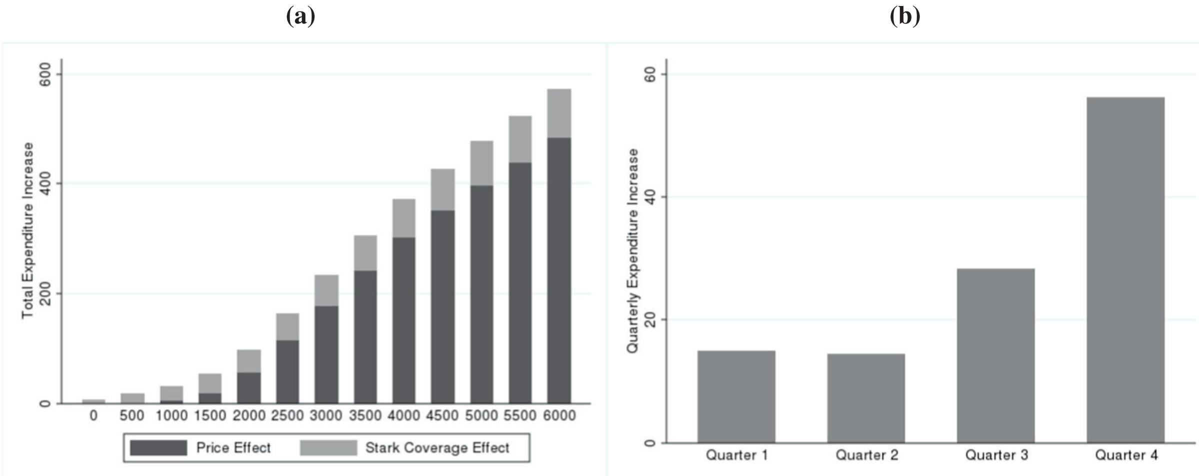

Fig. 4. Heterogeneity in effect of filling in the donut hole.

individuals who end the year well away from the kink and who would accordingly not exhibit bunching behavior in analyses that focus on the kink. For example, observations for enrollees ending a given year at a spending level within $200 of the donut threshold appropriate for that year account for only $16 (14%) of the $118 average effect of filling in the donut hole.

The right panel of Fig. 4 plots the mean increase separately by quarter. As implied by our estimate of the hyperbolic discount factor $\beta = 0.32$, individuals are much less responsive to filling in the donut hole at the beginning of the year than they are at the end of the year.

7. Discussion

We examine prescription drug consumption in the context of Medicare Part D, an insurance program in which enrollees are exposed to substantial cost-sharing incentives and in which nonlinear, complex price schedules are found to lead to additional price responses beyond those anticipated by the designers.

Our identification strategy allows us to estimate static and dynamic price elasticities using consumption responses to price variation in multiple regions of the nonlinear budget set. Part D enrollees are very unlikely to switch plans between years, giving us a plausibly exogenous source of identifying variation in the form of year-to-year changes in plan generosity for individuals already enrolled in Part D plans. Demand models under nonlinear budget sets can be challenging to estimate; in the absence of enough variation to estimate fully nonparametric responses, the solution is typically to specify a structural model of consumer optimization. In order to accommodate multiple behavioral models of consumption suggested by the nonlinear budget set literature in other contexts, we begin by using linear methods to estimate reduced form consumption responses to prices throughout the budget set. We focus on individuals whose marginal prices are highly likely to be in the interior of budget set segments in order to minimize the impact of nonlinear responses; analysis in the Appendix demonstrates that our results are robust to further sample restrictions and more complex modeling. We use the reduced form patterns to estimate a structural model with imperfect forward-looking behavior and price salience effects, and demonstrate that the model fits the data well; it outperforms both a fully rational model in which consumers respond to marginal price, as well as an alternative behavioral model in which consumers respond to average price. Notably, the model performs quite well outside the regression sample, which allows us to simulate consumption responses to counterfactual price schedules for enrollees throughout the spending distribution, including those near the donut hole kink.

We find that, while enrollees' static price elasticities are of a similar magnitude to estimates in the prior literature, Part D enrollees also exhibit certain behavioral biases in their consumption patterns related to the structure of Part D cost-sharing. In particular, we demonstrate evidence of imperfectly forward-looking behavior, in that enrollees are much more responsive to cost-sharing in current periods than in future periods. We also find that enrollees respond to more salient plan benefit changes, such as addition or removal of entire categories of drugs from donut hole coverage, beyond how those changes impact enrollees' actual out-of-pocket prices. Given that rational optimization of consumption in a setting such as Part D requires a complicated calculation with many inputs, these results may not be surprising.

The results described above yield some striking insights into insurance enrollees' decision-making, and carry important implications for policy. We find that, all else equal, filling in the Part D donut hole will increase spending by $114, or 6%, for the average enrollee in our sample. Over 30% of this effect is due to salience and, accordingly, impacts even relatively low-spending Part D enrollees. The remainder of the effect occurs primarily at the end of the year, due to imperfect forward-looking behavior, and falls disproportionately on higher-spending enrollees who are more likely to enter the donut hole.

Our findings suggest that prescription drug plan designers must carefully account for consumers' dynamic incentives and understanding of complex price schedules. The welfare implications of the current Part D plan design will be a function of the price responses documented here as well as the health impacts of altering drug consumption in response to prices and the overall program costs in- and outside Part D. We leave these topics for future research. This paper also developed a useful methodology for analyzing consumption in complex, nonlinear environments by estimating structural parameters using variation in linear regions of the budget set; we hope to extend the methodology to other applications, such as income taxation, in future work.

Appendix A

A.1. Price and instrument construction

In order to illustrate how our prices and price instruments are calculated, consider the following example. Suppose that, in 2006, the individual in question takes two drugs monthly, drug X and drug Y; in 2007, the individual also takes drug Z. In 2006, the individual is enrolled in plan A; in 2007, she switches to plan B. The retail prices and cost-sharing for Plans A and B, drugs X, Y, and Z, and years 2006 and 2007 are shown in Appendix Table A1. As we see in the Table, plan A has coverage of generics (drug Y only) in the donut hole in both years, while drug B has no donut hole coverage in either year. In general, both retail prices and copays are different across plans for each drug and across years for each plan-drug.

Table A1
Retail prices and out-of-pocket costs for example plans and drugs.

	2006 Retail and out-of-pocket prices					2007 Retail and out-of-pocket prices				
	Plan A, 2006					Plan A, 2007				
Drug	*Retail price*	*Deductible*	*ICR*	*Donut hole*	*Catastrophic*	*Retail price*	*Deductible*	*ICR*	*Donut hole*	*Catastrophic*
X	100.00	100.00	30.00	100.00	5.00	110.00	110.00	31.00	110.00	5.50
Y	30.00	30.00	10.00	10.00	2.00	32.00	32.00	12.00	12.00	2.15
Z	150.00	150.00	45.00	150.00	7.50	160.00	160.00	45.00	160.00	8.00
	Plan B, 2006					Plan B, 2007				
	Retail price	*Deductible*	*ICR*	*Donut hole*	*Catastrophic*	*Retail price*	*Deductible*	*ICR*	*Donut hole*	*Catastrophic*
X	115.00	115.00	40.00	115.00	5.75	117.00	117.00	40.00	117.00	5.85
Y	25.00	25.00	12.00	25.00	2.00	27.00	27.00	8.00	27.00	2.15
Z	130.00	130.00	50.00	130.00	6.50	130.00	130.00	55.00	130.00	6.50

The corresponding phase-year-specific prices and instruments are shown for each drug and on average across all drugs the individual takes in Appendix Table A2. Recall that our procedure requires a single price and instrument for each individual-year. For the "actual" prices, we aggregate the prices in plan A in 2006 using 2006 weights to obtain the 2006 average prices (e.g., the ICR price in 2006 is $20, the average copay in plan A across drugs X and Y) and we aggregate the prices in plan B in 2007 using 2007 weights to obtain the 2007 average prices (e.g., the donut price in 2007 is $91.33, the average copay in plan B across drugs X, Y, and Z). However, to obtain the instruments, we hold plan choice, retail price, and consumption weights fixed at 2006 values. Hence, the price instrument in 2006 is the same as the actual price in 2006 in each phase, and the 2007 price instruments differ from the 2006 price instruments only insofar as plan A's generosity changed between years 2006 and 2007 holding retail price fixed. In this example, there is no deductible price change in the instrument for any drug or on average, the 2006–2007 price change in the ICR equals the change in plan A's copays 2006–2007, and the donut price only changes for drug Y, the drug with some donut coverage in plan A.

Table A2
Average prices and price instruments for example individual.

		2006 Retail and out-of-pocket prices					2007 Retail and out-of-pocket prices				
"Actual" prices	Drug	*Quantity*	*Deductible*	*ICR*	*Donut hole*	*Catastrophic*	*Quantity*	*Deductible*	*ICR*	*Donut hole*	*Catastrophic*
	X	30.00	100.00	30.00	100.00	5.00	30.00	117.00	40.00	117.00	5.85
	Y	30.00	30.00	10.00	10.00	2.00	30.00	27.00	8.00	27.00	2.15
	Z	0.00	150.00	45.00	150.00	7.50	30.00	130.00	55.00	130.00	6.50
	Average		65.00	20.00	55.00	3.50		91.33	34.33	91.33	4.83
Price instruments		*Quantity*	*Deductible*	*ICR*	*Donut hole*	*Catastrophic*	*Quantity*	*Deductible*	*ICR*	*Donut hole*	*Catastrophic*
	X	30.00	100.00	30.00	100.00	5.00	30.00	100.00	31.00	100.00	5.35
	Y	30.00	30.00	10.00	10.00	2.00	30.00	30.00	12.00	12.00	2.15
	Z	0.00	150.00	45.00	150.00	7.50	0.00	150.00	45.00	150.00	7.50
	Average		65.00	20.00	55.00	3.50		65.00	21.50	56.00	3.75

A.2. Sample restrictions

Appendix Table A3 below describes our sample restrictions. The first column describes the full set of all unsubsidized individuals continuously enrolled in standalone PDPs for two years of any year pair 2006–7 through 2008–9. The second column describes the sample after removing individuals with no claims in either year of a year pair. The third column describes the set of high- and low-spending individuals whom we rely on for our reduced form and structural models; these individuals are highly likely to end the year in the interior of a budget set segment. Finally, for the sake of contrast, the fourth column describes individuals ending the year near the budget set kink. These are the individuals whose data are relied upon in "bunching" estimation strategies; for example, Einav et al. (2017) focus on a window of $200 around the kink.

Table A3
Sample restrictions and comparison.

	Full continuously- enrolled Part D sample		Sample with Part D claims in Year 1 + Year 2		Low/high groups		Near kink sample	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Num. beneficiaries	2,917,997		2,707,514		1,633,405		267,372	
Age	75	7	75	7	75	7	75	7
% White	95%	22%	95%	22%	95%	22%	95%	21%
% Female	62%	49%	63%	48%	62%	48%	62%	49%
Part D expenditure	$1869	$2421	$2009	$2456	$1315	$1339	$2417	$144
Part D days supply	1158	872	1245	844	988	759	1,495	598
Hypertension	38%	49%	40%	49%	38%	48%	44%	50%
Hypercholesterolemia	16%	37%	17%	37%	16%	37%	19%	39%
Diabetes	14%	35%	15%	36%	13%	33%	19%	39%
Arthritis	10%	30%	11%	31%	10%	30%	12%	33%
Any chronic	62%	49%	65%	48%	61%	49%	71%	45%
Parts A+B expenditure	$7353	$14,656	$7751	$14,986	$6523	$13,573	$8890	$15,419

Notes: Enrollee summary statistics for several different samples. The first sample is a full 20% random sample of unsubsidized, elderly Part D enrollees enrolled continuously in standalone PDPs throughout both years of any year pair 2006–2007 through 2008–2009. Medicare Advantage and employer-sponsored enrollees excluded. The second sample further excludes enrollees with zero claims in either year of the year pair. The third sample limits to enrollees in the "Low" and "High" spending groups, as defined in text. The fourth sample includes only individuals ending the first year of each year pair with total spending within $200 of the standard plan donut threshold.

While demographic characteristics are strikingly similar across all four samples, the heavy-tailed nature of health care utilization and spending implies that the samples differ substantially in these terms. The high/low and kink samples are about equally dissimilar from the full sample in terms of year 1 Part D expenditure (30% higher and 30% lower, respectively); however, along other dimensions of illness severity, the high/low sample is closer than the kink sample. The high/low sample has 11% lower Parts A and B spending, and 15% lower days supply, than the full sample; on these dimensions, the kink sample is 20–30% higher. The high/low sample also matches the distribution of chronic conditions in the full sample much more closely than the kink sample.

A.3. Indirect identification tests

As discussed in Section 4.2, our identification relies on changes in plans' cost-sharing parameters in different segments of the budget set. Our exclusion restriction would be violated if individuals choose their initial plans based on anticipated changes in plan cost-sharing the next year, or if changes in plan generosity are correlated with characteristics of enrollees that are unobserved to the econometrician and that predict differential trends in consumption.

In order to examine these potential violations, we perform two sets of analyses. First, we follow Abaluck and Gruber (2011) and estimate a model of plan choice among Part D enrollees. Using a 2% random sample of beneficiaries in the final choice sample in Abaluck and Gruber (2011), we estimate choice as a function of the following *current year* variables: annual premium, deductible and donut hole coverage, number of top 100 drugs on formulary, a normalized measure of plan quality, a summary cost-sharing measure, predicted out of pocket cost, variance in predicted out of pocket cost, and a dummy for the plan being chosen in the previous year (to capture inertia). We also add to the regression a measure of the beneficiary's predicted out of pocket cost in the *following* year. The results are shown in the left panel of Appendix Table A4, which displays the estimated coefficients on annual premium, current year out of pocket cost, and future out of pocket cost. While the estimate for future out of pocket cost is statistically significant, it is an order of magnitude smaller than the current year coefficients of interest and is thus economically very small. To put the number in perspective, a $100 increase in future out of pocket cost has the same effect on plan choice as a $4 increase in current year annual premium. We consider this effect of future costs on choice to be unlikely to meaningfully bias our results.

Table A4
Plan choice and pre-trend regressions.

	Plan choice model		Pre-trends regression						
				Quantity		Quantity		Quantity	
	Year 2 - Year 1			Year 2 - Year 1		Year 1 - Year 0		Year 1 - Year 0	
Characteristic	Coef	SE	Price	Coef	SE	Coef	SE	Coef	SE
Current premium	−0.0061	0.0005**	Average price Year 2 - Year 1	−0.058	0.010**	−0.009	0.013	0.023	0.026
Current OOP	−0.0031	0.0002**							
Future OOP	−0.0003	0.0000**							
			Control for current price change?	n/a		Yes		Yes, IV	
N enrollee-years	52,611		N enrollee-years	485,807		485,807		485,807	

Notes: Left panel: plan choice regression using 2% sample from Abaluck and Gruber (2011), enrollees with two consecutive years' enrollment only. Right panel: consumption change regression using full sample described in Table 1, enrollees with three years of continuous enrollment only. Superscript (**) indicates significance at the 1% level; superscript (*) indicates significance at the 5% level.

Though we cannot test directly for latent differential trends in consumption among individuals whose cost-sharing parameters change between years, we can examine whether our estimated effects may be a continuation of differential pre-trends. To explore this, we limit our sample to enrollees with continuous plan enrollment in three years between 2006 and 2009 and for whom prices *did not change* between their first two years of enrollment (i.e., the price change from year 0 to year 1 was below 5%), and we run the following specifications:

$$log(q_{i,2}) - log(q_{i,1}) = \alpha_1 + \alpha_2^{curr}(log(p_{j,2}) - log(p_{j,1}))$$
$$log(q_{i,1}) - log(q_{i,0}) = \alpha_1 + \alpha_2^{curr}(log(p_{j,2}) - log(p_{j,1})).$$

In each regression, for the sake of parsimony, we analyze the effect of a change in average price (weighting each budget set segment by the proportion of quantity consumed in that segment) on the year to year change in quantity consumed. We control for all of the same covariates as in our main specifications. In the first regression, the first column in the right panel of Appendix Table A4, we see that a 10% change in average price between year 1 and year 2 is associated with a 0.6% decrease in quantity consumed between year 1 and year 2. This is roughly similar to our results in Section 5. However, in the second column, we see that a 10% change in average price between year 1 and year 2 is associated with only a 0.09% decrease in quantity consumed between year 0 and year 1, and this decrease is not statistically significant.[43] Given these results, we find it unlikely that the estimated effects of price changes on quantity changes are simply due to differential trends in consumption.

A.4. Comparison of OLS, reduced form, and IV regressions

Appendix Table A5 shows a comparison of the ordinary least squares, reduced form, first stage, and instrumental variables regressions of log quarterly change in consumption on log price (or instrumental variable price) change in the ICR and donut regions of the budget set. The first column estimates the OLS version of Eq. (9). The second column estimates the reduced form Eq. (24):

$$\begin{aligned} log(q_{it,y}) - log(q_{it,y-1}) = \alpha_{1t}^{RF} &+ \alpha_{ICR,ty}^{RF}\left(log\left(p_{ICR,j,y}^{IV}\right) - log\left(p_{ICR,j,y-1}^{IV}\right)\right) \\ &+ \alpha_{Donut,ty}^{RF}\left(log\left(p_{Donut,j,y}^{IV}\right) - log\left(p_{Donut,j,y-1}^{IV}\right)\right) \\ &+ X_{ity}\phi_{ty}^{RF} + u_{ity}^{RF}. \end{aligned} \tag{24}$$

The third and fourth columns show the first stage estimates of Eqs. (10) and (11), with the ICR and donut price changes as the dependent variables, respectively. Finally, the IV estimates of Eq. (9) are in the fifth column. In the main text, these estimates are also shown in Tables 3 and 4.

Table A5
Results of quarterly ICR and donut price regressions – Low-spending and high-spending groups.

	Period	Price/IV price	Ordinary least squares		Reduced form		First stage (ICR)		First stage (Donut)		Instrumental variables	
			Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Low-spending	Q1	ICR	0.008	0.001**	−0.023	0.003**	0.407	0.013**	−0.135	0.008**	−0.051	0.006**
	Q1	Donut	0.036	0.001**	0.006	0.006	0.142	0.018**	0.748	0.029**	0.023	0.008**
	Q2	ICR	−0.009	0.002**	−0.022	0.003**	0.409	0.013**	−0.133	0.008**	−0.041	0.007**
	Q2	Donut	0.062	0.002**	0.018	0.007**	0.141	0.018**	0.746	0.029**	0.026	0.008**
	Q3	ICR	−0.016	0.002**	−0.019	0.003**	0.409	0.013**	−0.132	0.008**	−0.043	0.007**
	Q3	Donut	0.077	0.002**	0.002	0.007	0.140	0.018**	0.745	0.029**	0.008	0.009
	Q4	ICR	−0.025	0.002**	−0.019	0.003**	0.403	0.014**	−0.140	0.009**	−0.051	0.009**
	Q4	Donut	0.074	0.002**	−0.015	0.007*	0.144	0.019**	0.750	0.029**	−0.011	0.011
High-spending	Q1	ICR	0.030	0.003**	−0.011	0.005*	0.479	0.009**	−0.094	0.008**	−0.034	0.010**
	Q1	Donut	−0.028	0.003**	−0.042	0.010**	0.091	0.017**	0.742	0.024**	−0.053	0.014**
	Q2	ICR	0.010	0.003**	−0.004	0.005	0.479	0.009**	−0.094	0.008**	−0.023	0.011*
	Q2	Donut	−0.048	0.003**	−0.056	0.011**	0.085	0.017**	0.735	0.023**	−0.073	0.015**
	Q3	ICR	−0.005	0.004	−0.003	0.006	0.481	0.009**	−0.091	0.008**	−0.023	0.012
	Q3	Donut	−0.092	0.003**	−0.065	0.013**	0.085	0.017**	0.734	0.023**	−0.086	0.018**
	Q4	ICR	0.006	0.004	0.011	0.007	0.478	0.010**	−0.094	0.008**	−0.008	0.014
	Q4	Donut	−0.155	0.004**	−0.124	0.015**	0.098	0.017**	0.751	0.024**	−0.164	0.025**

Notes: Results of quarterly OLS, reduced form, first stage, and instrumental variables regressions of log consumption change on log change in ICR and donut prices (or, in the reduced form and first stage cases, the log change in ICR and donut IV prices), separately for low-spending individuals and high-spending individuals. Individuals with positive consumption in each quarter only. All years pooled. N=919,650 for low-spending group; N=294,898 for high-spending group. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

The first stage estimates are fairly consistent across quarter and spending group: the actual ICR price change has a strong positive association with the ICR price change IV, and the actual donut price change has a strong positive association with the donut price change IV. The coefficients are approximately 0.4 and 0.75, respectively.

The patterns in the IV results generally track the patterns in the reduced form results across spending group, coverage phase (ICR vs. donut), and quarter combinations. E.g., in both the reduced form and IV results, low-spenders' response to the ICR price change is negative and significant, but flat over time, while their response to the donut hole price change is small and positive. Similarly, the reduced form and IV

[43] In the right panel of Appendix Table A4, we instrument for average price change for each year pair using the change that would result holding consumption weights, enrollment, and retail prices fixed at their year 1 levels. In the second column, we control for the price change IV between year 0 and year 1. In the third column, we instrument for the price change between year 0 and year 1 using that year pair's average price IV. The results are not statistically significantly different depending on how we control for the first year pair's price.

results show that high-spenders' response to the ICR price change is low and decreasing in magnitude (and, conversely, their response to the donut price change is steeply increasing in magnitude) over the course of the year, as enrollees exit the ICR and enter the donut.

Both the reduced form and IV results deviate somewhat from the OLS results, which are often positive, contrary to expectations about downward-sloping demand curves (a significant exception is high-spenders' response to the donut price change). This discrepancy between the OLS and IV results is likely generated by the fact that, within each year pair over which differences in prices and consumption are taken, the instrumental variables prices for both years hold consumption weights and retail prices fixed at the first year's levels. The OLS approach, in not doing so, permits an endogeneity problem wherein, for example, enrollees experiencing upward shocks in health care needs require more expensive drugs, inducing a positive association between the year-to-year consumption change and price change.

A.5. Categorical donut hole coverage

As noted in the text, some of the donut hole coverage variation observed in our sample pertains to broad categories of drugs and would be easily understood by enrollees, but the sample also includes many changes that would be difficult for enrollees to translate into prices.

In 2006–7, all donut hole coverage changes are of the former, "stark" variety, in that they entail plans adding or dropping an entire category or more of drugs to the donut hole coverage; for example, consider Appendix Table 6a, which shows the count of sample enrollees in 2006–7 by 2006 gap coverage (across rows) and 2007 gap coverage (across columns). The Table shows that 35,325 individuals were in plans with no donut hole coverage in 2006 and generic donut hole coverage in 2007 – adding generic coverage implies large average price decreases for the donut hole of about $10 per 30-day supply. These decreases would be simple for individuals to understand given knowledge of the prices they face for branded and generic drugs in the ICR and an understanding of which drugs are generic. In contrast, some plans in 2007–8 and even more in 2008–9 had slight alterations in their coverage in the donut hole which did not entail large average price changes and which were not generally easily understandable; see the count of individuals according to year 1 and year 2 gap coverage in Appendix Table 6b and Appendix Table 6c. For example, 26,551 enrollees changed from "All Generic" coverage in 2008 to "Many Generic" coverage in 2009. This did not serve to universally increase average prices in the donut hole – some plans still decreased copays for covered generics while removing coverage for others – and the average price increase across plans was small, around $0.56. Further, this type of coverage change would require a more complicated calculation for individuals to respond to it than a stark coverage change such as removing/adding coverage for an entire class of drugs – it is arguably surely easier for enrollees to identify which drugs are branded and generic than to identify which generic drugs are "Many Generic" or which branded drugs are "Few Brand."[44]

Table 6a
Enrollment by donut hole coverage changes, 2006–7.

		2007 Gap coverage, 2006–7 sample			
		Brand & gen	Generics, pref. brands	Generic	None
2006 Gap coverage, 2006–7 sample	Brand & gen	7,832	18	12,895	6,773
	Generic	152	132	31,172	3,093
	None	1,851	116	35,325	352,273

Notes: Count of sample enrollees with given gap coverage in 2006 and 2007 chosen plan(s). 2006 gap coverage designated by row value; 2007 gap coverage designated by column value.

Table 6b
Enrollment by donut hole coverage changes, 2007–8.

		2008 Gap coverage, 2007–8 sample				
		Some generics	All generics and some brands	Generics, pref. brands	Generic	None
2007 Gap coverage, 2007–8 sample	Brand & gen	89	1	666	7556	7746
	Generics, pref. brands	1	0	392	7	136
	Generic	27,909	19	62,653	52,386	24,629
	None	218	9	2690	5211	934,364

Notes: Count of sample enrollees with given gap coverage in 2007 and 2008 chosen plan(s). 2007 gap coverage designated by row value; 2008 gap coverage designated by column value.

Table 6c
Enrollment by donut hole coverage changes, 2008–9.

		2009 Gap coverage, 2008–9 sample					
		Some generics	All generics	All generics and few brands	Many generics	Many generics and few brands	None
2008 Gap coverage, 2008–9 sample	Some generics	985	48	0	25,710	0	1651
	All preferred generics	136	252	14	55,588	384	6336
	All generics	701	30,506	44	26,551	0	7230
	All generics and some brands	28	0	0	0	0	13
	None	3099	775	55	2452	2	966,640

Notes: Count of sample enrollees with given gap coverage in 2008 and 2009 chosen plan(s). 2008 gap coverage designated by row value; 2009 gap coverage designated by column value.

[44] The fact that essentially none of the donut coverage variation in 2008–9 is of this stark form (in contrast to 2006–7 and 2007–8) may account for the observation in Table 4 that end of year marginal price responses are somewhat smaller at the end of the year for the 2008–9 sample than for the 2006–7 and 2007–8 samples.

In incorporating donut coverage variation in our sample into our regression analysis, we focus on the sort of "stark" variation observed primarily in 2006–7 and 2007–8, as it maps more clearly into price changes and is more likely to be understood by enrollees. Sensitivity analysis where we include separate coverage change variables for "stark" coverage changes and for "non-stark" coverage changes leave the included variables unchanged – the "non-stark" variables capturing subtle donut hole price changes are not statistically significant.

A.6. Taylor expansion in structural model

Letting $\mathbf{z_y}$ be the vector of year y prices, we linearize the specification in Eq. (12) around $\mathbf{z}_1$. We use the following Taylor expansion:

$$\begin{aligned} q_{it,2}(\mathbf{z_2}) = f(\mathbf{z_2}) &= q_{it,1} * exp\left(\alpha_{1t} + X_{it1}\phi_t + \alpha_{ICR,t} * \left(\frac{log(p_{ICR,j,2})}{log(p_{ICR,j,1})}\right)\right) \\ &* q_{it,1} * exp\left(\alpha_{Donut,t} * \left(\frac{log(p_{Donut,j,2})}{log(p_{Donut,j,2})}\right)\right) \\ &* q_{it,1} * exp\left(\theta_{Ded,t} * \Delta\{Ded_i\} + \theta_{Stark,t} * \Delta\{Stark_i\}\right) \\ &* q_{it,1} * exp(u_{it,2}) \\ &\cong f(\mathbf{z_1}) + \left(\frac{\partial f(\mathbf{z_1})}{\partial p_{ICR}}\right)(p_{ICR,j,2} - p_{ICR,j,1}) + \left(\frac{\partial f(\mathbf{z_1})}{\partial p_{Don}}\right)(p_{Don,j,2} - p_{Don,j,1}) \\ &+ \left(\frac{\partial f(\mathbf{z_1})}{\partial \Delta\{Ded\}}\right)\Delta\{Ded_i\} + \left(\frac{\partial f(\mathbf{z_1})}{\partial \Delta\{Stark\}}\right)\Delta\{Stark_i\}. \end{aligned}$$

The Taylor expansion yields the following form for year 2 consumption:

$$\begin{aligned} q_{it,2}(\mathbf{z_2}) &\cong q_{it,1} * \exp\left(\alpha_{1t} + X_{it1}\phi_t + u_{it,2}\right) \\ &+ q_{it,1} * exp\left(\alpha_{1t} + X_{it1}\phi_t + u_{it,2}\right) * \left(\frac{p_{ICR,j,2} - p_{ICR,j,1}}{p_{ICR,j,1}}\right) * \alpha_{ICR,t} \\ &+ q_{it,1} * exp\left(\alpha_{1t} + X_{it1}\phi_t + u_{it,2}\right) * \left(\frac{p_{Don,j,2} - p_{Don,j,1}}{p_{Don,j,1}}\right) * \alpha_{Don,t} \\ &+ q_{it,1} * exp\left(\alpha_{1t} + X_{it1}\phi_t + u_{it,2}\right) * \left(\Delta\{Ded_i\} * \theta_{Ded,t} + \Delta\{Stark_i\} * \theta_{Stark,t}\right) \\ &= q_{it,1} * exp\left(\alpha_{1t} + X_{it1}\phi_t + u_{it,2}\right)(1 - \alpha_{ICR,t} - \alpha_{Don,t}) \\ &+ \left(\frac{q_{i,1}}{p_{ICR,j,1}}\right) * exp\left(\alpha_{1t} + X_{it1}\phi_t + u_{it,2}\right) * p_{ICR,j,2} * \alpha_{ICR,t} \\ &+ \left(\frac{q_{i,1}}{p_{Don,j,1}}\right) * exp\left(\alpha_{1t} + X_{it1}\phi_t + u_{it,2}\right) * p_{Don,j,2} * \alpha_{Don,t} \\ &+ q_{it,1} * exp\left(\alpha_{1t} + X_{it1}\phi_t + u_{it,2}\right) * \left(\Delta\{Ded_i\} * \theta_{Ded,t} + \Delta\{Stark_i\} * \theta_{Stark,t}\right). \end{aligned}$$

A.7. Robustness

In this Section, we explore the sensitivity of our methodology to the assumptions that enrollees in our estimation sample have no uncertainty about marginal coverage phase and that current and marginal coverage phase assignments are fixed and exogenous with respect to out-of-pocket prices.

A.7.1. Limited uncertainty about marginal coverage phase

In the regression estimations used to obtain our structural parameter estimates, we noted that our restriction to "low" and "high" spending individuals does not guarantee that those individuals have no uncertainty about marginal coverage phase. "Low" spending individuals cross the year 2 donut threshold 3% of the time; "high" spending individuals do *not* cross the year 2 donut threshold 14% of the time. In this Section, we investigate sensitivity to classification of the low and high-spending individuals to determine whether uncertainty or switching behavior is likely to bias our results. In order to perform this check, we first regress an indicator for ending the year in the "appropriate" (ICR for low-spending individuals, donut for high-spending individuals) coverage phase on all demographic and spending controls. We reduce the "error", as defined by the individual ending the year on the wrong side of the donut threshold, by 50% and by 25% by restricting the sample based on the predicted probability of ending in the appropriate marginal coverage phase, and display the results of the quarterly regression for these samples.

The results are shown in Appendix Table A7. Of the 48 coefficients estimated, two are statistically significantly different than in the baseline sample – the deductible change response is larger in magnitude in the restricted samples, significantly so in the 25% restricted sample for Q2, and the Q2 ICR response by low-spending individuals is significantly smaller in the 50% restricted sample. Otherwise, the point estimates are similar in magnitude and exhibit similar dynamic patterns – the marginal price coefficients in the 25% restricted sample are slightly smaller in magnitude, while the marginal price coefficients in the 50% restricted sample are sometimes smaller, sometimes larger, than in the baseline sample.

Table A7
Results of quarterly ICR and donut price regressions, with stark donut coverage and deductible variables – High- and low-spending enrollees, pooled years, baseline and restricted samples.

		Baseline thresholds [N=1,214,548]				Pred (Err. reduce 25%) [N=988,123]				Pred (Err. reduce 50%) [N=700,114]			
		Low-spending enrollee response		High-spending enrollee response		Low-spending enrollee response		High-spending enrollee response		Low-spending enrollee response		High-spending enrollee response	
Period	Price	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Q1	ICR	−0.056	0.005 **	−0.022	0.007 **	−0.056	0.006 **	−0.019	0.009 *	−0.049	0.007 **	−0.023	0.013
Q1	Donut	0.030	0.007 **	−0.040	0.009 **	0.027	0.007 **	−0.032	0.011 **	0.021	0.009 *	−0.031	0.017
Q1	Ded. Chg	−0.035	0.005**	−0.035	0.005 **	−0.046	0.006 **	−0.046	0.006 **	−0.047	0.007 **	−0.047	0.007 **
Q1	Stark	−0.015	0.006 *	−0.015	0.006 *	−0.026	0.012 *	−0.026	0.012 *	−0.009	0.015	−0.009	0.015
Q2	ICR	−0.051	0.005 **	−0.007	0.007	−0.044	0.006 **	−0.002	0.009	−0.029	0.007 **	0.011	0.015
Q2	Donut	0.029	0.007 **	−0.051	0.015 **	0.030	0.008 **	−0.029	0.016	0.025	0.009 **	−0.029	0.023
Q2	Ded. Chg	−0.009	0.005	−0.009	0.005	−0.027	0.006 **	−0.027	0.006 **	−0.014	0.007 *	−0.014	0.007*
Q2	Stark	0.001	0.008	0.001	0.008	0.010	0.011	0.010	0.011	0.011	0.015	0.011	0.015
Q3	ICR	−0.048	0.005 **	0.002	0.008	−0.038	0.006 **	0.016	0.010	−0.036	0.007 **	0.008	0.016
Q3	Donut	0.014	0.007	−0.091	0.014 **	0.011	0.008	−0.075	0.015 **	0.001	0.010	−0.085	0.020 **
Q3	Ded. Chg	−0.008	0.005	−0.008	0.005	−0.016	0.006 **	−0.016	0.006 **	−0.018	0.008 *	−0.018	0.008 *
Q3	Stark	0.000	0.007	0.000	0.007	−0.005	0.011	−0.005	0.011	−0.014	0.014	−0.014	0.014
Q4	ICR	−0.062	0.007 **	0.032	0.010 **	−0.054	0.008 **	0.050	0.013 **	−0.052	0.009 **	0.050	0.019 **
Q4	Donut	0.006	0.010	−0.183	0.017 **	0.007	0.011	−0.172	0.026 **	0.007	0.012	−0.184	0.039 **
Q4	Ded. Chg	−0.004	0.005	−0.004	0.005	−0.015	0.007 *	−0.015	0.007 *	−0.011	0.008	−0.011	0.008
Q4	Stark	−0.041	0.009 **	−0.041	0.009 **	−0.058	0.014**	−0.058	0.014 **	−0.071	0.017 **	−0.071	0.017**

Notes: Results of quarterly regressions of log consumption change on log change in ICR and donut prices, as well as changes in stark gap coverage and deductible coverage, main regression estimation sample ("Baseline") vs. more restricted samples. The sample denoted "Error reduce 25%" uses the results of a regression of an indicator for ending the year in the appropriate coverage phase on year 1 demographic and spending variables to reduce the "Error" rate (defined as ending the year on the wrong side of the donut threshold) by 25%. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

On balance, the results of this specification check indicate that our results are not sensitive to further restriction of our sample to reduce marginal coverage phase switching behavior.

A.7.2. Allowing current and marginal coverage phase to respond to out-of-pocket prices

Recall the derivation of the dynamic structural model in Section 3.1. The solution for optimal consumption shown in Eq. (2) assumed that individuals end each period in the interior of a coverage phase (i.e., no bunching), and also that current and marginal coverage phases do not respond to out-of-pocket prices (i.e., no switching). The restriction of the estimation sample to individuals ending year 1 well away from the donut threshold is intended to limit violations of these assumptions, and the evidence in Section A.7.1 above provides reassurance on this point. In this Section, we estimate a richer model that allows for both bunching and switching, and find ultimately that our results are unaffected.

Consider a more general version of the model in Section 3.1, where consumers choose consumption in each period according to the following value functions:

$$V_T(X_T,A_T) = W_T(X_T,A_T) = max_q u(q,A_T) - E^{OOP}(X_T,q)$$
$$V_t(X_t,A_t) = max_q u(q,A_t) - E^{OOP}(X_t,q) + \beta \int W_{t+1}(X_t + q * R, A_{t+1}) dF(A_{t+1}) \forall t < T$$
$$W_t(X_t,A_t) = max_q u(q,A_t) - E^{OOP}(X_t,q) + \int W_{t+1}(X_t + q * R, A_{t+1}) dF(A_{t+1}).$$

Both V_t and W_t are continuous everywhere and differentiable everywhere except $X_t = \bar{X}$. A_t denotes a set of parameters impacting the utility function in period t (e.g., A_t could scale marginal utility in period t). Denote $u_t(q) = u(q, A_t)$. We allow for uncertainty regarding prescription drug needs in future periods using this term. For the sake of exposition, we begin by assuming that there is a single coverage phase kink, a convex kink at $\bar{X}$ where the out-of-pocket price changes from p_1 to $p_2 > p_1$. If an individual has spent X_t on drugs up until period t and purchases q_t units in period t, her period t out-of-pocket expenditure will be:

$$E^{OOP}(X_t,q_t) = \begin{cases} p_1 * q_t & \text{if } X_t + R * q_t \le \bar{X} \\ p_1 * \frac{\bar{X}-X_t}{R} + p_2 * \left(q_t - \frac{\bar{X}-X_t}{R}\right) & \text{if } X_t + R * q_t > \bar{X} \text{ and } X_t \le \bar{X} \\ p_2 * q_t & \text{if } X_t > \bar{X}. \end{cases}$$

We will then generalize the solution to allow for deductible and catastrophic coverage phases in addition to the ICR and donut.

Consider first period T; at the beginning of period T, all uncertainty regarding prescription drug needs for the year will have been resolved. If $X_T \ge \bar{X}$, then the consumer faces a linear price of p_2 per unit and optimal consumption will equal $q_T^* = u_T'^{-1}(p_2)$. If $X_T < \bar{X}$, then the solution will be the piecewise nonlinear budget set solution as in Eq. (1):

$$q_T^*(X_T) = \begin{cases} u_T'^{-1}(p_1) & \text{if } u_T'\left(\frac{\bar{X}-X_T}{R}\right) \le p_1 \text{ and } X_T \le \bar{X} \\ \frac{\bar{X}-X_T}{R} & \text{if } p_1 < u_T'\left(\frac{\bar{X}-X_T}{R}\right) \le p_2 \text{ and } X_T \le \bar{X} \\ u_T'^{-1}(p_2) & \text{if } u_T'\left(\frac{\bar{X}-X_T}{R}\right) > p_2 \text{ or } X_T > \bar{X}. \end{cases}$$

Note that there is some probability of bunching in the final period if $X_T < \bar{X}$. Denote $\tilde{q}_t^c = u_t'^{-1}(p_c)$.

Next consider period $T-1$. If $X_{T-1} \geq \bar{X}$, then the consumer faces a linear price p_2 in the remaining period and $W_T'(X, A_T) = 0$ for all A_T. The solution will be $q_{T-1}^* = \tilde{q}_{T-1}^2$. If $X_{T-1} < \bar{X}$, then we must consider three cases. If the individual consumes past the coverage threshold in period $T-1$, then we again have $W_T'(X, A_T) = 0$ for all A_T and the solution will be $q_{T-1}^* = \tilde{q}_{T-1}^2$. If the individual bunches in period $T-1$, then mechanically $q_{T-1}^* = (\bar{X} - X_{T-1})/R$. Finally, if the individual remains in the ICR in period $T-1$, then we must in turn consider three possibilities for period T consumption behavior – the case where the individual remains in the ICR in period T as well: $\{stay_T\}$; the case where the individual bunches in period T: $\{bunch_T\}$; and the case where the individual crosses the threshold in period T: $\{cross_T\}$. The objective function conditional on remaining in the ICR in period $T-1$ is:

$$\begin{aligned} & max_q u_{T-1}(q) - p_1 * q \\ & \quad + \beta * Pr\{stay_T\} * \mathbb{E}(W_T(X_{T-1}, A_T) | stay_T) \\ & \quad + \beta * Pr\{bunch_T\} * \mathbb{E}(W_T(X_{T-1}, A_T) | bunch_T) \\ & \quad + \beta * Pr\{cross_T\} * \mathbb{E}(W_T(X_{T-1}, A_T) | cross_T). \end{aligned}$$

Substituting in the results from above, we have:

$$\begin{aligned} & max_q u_{T-1}(q) - p_1 * q \\ & \quad + Pr\{stay_T\} * \beta * \int \left(u_T\left(\tilde{q}_T^1\right) - p_1 * \tilde{q}_T^1\right) dF(A_T) \\ & \quad + Pr\{bunch_T\} * \beta * \int \left(u_T\left(\frac{\bar{X} - X_{T-1}}{R} - q\right) - p_1 * \left(\frac{\bar{X} - X_{T-1}}{R} - q\right)\right) dF(A_T) \\ & \quad + Pr\{cross_T\} * \beta * \int \left(u_T(\tilde{q}_T^2) - p_1 * \left(\frac{\bar{X} - X_{T-1}}{R} - q\right) - p_2 * \left(\tilde{q}_T^2 - \frac{\bar{X} - X_{T-1}}{R} + q\right)\right) dF(A_T). \end{aligned}$$

By the envelope theorem, this implies the following first-order condition:

$$\begin{aligned} u_{T-1}'(q) &= p_1 + Pr\{bunch_T\} * \beta * \int \left(u_T'\left(\frac{\bar{X} - X_{T-1}}{R} - q\right) - p_1\right) dF(A_T) \\ & \quad + Pr\{cross_T\} * \beta * (p_2 - p_1) \\ &= (1 - \beta * Pr\{cross_T\}) * p_1 + Pr\{cross_T\} * \beta * p_2 \\ & \quad + \beta * Pr\{bunch_T\} * \mathbb{E}\left(u_T'\left(\frac{\bar{X} - X_{T-1}}{R} - q\right) - p_1\right). \end{aligned}$$

Let $q_{T-1}^S(X_{T-1})$ be the unique solution to the above equation. To determine the region of A_{T-1} for which the individual will remain in the ICR ($stay_{T-1}$), bunch at the donut kink ($bunch_{T-1}$), or cross into the donut hole ($cross_{T-1}$), we consider the limits of the first-order condition as q approaches $(\bar{X} - X_{T-1})/R$ from the left-hand side and right-hand side. The limit of the period $T-1$ first-order condition above as we approach the kink from the left-hand side is:

$$u_{T-1}'(q) - (1 - \beta) * p_1 - \beta * p_2,$$

as the probability of entering the donut in period T goes to one (and, accordingly, as $Pr\{bunch_T\}$ goes to zero). The limit of the period $T-1$ first-order condition as we approach the kink from the right-hand side is simply:

$$u_{T-1}'(q) - p_2$$

since $W_T' = 0$ for all q such that $X_{T-1} + R * q > \bar{X}$. Then optimal consumption in period $T-1$ will be:

$$q_{T-1}^*(X_{T-1}) = \begin{cases} q_{T-1}^S(X_{T-1}) & \text{if } u_{T-1}'\left(\frac{\bar{X} - X_{T-1}}{R}\right) \leq (1 - \beta) * p_1 + \beta * p_2 \text{ and } X_{T-1} \leq \bar{X} \\ \frac{\bar{X} - X_{T-1}}{R} & \text{if } (1 - \beta) * p_1 + \beta * p_2 < u_T'\left(\frac{\bar{X} - X_{T-1}}{R}\right) \leq p_2 \text{ and } X_{T-1} \leq \bar{X} \\ \tilde{q}_{T-1}^2 & \text{if } u_T'\left(\frac{\bar{X} - X_{T-1}}{R}\right) > p_2 \text{ or } X_{T-1} > \bar{X} \end{cases}.$$

This analysis yields the following general solution.

Theorem 1. *For any period $t < T$, the optimal consumption path will be:*

$$q_t^*(X_t) = \begin{cases} q_t^S(X_t) & \text{if } X_t \leq \bar{X} \text{ and } u_t'\left(\frac{\bar{X} - X_t}{R}\right) \leq (1 - \beta) * p_1 + \beta * p_2 \\ \frac{\bar{X} - X_t}{R} & \text{if } X_t \leq \bar{X} \text{ and } (1 - \beta) * p_1 + \beta * p_2 < u_t'\left(\frac{\bar{X} - X_t}{R}\right) \leq p_2 \\ \tilde{q}_t^2 & \text{if } X_t > \bar{X} \text{ or } u_t'\left(\frac{\bar{X} - X_t}{R}\right) > p_2 \end{cases}.$$

with $q_t^S(X_t)$ defined by

$$u_t'\left(q_t^S(X_t)\right) = \left(1 - \beta * \sum_{i=t+1}^{T} Pr\{cross_i|t\}\right) * p_1 + \beta * \sum_{i=t+1}^{T} Pr\{cross_i|t\} * p_2 + \beta * \sum_{i=t+1}^{T} Pr\{bunch_i|t\} * E\left(u_i'\left(\frac{\bar{X} - X_{i-1}}{R} - q_{i-1}\right) - p_1|t\right).$$

Proof. We prove this result by induction. Suppose that the result holds in $t+1$, and consider the optimal consumption for period t. If $X_t \geq \bar{X}$, then as usual the consumer faces a linear price p_2 in all remaining periods and $W'_{t+1}(X, A_{t+1}) = 0$ for all A_{t+1}; the solution will be $q_t^* = u_t'^{-1}(p_2) = \tilde{q}_t^2$. If $X_t < \bar{X}$, then we must consider three cases. If the individual consumes past the coverage threshold in period t, then we again have $W'_{t+1}(X, A_{t+1}) = 0$ for all A_{t+1} and the solution will be $q_t^* = \tilde{q}_t^2$. If the individual bunches in period t, then mechanically $q_t^* = (\bar{X} - X_t)/R$. If, however, the individual remains in the ICR in period t, then we must consider the three potential outcomes in period $t+1$: $\{stay_{t+1}\}$, $\{bunch_{t+1}\}$, and $\{cross_{t+1}\}$. The objective function is:

$$\begin{aligned} &max_q u_t(q) - p_1 * q \\ &+\beta * Pr\{stay_{t+1}|t\} * E\left(W_{t+1}(X_t + R * q, A_{t+1})|stay_{t+1}\right) \\ &+\beta * Pr\{bunch_{t+1}|t\} * E\left(W_{t+1}(X_t + R * q, A_{t+1})|bunch_{t+1}\right) \\ &+\beta * Pr\{cross_{t+1}|t\} * E\left(W_{t+1}(X_t + R * q, A_{t+1})|cross_{t+1}\right). \end{aligned}$$

Again applying the envelope theorem, the first-order condition for this objective function is:

$$\begin{aligned} u_t'(q) = p_1 &- \beta * Pr\{stay_{t+1}|t\} * E\left(W'_{t+1}(X_t + R * q, A_{t+1})|stay_{t+1}\right) \\ &- \beta * Pr\{bunch_{t+1}|t\} * E\left(W'_{t+1}(X_t + R * q, A_{t+1})|bunch_{t+1}\right) \\ &- \beta * Pr\{cross_{t+1}|t\} * E\left(W'_{t+1}(X_t + R * q, A_{t+1})|cross_{t+1}\right). \end{aligned} \tag{25}$$

If the individual either crosses or bunches in period $t+1$, then $\mathbb{E}(W_i'(X_i, A_i)|t) = 0$ for all $i > t+1$. Accordingly, if $\{cross_{t+1}\} = 1$, then

$$\mathbb{E}\left(W'_{t+1}(X_t + R * q, A_{t+1})|cross_{t+1}\right) = -(p_2 - p_1) \tag{26}$$

and if $\{bunch_{t+1}\} = 1$, then

$$\mathbb{E}\left(W'_{t+1}(X_t + R * q, A_{t+1})|bunch_{t+1}\right) = -E\left(u'_{t+1}\left(\frac{\bar{X} - X_{t+1}}{R}\right)|t\right) + p_1. \tag{27}$$

Finally, if $\{stay_{t+1}\} = 1$, then by the inductive hypothesis we have

$$\begin{aligned} &\mathbb{E}\left(W'_{t+1}(X_t + R * q, A_{t+1})|t+1, stay_{t+1}\right) \\ &= \mathbb{E}\left(\sum_{i=t+2}^{T} Pr\{bunch_i|t+1, stay_{t+1}\} * \mathbb{E}\left(-u_i'\left(\frac{\bar{X} - X_i}{R}\right) + p_1|t+1, stay_{t+1}\right)|t\right) \\ &\quad +\mathbb{E}\left(\sum_{i=t+2}^{T} Pr\{cross_i|t+1, stay_{t+1}\} * (p_1 - p_2)|t\right). \end{aligned} \tag{28}$$

Substituting Eqs. (26), (27), and (28) into Eq. (25), and applying the law of iterated expectations, we obtain

$$\begin{aligned} u_t'(q) = p_1 &- \beta * \mathbb{E}\left(\sum_{i=t+2}^{T} Pr\{bunch_i\} * \left(-u_i'\left(\frac{\bar{X} - X_i}{R}\right) + p_1\right)|t\right) \\ &- \beta * \mathbb{E}\left(\sum_{i=t+2}^{T} Pr\{cross_i\} * (p_1 - p_2)|t\right) \\ &+ \beta * Pr\{bunch_{t+1}|t\} * \left(\mathbb{E}\left(u'_{t+1}\left(\frac{\bar{X} - X_{t+1}}{R}\right)|t\right) - p_1\right) \\ &+ \beta * Pr\{cross_{t+1}|t\} * (p_2 - p_1). \end{aligned}$$

The second and third lines follow from the fact that $Pr\{stay_{t+1}|t\} * Pr\{bunch_i|stay_{t+1},t\} = Pr\{stay_{t+1} \cap bunch_i|t\} = Pr\{stay_{t+1}|bunch_i,t\} * Pr\{bunch_i|t\} = Pr\{bunch_i|t\}$and similarly $Pr\{stay_{t+1}|t\} * Pr\{cross_i|stay_{t+1},t\} = Pr\{cross_i|t\}$ for all $i > t + 1$ (by definition, one cannot cross or bunch in any period unless they have "stayed" in all previous periods). This expression simplifies to

$$u_t'(q) = p_1 + \beta * \mathbb{E}\left(\sum_{i=t+1}^{T} Pr\{bunch_i|t\} * \left(u_i'\left(\frac{\bar{X} - X_i}{R}\right) - p_1\right)|t\right)$$
$$+ \beta * \sum_{i=t+1}^{T} Pr\{cross_i|t\} * (p_2 - p_1).$$

This implies that the optimal consumption level if the individual remains in the ICR in period t will be $q_t^S(X_t)$ such that

$$u_t'\left(q_t^S(X_t)\right) = \left(1 - \beta * \sum_{i=t+1}^{T} Pr\{cross_i|t\}\right) * p_1 + \beta * \sum_{i=t+1}^{T} Pr\{cross_i|t\} * p_2$$
$$+ \beta * \sum_{i=t+1}^{T} Pr\{bunch_i|t\} * E\left(u_i'\left(\frac{\bar{X} - X_{i-1}}{R} - q_{i-1}\right) - p_1|t\right)$$

which, in turn, implies the full solution[45]

$$q_t^*(X_t) = \begin{cases} q_t^S(X_t) & if\ X_t \le \bar{X}\ \ and\ \ u_t'\left(\frac{\bar{X}-X_t}{R}\right) \le (1-\beta) * p_1 + \beta * p_2 \\ \frac{\bar{X}-X_t}{R} & if\ \ X_t \le \bar{X}\ \ and\ \ (1-\beta) * p_1 + \beta * p_2 < u_t'\left(\frac{\bar{X}-X_t}{R}\right) \le p_2 \\ \tilde{q}_t^2 & if\ \ X_t > \bar{X}\ \ or\ \ u_t'\left(\frac{\bar{X}-X_t}{R}\right) > p_2 \end{cases}.$$

Thus completing the proof. □

We now describe how we generalize this model to allow for the deductible and catastrophic coverage phases, estimate the richer model on our sample of low- and high-spending individuals, and use the resulting parameters to simulate consumption for the full sample (including individuals near the donut hole kink) as in Section 6. While our model in principle can allow for bunching, in practice the probability of bunching in the sample of low- and high-spending individuals is so low ($< 0.1\%$ according any model of expectations) that we omit this term.[46] We then have

$$q_t^*(X_t) = \begin{cases} q_t^S(X_t) & \text{if } X_t \le \bar{X} \text{ and } u_t'\left(\frac{\bar{X}-X_t}{R}\right) \le (1-\beta) * p_1 + \beta * p_2 \\ \tilde{q}_t^2 & \text{if } X_t > \bar{X} \text{ or } u_t'\left(\frac{\bar{X}-X_t}{R}\right) > p_2 \end{cases}$$

with $q_t^S(X_t)$ such that

$$u_t'\left(q_t^S(X_t)\right) = \left(1 - \beta * \sum_{i=t+1}^{T} Pr\{cross_i|t\}\right) * p_1 + \beta * \sum_{i=t+1}^{T} Pr\{cross_i|t\} * p_2.$$

It is useful to rewrite this solution such that marginal utility is equal to the appropriate virtual price given the coverage phase in which the individual ends the current period and the year:

$$u'\left(q_t^*\right) = Pr\{ICR_t\}\left[(1 - \beta * Pr\{donut_T|ICR_t\}) * p_1 + \beta * Pr\{donut_T|ICR_t\} * p_2\right]$$
$$+ Pr\{donut_t\} * p_2$$

[45] We apply the same reasoning as in the above analysis for $T - 1$ to define the ranges of $u_t'(.)$ such that the individual will choose to stay, bunch, or cross in period t.
[46] The estimation strategy below can be modified to allow for bunching by estimating an auxiliary model of bunching as a function of prices.

where, in this notation, $Pr\{c_t\}$ is the probability that the individual ends period t in coverage phase c. Again, all bunching terms are omitted. The deductible and catastrophic phases are conceptually straightforward to incorporate into our model because they are convex kinks and thus do not permit bunching. Repeating the above model derivation allowing for these additional coverage phases yields:

$$\begin{aligned} u'(q_t^*) = & Pr\{ded_t\} * (1 - \beta * (Pr\{ICR_T|ded_t\} + Pr\{don_T|ded_t\} + Pr\{cat_T|ded_t\} \ * p_0 \\ & + Pr\{ded_t\} * \beta * [Pr\{ICR_T|ded_t\} * p_1 + Pr\{don_T|ded_t\} * p_2 + Pr\{cat_T|ded_t\} * p_3] \\ & + Pr\{ICR_t\} * (1 - \beta * (Pr\{don_T|ICR_t\} + Pr\{cat_T|ICR_t\})) * p_1 \\ & + Pr\{ICR_t\} * \beta * [Pr\{don_T|ICR_t\} * p_2 + Pr\{cat_T|ICR_t\} * p_3] \\ & + Pr\{don_t\} * ((1 - \beta * Pr\{cat_T|don_t\}) * p_2 + \beta * Pr\{cat_T|don_t\} * p_3) + Pr\{cat_t\} * p_3. \end{aligned}$$

This expression simplifies to our result from Eq. (2): $q^* = u'^{-1}((1-\beta) * CP + \beta * MP)$. However, it provides greater detail on where each individual price enters the optimal consumption function. Suppose, as in our structural modeling exercise above, that $u(q)$ is quadratic in q. We can then rearrange and differentiate to obtain:

$$\begin{aligned} \frac{\partial q_t^*}{\partial p_1} = & \eta * \frac{\partial Pr\{ded_t\} * (1 - \beta * (Pr\{ICR_T|ded_t\} + Pr\{don_T|ded_t\} + Pr\{cat_T|ded_t\}))}{\partial p_1} * p_0 \\ & + \eta * \beta * \left[\frac{\partial Pr\{ded_t \cap ICR_T\}}{\partial p_1} * p_1 + \frac{\partial Pr\{ded_t \cap don_T\}}{\partial p_1} * p_2 + \frac{\partial Pr\{ded_t \cap cat_T\}}{\partial p_1} * p_3 \right] \\ & + \eta * \left[\frac{\partial Pr\{ICR_t\}}{\partial p_1} - \beta * \left(\frac{\partial Pr\{ICR_t \cap don_T\}}{\partial p_1} + \frac{\partial Pr\{ICR_t \cap cat_T\}}{\partial p_1} \right) \right] * p_1 \\ & + \eta * \beta * \left[\frac{\partial Pr\{ICR_t \cap don_T\}}{\partial p_1} * p_2 + \frac{\partial Pr\{ICR_t \cap cat_T\}}{\partial p_1} * p_3 \right] \\ & + \eta * \left(\frac{\partial Pr\{don_t\}}{\partial p_1} - \beta * \frac{\partial Pr\{don_t \cap cat_T\}}{\partial p_1} \right) * p_2 + \eta * \beta * \frac{\partial Pr\{don_t \cap cat_T\}}{\partial p_1} * p_3 \\ & + \eta * \frac{\partial Pr\{cat_t\}}{\partial p_1} * p_3 \\ & + \eta * \beta * Pr\{ded_t \cap ICR_T\} \\ & + \eta * Pr\{ICR_t\} * (1 - \beta * (Pr\{don_T|ICR_t\} + Pr\{cat_T|ICR_t\})) \end{aligned} \tag{29}$$

and similarly for $\partial q_t^*/\partial p_2$. The key distinction between this specification and the one estimated in Section 6 is that the current, richer specification allows for endogenous coverage phase switching, as each coverage phase probability term is allowed to respond endogenously to coverage phase prices.

In order to estimate this richer model, we use the following procedure: (1) we estimate how each coverage phase probability term (e.g., $Pr\{ICR_T|ded_t\}$, $Pr\{don_t\}$) responds to p_1 and p_2. We use the same exact regression specification and controls as in Section 3, but use the relevant probability term as the left-hand-side variable. The probability term is calculated assuming perfect foresight as our model of expectations, as in Section 6. Next, we (2) perform the GMM estimation as in Section 6, but instead of using Eqs. (4) and (5) in the GMM objective function directly, we replace the left-hand sides of Eqs. (4) and (5) with Eq. (29) and its analog for the donut price. We use this richer specification as the GMM objective function to recover η and β. Finally (3), using the new estimates of η and β, we simulate consumption for all sample individuals and compare to observed consumption.[47]

The results of this procedure are overall quite similar to the results when we do not allow for endogenous coverage phase switching. The estimates of η and β are not statistically significantly different from the baseline estimates in Section 6 (the richer model yields $\beta = 0.313$ and $\bar{\eta} = -1.77$, whereas the simpler model estimates were $\beta = 0.312$ (SE=0.08) and $\bar{\eta} = -1.66$ (SE=0.15)). The comparison of actual to simulated spending are shown, for each model, in Appendix Table A8. If anything, the simulation of consumption both in- and outside the regression sample performs slightly *worse* than the simpler approach – the mean squared error with the richer specification is 1% higher in the regression sample, and 5% higher in the sample of individuals near the donut threshold.

Table A8
Comparison of actual to predicted spending – Basic structural model and richer specification.

	Baseline structural model comparison				Richer structural model comparison			
	Actual	Predicted			Actual	Predicted		
		Mean	% diff	MSE		Mean	% diff	MSE
In-sample (Low/high spending enrollees)	1718.2	1703.6	−0.85%	996,251	1718.2	1694.2	−1.40%	1,006,810
Out-of-sample (Medium spending enrollees)	2324.1	2328.3	0.18%	990,186	2324.1	2316.3	−0.34%	1,035,317
All spending groups	1972.5	1965.7	−0.34%	993,706	1972.5	1955.3	−0.87%	1,018,774

Notes: N=1,960,008. Comparison of actual and simulated spending, baseline structural model from Section 3 vs. richer structural model from Appendix G.2. In each comparison, data are shown only for individuals within the 1st to 99th percentiles of the distribution of the predicted trend in consumption.

[47] Note that, whether we use the simple GMM procedure in Section 6 or the richer GMM procedure outlined here to recover η and β, the simulations of consumption in- and outside the regression sample always allow for bunching and switching behavior.

Overall, this analysis demonstrates that the simple approach, which restricts the sample to individuals on the linear portions of the budget set and ignores bunching and switching behavior, performs as well as more complex nonlinear models of consumption.

A.7.3. Alternative construction of IV

Here, we separately present results based on our preferred IV, in which we construct instruments for price holding the plan of enrollment fixed across adjacent years at the first year's chosen plan, and compare them to an alternative IV that allows for switching of plans between years. The key identifying assumption of our preferred IV is that, conditional on the plan an individual initially chooses, changes in cost-sharing between years are exogenous with respect to trends in consumption. Given the extremely high degree of inertia observed in the population of interest, we consider this to be the optimal strategy for determining how the average (inertial) Part D enrollee would respond to future generosity changes. However, there is a small population of enrollees that do switch plans between years, and one might be concerned about our estimates' reliance on the enrollees known to be most inertial. In order to demonstrate the sensitivity of our estimates to this identification approach, we show two alternative analyses below.

First, we show robustness analyses in which we construct our instruments without holding plan enrollment fixed. That is, price in year $t+1$ for year pair $(t, t+1)$ is a function of year $t+1$ plan characteristics, year t consumption weights, and year t drug prices. As seen in the below, the alternative construction of our instrument does not impact our coefficient estimates in any economically meaningful way. If anything, the donut price coefficient looks slightly better under the alternative IV, but we see no strong argument that the alternative estimates are more representative of the truth and thus continue relying on the preferred IV throughout the paper.[48]

Table A9
Results of full year ICR and donut price regressions, with stark donut coverage and deductible variables – All enrollees.

	Original IV		IV with switching	
Price	Coef	SE	Coef	SE
ICR	−0.078	0.006**	−0.072	0.005**
Donut	0.021	0.009*	−0.041	0.010**
Ded. Chg	−0.039	0.006**	−0.039	0.006**
Stark	−0.043	0.008**	−0.023	0.010*

Notes: Results of full year regressions of log consumption change on log change in ICR and donut prices, as well as changes in stark gap coverage and deductible coverage. Comparison of "inertial IV" to IV that allows for individuals' plan switching to impact price. Results pooled across full sample of 2.7 million enrollees. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

Second, we below show results where we use the preferred IV, but eliminate those who switch plans between years from the sample. The preferred ("inertial") IV approach is structured to explicitly account for the possibility of endogenous plan switching, and if endogenous plan switching does indeed occur, this sampling restriction would clearly be inappropriate. However, we see in Appendix Table A10 that the results are essentially identical, consistent with the fact that the vast majority of enrollees are indeed inertial.

Table A10
Results of full year ICR (and donut) price regressions, with stark donut coverage and deductible variables – All enrollees vs. non-switchers only.

	Full sample, all years pooled			
	All N=2,707,315		Non-switchers N=2,435,952	
Price	Coef	SE	Coef	SE
ICR	−0.078	0.006**	−0.077	0.006**
Donut	0.021	0.009*	0.022	0.009*
Ded. Chg	−0.039	0.006**	−0.036	0.005**
Stark	−0.043	0.008**	−0.029	0.008**

Notes: Results of full year regressions of log consumption change on log change in ICR and donut prices, as well as changes in stark gap coverage and deductible coverage. Full sample of enrollees included. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

A.7.4. Robustness of "stark" effect

Here, we analyze the robustness of our choice to model two distinct but closely related price responses: the response to the individual-specific price in the donut hole, and the response to categorical changes in donut hole coverage. This analysis is intended to address concerns that the patterns presented in our preferred model are driven in part by collinearity between these two variables. Reassuringly, as we see in Appendix Table A11 below, the consumption response to the "stark" variable is similar regardless of whether the specification includes a separate control for donut price.

[48] For the sake of brevity, these results pool estimates across both high- and low-spending enrollees and focus on the full year, rather than quarterly, price response. The quarterly price responses, separated out by spending group, show the same basic finding: that the results are not meaningfully changed if we use the alternative IV.

Table A11
Results of full year ICR and donut price regressions, with stark donut coverage and deductible variables – All enrollees.

	Full sample, all years pooled			
	Including donut		Excluding donut	
Price	Coef	SE	Coef	SE
ICR	−0.078	0.006**	−0.079	0.006**
Donut	0.021	0.009*		
Ded. Chg	−0.039	0.006**	−0.040	0.006**
Stark	−0.043	0.008**	−0.037	0.008*

Notes: Results of full year regressions of log consumption change on log change in ICR and donut prices, as well as changes in stark gap coverage and deductible coverage. Comparison of "inertial IV" to IV that allows for individuals' plan switching to impact price. Results pooled across full sample of 2.7 million enrollees. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

A.8. Additional tables

Table A12
Results of quarterly ICR and donut price regressions, with stark donut coverage and deductible variables – High- and low-spending enrollees, pooled all years, pooled vs. separate regressions.

		Pooled regression				Separate regressions			
		Low-spending enrollee response		High-spending enrollee response		Low-spending enrollee response		High-spending enrollee response	
Period	Price	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Q1	ICR	−0.054	0.006**	−0.027	0.008**	−0.062	0.006**	−0.026	0.008**
Q1	Donut	0.023	0.007**	−0.045	0.010**	0.035	0.008**	−0.061	0.012**
Q1	Ded. Chg	−0.048	0.006**	−0.048	0.006**	−0.033	0.005**	−0.042	0.006**
Q1	Start	−0.016	0.009	−0.016	0.009	−0.025	0.007**	−0.003	0.008
Q2	ICR	−0.046	0.006**	−0.010	0.009	−0.054	0.006**	−0.016	0.008
Q2	Donut	0.029	0.007**	−0.037	0.014**	0.034	0.007**	−0.088	0.014**
Q2	Ded. Chg	−0.024	0.005**	−0.024	0.005**	−0.014	0.006*	−0.004	0.007
Q2	Start	0.000	0.009	0.000	0.009	−0.003	0.008	0.006	0.011
Q3	ICR	−0.045	0.006**	−0.003	0.009	−0.047	0.006**	−0.015	0.009
Q3	Donut	0.015	0.008	−0.076	0.013**	0.010	0.008	−0.109	0.015**
Q3	Ded. Chg	−0.016	0.006**	−0.016	0.006**	−0.011	0.006	−0.004	0.008
Q3	Start	−0.008	0.009	−0.008	0.009	0.006	0.008	−0.011	0.010
Q4	ICR	−0.060	0.008**	0.022	0.012	−0.060	0.008**	0.009	0.011
Q4	Donut	0.008	0.010	−0.162	0.017**	0.001	0.010	−0.204	0.020**
Q4	Ded. Chg	−0.012	0.007	−0.012	0.007	−0.005	0.006	−0.005	0.010
Q4	Start	−0.047	0.011**	−0.047	0.011**	−0.046	0.010**	−0.030	0.013*

Notes: Results of quarterly regressions of log consumption change on log change in ICR and donut prices, as well as changes in stark gap coverage and deductible coverage. Results with pooled vs. separate Ded. Chg. and Stark coefficients. Superscript
** indicates significance at the 1% level; superscript
* indicates significance at the 5% level.

Table A13
Estimated structural model parameters, pooled all years – By chronic condition.

	Full sample [N=1,985,127]		Chronic (All) [N=1,349,635]		Non-chronic [N=633,922]	
Description	Structural parameter	Mean estimate	Structural parameter	Mean estimate	Structural parameter	Mean estimate
Days supply ($P = 0$)	γ	374.910	γ	392.532	γ	333.782
Myopia	β	0.312	β	0.339	β	0.284
Marginal price effect	η	−1.663	η	−1.824	η	−1.442
Deductible effect	κ_{deduct}	−8.915	κ_{deduct}	−9.203	κ_{deduct}	−7.759
Stark gap effect	κ_{stark}	−6.409	κ_{stark}	−5.847	κ_{stark}	−13.727
Implied elasticity	ε	−0.126	ε	−0.136	ε	−0.110

	Hypertension [N=857,908]		Hypercholesterolemia [N=344,792]		Diabetes [N=316,608]	
Description	Structural parameter	Mean estimate	Structural parameter	Mean estimate	Structural parameter	Mean estimate
Days supply ($P = 0$)	γ	395.156	γ	364.726	γ	482.269
Myopia	β	0.375	β	0.455	β	0.416
Marginal price effect	η	−1.795	η	−1.568	η	−1.952
Deductible effect	κ_{deduct}	−10.321	κ_{deduct}	−13.019	κ_{deduct}	−9.673
Stark gap effect	κ_{stark}	−5.219	κ_{stark}	−12.644	κ_{stark}	−3.122
Implied elasticity	ε	−0.127	ε	−0.134	ε	−0.135

Notes: Authors' calculations. Data are shown only forindividuals within the 1st to 99th percentiles of thedistribution of the predicted trend in consumption. All parameters shown are averages except for the hyperbolic discount factor. Following Goldman et al. (2004), chronic illnesses are identified using diagnosis codes from the individuals' medical claims histories.

Table A14
Results of quarterly ICR and donut price regressions, with stark donut coverage and deductible variables – Low- and high-spending enrollees.

		All years pooled				2006–7				2007–8				2008–9			
		Enrollee response		Enrollee response		Enrollee response		Enrollee response		Enrollee response		Enrollee response		Enrollee response		Enrollee response	
Period	Price	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Q1	ICR	−0.053	0.006**	−0.026	0.008**	−0.069	0.013**	−0.012	0.016	−0.039	0.010**	0.012	0.012	−0.061	0.008**	−0.067	0.012**
Q1	Donut	0.019	0.007**	−0.048	0.010**	−0.008	0.020	−0.050	0.019**	0.025	0.010*	−0.037	0.013**	0.023	0.009*	−0.060	0.024*
Q1	Ded. Chg	−0.049	0.005**	−0.049	0.005**	−0.031	0.013*	−0.031	0.013*	−0.059	0.009**	−0.059	0.009**	−0.035	0.008**	−0.035	0.008**
Q1	Start	−0.014	0.009	−0.014	0.009	−0.004	0.011	−0.004	0.011	−0.027	0.014	−0.027	0.014	0.031	0.034	0.031	0.034
Q2	ICR	−0.045	0.006**	−0.011	0.009	−0.048	0.017**	−0.020	0.018	−0.032	0.010**	0.007	0.013	−0.052	0.008**	−0.024	0.013
Q2	Donut	0.026	0.007**	−0.044	0.013**	0.028	0.023	−0.040	0.022	0.017	0.011	−0.041	0.022	0.032	0.011**	−0.073	0.026**
Q2	Ded. Chg	−0.024	0.005**	−0.024	0.005**	0.003	0.014	0.003	0.014	−0.028	0.008**	−0.028	0.008**	−0.029	0.008**	−0.029	0.008**
Q2	Start	0.002	0.009	0.002	0.009	0.004	0.012	0.004	0.012	−0.013	0.013	−0.013	0.013	0.027	0.040	0.027	0.040
Q3	ICR	−0.043	0.006**	−0.003	0.009	−0.049	0.015**	−0.021	0.019	−0.035	0.011**	0.001	0.015	−0.047	0.008**	−0.003	0.014
Q3	Donut	0.012	0.008	−0.083	0.012**	0.020	0.023	−0.092	0.024**	0.000	0.014	−0.088	0.017**	0.018	0.011	−0.057	0.022*
Q3	Ded. Chg	−0.016	0.006**	−0.016	0.006**	0.001	0.017	0.001	0.017	−0.028	0.010**	−0.028	0.010**	−0.016	0.009	−0.016	0.009
Q3	Start	−0.007	0.009	−0.007	0.009	−0.002	0.013	−0.002	0.013	−0.014	0.014	−0.014	0.014	0.028	0.023	0.028	0.023
Q4	ICR	−0.056	0.008**	0.024	0.012*	−0.035	0.016*	0.027	0.023	−0.052	0.015**	0.013	0.018	−0.066	0.010**	0.034	0.017*
Q4	Donut	0.003	0.010	−0.168	0.016**	−0.008	0.025	−0.178	0.023**	0.003	0.017	−0.172	0.032**	0.007	0.015	−0.165	0.032**
Q4	Ded. Chg	−0.012	0.007	−0.012	0.007	0.026	0.018	0.026	0.018	−0.026	0.011*	−0.026	0.011*	−0.012	0.009	−0.012	0.009
Q4	Start	−0.047	0.011**	−0.047	0.011**	−0.056	0.013**	−0.056	0.013**	−0.023	0.018	−0.023	0.018	−0.010	0.031	−0.010	0.031
N		1,214,548				189,610				515,506				509,432			

Notes: Results of quarterly regressions of log consumption change on log change in ICR and donut prices, as well as changes in stark gap coverage and deductible coverage. Regression for high- and low-spending enrollees, with separate ICR and donut responses for each spending group. "Stark" gap coverage and deductible change response held fixed across spending groups. Superscript

** indicates significance at the 1% level; superscript

* indicates significance at the 5% level.

References

Abaluck, J., Gruber, J., 2011. Heterogeneity in choice inconsistencies among the elderly: evidence from plan choice in the medicare Part D program. Am. Econ. Rev. 101 (3), 377–381. May.

Abaluck, J., Gruber, J., 2016. Evolving choice inconsistencies in choice of prescription drug insurance. Am. Econ. Rev. 106 (8), 2145–2184. August.

Agarwal, S., Driscoll, J., Gabaix, X., Laibson, D., 2009. The Age of Reason: Financial Decisions Over the Life-cycle with Implications for Regulation. Brookings Papers on Economic Activity, October.

Aron-Dine, A., Einav, L., Finkelstein, A., Cullen, M., 2015. Moral hazard in health insurance: do dynamic incentives matter? Rev. Econ. Stat. 97 (4), 725–741. October.

Blomquist, W., Newey, W., 2002. Nonparametric estimation with nonlinear budget sets. Econometrica J. Econ. Soc. 70 (6), 2455–2480. November.

Brot-Goldberg, Z., Chandra, A., Handel, B., Kolstad, J., 2015. What Does a Deductible Do? the Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics. Mimeo Available at http://faculty.haas.berkeley.edu/jkolstad/Utilization_BCHK_Web.pdf.

Centers for Medicare and Medicaid Services (CMS), 2015. Closing the Coverage Gap - Medicare Prescription Drugs are Becoming More Affordable. CMS Product No. 11493. January, https://www.medicare.gov/Pubs/pdf/11493.pdf. Available at.

Chandra, A., Gruber, J., McKnight, R., 2010. Patient cost-sharing and hospitalization offsets in the elderly. Am. Econ. Rev. 100 (1), 193–213. March.

Chetty, R., Looney, A., Kroft, K., 2009. Salience and taxation: theory and evidence. Am. Econ. Rev. 99 (4), 1145–1177.

Dalton, C. M., Gowrisankaran, G., Town, R., 2015. Myopia and Complex Dynamic Incentives: Evidence from Medicare Part D. Mimeo Available at http://users.wfu.edu/daltonc/docs/Research/Dalton_Gowrisankaran_Town_PartD.pdf.

Duan, N., 1983. Smearing estimate: a nonparametric retransformation method. J. Am. Stat. Assoc. 78 (383), 605–610.

Duggan, M., Healy, P., Morton, F., 2008. Providing prescription drug coverage to the elderly: America's experiment with Medicare Part D.. J. Econ. Perspect. Fall 22 (4), 69–92. 24.

Einav, L., Finkelstein, A., Schrimpf, P., 2015. The response of drug expenditures to Non-Linear contract design: evidence from Medicare Part D. Q. J. Econ. 130 (2), 841–899. May.

Einav, L., Finkelstein, A., Schrimpf, P., 2017. Bunching at the kink: implications for spending responses to health insurance contracts. J. Public Econ. 146, 27–40.

Ericson, K., 2014. Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange. Am. Econ. J. Econ. Pol. 6 (1), 38–64. February.

Feldman, N., Katuscak, P., Kawano, L. 2015. Taxpayer Confusion: Evidence from the Child Tax Credit. Mimeo. July. Available at SSRN http://ssrn.com/abstract=2296542. accessed August 27, 2015

Fratiglioni, L., De Ronchi, D., Aguero-Torres, H., 1999. Worldwide prevalence and incidence of dementia. Drugs Aging 15 (5), 365–375. November.

Goldman, D., Joyce, G., Zheng, Y., 2007. Prescription drug cost sharing. J. Am. Med. Assoc. 298 (1), 61–69. July.

Goldman, D.P., Joyce, G.F., Escarce, J.J., Pace, J.E., Colomom, M.D., Laouri, M., Landsman, P.B., Teutsch, S.M., 2004. Pharmacy benefits and the use of drugs by the chronically ill. J. Am. Med. Assoc. 291, 2344-23501.

Handel, B., Kolstad, J., 2015. Health insurance for "Humans": information frictions, plan choice, and consumer welfare. Am. Econ. Rev. 105 (8), 2449–2500. August.

Hausman, J., 1985. The econometrics of nonlinear budget sets. Econometrica J. Econ. Soc. 53 (6), 1255–1282. November.

Heiss, F., McFadden, D., Winter, J., 2006. Who failed to enroll in Medicare Part D, and why? Health Aff. 25 (5), w344–354. September–October.

Ho, K., Hogan, J., Morton, F.S., 2015. The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program. National Bureau of Economic Research Working Paper No. 21028. March,

Ito, K., 2014. Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. Am. Econ. Rev. 104 (2), 537–563. February.

Ketcham, J., Lucarelli, C., Miravete, E., Roebuck, M., 2012. Sinking, swimming, or learning to swim in Medicare Part D. Am. Econ. Rev. 102 (6), 2639–2673. October.

Ketcham, J., Simon, K., 2008. Medicare Part D's effects on elderly patients' drug costs and utilization. Am. J. Manag. Care 14 (11 Suppl), 14–22.

Kling, J., Mullainathan, S., Shafir, E., Vermeulen, L., Wrobel, M. 2008. Misperception in Choosing Medicare Drug Plans. Unpublished manuscript. November.

Kowalski, A., 2015. Estimating the Tradeoff between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance. National Bureau of Economic Research Working Paper No. 18108. July,

Laibson, D. Repetto, A., Tobacman, J., 2015. Estimating Discount Functions with Consumption Choices over the Lifecycle. Mimeo. July. Available at http://assets.wharton.upenn.edu/tobacman/papers/Estimating%20Discount%20Functions.pdf.

Lichtenberg, F., Sun, S., 2007. The impact of Medicare Part D on prescription drug use by the elderly. Health Aff. 26 (6), 1735–1744. November-December.

Liebman, J.B., Zeckhauser, R.J., 2004. Schmeduling. October.

Loewenstein, G., et al. 2013. Consumers' misunderstanding of health insurance. J. Health Econ. 32, 850–862.

Manning, W., Newhouse, J., Duan, N., Keeler, E., Leibowitz, A., 1987. Health insurance and the demand for medical care: evidence from a randomized experiment. Am. Econ. Rev. 77 (3), 251–277. June.

Salthouse, T., 2004. What and when of cognitive aging. Curr. Dir. Psychol. Sci. 13 (4), 140–144.

Saez, E., 2010. Do taxpayers bunch at kink points?. Am. Econ. J. Econ. Pol. 2 (3), 180–212.

Saez, E., Slemrod, J., Giertz, S.H., 2012. The elasticity of taxable income with respect to marginal tax rates: a critical review. J. Econ. Lit. 50 (1), 3–50. March.

Yin, W., Basu, A., Zhang, J., Rabbani, A., Meltzer, D., Alexander, C., 2008. The effect of the medicare Part D prescription benefit on drug utilization and expenditures. Ann. Intern. Med. 148, 169–177.