

Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D

Christina M. Dalton* Gautam Gowrisankaran[†] Robert Town[‡]

May 29, 2017

Abstract

The standard Medicare Part D drug insurance contract is nonlinear—with reduced subsidies in a coverage gap—resulting in a dynamic purchase problem. We consider enrollees who arrived near the gap early in the year and show that they should expect to enter the gap with high probability, implying that, under the neoclassical model, the gap should impact them very little. We find that these enrollees have flat spending in a period before the doughnut hole and a large spending drop in the gap, providing evidence against the neoclassical model. We structurally estimate behavioral dynamic drug purchase models and find that a price salience model where enrollees do not incorporate future prices into their decision making at all fits the data best. For a nationally representative sample, full price salience would decrease enrollee spending by 31%. Entirely eliminating the gap would increase insurer spending 27%, compared to 7% for generic-drug-only gap coverage.

JEL Codes: I13, I18, D03, L88

Keywords: nonlinear prices, cost sharing, doughnut hole, discontinuity

We have received helpful comments from Jason Abaluck, David Bradford, Chris Conlon, Juan Esteban Carranza, Aureo de Paula, Pierre Dubois, Martin Dufwenberg, Liran Einav, Dave Frisvold, Guido Lorenzoni, Carlos Noton, Matthew Perri, Asaf Plan, Mary Schroeder, Marciano Siniscalchi, Changcheng Song, Ashley Swanson, Bill Vogt, Glen Weyl, Tiemen Woutersen, and seminar participants at numerous institutions. We thank Doug Mager at Express Scripts for data provision and Amanda Starc for data assistance. Nora Becker, Emma Dean, Mike Kofoed, Tola Kokoza, and Sanguk Nam provided excellent research assistance. Gowrisankaran acknowledges research support from the Center for Management Innovations in Healthcare at the University of Arizona. A previous version of the paper was distributed under the title “Myopia and Complex Dynamic Incentives: Evidence from Medicare Part D.”

*Wake Forest University

[†]University of Arizona, HEC Montreal, and NBER

[‡]University of Texas and NBER

1 Introduction

In 2006, the U.S. added a new entitlement to the Medicare program, Part D, which offered prescription drug coverage to enrollees on top of the original entitlements of hospital (Part A) and physician/outpatient services (Part B). Part D, which was the largest benefit change to Medicare since its introduction in 1966, has proven very popular with Medicare enrollees.¹ Despite its popularity, the program nonetheless has its critics. Perhaps the biggest criticism of Part D is its nonlinear price schedule. Enrollees with a standard Part D benefit faced modest out-of-pocket expenditures in the initial coverage region until their accrued total year-to-date drug spending placed them in the coverage gap—also called the “doughnut hole.” Once in the doughnut hole, the enrollee paid the full price of all drugs until reaching the catastrophic region. As shown in Figure 1, in 2008, the year of our data, the gap began at \$2,510 in total drug spending and did not end until \$4,050 in out-of-pocket expenditures, which corresponds to a mean of \$5,932.50 in total drug spending.²

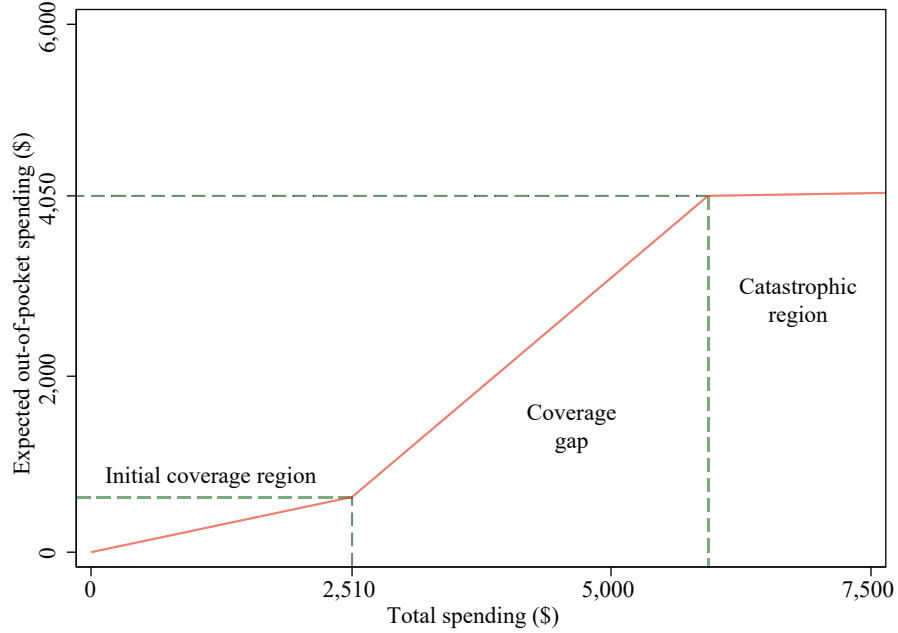
With a nonlinear price schedule, a rational dynamically-optimizing enrollee must forecast her future expenditures when making prescription purchase decisions. For instance, if she is currently in the initial coverage region but forecasts that she will end the year in the doughnut hole, then she would want to account for the higher future price, which would likely make her choose cheaper or fewer drugs than otherwise. If enrollees do not act as neoclassical dynamic optimizers in the presence of nonlinear insurance contracts, such contracts can create a welfare loss from “behavioral hazard,” defined as sub-optimal behavior resulting from mistakes or non-neoclassical biases (Baicker et al., 2012).

Understanding the importance of behavioral hazard in Part D is important because some studies find that Part D enrollees do not act fully rationally in their choice of Part D health plans (Abaluck and Gruber, 2011, 2013; Ho et al., 2014; Heiss et al., 2010; Schroeder et al.,

¹The program enrolled over 38 million (or 68%) of Medicare beneficiaries in 2013 (Medpac, 2014). Evidence indicates that Part D lowered Medicare beneficiaries’ out-of-pocket costs while increasing prescription drug consumption (Yin et al., 2008; Zhang et al., 2009; Lichtenberg and Sun, 2007; Ketcham and Simon, 2008).

²The mean coinsurance rates are 25% in the initial coverage region and 2% in the catastrophic region. The 25% rate implies that the initial coverage region has mean out-of-pocket expenditures of \$627.50. Thus, the coverage gap ends after a mean of \$3,422.50 in further out-of-pocket/total expenditures, for a combined \$5,932.50 in total drug spending.

Figure 1: Coverage by region in 2008 with standard Medicare Part D plans



2014),³ while other studies find that enrollees are, at least in part, rational in their Part D plan choice (Ketcham et al., 2012). Moreover, although the doughnut hole is specific to Part D, most health insurance plans have nonlinear aspects, such as out-of-pocket maxima and deductibles, implying that behavioral hazard is potentially important in many healthcare contexts.⁴ Finally, nonlinear contracts such as high-deductible health plans are likely to increase in the U.S. and other countries as a way to contain increasing costs.

This paper has two goals. The first is to test whether the behavior of Part D enrollees in their purchase of prescription drugs meaningfully deviates from the predictions of the neoclassical model of dynamically optimization. We develop tests that avoid several selection issues that often make such inference challenging. The second is to identify the sources and magnitudes of any behavioral hazard and how they affect counterfactual policy outcomes.

³Also consistent with behavioral hazard, critics of Part D point to the possibility that the doughnut hole may lead to adverse health consequences (Liu et al., 2011).

⁴This point that has been recognized since at least the RAND Health Insurance Experiment, which found that utilization increased once enrollees hit their out-of-pocket maxima (Newhouse, 1993).

We proceed by constructing two behavioral dynamic models of drug purchases: quasi-hyperbolic discounting (Laibson, 1997; Phelps and Pollak, 1968; Strotz, 1956) and price salience (Chetty et al., 2009; Bordalo et al., 2012). The neoclassical model is a limiting case for both models. For both models, we derive and/or compute the implications for drug purchases in the face of nonlinear insurance contracts. We use the implications of these models and a discontinuity design to test for deviations from the neoclassical model and provide evidence that enrollees’ drug consumption behavior deviates from its predictions but can be explained by behavioral models. We then structurally estimate the parameters of both behavioral models. Using the estimated structural model, we obtain inference on which behavioral model can best explain purchase patterns, the importance of behavioral hazard, and the impact of policies such as eliminating the coverage gap.

We believe that our tests of the neoclassical model and estimation framework may be useful more broadly. In particular, there has been substantial recent interest in understanding the implications of nonlinear pricing in a variety of sectors, with many papers rejecting the predictions of the neoclassical model.⁵ We contribute to this literature by developing new tests of the neoclassical model—which are not vulnerable to many important selection issues—and a framework to structurally estimate both price salience and quasi-hyperbolic discounting.

Both of our behavioral models (as well as the limiting neoclassical model) consider a Part D enrollee’s drug purchase decisions within a calendar year. Each week, the enrollee faces a distribution of possible health shocks and, for each shock, chooses one of a number of drug treatments, or no treatment. Future weeks are discounted with the weekly discount factor δ . The drug choice decision is dynamic because purchasing a drug in the initial coverage region moves the enrollee closer to the coverage gap. With our first behavioral model of quasi-hyperbolic discounting, the enrollee or her physician discounts future health expenditures in

⁵Brot-Goldberg et al. (2015) find that employees who were forced into a high deductible health insurance plan significantly reduced healthcare expenditures even when they would not reduce out-of-pocket expenditures from this decision. Ito (2014) shows that enrollees respond to average electricity prices, even though the neoclassical model implies that people should respond to marginal prices. Grubb and Osborne (2014) finds that consumers exhibit a range of biases in nonlinear cellular phone contracts. Finally, Nevo et al. (2016) model forward-looking consumers faced with nonlinear broadband internet contracts.

the current week with the factor β , in one week with the factor $\beta\delta$, in two weeks with the factor $\beta\delta^2$, etc. A quasi-hyperbolic discounter with $\beta < 1$ is myopic: she would make different tradeoffs at time t between utility at times $t+1$ and $t+2$ than she would make upon reaching time $t+1$.⁶ Our second behavioral model, price salience, specifies that any decision that the enrollee and her physician make in the initial coverage region only incorporates the possibility of a price change in the doughnut hole with probability σ . Doughnut hole prices become fully salient during the first purchase decision made *after* arriving inside the coverage gap. A value of $\sigma < 1$ implies that doughnut hole prices are less than fully salient. The two behavioral models predict different timings of when the coverage gap prices are fully internalized and as a consequence (and as long as $\beta < 1$ or $\sigma < 1$) imply different consumption dynamics as enrollees approach and enter the doughnut hole. For $\beta = \sigma = 1$, both behavioral models are equivalent to each other and to geometric discounting with full salience. We define the neoclassical model to be the geometric discounting model with δ close to 1 at an annual level.⁷

In the neoclassical model, drug purchase decisions depend largely on the distribution of coverage regions where the individual expects to end the year. To see this, consider an extra drug purchase in the initial coverage region for an enrollee who expects to end the year in the coverage gap. This extra purchase results in some later purchase(s) no longer having an insurance subsidy, implying that the total extra price will be roughly the full price rather than the price with insurance. This makes robust tests of the neoclassical model challenging, generally requiring an estimation of the expected distribution of the coverage regions where the individual expects to end the year, made at each potential purchase point in the sample.

Our innovation is to consider enrollees who have reached \$2,000 in total spending early in the year. Since these enrollees have reached *near* the coverage gap start of \$2,510 early in the year, we hypothesize, and then verify, that they will enter into the coverage gap with near certainty and leave with low probability. Thus, we can approximate rational expectations

⁶We estimate both specifications where the quasi-hyperbolic discounters are sophisticated and naïve about their future behavior.

⁷Simple economic theories imply that people should only save money if their discount factor is greater than the real interest rate. Since savings do occur, these theories and the low observed real interest rates imply that discount factors are close to 1 at an annual level.

with the simple assumption that the enrollee will end the year in the gap with certainty. Moreover, since these enrollees will end the year in the gap with high probability, Part D insurance is very close to a fixed subsidy for these enrollees under the neoclassical model. We show that this implies that, under the neoclassical model, there should be little or no drop in prescription drug purchases upon entering the doughnut hole. In contrast, under either behavioral model, because enrollees do not fully account for the prices that they will pay in the coverage gap, purchases will be flat away from the doughnut hole, drop on approach into the doughnut hole, and again be flat inside the doughnut hole. Finally, for the geometric discounting model with a low but positive δ , purchase probabilities should drop throughout the initial coverage region.

We test the predictions of the neoclassical model by examining whether there are drops in spending upon reaching the doughnut hole for the set of enrollees noted above. We further test for geometric discounting with a low discount factor versus the behavioral models by evaluating whether purchases are flat in a period before the doughnut hole. Finally, since the two behavioral models have different predictions as to when doughnut hole prices start to affect behavior, our structural estimation identifies the most appropriate behavioral model by evaluating which estimated structural model fits the data best on this dimension.

Our empirical work is based on 2008 Medicare Part D administrative claims data from a large pharmacy benefit manager. Using the subset of enrollees who arrive near the doughnut hole early in the year, we estimate weekly spending as a function of individual fixed effects and an indicator for being in the coverage gap. Consistent with the predictions of the behavioral models, we find that drug purchases are flat in a region before the doughnut hole and drop significantly and sharply upon reaching the doughnut hole, with mean total drug expenditures falling by 28% and the number of filled prescriptions falling by 21%. Thus, we find violations of the neoclassical model.

We identify the sources and magnitudes of behavioral hazard by structurally estimating the parameters of our models for the quasi-hyperbolic discounting and price salience models using a nested fixed-point maximum likelihood estimation and the the same subset of enrollees. While versions of the quasi-hyperbolic discounting model have been previously

estimated (e.g. Fang and Wang, 2013), to our knowledge, this is the first paper to estimate a structural dynamic model of price salience. The parameters of the structural models are price elasticity parameters, fixed effects for each drug, the geometric discount factor δ , and the behavioral parameter β/σ .⁸ We show that we can identify the discount factor and behavioral parameter given sufficient variation in drug attributes.

Our structural estimation splits our sample into subsamples based on an ex-ante measure of expected pharmacy expenditures. For each of the three subsamples, we can reject $\beta/\sigma > 0$ for both models. The price salience model fits the data better, with a much higher estimated likelihood. The reason is because the quasi-hyperbolic discounting model cannot explain the sharpness of the drop in drug spending, even with $\beta = 0$, which has the sharpest spending drop. These findings imply that future doughnut hole prices are not at all salient when in the initial coverage region. Alternately put, enrollees in our sample appear not to take future coverage gap prices into account at all in their choices of drugs.

Using our structural estimates, we examine behavioral and policy counterfactuals for a nationally representative sample.⁹ To isolate the importance of price salience, we examine how prescription purchase behavior changes under the neoclassical model, using an annual discount factor of 0.95. Neoclassical optimization would cause enrollees to reduce their spending by 31%, with total prescription drug spending dropping by 15%. In contrast, eliminating drug insurance would lower total prescription drug spending by 35%, implying that both behavioral hazard and drug insurance are important in this market.

Our policy counterfactuals examine the elimination of the doughnut hole as mandated by the 2010 Affordable Care Act. We find that eliminating the doughnut hole would increase total spending by 10% and insurer spending by 27%, implying a substantial cost to the government. Coinsurance would have to increase to 37% from the current average of 25% to implement a revenue neutral insurance scheme without the doughnut hole. Providing doughnut hole coverage for generic drugs only would increase insurer spending by only 7%.

⁸We use “ β/σ ” throughout the paper to mean “ β in the quasi-hyperbolic discounting model and σ in the price salience model.”

⁹The sample is composed of a mix of the estimation sample and others in our claims data, with the mix chosen to ensure that the percent of enrollees reaching the doughnut hole is equal to the population average.

Our paper is most closely related to the works of Einav et al. (2015) and Abaluck et al. (2015) who both also consider the implications of benefit design for Medicare Part D. We develop complementary tests to Einav et al. (2015): we test for violations of the neoclassical model by evaluating whether there are changes in behavior upon crossing into the doughnut hole when the neoclassical model predicts none, while Einav et al. tests for the presence of forward-looking behavior by evaluating whether there are changes in behavior when predicted by the neoclassical model (in their case, across enrollees joining Part D plans with deductibles at different points of the year). Our tests avoid selection issues that may be present in other studies by comparing the same individuals at different points in time. Einav et al. also estimate a structural, dynamic model and find that the weekly discount factor is $\delta = 0.96$, implying an annualized discount factor of 0.12; our framework provides a behavioral explanation for our findings and can reject the geometric discounting model with a low but positive δ . Our structural estimation also builds on Einav et al. by developing a modeling framework for drug choices that is more similar to a standard dynamic multinomial choice models and by providing results on identification for this type of model. Abaluck et al. (2015) use a very different identification strategy based on the assumption that changes in plan benefits are exogenous and do not result in enrollee selection due to plan stickiness. Using this assumption, they develop a simpler structural model of drug choice that abstracts away from the fact that enrollees may not fully know their health shocks requiring drug purchases at the beginning of the year. They also find that price salience plays an important role in explaining deviations from the neoclassical model. Finally, our structural model of quasi-hyperbolic discounting builds on Fang and Wang (2013) and Chung et al. (2013).

The paper proceeds as follows. Section 2 provides our model. Section 3 describes our data. Section 4 presents evidence based on the discontinuity near the doughnut hole. Section 5 describes the econometrics of our structural model. Section 6 provides results and counterfactuals, and Section 7 concludes.

2 Model

2.1 Overview

We develop a dynamic framework to study the drug purchase decisions of a Medicare Part D enrollee within a calendar year.¹⁰ We consider two behavioral models as well as the limiting case of the geometric discounting model. Our first behavioral model allows enrollees to have time-inconsistent or myopic preferences that satisfy quasi-hyperbolic discounting (Laibson, 1997; Phelps and Pollak, 1968; Strotz, 1956). In this model, enrollees are present-biased and discount the future more than would geometric discounters. Our second behavioral model allows future doughnut hole prices to lack full “salience” (Chetty et al., 2009; Bordalo et al., 2012; Abaluck et al., 2015). In this model, the enrollee does not pay full attention to the fact that prices will change in the future. The two explanations differ in the underlying causes of the departure from neoclassical behavior. Moreover, as we formalize below, the two models imply different purchase patterns near the coverage gap start, thereby allowing our estimation to evaluate the sources of any deviations from the neoclassical model.¹¹

We now explain the choice framework of our model. A period in our model is a week, starting with Sunday.¹² Future weeks are discounted with the weekly (geometric) discount factor δ . Each week, the enrollee is faced with a number, zero or more, of health shocks. A health shock is defined by a unique set of drugs (a “drug class”) that can be used as treatments for that shock. An example of a health shock is “conditions treated with calcium channel blockers,” which has an accompanying drug class of “calcium channel blockers.” An example of a calcium channel blocker is Cardizem (diltiazem hydrochloride) in tablet form; our uniqueness assumption implies that this drug is not in any other drug class. Upon receiving a health shock, the enrollee makes a discrete choice of one of the drugs in the

¹⁰Section 5 discusses estimation of the model which involves aggregation across enrollees.

¹¹A previous working paper version of this paper only allowed for quasi-hyperbolic discounting. The current model generalizes the earlier version by considering both price salience and time-inconsistent preferences.

¹²Our empirical analysis uses the enrollee/week as the unit of observation. A longer time interval, such as a month, would reduce information through aggregation, while a shorter time interval, such as a day, may have noisy outcomes because a typical enrollee will fill zero prescriptions on most days. We chose an interval of a week as a balance between these two constraints.

accompanying drug class, or the outside option, which consists of no drug treatment. It is important to model the outside option because individuals may substitute away from drug purchases when in the doughnut hole.

The enrollee’s number of health shocks in each week is distributed *i.i.d.* categorical (equivalently, multinomial with one trial), with a minimum of 0 and a maximum of N shocks. An enrollee who receives a health shock does not know how many more health shocks she will receive in the week, although she does know the parameters of her categorical distribution, and hence her conditional distribution of additional shocks. Each health shock is an *i.i.d.* draw from the enrollee’s drug class distribution.¹³ Because the distribution of drug classes is specific to an enrollee, our model is consistent with within-enrollee correlations of drug classes, as would occur with a chronic disease. For instance, some enrollees might have type II diabetes, and those enrollees would draw from a drug class distribution that includes insulin sensitizers, while other individuals would not have type II diabetes and hence would draw from a distribution that excludes this drug class.¹⁴

The enrollee’s decision problem is dynamic because each drug purchase brings her closer to the next phase of her nonlinear insurance scheme (i.e., the coverage gap if in the initial coverage region), and purchasing an expensive drug brings her relatively closer than purchasing a cheaper one. The quasi-hyperbolic discounting model specifies that the enrollee discounts a future event $t \geq 0$ weeks in the future with factor $\beta\delta^t$. We estimate two variants of the quasi-hyperbolic discounting model (Strotz, 1956; Fang and Wang, 2013). Under the “sophisticates” model, the enrollee knows that in the future she will continue to act as a quasi-hyperbolic discounter. Under the “naïfs” model, the enrollee believes that she will follow the geometric discounting model in future drug purchase decisions. Both variants with $\beta = 1$ are equivalent to the geometric discounting model.

The price salience model focuses on the information that the enrollee uses to make her drug purchase decision. We specify that the enrollee—or her physician acting as her agent—

¹³We model multiple potential drug purchases within a week in this way in order to leverage standard discrete choice multinomial logit models for each individual purchase decision.

¹⁴Our structural estimation stratifies enrollees into groups based on health risks and allows for each group to draw from different health shock distributions.

makes her drug purchase decision prior to the point of sale, e.g., in the physician’s office or at home before going to fill a prescription when her current supply runs out.¹⁵ At the decision point, the enrollee is aware of the drug prices in the coverage region of her last purchase, but is not necessarily fully salient about future prices. We assume that the enrollee in the initial coverage region assesses a probability σ that there is a future coverage region, with this probability changing to 1 only after the individual has made a purchase that brings her into the gap. In other words, with $\sigma < 1$, the first purchase decision made with full salience about the doughnut hole prices will be the first one made after \$2,510 or more in total expenditures. Note that $\sigma = 1$ is equivalent to the geometric discounting model.

2.2 Enrollee optimization

We first introduce some additional notation and then formally define enrollee preferences. We represent the categorical distribution of the number of health shocks via conditional probabilities: let Q_n , for $n = 0, \dots, N$, denote the conditional probability of having another health shock given that n have already occurred in the current week. Note that $Q_N = 0$. At the n th drug purchase decision node in any week, the enrollee’s information regarding the number of future health shocks that she will receive in the week is given by $Q_n, \dots, Q_N$.

Let H denote the number of drug classes (or health shocks). We assume that health shock $h \in \{1, \dots, H\}$ occurs with probability P_h . For each h , denote the prescription drugs that can be used for treatment by $j = 1, \dots, J_h$. For each h and j , let p_{hj} denote the full price and oop_{hj} denote the out-of-pocket price when inside the initial coverage region. Each h also has a baseline health cost c_h that applies equally to all treatment options.¹⁶

At each state, the enrollee maximizes the expected discounted value of her perceived flow utility subject to her behavioral biases regarding the valuation of future states, the salience of price changes, and expectations regarding her future behavior. We now discuss

¹⁵This is similar to other empirical specifications. For instance, Chetty et al. (2009)’s estimation is based on the idea that purchase decisions for grocery store items are made at the place where items are displayed and not at the point of sale.

¹⁶Since c_h affects all options equally, it does not affect choices and is not identified. Since our counterfactuals never alter the distribution of health shocks, our counterfactual conclusions are not affected by the fact that we cannot estimate it.

the components of the perceived flow utility for the different models. Appendix B details the behavioral dynamic optimization problems.

The perceived flow utility from drug j for drug class h is additive in: (1) the fixed utility from treatment, ϕ_{hj} , which is a parameter to estimate; (2) the disutility from the current perceived price of the drug, which we detail below;¹⁷ and (3) an unobservable component ε_{hj} , which is distributed type 1 extreme value, *i.i.d.* across health shocks and individuals. We assume that current, but not future, values of ε_{hj} are known to the individual when making her choice decision. For all h , denote the outside option as good 0. We assume that $p_{h0} = oop_{h0} = \phi_{h0} = 0$ and that its flow utility is ε_{h0} .

Our estimation focuses on enrollees who have spent to near the start of the coverage gap early in the year. Given this, our tests of the neoclassical model and estimation of the structural parameters are based on:

Assumption 1. *With probability 1, enrollees in our sample expect that, even if they change their purchase for any one health shock:*

- (a) *they will reach the doughnut hole start of \$2,510 in total spending, and*
- (b) *they will not reach the sample minimum catastrophic region start.*

Given Assumption 1, we can treat the doughnut hole as an absorbing state which will always be reached. This allows us to construct simple tests of the neoclassical model. It also simplifies the dynamic decision problem, as we do not have to account for the week of the year and can instead exposit the problem as an infinite horizon Bellman equation. This in turn reduces the computational burden of our estimation algorithm.¹⁸ The state space at the time of a drug purchase then consists of four elements, with a typical state written as $(m, n, h, \vec{\varepsilon})$: m indicates the monetary distance to the doughnut hole at the start of a given purchase decision;¹⁹ $n \in \{0, \dots, N - 1\}$ indicates the number of previous health shocks

¹⁷Our inclusion of this price term in the flow utility is equivalent to there being a money good with utility equal to the negative of this term.

¹⁸Our counterfactuals do not impose Assumption 1 because it may not be close to accurate under counterfactual benefit structures. The state space for the counterfactuals explicitly incorporates a 52-week year and the three coverage zones (initial coverage region, coverage gap, and catastrophic coverage).

¹⁹For instance, if the individual had already spent \$2,350, then $m = \$2,510 - \$2,350 = \$160$.

during the week; h is the drug class of the shock; and $\vec{\varepsilon} \equiv (\varepsilon_{h0}, \dots, \varepsilon_{hJ_h})$ is the vector of unobservables. Let $s(m, n, h, j)$, $j = 0, \dots, J_h$ denote the probability of purchase of drug j for a given set of state variables (m, n, h) , integrating over $\vec{\varepsilon}$.

Let $p^{eff}(m, p_{hj}, oop_{hj})$ be the effective price perceived by the enrollee. When price is fully salient as in the quasi-hyperbolic discounting model, we can write p^{eff} as:

$$p^{eff}(m, p_{hj}, oop_{hj}) = \begin{cases} p_{hj}, & \text{if } 0 \leq m < oop_{hj} \\ oop_{hj} + p_{hj} - m, & \text{if } oop_{hj} \leq m < p_{hj} \\ oop_{hj}, & \text{if } p_{hj} \leq m. \end{cases} \quad (1)$$

In (1), the first line pertains to the enrollee who has to pay the full price because she is either already completely in the coverage gap or would be completely inside after paying her out-of-pocket price. The second line considers the intermediate case where the purchase would move the enrollee into the coverage gap at some point after she pays the out-of-pocket price. The last line considers the enrollee who is completely in the initial coverage region, even after the current purchase. The first and second lines reflect Part D rules which specify that, when a purchase moves the enrollee into the coverage gap, the enrollee pays the out-of-pocket price, the insurer pays any remaining amount until total spending reaches the coverage gap start, and the enrollee also pays the final remaining amount.

For the price salience model, p^{eff} satisfies:

$$p^{eff}(m, p_{hj}, oop_{hj}) = \begin{cases} p_{hj}, & \text{if } m = 0 \\ \sigma p_{hj} + (1 - \sigma) oop_{hj}, & \text{if } 0 < m < oop_{hj} \\ \sigma(oop_{hj} + p_{hj} - m) + (1 - \sigma) oop_{hj}, & \text{if } oop_{hj} \leq m < p_{hj} \\ oop_{hj}, & \text{if } p_{hj} \leq m. \end{cases} \quad (2)$$

In (2), the first and last line are the same as in (1). However, the middle two lines, which consider the intermediate cases where the purchase would move the enrollee into the coverage gap, are different. In these cases, with probability $1 - \sigma$, the enrollee perceives that prices are

simply the out-of-pocket prices, since her drug purchase decision was made in the physician’s office where these prices were not yet fully salient. But, with probability σ , the doughnut hole prices are salient and the individual pays the prices from (1).

Finally, let the function $\alpha(p)$ denote the disutility from current perceived spending level p . In order to flexibly capture the different impacts of price on decisions, our estimation allows $\alpha(\cdot)$ to be a linear spline, which nests the case of a linear price coefficient. Note that the current disutility from perceived price in our behavioral models will satisfy $\alpha(p^{eff}(m, p_{hj}, oop_{hj}))$.

Note that the price salience model is very similar in its implications to the quasi-hyperbolic discounting sophisticates model, but not to the naïfs model. With limited salience, the enrollee believes that at any future pre-doughnut hole state she will still perceive a salience probability of σ . This is similar to the sophisticate who believes that she will continue to act as a quasi-hyperbolic discounter in the future. Naïfs believe that they will act as geometric discounters in the future, which leads to different purchase decisions. The one difference between the price salience and sophisticates model is in p^{eff} for drugs that move the enrollee into the doughnut hole, which are the two middle cases in (2).²⁰

2.3 Testable Implications of the Model

This subsection discusses testable implications of our model that allow us to distinguish between neoclassical dynamically optimizing enrollees and other models. We focus on enrollees who have spent \$2,000 or more early in the year and hence impose Assumption 1 throughout.

Our main insight is that, under the neoclassical model—i.e., the geometric discounting model (where all prices are fully salient) and with an annual discount factor close to 1—enrollees will act approximately the same before and after the doughnut hole. This is not true for the quasi-hyperbolic discounting (with $\beta < 1$) and price salience models (with $\sigma < 1$). Intuitively, if neoclassical enrollees perceive that they will end the year inside the doughnut hole, they will always obtain the full insurance subsidy for the initial coverage region, and

²⁰It is also possible to define the drug purchase decision as occurring at (instead of prior to) the point of sale, in which case the salience model would be identical to the sophisticate model, except for the relabeling of the parameter β to σ .

hence will not change their behavior upon crossing into the doughnut hole. Hence, for these enrollees, Part D insurance is very similar to a lump-sum check for the insured amount. Formally, we can show that there is no change in behavior upon crossing into the doughnut hole, for the case where every drug has the same full and out-of-pocket prices and $\delta = 1$:

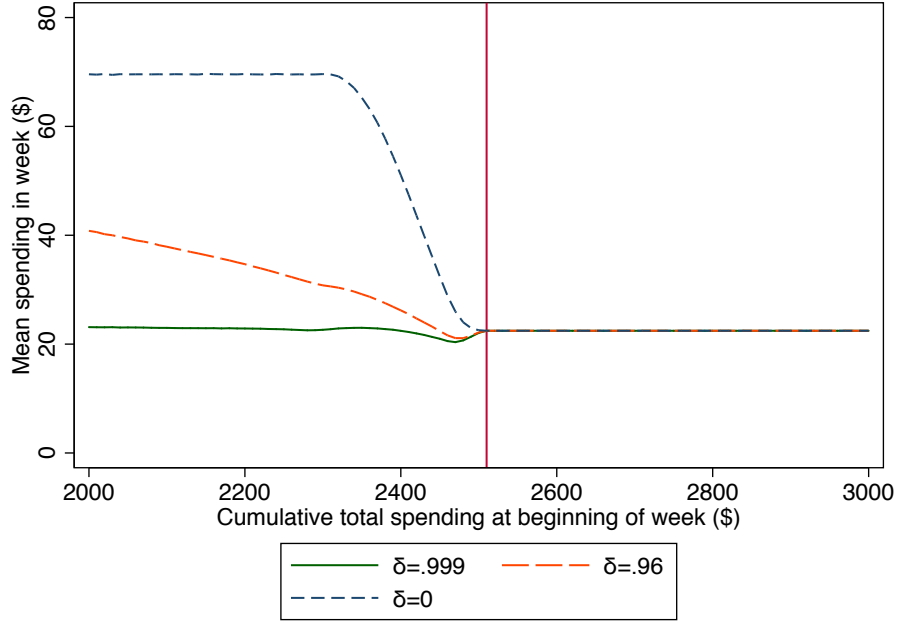
Proposition 1. *Consider a neoclassical dynamically-optimizing Part D enrollee for whom Assumption 1 holds and for whom $\beta = \delta = 1$. Suppose further that there is a common full price $\bar{p}$ and out-of-pocket price $\overline{op}$ that is charged for every (inside-good) drug and that price disutility is linear so that $\alpha(p) \equiv \bar{\alpha}p$. Then, the purchase probability of each hj is the same across ex ante states, i.e., for all m, n, m', n', h, j , $s(m, n, h, j) = s(m', n', h, j)$.*

(Proofs of propositions are in Appendix C.)

We note two points about Proposition 1. First, the proposition considers the case where all drugs have the same total and out-of-pocket prices. If there were variation in prices, then enrollees might change their behavior before and after the doughnut hole, because the doughnut hole start is based on total spending and not out-of-pocket spending. For instance, if one drug has a higher out-of-pocket price relative to the full price than a second one, then the enrollee would substitute towards the first drug when in the doughnut hole. Overall, though, we would expect such substitution to not affect the basic testable implication of the proposition, which is that, for this sample, crossing into the doughnut hole should not reduce spending. Second, while Proposition 1 considers the case of $\delta = 1$, we expect the results to be approximately true for δ close to 1.

To provide further insight as to the role of δ in the geometric discounting model in affecting dynamic drug consumption patterns we simulate the model for different values of δ . Figure 2 reports simulated mean total spending per week across discount factors as a function of the cumulative total spending at the beginning of the week. We report simulations for three discount factors δ : 0.999, which corresponds to an annualized 5% discount; 0.96, which is the weekly discount factor estimated by Einav et al. (2015) and corresponds to an annual discount factor of 0.12; and 0, the case of perfect myopia. We calculate dynamically-optimizing decision-making for enrollees and then simulate weekly spending in the figure.

Figure 2: Simulated drug spending for the geometric model across discount factors



Enrollees in the simulation all have one health shock each week and each health shock is drawn with equal probability from one of 20 drug class distributions, each with one drug.²¹

Figure 2 shows that mean weekly spending with $\delta = 0.999$ is flat before and after the doughnut hole. This occurs even though there are different priced drugs in our sample, generalizing Proposition 1.²² With $\delta = 0.96$, spending decreases throughout the initial coverage region and then is flat inside the doughnut hole. The reason for the sustained decrease is that the time value of money drives the drop in spending: with a 20% coinsurance, a foregone \$80 purchase with \$2,300 in total spending would result in \$20 in immediate savings and \$60 in savings discounted by the time until the enrollee expects to cross into the doughnut hole. The same foregone purchase with \$2,100 in total spending would have the \$60 in savings dis-

²¹Drug 1 has price $p = \$10$ and quality $\phi = 0.1$; drug 2 has price \$20 and quality 0.2. Other drugs follow the same pattern until drug 20, which has price \$200 and quality 2.0. Out-of-pocket prices *oop* are always 25% of total price. These ranges of prices are roughly similar to the sample.

²²The slight dip before the doughnut hole is due to the peculiarities of Part D coverage around the doughnut hole, as reflected in (1) and the discussion surrounding it, whereby cheaper drugs are insured at a higher rate than more expensive ones right before the doughnut hole.

counted more because the time until the expected crossing is longer. With $\delta = 0$, spending is flat in the pre-doughnut-hole region since discounted savings are worth nothing.

Now we consider spending given the behavioral models. Both behavioral models result in the future effectively being discounted but in a different way than for the geometric discounting model. With $\delta = 1$, in the quasi-hyperbolic discounting model, *all* future purchase occasions are discounted by the same β . In the price salience model, future doughnut hole prices are salient with the *same* probability σ . This suggests that the model can predict flat spending before and after the doughnut hole but a drop in spending upon reaching the doughnut hole. We formalize:

Proposition 2. *Consider a Part D enrollee for whom Assumption 1 holds and for whom $\delta = 1$. Suppose further that there is a common full price $\bar{p}$ and out-of-pocket price $\overline{op}$ that is charged for every (inside-good) drug and that price disutility is linear so that $\alpha(p) \equiv \bar{\alpha}p$. Finally, assume that there is a unique solution to the ex ante value functions for the behavioral models. Then, for any h and j ,*

- (a) *at the doughnut hole: under the sophisticates or naïfs quasi-hyperbolic discounting model with $\beta < 1$ or the price salience model with $\sigma < 1$, $s(0, n, h, j)$ will be equal to its value under the neoclassical model for all n, h, j ;*
- (b) *away from the doughnut hole: under the price salience model with $\sigma < 1$ or the sophisticates or naïfs quasi-hyperbolic discounting model with $\beta < 1$, $s(m, n, h, j) = s(m', n', h, j) > s(0, n'', h, j)$ if $m, m' \geq \bar{p}$ and for all n, n', n'', h, j ; and*
- (c) *across models: the purchase probabilities $s(m, n, h, j)$ will be the same for the sophisticates quasi-hyperbolic discounting model as for the price salience model and higher than for the quasi-hyperbolic discounting naïfs model if $m \geq \bar{p}$ and $0 < \beta = \sigma < 1$ and for all m, n, h and for $j = 1, \dots, J_h$.*

Proposition 2 shows that enrollees will purchase the same amount in every period when completely before the doughnut hole. Similarly, they will consume the same amount in each period when inside the doughnut hole. Importantly, however, the within doughnut hole

consumption will be strictly lower than the outside doughnut hole consumption. The logic for this is that, unlike in the neoclassical model, the decision process is now different before and inside the doughnut hole. In the initial coverage region, the quasi-hyperbolic discounter knows that she will essentially have to repay the insurance subsidy by moving one purchase into the doughnut hole, but that repayment is discounted with a factor β . The enrollee in the price salience model only considers that the repayment will occur with probability σ , thereby generating an analogous result. The fact that the effective discount of this repayment is always β/σ , regardless of how far the individual is from the coverage gap start, is what generates the result that spending is flat before the doughnut hole. Naïfs spend less than sophisticates in the pre-doughnut-hole region because naïfs expect that their future selves will make the most responsible choices possible, which raises the value in saving for the future.

Figure 3: Simulated drug spending for different behavioral models

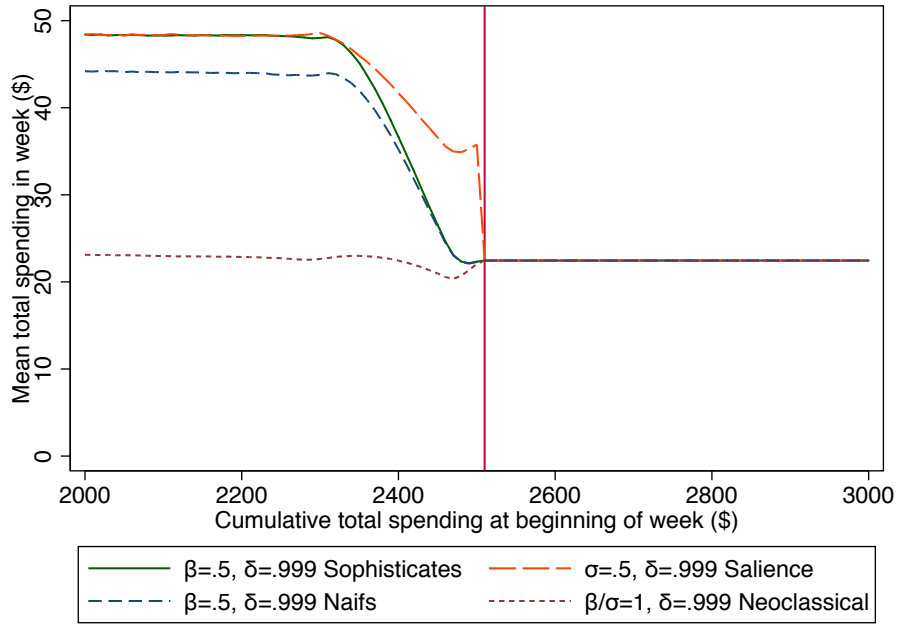


Figure 3 shows simulation evidence for the same set of flow utility parameters as in Figure 2 but now across behavioral models, setting $\delta = 0.999$ throughout. The figure displays results from the two quasi-hyperbolic discounting models with $\beta = 0.5$, from the salience

model with $\sigma = 0.5$, and also repeats the neoclassical case of $\beta = 1$ from Figure 2.

The figure shows that the same results from Proposition 2 are approximately true here. In particular, the three behavioral models all show virtually flat mean spending per week when the cumulative spending is less than \$2,310 (up to which even the most expensive drug would not move the enrollee into the doughnut hole). The sophisticates and price salience models generate virtually the same expected spending in the pre-\$2,310 region while the naïfs model shows lower spending. Note also that the behavioral models have different predictions from the geometric model with the low weekly discount factor of $\delta = 0.96$. Under the behavioral models, spending is flat until reaching a drug that could move the individual into the doughnut hole while under the low geometric discount factor model, spending decreases continuously from the beginning of the sample.

Importantly, the price salience model differs from the sophisticates model at the point of entry into the doughnut hole. Under the price salience model, enrollees are not fully aware of the doughnut hole prices until after the purchase that moves them into the doughnut hole, while the quasi-hyperbolic discounter makes decisions based on the price at the point of sale. Thus, as shown in the figure, the sophisticate will have lower spending than the enrollee with price salience in the region between \$2,310 and \$2,510. In the limiting case of $\sigma = 0$, under the price salience model, the enrollee would not lower her spending at all in this region, (given that there is only one health shock per week). This difference between the two models near the doughnut hole can identify which behavioral model is accurate.

Combining the insights from the propositions and the figures, the testable implications of our model are:

1. We can test for deviations from the neoclassical model by evaluating whether there is a significant drop in spending at the doughnut hole.
2. We can test for deviations from a geometric model with low δ by examining whether there is a region before the doughnut hole where spending is flat.
3. The price salience and sophisticates models have similar implications for drug purchases away from the coverage gap but the price salience model has higher spending

immediately before the doughnut hole, generating a steeper decline at the gap start.

4. Conditioning on other parameters, the naïfs model with $0 < \beta < 1$ has less spending before the coverage gap than the price salience or sophisticates models.

We test implications 1 and 2 in Section 4 and our structural estimation results in Section 6.1 are identified by implications 3 and 4.

3 Data

For our analysis, we rely on a proprietary claims-level dataset of employer-sponsored Part D plans in 2008, the third year of the program. The data come from the pharmacy benefits manager Express Scripts, which managed Medicare Part D benefits for approximately 30 different employer-sponsored Medicare Part D plans with a total of 100,000 enrollees. The plans were offered to eligible employees and retirees as part of their benefits. Employers receive subsidies from Medicare in exchange for providing these plans to their employees. We believe that enrollees in employer-sponsored Part D plans have, on average, higher income than typical Part D enrollees, and hence are less likely to be liquidity constrained. The employer-sponsored Part D market constituted nearly 7 million enrollees or 15 percent of Part D enrollment in 2008 (Medpac, 2009, p. 282).

The data contain all claims made by an enrollee in the year 2008 for each plan. For each claim, we have plan and patient identifiers, the age (at the fill date) and gender of the patient, the date the prescription was filled, the total price of the drug, the amount paid by the patient, the national drug code (a unique identifier for each drug), the pill name, the drug type (e.g., tablet, cream, etc.), the most common indicator of the drug (e.g., skin conditions, diabetes, infections, etc.), the dispensed quantity of the drug, and an indicator for whether the drug is generic or branded. We keep only individuals who are 65 or older at the time that they fill their first prescription.

Each of the employers offered multiple plans, each with different coverage structures. Our base analysis uses data from five Express Scripts plans. We chose these plans because (1)

they have a coverage gap that starts at exactly \$2,510 in total expenditures and ends at greater than \$4,000 in out-of-pocket expenditures; (2) there is no insurance in the coverage gap; and (3) the employers that offer these plans allowed us to use their data. We also include falsification evidence from a sixth plan which has the coverage gap start at a higher spending level.

Table 1: Plan characteristics and enrollment

	Plan					
	A	B	C	D	E	F
Employer	1	1	1	2	2	3
% of employees from employer	26	45	9	79	21	46
Deductible (\$)	275	100	100	0	200	0
Doughnut hole start (total \$)	2,510	2,510	2,510	2,510	2,510	4,000
Catastrophic start (out-of-pocket \$)	4,050	4,050	4,050	4,010	4,010	4,050
Total enrollment	7,541	12,858	2,431	4,062	1,058	35,395
% hitting \$2,510	20	13	16	16	13	20
% hitting catastrophic coverage	2	1	1	1	1	0
Estimation sample:						
Enrollment	620	644	126	304	49	2,981
% hitting \$2,510	96	94	95	97	94	97
% hitting catastrophic	11	6	9	10	12	0
Mean total spending (\$)	4,284	3,867	4,009	4,246	3,974	4,072
Mean out-of-pocket (\$)	2,373	2,010	2,125	2,045	2,071	1,026
Mean age	74	73	73	75	75	78
Percent female	62	58	53	62	59	64
Mean ACG score	1.04	1.17	1.18	0.91	1.07	0.67

Note: plan A provides generic coverage in deductible region; Plan F used for falsification exercise only and also provides generic coverage in doughnut hole.

Table 1 displays the characteristics of the six plans that we consider. The plans represent three different employers; plan and employer identities are masked. We consider all covered individuals at employer 2 and the majority of covered individuals at employer 1 (with the other covered individuals at this employer choosing plans with different coverage gap regions or some insurance in the coverage gap). Importantly, the fact that each covered individual could choose from only similar plans minimizes the selection issues across plans that one might observe in non-employer-sponsored Part D coverage.

Four of the five plans in our base analysis have a deductible. All deductibles take relatively

low values of \$275 or less. Each plan features a tiered drug copayment structure, with higher copays for brand and specialty drugs, and reduced copays for the use of mail-order pharmacies. By construction, the coverage gap start is the same across the base plans and the coverage gap end spending levels are similar. All six plans include generous coverage in the catastrophic region. Table 1 also lists summary statistics on plan enrollment. The five base plans cover a total of 27,950 individuals.

Our base estimation sample consists of all enrollees who start a week between Sunday, March 30 and Sunday, July 20, 2008 with total spending in the range [\$2000, \$2,510). We chose these dates and this range of spending to be in the part of the year where enrollees are not yet in the doughnut hole but should perceive that they will end the year in the doughnut hole with very high probability under the neoclassical model. This sample contains 1,743 enrollees distributed across the five plans in our sample. Between 94 and 97 percent of the enrollees in the estimation sample hit the coverage gap during the year, reflected in a mean total spending levels of approximately \$4,000 across the plans. The mean percent hitting the catastrophic coverage region ranges from 6 to 12 percent, reflected in mean out-of-pocket spending levels of approximately \$2,200 across plans, or about 55 percent of the value necessary to hit the start of catastrophic coverage.

The falsification plan F has the coverage gap start at \$4,000 in total spending, a much higher level than for the base plans. Its enrollees are older and disproportionately female relative to the plans in our base analysis sample. It also provides generic drug coverage during its coverage gap. Very few of its enrollees hit the catastrophic coverage region, due to the fact that they require much higher *total* spending to reach it.

Using our database of claims, we first drop claims for drugs which we believe are not in the formulary. Drugs that are not in the formulary are sometimes reported to the insurance company by the enrollee but do not count towards spending for purposes of determining if the enrollee is in the coverage gap or catastrophic coverage regions. We assume that any claim in the initial coverage region for which the total price is \$100 or higher and the out-of-pocket price is the same as the total price reflects a drug that is not in the formulary.²³ We

²³We also drop one claim with a quantity-filled entry of over 1 million.

then calculate the dollars until the doughnut hole (m) for each prescription by tabulating the spending up to this point during the year.²⁴

We merge our claims data with data on the expected pharmacy claims cost for each patient, based on their claims from before our sample period. Specifically, we use claims from Jan. 1, 2008 to Mar. 29, 2008 to construct the Johns Hopkins Adjusted Clinical Group (ACG) Version 10.0 score for each enrollee. The ACG score is meant to predict the drug expenditures over the following one-year period. We use the ACG scores to define groups for the structural analysis and then estimate separate coefficients for each group. ACG scores have been widely used to predict future health expenditures in the health economics and health services literature (see, e.g., Handel, 2013; Gowrisankaran et al., 2013). Table 1 shows that the base plans have mean ACG scores which are similar to the over-65 population mean score of 1; the falsification plan has a somewhat lower mean score.

Our analysis classifies each drug into a unique drug class meant to capture the function of the drug. We had the drug class coding performed by a clinically trained research assistant using the pill name, drug type (e.g., tablet or cream), most common indication, and national drug code. We classified drugs on the basis of function rather than the diseases they treat because we believe that drug function is the relevant attribute for a choice model. Thus, even though both calcium channel blockers and renin-angiotensin system blockers are used to treat hypertension, we treat them as separate drug classes because their mechanisms are separate.

Table 2 lists the drug classes with the most claims in our estimation sample. Approximately 9 percent of the claims were for cholesterol-lowering (antihyperlipidemic) drugs. The next most common categories include blood pressure medicines, opioids, and antidepressants.²⁵

²⁴There is some ambiguity of the order of claims if there are multiple claims filled on the same date for a given enrollee. For such multiple claims, we assume that the claims are filled in increasing order of out-of-pocket price. For multiple claims for an enrollee on a given date with the same out-of-pocket price, we use the order specified in the database that we received from Express Scripts.

²⁵Table A1 in Appendix A provides details on the ten most common drugs purchased.

Table 2: Most common drug classes in base estimation sample

Drug class	Number Rx	% of obs.	Most common Rx
Cholesterol Lowering	2,143	9.4	Simvastatin
Renin-Angiotensin System Blocker	1,814	7.9	Lisinopril
Beta-Blocker	1,259	5.5	Metoprolol
Opioid	1,200	5.2	Hydrocodon
Antidepressant	1,190	5.2	Sertraline
Diuretic	1,183	5.2	Furosemide
Calcium Channel Blocker	933	4.1	Amlodipine
Insulin Sensitizer	792	3.5	Metformin
Gastroesophageal Reflux & Peptic Ulcer	778	3.4	Omeprazole
Hypothyroidism	774	3.4	Levothyroxine

4 Evidence from Discontinuity Near Doughnut Hole

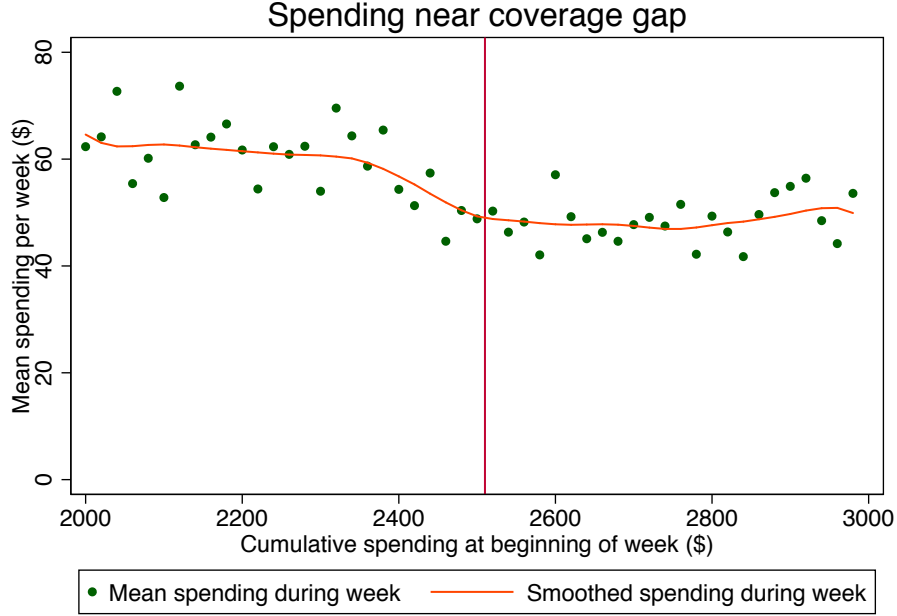
This section presents evidence on whether individuals act in a way that is consistent with the neoclassical model, with geometric discounting with a low but positive discount factor, or with our behavioral models. We base our evidence on the testable implications of the model developed in Section 2.3. We perform a series of discontinuity-based analyses that all use our analysis sample of enrollees who arrived near the doughnut hole in the middle of the year. Our analyses are similar to a standard regression discontinuity framework. However, while regression discontinuity analyses typically consider different individuals near a breakpoint, we consider the same individual immediately before and after reaching the coverage gap.

Specifically, the unit of observation for each regression is an enrollee observed over a week. Enrollees are in the estimation sample from the first week with starting expenditures of over \$2,000 until the last week with starting expenditures of less than \$3,000, or the end of the year if it comes first.

We start by graphing mean weekly spending levels and non-parametric regressions of these levels. Figure 4 plots mean total drug spending by \$20 increments of beginning-of-week cumulative spending and a kernel smoothed “lowess” regression of mean total drug spending on beginning-of-week cumulative spending.²⁶ The mean total drug spending shows little

²⁶We use a bandwidth of 0.3 for these regressions.

Figure 4: Spending near coverage gap for base estimation sample



change in spending over the range \$2,000-2,380 in beginning-of-week cumulative spending. Mean spending then drops until the doughnut hole and remains roughly constant until the highest cumulative spending level.

Note that week observations that are near the doughnut hole but not yet in the doughnut hole may move the individual into the doughnut hole, either because of an expensive drug or because of multiple drugs. Thus, the fact that spending starts to drop slightly before the doughnut hole does not necessarily indicate that individuals are forward-looking. In contrast, the flat spending in the \$2,000-2,380 range and the flat but lower spending in the doughnut hole range is a pattern that is consistent with quasi-hyperbolic discounting or limited price salience but not geometric discounting with $\delta > 0$, as in Figure 2.²⁷

Figure A2 in Appendix A provides a falsification exercise on Plan F, which had a coverage gap that started at \$4,000 in total spending. We report the same plots on this plan as on our base sample. We find very different results: there is no drop in spending upon reaching

²⁷Figure A1 in Appendix A displays the analogous figure to Figure 4 for the catastrophic zone. The catastrophic sample size is small and so the impact of entering the catastrophic zone on spending is imprecise.

\$2,510 in total spending. This result allows us to rule out that our results are due to the drop in spending when hitting \$2,510 in our sample being coincident with a medical condition, such as the seasonal onset of a disease. Thus, the figure supports the conclusion that the drop in spending is due to the coverage gap itself.

Having shown visually that there is flat spending in a region before the doughnut hole and a drop in spending at the coverage gap start, we now examine the data in more detail with linear regressions. Our linear regression specifications follow:

$$Y_{it} = FE_i + \lambda_1 1\{0 < m_{it0} \leq \$110\} + \lambda_2 1\{m_{it0} = 0\} + v_{it}, \quad (3)$$

where m_{it0} is the beginning-of-week spending left until the doughnut hole, FE_i are enrollee fixed effects, λ_1 is the coefficient on an indicator for being above \$2,400 in spending (within \$110 of the doughnut hole) and λ_2 is the coefficient on an indicator for being in the doughnut hole, which implies starting the week with at least \$2,510 in expenditures. We examine a number of different dependent variables Y_{it} , including total prescription drug expenditures, branded drug expenditures, and number of prescriptions filled. The λ_1 coefficient captures the fact that observations that are near the doughnut hole but not yet in the doughnut hole may move the individual into the doughnut hole.

By selecting a small region around the doughnut hole, we are comparing the same individual at similar points in the year but faced with different current prices. This minimizes the possibility that factors other than the presence of the doughnut hole might be influencing our findings. By including individual fixed effects, we are further controlling for individual differences at different points in our sample, i.e. the possibility that more severely ill individuals show up more in the region after the doughnut hole.

Our first set of linear regression findings are reported in Table 3.²⁸ We find sharp drops in most measures of prescription drug use. Supporting the results in Figure 4, total drug spending dropped by \$18 from a baseline of \$62. The number of prescriptions fell by 21% from a baseline mean of 0.84 per week. Branded prescriptions fell more than generic prescriptions:

²⁸In the interest of brevity, we do not report either the enrollee fixed effects or λ_1 values in our tables.

27% versus 19%. Similarly, expensive prescriptions – those with a total price of \$150 or more – fell by 27% while inexpensive ones – those under \$50 – had no significant drop. The mean total price of a prescription fell by 12% from a baseline level of \$80. All effects, except for those on the number of inexpensive prescriptions, are statistically significant. Not reported in the table, the indicators for weeks that start with \$2,400 to \$2,509 in total spending are generally significantly negative and much smaller than the reported coverage gap indicators.

These results paint a picture of enrollees who react strongly to being in the doughnut hole. As discussed in Section 2.3, the interpretation of this result is that individuals have either a β/σ or a δ that is substantially less than one: they are not acting as neoclassical agents in the dynamics of their drug purchase decisions.

Table 3: Behavior for sample arriving near coverage gap

Dependent variable:	Mean value before \$2,400	Beginning of week spending in: \$2,510 - 2,999	<i>N</i>
Mean spending in week	61.97	−17.46** (1.38)	28,543
Mean price per Rx	79.47	−9.77** (1.37)	10,846
Number of Rxs	0.84	−0.18** (0.02)	28,543
Number of branded Rxs	0.30	−0.08** (0.01)	28,543
Number of generic Rxs	0.54	−0.10** (0.01)	28,543
Expensive Rxs	0.12	−0.04** (0.00)	28,543
Medium Rxs	0.23	−0.06** (0.01)	28,543
Inexpensive Rxs	1.10	−0.01 (0.01)	28,543

Note: standard errors are in parentheses. ‘***’ denotes significance at the 1% level and ‘*’ at the 5% level. Each row represents one regression. All regressions also include enrollee fixed effects and an indicator for beginning-of-week spending between \$2,400 and \$2,509, and cluster standard errors at the enrollee level. An observation is an enrollee/week for an enrollee in the base estimation sample and beginning-of-week spending $\geq$ \$2,000 and $<$ \$3,000. Inexpensive Rxs are less than \$50 and expensive ones are \$150 or more.

Next, Table 4 provides evidence on whether drug spending is downward sloped in all regions before the doughnut hole, as predicted by the geometric model with a low but positive discount factor (e.g. Einav et al., 2015), but not by the behavioral models. We perform the same regressions as in Table 3 but with the addition of an extra regressor, which measures the change in spending in the region \$2,200 to \$2,399. Thus, the excluded region is now \$2,000 to \$2,199. Supporting the results in Figure 4 again, there is no significant effect of total spending in the \$2,200 to \$2,399 range. The implication is that, while spending before

the doughnut hole is higher than in the doughnut hole, the increment does not grow as one moves further back, inconsistent with the geometric model with a low but positive discount factor but consistent with the predictions of the behavioral models.

Table 4: Behavior near coverage gap with variation in pre-coverage gap region

Dependent variable:	Mean value before \$2,400	Beginning of week spending in:		<i>N</i>
		\$2,510 - 2,999	\$2,200 - 2,399	
Mean spending in week	61.97	−17.79** (1.76)	−0.68 (2.25)	28,543
Mean price per Rx	79.47	−8.97** (1.72)	1.64 (2.13)	10,846
Number of Rxs	0.84	−0.20** (0.02)	−0.03 (0.03)	28,543
Number of branded Rxs	0.30	−0.08** (0.01)	0.01 (0.01)	28,543
Number of generic Rxs	0.54	−0.12** (0.02)	−0.04* (0.02)	28,543
Expensive Rxs	0.12	−0.04** (0.01)	−0.00 (0.01)	28,543
Medium Rxs	0.23	−0.06** (0.01)	0.00 (0.01)	28,543
Inexpensive Rxs	1.10	−0.02* (0.01)	−0.01 (0.02)	28,543

Note: standard errors in parentheses. ‘***’ denotes significance at the 1% level and ‘**’ at the 5% level. Each row represents one regression. All regressions also include enrollee fixed effects and an indicator for beginning-of-week spending between \$2,400 and \$2,509, and cluster standard errors at the enrollee level. An observation is an enrollee/week for an enrollee in the base estimation sample and beginning-of-week spending $\geq$ \$2,000 and $<$ \$3,000. Inexpensive Rxs are less than \$50 and expensive ones are \$150 or more.

Finally, Table A2 in Appendix A provides evidence on the five drug classes which have the largest drops in prescriptions upon entering the doughnut hole and the five with the largest increases in prescriptions. Here, we perform similar regressions to Table 3 but with the number of prescriptions in the drug class as the dependent variable. We then report the drug classes with the biggest and smallest coefficients on the spending drop in the doughnut hole region. The five drug classes with the biggest drops in prescriptions are also among the ten most common drug classes, as reported in Table 2. Indeed, the only one of the top five drug classes that does not have a drop that is also in the top five is opioids. The five drug classes with the biggest increases in prescriptions upon entering the doughnut hole are all drug classes with very few prescriptions (and the coefficients are all insignificant). Overall, this table shows that the percentage drops in prescriptions are similar across most drug classes. This finding is also consistent with Chandra et al. (2010) who find similar demand responses to increased cost-sharing across drug categories. Appendix D considers, and eliminates, a number of other threats to our identification of our results rejecting the

neoclassical model and geometric model with a low but positive discount factor.

5 Econometrics of the Structural Model

5.1 Estimation

We structurally estimate the model developed in Section 2. Our estimation partitions enrollees into groups $g = 1, \dots, G$ based on their ACG score, with separate parameters by group. We assume that Q_n (the probability of further health shocks), N (the maximum number of health shocks), and P_h (the probability of each health shock) vary across groups. Our data include 8 discrete ACG score groups. Table A3 in Appendix A provides details on the enrollees by group.

Our data do not allow us to directly estimate P_h and Q_n since we do not know when enrollees have a health shock but choose the outside good. Rather than attempting to identify these parameters from our estimation sample, we estimate them from the same enrollees, observed earlier in the year. Specifically, we assume that enrollees in our estimation sample will always choose an inside drug in the months before they enter our assumption sample, with the logic being that the doughnut hole is sufficiently far away. Thus, we estimate P_h and Q_n for each group from the weekly drug purchases for enrollees in our estimation sample in that group using their purchases measured from the first week at which they start after the deductible region (conservatively defined as \$300 in total spending) until the last week before they enter our sample (which starts at \$2,000 in total spending).

We estimate a separate P_h and Q_n distribution for each group g . In addition, as noted in Table A3, we allow the other parameters to vary in three sets: the lowest, highest, and all other ACG scores. For each estimation, we lump together drug classes with fewer than 100 prescriptions filled for the estimation sample over the entire year and in a class called “Other.” We also lump together drugs within a drug class as “Other” until such point as every drug has at least 50 prescriptions filled over the entire year. We make these simplifications for computational tractability, since our estimation has fixed effects for each drug and requires

an accurate estimation of the probability that each drug class occurs.

Our basic approach to estimation is maximum likelihood estimation with a nested fixed point algorithm: for any parameter vector, we solve for agents' dynamically optimal decisions, and then define the likelihood function based on s , the predicted shares at the optimum. The model is an optimal stopping problem (where stopping indicates a drug purchase) with many options (where an option is a particular drug). In this way, the problem is similar to Rust (1987)'s classic paper on optimal stopping and also to more recent work that combines optimal stopping decisions with a multinomial choice (see, for instance, Melnikov, 2013; Hendel and Nevo, 2006; Gowrisankaran and Rysman, 2012).

Our framework differs from these models in that we do not observe all health shocks: we only observe health shocks when the individual chooses to purchase a drug rather than the outside option. Moreover, a large part of our identification will come from people choosing not to purchase drugs as they approach or are in the doughnut hole. Thus, we develop methods that allow us to integrate in closed form over the shocks at which the individual chooses a drug, which makes this estimator computationally tractable.²⁹ Appendix B provides details on the likelihood function.

Finally, note that we estimate over 200 parameters, mostly drug fixed effects ϕ . It can be difficult to estimate structural, dynamic models with this many parameters. Fortunately, with the exception of the discount / salience effects, our estimation is similar to a multinomial logit model, which has a well-behaved likelihood. We estimate the model by performing a grid search over β/σ and δ and then using a derivative-based search for all other parameters, given each value of β/σ and δ .³⁰ Not reported in the paper, we also performed Monte Carlo simulations to verify the accuracy of the code and power of the estimator.

²⁹We also cannot easily use the computationally advantageous conditional choice probability estimators initially proposed by Hotz and Miller (1993). These estimators rely on observing all serially correlated state variables which is not the case in our setting. Specifically, we do not observe the state variable n , which is the purchase occasion within the week, because we do not observe the outside option purchase. Moreover, a high n for one drug purchase is positively correlated with a high n for the next drug purchase.

³⁰We also sped up computation by using parallel computation methods and by using the structure of the problem, where the doughnut hole is an absorbing state without any dynamic behavior, to simplify the value function calculation.

5.2 Identification

The parameters that we seek to identify from our structural likelihood estimation are the fixed utility from treatment parameters ϕ , the price elasticity parameters of $\alpha(\cdot)$, δ , and β/σ . In dynamic discrete choice models, an exclusion restriction can be used to identify both δ and choice-specific value functions (Magnac and Thesmar, 2002; Fang and Wang, 2013). In our case, the variability of drug prices near the doughnut hole provides such an exclusion restriction. Intuitively, consider the geometric discounting case and suppose that one drug has a \$25 copay and a \$100 full price while a second drug has a \$10 copay and a \$40 full price. Then, from equation (1), at a state that is $m = \$20$ dollars from the doughnut hole, there is no insurance subsidy for drug 1 but there is \$10 in insurance subsidy for drug 2. Hence, the utility from purchasing drug 1 is the same as inside the doughnut hole, which provides an exclusion restriction and allows us to identify δ in the geometric model.

While Magnac and Thesmar (2002) focus on the identification of the expected discounted utility for each choice at each state, we are interested in decomposing these effects into the structural parameters noted above. In the geometric discounting model, once we have identified δ , we can identify the parameters of $\alpha(\cdot)$. This is because the doughnut hole provides variation in prices that is different across different drugs. Thus, in the above example, the expected discounted utility for drug 2 at $m = \$20$ can be obtained from the above exclusion restriction. The difference in expected discounted utility for drug 2 at $m = \$20$ relative to at $m = \$0$ then identifies the parameters of $\alpha(\cdot)$. Finally, the fact that the doughnut hole is modeled as an absorbing state and hence has no relevant dynamics allows us to identify ϕ from the market share of a product net of the price disutility.

We next discuss how to identify β/σ . A behavioral economics literature has shown, in a setting where per-period utility is known, that one can identify β as the ratio of time t tradeoffs between t and $t + 1$ purchases to time t tradeoff between purchases at $t + 1$ and $t + 2$ (Laibson, 1997). In our context, we can identify β/σ using states with two remaining purchase occasions with an insurance subsidy. The reason for this is that the first of these two purchase occasions has implications that are two purchase occasions in the future, which

implies that there is a relevant tradeoff between $t + 1$ and $t + 2$, while the second occasion only has implications one purchase in the future.

We offer a formal identification result, which uses the above intuition:

Proposition 3. *Let Assumption 1 hold. Assume that there is exactly one health shock per week, and that there is one drug class. Assume further that there is sufficient price variation across drugs such that for some drug k with the lowest out-of-pocket price, $p_1, \dots, p_J > oop_k$, and for some drug l , $oop_l > oop_k$. In addition, assume that the price disutility is linear so that $\alpha(p) \equiv \bar{\alpha}p$. Finally, assume that the set of drugs that can be purchased has enough price variation that all states m can be reached. Then, the geometric discounting model (with full price salience) is identified if $\delta > 0$. Furthermore, each of the three behavioral models—quasi-hyperbolic discounting naïfs and sophisticates, price salience—is identified if $\beta/\sigma, \delta > 0$ and full rank conditions hold.*

For tractability, Proposition 3 imposes a number of assumptions—such as the presence of only one drug class and only one shock per week—but more complex environments should yield more identifying variation. Note also that the proposition did not consider identification when $\beta/\sigma = 0$, because δ is not identified in this case (since any future state will not affect current decisions). However, Proposition 3 can be modified to show that, conditioning on δ , β/σ is identified even when equal to 0. Since δ does not affect behavior with $\beta/\sigma = 0$, this then shows that β/σ is identified even when equal to 0. Finally, note that we did not formally consider the identification of the different behavioral models. However, from our evidence in Section 2, intuitively, the steepness of the slope near the doughnut hole will identify the different models.

More generally, our identification leverages the heterogeneity of prices across drugs and drugs classes and responses to this heterogeneity. Our overall takeaway is that to identify discount factors from administrative data such as ours, it is necessary to have variation in prices across drugs. Moreover, to accurately identify the behavioral parameters, we need to concurrently identify price elasticity parameters, implying that an accurate specification of a choice model is important.

6 Structural Estimation Results and Counterfactuals

6.1 Estimation Results

Our structural estimation stratifies the sample of patients in Section 4 by ACG score and performs the estimation on the three separate samples. For each sample, we estimate the quasi-hyperbolic discounting with naïfs and sophisticates and the price salience model.

Table 5: Main results of structural estimation

Model:	Quasi-hyperbolic discounting: naïfs	Quasi-hyperbolic discounting: sophisticates	Price salience
Price spline $< \$20$	-0.116^{**} (0.006)	-0.116^{**} (0.006)	-0.148^{**} (0.007)
Price spline $\in [\$20, \$50)$	-0.012^{**} (0.002)	-0.012^{**} (0.002)	-0.014^{**} (0.002)
Price spline $\in [\$50, \$150)$	-0.013^{**} (0.008)	-0.013^{**} (0.008)	-0.018^{**} (0.0009)
Price spline $\geq \$150$,	-0.006^{**} (0.001)	-0.006^{**} (0.001)	-0.003^* (0.001)
Behavioral parameter: β/σ	0 (—)	0 (—)	0 (—)
Discount factor: δ	—	—	—
log L	−95,594.6	−95,594.6	−94,456.7
log L $\beta/\sigma = 0.1, \delta = 0.1$	−95,604.8	−95,604.5	−95,462.0
P-value for LM test	0.00	0.00	0.00
log L $\beta/\sigma = 0.1, \delta = 0.4$	−95,604.8	−95,604.5	−95,462.0
P-value for LM test	0.00	0.00	0.00
log L at $\beta/\sigma = 0.1, \delta = 0.995$	− 95,619.6	−95,615.7	−95,471.6
P-value for LM test	0.00	0.00	0.00
log L at $\beta/\sigma = 0.3, \delta = 0.995$	−95,672.5	−95,563.8	−95,532.9
P-value for LM test	0.00	0.00	0.00
Number of drug classes H		60	
Number of drug fixed effects ϕ		245	
N		18,897	

Note: standard errors reported in parentheses do not account for variance in β/σ or δ . ‘***’ denotes significance at the 1% level and ‘**’ at the 5% level. An observation is an enrollee/week for an enrollee in the base estimation sample and beginning-of-week spending $\geq \$2,000$ and $< \$3,000$, with a middle ACG score. Each column displays the results from the maximum likelihood estimation for one model. Reported price coefficients are $-\alpha(\cdot)$; all prices affect utility negatively. All specifications also include fixed effects ϕ for each drug. LM tests are for the restrictions on β/σ and δ .

Table 5 reports results for the middle ACG scores. We find complete myopia or lack of price salience, that $\beta = 0$ for the quasi-hyperbolic discounting models and $\sigma = 0$ for the price salience model. With $\beta = 0$, the implications of the naïfs and sophisticates model are

identical. Since δ is not identified when $\beta/\sigma = 0$, we do not report δ .

We cannot compute a standard error for β/σ given our estimated parameters, because they are not on the interior of the parameter space. Instead, we performed Lagrange multiplier tests on the restricted model with fixed δ and β/σ (Newey and McFadden, 1994), over a grid of these values. We reject all values of $\beta/\sigma > 0$ and $\delta > 0$ that we tested. Table 5 provides test statistics for selected values of these parameters.

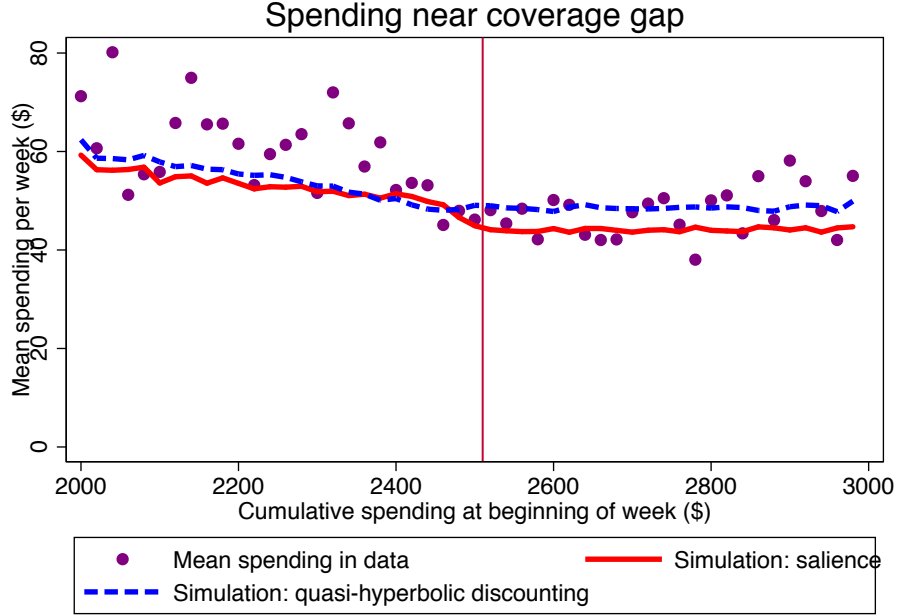
We next turn to model selection. Here we find that the price salience model fits the data better than the quasi-hyperbolic discounting model, with a log likelihood that is 137.9 points higher. Using our estimated parameters at $\beta/\sigma = 0$, we evaluated the likelihood of a mixture of the two models, to test between the two models. We find that a mixture of 9% quasi-hyperbolic discounting and 91% price salience fits the data the best. Again using a Lagrange multiplier test, we reject any mixture if and only if the model specifies 19% or more quasi-hyperbolic discounting, at the 5% significance level.

To provide graphical evidence on the fit of our models, Figure 5 reports mean spending in the data and from equilibrium simulations of both estimated models. The simulations use the same empirical distribution of health shocks as does our estimation. The figure shows that the salience model follows the pattern of relatively constant spending inside the initial coverage region and a steep drop in spending in weeks that start right before the doughnut hole. If some groups in our estimation sample were salient about the doughnut hole prices when in the initial coverage region, we would expect to see a more gradual drop in spending before the doughnut hole than predicted by the estimated model. In addition, the quasi-hyperbolic discounting model, even at its extreme of $\beta = 0$, predicts too early and a not steep enough spending decline. Thus, by all metrics that we examined, the price salience model fits the data better than the quasi-hyperbolic discounting model.

Table A4 in Appendix A reports results for other ACG scores, which are very similar to the base results.³¹ Overall, our estimates of $\sigma = 0$ are consistent with our reduced form

³¹Not reported in the paper, we also estimated the model with stratification both by ACG core and by whether the patient had purchased an insulin sensitizer or cholesterol-lowering drug. The results are similar across these groups, suggesting that our assumptions on the arrival of disease shocks are not overly influencing our results.

Figure 5: Fit of quasi-hyperbolic discounting and price salience models



findings that there is a large drop in drug spending after the start of the doughnut hole. They imply that enrollees in the initial coverage region were not taking into account the fact that they will face doughnut hole prices in the future when making their drug purchase decisions in the initial coverage region.

Finally, we turn to our estimates of the price spline coefficients. These coefficients are all negative and statistically significant implying that enrollees value price negatively at all ranges for all three models. However, it appears that enrollees care far less about price for higher-priced drugs than for lower-priced ones. In order to further understand our price coefficients, we used our estimated parameters to simulate the impact of a 1% increase in all drug prices and out-of-pocket prices on the expected number of drugs purchased by enrollees in our sample over the entire year 2008. We find that the 1% price increase would lead to a 0.54% decrease in the base-price-weighted number of drugs purchased in 2008 for enrollees in our estimation sample. Comparing our elasticity of -0.54 to analogous numbers from

the literature,³² Abaluck et al. (2015) estimate a Medicare Part D elasticity of -0.09 , Einav et al. (2015) estimate -0.50 , and Ketcham and Simon (2008) estimate -0.22 . Karaca-Mandic et al. (2013) estimate an elasticity of adherence for statin drugs of -0.95 . Thus, our elasticity numbers are in the middle of the range reported by the literature.

6.2 Counterfactuals

We now consider counterfactuals as to enrollee preferences and insurance environments. We use enrollees and estimates for the price salience model and the middle ACG scores (the last column of Table 5), but the results are very similar across ACG scores.

Our counterfactuals modify our structural estimation framework in two ways. First, since our estimation sample pertains to a selected set of enrollees who reached a high spending level early in the year, we create a nationally representative sample by taking a convex combination of enrollees in our estimation sample and enrollees in the same plans who are not in our estimation sample. The combination is chosen so that 33% of enrollees reach the coverage gap after 52 weeks, the same as the aggregate figure for 2008.³³ Second, we compute a 52-week model, where we model both the doughnut hole and the catastrophic coverage region, instead of an infinite horizon model with the doughnut hole as an absorbing state. The reason is that individuals in counterfactual environments may frequently not reach the doughnut hole, unlike in our base estimation.

We start by examining the relative importance of behavioral hazard to drug insurance. Here, behavioral hazard is the extent to which the lack of salience about future drug prices affects purchase decisions.³⁴ To quantify behavioral hazard, we compare the baseline Part D program to the neoclassical model, which we define here as geometric discounting with $\delta = 0.999$ at the weekly level (or 0.95 at an annual level). We examine the importance of drug insurance by comparing the baseline to the case without insurance. All cases report

³²Using the nationally representative sample described in Section 6.2 below, we find an elasticity of -0.38 .

³³Our counterfactual sample draws 31.8% from our estimation sample, with the remainder from other enrollees. For each ACG score, we estimate different distributions for the parameters on health shocks (Q_n) and drug class probabilities (P_h) for the estimation and non-estimation samples.

³⁴This is a relatively simple view of behavioral hazard, focusing only on overconsumption of drugs, and not on substitution to non-pharmaceutical spending.

enrollee welfare using $\sigma = 1$ and an annualized 95% discount factor.³⁵

Table 6: Relative impact of behavioral hazard and drug insurance

Statistic (per week)	Baseline: $\sigma = 0$ future not salient Case 1	Neoclassical model (no behavioral hazard) Case 2	No drug insurance Case 3
Number of Rxs	0.59	0.57	0.42
Number of branded Rxs	0.16	0.13	0.11
Number of generic Rxs	0.35	0.38	0.25
Expensive Rxs	0.08	0.06	0.05
Medium Rxs	0.17	0.15	0.13
Inexpensive Rxs	0.33	0.36	0.25
Enrollee spending (\$)	15.82	10.91	26.61
Insurer spending (\$)	25.39	24.17	0.00
Total spending (\$)	41.20	35.08	26.61
Enrollee welfare	1.21	1.27	0.63

Note: simulations use estimated parameters from Table 5 column 3. Inexpensive Rxs are less than \$50 and expensive ones are \$150 or more. Simulations are performed for 52 weeks starting enrollees at \$0 in expenditures and use a mix of the estimation sample and other enrollees in same plans so that 33% reach the doughnut hole in the base case. Geometric discounting case uses an annualized discount factor of 95%.

The results, in Table 6, show that the neoclassical model (Case 2) would cause a 31% drop in weekly enrollee prescription drug spending and a 15% drop in total drug expenditures relative to our estimated baseline with $\sigma = 0$ (Case 1). However, there is little difference in the number of prescriptions drugs between the two scenarios. Instead, there is a significant change in the composition of drugs consumed. There is a 25% drop in prescriptions for expensive drugs with substitution towards the most inexpensive. This substitution effect is also apparent in the increase in the number of generic drugs under the neoclassical model. Interestingly, there is a small decrease in insurer expenditures in moving to geometric discounting, as enrollees substitute to drugs which are cheaper for themselves and also for the insurers. Comparing the baseline to the case without insurance (Case 3), we find that eliminating drug insurance would cause drug expenditures to drop by 35%. Thus, both behavioral hazard and drug insurance are important in affecting drug spending.

³⁵Note that there is no one definition of welfare in models with behavioral hazard. Moreover, our welfare effects do not account for substitute therapies to drugs. For these reasons, we report, but do not focus on, welfare effects.

Table 7: Impact of filling the doughnut hole

Statistic (per week)	Baseline: $\sigma = 0$ future not salient Policy 1	No doughnut hole Policy 2	No doughnut hole with constant insurer spending Policy 3	No doughnut hole for generics only Policy 4
Number of Rxs	0.59	0.63	0.62	0.61
Number of branded Rxs	0.16	0.18	0.17	0.16
Number of generic Rxs	0.35	0.37	0.37	0.39
Expensive Rxs	0.08	0.10	0.08	0.08
Medium Rxs	0.17	0.18	0.16	0.17
Inexpensive Rxs	0.33	0.35	0.38	0.36
Enrollee spending (\$)	15.71	12.83	14.88	14.25
Insurer spending (\$)	25.55	32.49	25.55	27.23
Total spending (\$)	41.26	45.32	40.43	41.48
Enrollee welfare	1.21	1.35	1.31	1.29
Coinsurance rate	from data	from data	37%	from data

Note: simulations use estimated parameters from Table 5 column 3. Inexpensive Rxs are less than \$50 and expensive ones are \$150 or more. Simulations are performed for 52 weeks starting enrollees at \$0 in expenditures and use a mix of the estimation sample and other enrollees in same plans so that 33% reach the doughnut hole in the base case.

We now examine the implications of counterfactual policies regarding eliminating the doughnut hole. Table 7 presents the results of the baseline (Policy 1) and three counterfactual policies. Policy 2 extends the initial coverage region out-of-pocket prices to the doughnut hole. Policy 3 also eliminates the doughnut hole but leaves insurance spending constant by setting the coinsurance to a constant fraction of the total price of the drug. Finally, Policy 4 removes the doughnut hole for generics only.

We find that removing the doughnut hole (Policy 2) results in the total number of prescriptions increasing 7% and total drug spending increasing 10%. Insurer drug spending would increase 27%. Enrollees consume more drugs and more expensive drugs. Einav et al. (2015) also estimate that removing the doughnut hole will increase pharmaceutical spending 10%, while Abaluck et al. (2015) estimate that figure to be 6%.

It is important to evaluate what might be the overall health consequences of removing the doughnut hole. We can provide some back-of-the-envelope calculations using Chandra et al. (2010), who estimate substitution between drug utilization and inpatient hospitalization.

Applying these estimates to the increase in drug consumption under Policy 2 implies that inpatient hospital admissions would decrease by 1.8% by eliminating the doughnut hole.³⁶

Under a linear contract with the same insurer cost (Policy 3), enrollees would face a 37% coinsurance rate. This is significantly higher than the current average 25% coinsurance. Enrollees consume more drugs but fewer expensive ones than in the baseline as the lack of price salience is no longer relevant since the contract is not dynamic. Not shown in the table, this contract also lowers the expected standard deviation of enrollee spending, to \$548 relative to \$625 under Policy 1. Finally, Policy 4, removing the doughnut hole for generics, yields a 9% decrease in enrollee spending and a 7% increase in insurer spending relative to the baseline. The end effect is that total spending is almost the same as in the baseline, as enrollees substitute to generic drugs and away from branded drugs.

7 Conclusion

The Medicare Part D program established an important prescription drug benefit, but one that required enrollees interested in optimizing their drug purchases to calculate an inherently dynamic problem, due to the coverage gap. We develop a dynamic behavioral modeling framework for complex insurance contracts which allows for quasi-hyperbolic discounting and price salience. Using the framework we provide a discontinuity-based test of the neoclassical model. A central challenge of estimating the impact of dynamic incentives on consumer behavior is selection: individuals compared across different settings may be different in dimensions that are often unobservable. Our test is based on examining how individuals who arrive near the doughnut hole early in the year change their behavior upon reaching the doughnut hole. It avoids selection issues by considering how a given enrollee changes her behavior within a relatively small time period.

We find strong evidence against the neoclassical model. Enrollees lower their prescription

³⁶Chandra et al. (2010) find that a drug use drop of 18.2% leads to an increase in hospitalizations of 5.4%. We derive our result by applying the resulting elasticity of 0.27 to our 6.8% increase in drug utilization. Our calculation assumes that all the offset in Chandra et al. (2010) is attributable to the decline in drug consumption and not the decline in outpatient visits.

drug purchases upon reaching the doughnut hole, with a disproportionate drop for drugs that cost over \$150 and branded drugs. Moreover, the data can reject a geometric discounting model with a low but positive discount factor because spending is flat in a region before the doughnut hole. Having established evidence against the geometric discounting model, we turn to structurally estimating the parameters of our model. Our modeling framework builds on standard industrial organization choice models, with a multinomial choice problem where enrollees face random disease shocks that require treatment by a particular drug class and then choose to purchase one of a number of drugs in that class or the outside option. The price elasticity parameters are separately identified from the geometric discount factor and the behavioral parameter by the fact that different drugs have different prices. We find that enrollees have significant price elasticities and that future prices are not at all salient.

Our structural estimation approach has several limitations. We do not allow for any medical dynamics to treatment; we do not measure substitute therapies to drugs; we do not model imperfect physician agency; and our arrival process for diseases is relatively simple. Nonetheless, we believe that our structural results are reasonable, given the large drop in spending shown by the discontinuity evidence.

Last, we examine the impact of counterfactual preferences and policies. We find that closing the doughnut hole would raise total spending 10% or necessitate a 37% coinsurance for budget balancing. Doughnut hole coverage for generics only would be much less expensive.

References

- Abaluck, J. and Gruber, J. (2011). Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program. *American Economic Review*, 101(4):1180–1210.
- Abaluck, J. and Gruber, J. (2013). Evolving choice inconsistencies in choice of prescription drug insurance. NBER Working Paper 19163.

- Abaluck, J., Gruber, J., and Swanson, A. (2015). Prescription drug utilization under Medicare Part D: A dynamic perspective. NBER Working Paper 20976.
- Baicker, K., Mullainathan, S., and Schwartzstein, J. (2012). Behavioral hazard in health insurance. NBER Working Paper 18468.
- Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285.
- Brot-Goldberg, Z., Chandra, A., Handal, B., and Kolstad, J. (2015). What does a deductible do? the impact of cost-sharing on health care prices, quantities and spending dynamics. NBER Working Paper 21632.
- Cabral, M. (2013). Claim timing and ex post adverse selection. Mimeo, University of Texas.
- Chandra, A., Gruber, J., and McKnight, R. (2010). Patient cost-sharing and hospitalization offsets in the elderly. *American Economic Review*, 100:193–213.
- Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *The American Economic Review*, 99(4):1145–1177.
- Chung, D. J., Steenburgh, T., and Sudhir, K. (2013). Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans. *Marketing Science*, 33(2):165–187.
- Einav, L., Finkelstein, A., and Schrimpf, P. (2015). The response of drug expenditures to contract design in Medicare Part D. *Quarterly Journal of Economics*. Forthcoming.
- Fang, H. and Wang, Y. (2013). Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. Mimeo, University of Pennsylvania.
- Gowrisankaran, G. (1999). Efficient representation of state spaces for some dynamic models. *Journal of Economic Dynamics and Control*, 23:1077–1099.

- Gowrisankaran, G., Norberg, K., Kymes, S., Chernew, M. E., Stwalley, D., Kemper, L., and Peck, W. (2013). A hospital systems wellness program linked to health plan enrollment cut hospitalizations but not overall costs. *Health Affairs*, 32(3):477–485.
- Gowrisankaran, G. and Rysman, M. (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy*, 120:1173–1219.
- Grubb, M. D. and Osborne, M. (2014). Biased beliefs, learning, and bill shock. *The American Economic Review*, 105(1):234–271.
- Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–2682.
- Heiss, F., McFadden, D., and Winter, J. (2010). Mind the gap! Consumer perceptions and choices in Medicare Part D prescription drug plans. In Wise, D., editor, *Research Findings in the Economics of Aging*. University of Chicago Press.
- Hendel, I. and Nevo, A. (2006). Measuring the implications of sales and consumer stockpiling behavior. *Econometrica*, 74:1637–1673.
- Ho, K., Hogan, J., and Morton, F. S. (2014). The impact of consumer inattention on insurer pricing in the Medicare Part D program. Mimeo, Columbia University.
- Hotz, J. V. and Miller, R. J. (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60:497–529.
- Ito, K. (2014). Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *The American Economic Review*, 104(2):537–563.
- Karaca-Mandic, P., Swenson, T., Abraham, J. M., and Kane, R. L. (2013). Association of Medicare Part D medication out-of-pocket costs with utilization of statin medications. *Health Services Research*, 48(4):1311–1333.
- Ketcham, J. and Simon, K. (2008). Medicare Part D’s effects on elderly patients’ drug costs and utilization. *American Journal of Managed Care*, 14(11 Suppl):SP14–22.

- Ketcham, J. D., Lucarelli, C., Miravete, E. J., and Roebuck, M. C. (2012). Sinking, swimming, or learning to swim in Medicare Part D. *American Economic Review*, 102(6):2639–2673.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2):443–478.
- Lichtenberg, F. and Sun, S. (2007). The impact of Medicare Part D on prescription drug use by the elderly. *Health Affairs*, 16:1735–1744.
- Liu, F., Alexander, C., Crawford, S., Hedeker, D., Pickard, S., and Walton, S. (2011). The impact of Medicare Part D on out-of-pocket costs for prescription drugs, medication utilization, health resource utilization, and preference-based health utility. *Health Services Research*, 46(4):1104–1123.
- Magnac, T. and Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816.
- Medpac (2009). Report to the Congress: Medicare payment policy. Technical report, Medicare Payment Advisory Commission.
- Medpac (2014). Report to the Congress: Medicare payment policy. Technical report, Medicare Payment Advisory Commission.
- Melnikov, O. (2013). Demand for differentiated products: The case of the U.S. computer printer market. *Economic Inquiry*, 51:1277–1298.
- Nevo, A., Turner, J. L., and Williams, J. W. (2016). Usage-based pricing and demand for residential broadband. *Econometrica*, 84(2):411–443.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.
- Newhouse, J. P. (1993). *Free for all? Lessons from the RAND health insurance experiment*. Harvard University Press.

- Phelps, E. and Pollak, R. (1968). On second-best national saving and game-equilibrium growth. *Review of Economic Studies*, 35:201–208.
- Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55:999–1033.
- Schroeder, M. C., Brooks, J. M., and Cook, E. A. (2014). Heart attacks and subsequent medication choices: the effect of Medicare Part D benefit phase and time of year. Mimeo, University of Iowa.
- Starc, A. and Town, R. J. (2016). Externalities and benefit design in health insurance.
- Stokey, N. L., Lucas, R. E., and Prescott, E. C. (1989). *Recursive methods in economic dynamics*. Harvard University Press.
- Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23:165–80.
- Yin, W., Basu, A., Zhang, J. X., Rabbani, A., Meltzer, D. O., and Alexander, C. (2008). The effect of the Medicare Part D prescription benefit on drug utilization and expenditures. *Annals of Internal Medicine*, 148:169–177.
- Zhang, Y., Donohue, J. M., Lave, J. R., O'Donnell, G., and Newhouse, J. P. (2009). The effect of Medicare Part D on drug and medical spending. *New England Journal of Medicine*, 361:52–61.

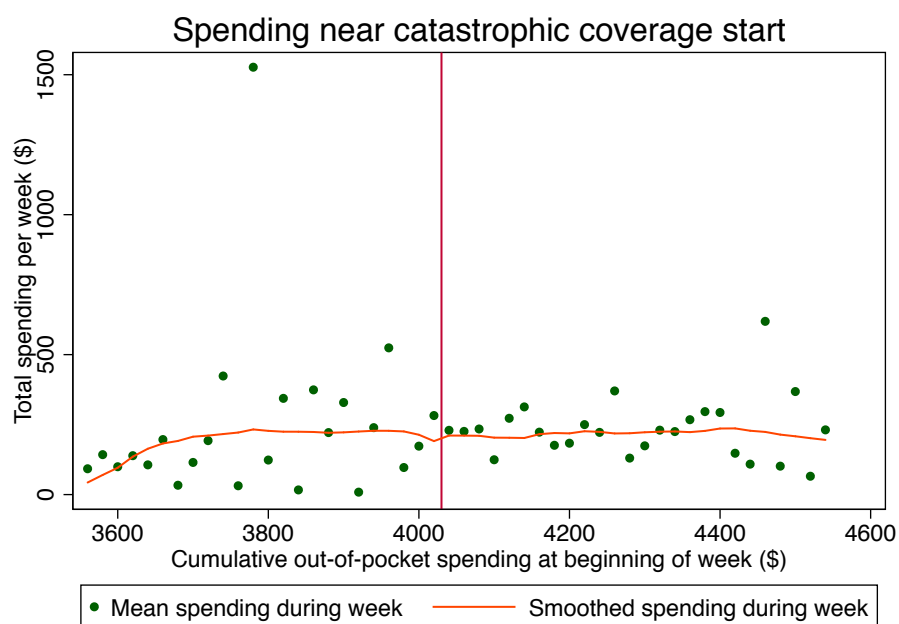
Appendix A: Extra Figures and Tables

Table A1: Most common drugs in base estimation sample

Drug name	Indication	Branded	Total price (\$)	Out of pocket price (\$)	Number of Rxs	% of obs.
Lisinopril	Renin-Angiotensin System Blocker	N	18.28	9.75	709	3.1
Metoprolol	Beta-Blocker	N	29.11	10.07	629	2.7
Simvastatin	Cholesterol Lowering	N	32.57	11.14	629	2.7
Hydrocodon	Opioid	N	21.44	7.88	609	2.7
Plavix	Antiplatelet	Y	169.55	40.47	594	2.6
Furosemide	Diuretic	N	8.16	6.79	575	2.5
Levothyroxine	Hypothyroidism	N	11.38	9.16	549	2.4
Metformin	Insulin Sensitizer	N	23.86	9.50	514	2.2
Amlodipine	Calcium Channel Blocker	N	52.16	10.93	496	2.2
Warfarin	Anticoagulant	N	16.21	8.50	339	1.5

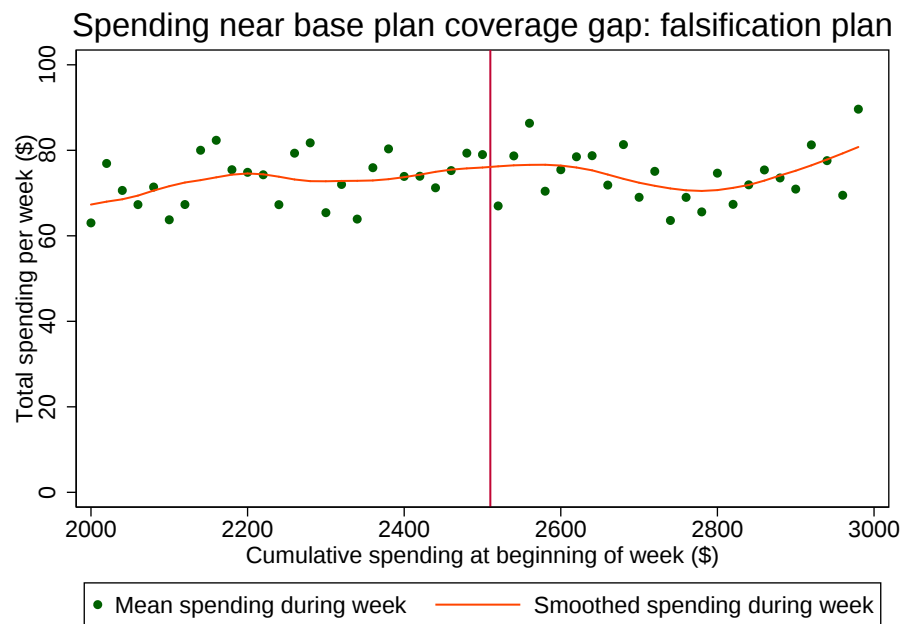
Note: reported total prices and out-of-pocket prices derived from authors' calculations.

Figure A1: Spending near catastrophic coverage start for base estimation sample



Note: figure is based on enrollees in plans A-E who start a week with \$3,550 to \$4,050 in out-of-pocket spending between Mar. 30 and Jul. 20, 2008.

Figure A2: Spending near \$2,510 for falsification plan F



Note: figure is based on analog to base estimation sample for plan F.

Table A2: Drug classes with largest and smallest spending changes near coverage gap

Dependent variable Number of Rxs for:	Mean value before \$2,400	Beginning of week spending in: \$2,510 - 2,999	<i>N</i>
Cholesterol Lowering	0.081	−0.0177** (0.0034)	28,543
Beta-Blocker	0.046	−0.0135** (0.0023)	28,543
Gastroesophageal Reflux & Pep- tic Ulcer	0.032	−0.0130** (0.0022)	28,543
Renin-Angiotensin System Blocker	0.065	−0.0120** (0.0029)	28,543
Antidepressant	0.045	−0.0102** (0.0024)	28,543
Anti-Glaucoma	0.010	0.0001 (0.0014)	28,543
Antidiarrheal	0.001	0.0002 (0.0004)	28,543
Diuretic & Renin-Angiotensin System Blocker	0.002	0.0003 (0.0005)	28,543
Folic Acid Antagonist Antibiotic	0.003	0.0005 (0.0008)	28,543
Antiarrhythmic	0.002	0.0007 (0.0005)	28,543

Note: standard errors in parentheses. ‘***’ denotes significance at the 1% level and ‘*’ at the 5% level. Each row represents one regression. All regressions also include enrollee fixed effects and an indicator for beginning-of-week spending between \$2,400 and \$2,509, and cluster standard errors at the enrollee level. An observation is an enrollee/week for an enrollee in the base estimation sample and beginning-of-week spending $\geq$ \$2,000 and $<$ \$3,000. Inexpensive Rxs are less than \$50 and expensive ones are \$150 or more.

Table A3: ACG scores by base estimation sample status

ACG score	Enrollees in base sample	Enrollees not in base sample	Maximum number of health shocks, <i>N</i>	Used in which estimator
0.000	3,287	296	8	Lowest ACG score
0.024	878	71	8	Middle ACG scores
0.260	2,265	203	8	Middle ACG scores
0.970	1,699	100	8	Middle ACG scores
1.043	3,192	207	8	Middle ACG scores
1.541	9,008	574	8	Middle ACG scores
1.753	2,659	167	8	Middle ACG scores
2.251	7,413	444	8	Highest ACG score

Note: our data contain only the 8 discrete ACG scores listed above.

Table A4: Robustness results of structural estimation: lowest and highest ACG scores

Estimation sample:	Lowest ACG	Lowest ACG	Highest ACG	Highest ACG
Model	Sophisticates	Price salience	Sophisticates	Price salience
Price < \$20	−0.101** (0.011)	−0.122** (0.012)	−0.141** (0.014)	−0.193** (0.018)
Price ∈ [\$20, \$50)	−0.015** (0.004)	−0.022** (0.004)	−0.025** (0.005)	−0.021** (0.005)
Price ∈ [\$50, \$150)	−0.014** (0.002)	−0.017** (0.002)	−0.017** (0.002)	−0.017** (0.002)
Price ≥ \$150	−0.003 (0.004)	−0.002 (0.004)	−0.002 (0.003)	−0.002 (0.002)
Behavioral β/σ	0 —	0 —	0 —	0 —
δ	—	—	—	—
log L	−27,244.4	−27,220.7	−19,558.9	−19,504.9
# drug classes	37	37	31	31
# drug FEs	120	120	94	94
N	4,692	4,692	4,958	4,958

Note: standard errors reported in parentheses do not account for variance in β/σ or δ . ‘***’ denotes significance at the 1% level and ‘**’ at the 5% level. An observation is an enrollee/week for an enrollee in the base estimation sample and beginning-of-week spending $\geq \$2,000$ and $< \$3,000$, with a middle ACG score. Each column displays the results from the maximum likelihood estimation for one model. Reported price coefficients are $-\alpha(\cdot)$; all prices affect utility negatively. All specifications also include fixed effects ϕ for each drug.

Figure A3: Histogram of total year drug spending for base estimation and full samples

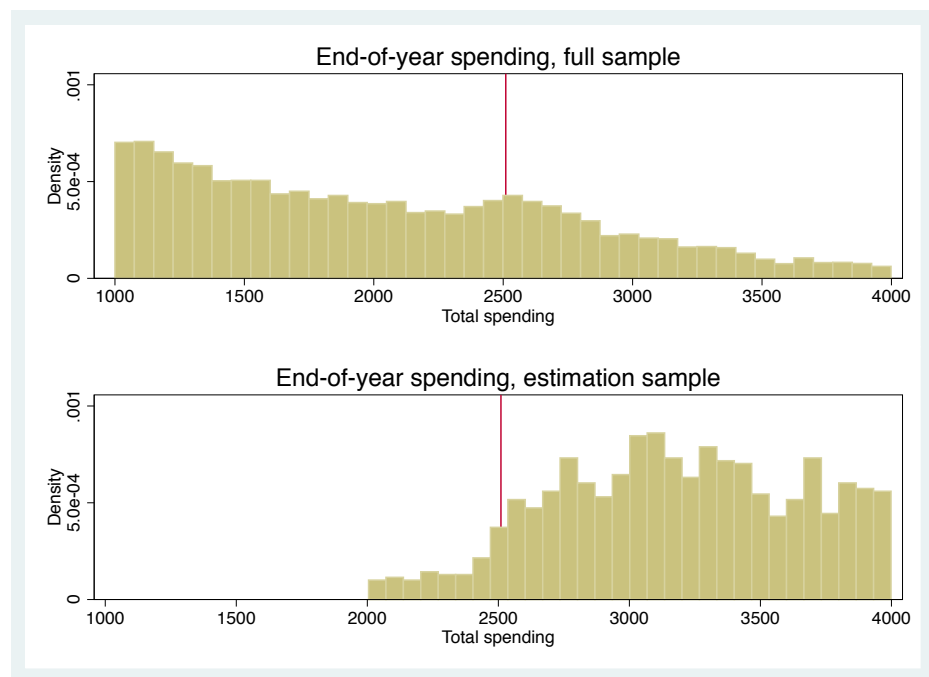


Figure A4: Information provided to Part D enrollees on distance to doughnut hole

STAGE 1 Yearly Deductible	(Because there is no deductible for this plan, this payment stage does not apply to you.)
You are in this stage:	
STAGE 2 Initial Coverage	- You begin in this payment stage when you fill your first prescription of the year. During this payment stage, the plan pays its share of the cost of your drugs and you (or others on your behalf) pay your share of the cost. - You generally stay in this stage until the amount of your year-to-date "total drug costs" reaches \$2,850.00. As of 08/30/2014, your year-to-date "total drug costs" was \$321.05. (See definitions in Section 3).
What happens next?	
Once you have an additional \$2,528.95 in "total drug costs," you move to the next payment stage (stage 3, Coverage Gap).	
STAGE 3 Coverage Gap	- During this payment stage, you (or others on your behalf) receive a discount on brand name drugs and you pay up to 72% of the costs of generic drugs. - You generally stay in this stage until the amount of your year-to-date "out-of-pocket costs" (see Section 3) reaches \$4,550.00. When this happens, you move to payment stage 4, Catastrophic Coverage.
STAGE 4 Catastrophic Coverage	- During this payment stage, the plan pays most of the cost for your covered drugs. - You generally stay in this stage for the rest of the plan year (through December 31, 2014).

Appendix B: Details of Dynamic Optimization Problems and Likelihood Function

We now exposit the enrollee's dynamic optimization problem, starting with the quasi-hyperbolic discounting model for sophisticates. Define the ex ante state (m, n) to be the state before the current health shock or $\bar{\varepsilon}$ are realized. We define the value function, $V(m, n)$, to be a function of the ex ante state. $V(m, n)$ is the value gross of β , with m dollars remaining until the doughnut hole and at the point in time where n health shocks have already occurred this week but before it is known whether the $n + 1$ th health shock will exist or what it will be. Using the value function, we can specify enrollee optimization. For any state $(m, n, h, \bar{\varepsilon})$, the enrollee's perceived utility from choice $j \in \{0, \dots, J_h\}$ can be written as:

$$\bar{u}_j(m, n, h) + \varepsilon_{hj} \equiv \phi_{hj} - \alpha(p^{eff}(m, p_{hj}, oop_{hj})) - c_h + \beta V(\max\{m - p_{hj}, 0\}, n + 1) + \varepsilon_{hj}. \quad (4)$$

Equation (4) states that the value of a choice is given by the current flow utility (the first, second, third, and fifth terms) plus the future value (the fourth term). The fourth term shows that the dynamic effect is that the purchase of drug j moves the individual closer to the doughnut hole by p_{hj} dollars. But, because the enrollee is a quasi-hyperbolic discounter, this term is discounted with factor β . We combine the first three terms of (4) into the mean utility, defined as $\bar{u}_j(m, n, h)$.

We now consider $s(m, n, h, j)$, the ex ante purchase probability at each state. Because the drug choice problem is equivalent to a standard logistic utility with mean utility $\bar{u}_j(m, n, h)$, $s(m, n, h, j)$ takes on a standard logit functional form:

$$s(m, n, h, j) = \frac{\exp(\bar{u}_j(m, n, h))}{\sum_{k=0}^{J_h} \exp(\bar{u}_k(m, n, h))}. \quad (5)$$

Finally, we exposit the value function:

$$V(m, n) = (1 - Q_n)\delta V(m, 0) + Q_n \sum_{h=1}^H P_h \sum_{j=0}^{J_h} s(m, n, h, j) \times \\ \left[\phi_{hj} - \alpha(p^{eff}(m, p_{hj}, oop_{hj}) - c_h) + V(\max\{m - p_{hj}, 0\}, n + 1) - \log s(m, n, h, j) + \gamma \right], \quad (6)$$

where γ is Euler's constant. Equation (6) evaluates, in turn, the two possibilities ex-ante to the health shock realization: first, that there are no more health shock in the week (which occurs with probability $1 - Q_n$), and second, that there are more health shocks (which occurs with probability Q_n). In the second case, the equation sums the utility over drug classes. Here, we cannot use the standard logit expression for utility because the individual is not necessarily making the optimizing choice given geometric discounting. The first three terms on the second line of (6) account for the expected future utility gross of ε_j . The final terms, $-\log s(m, n, h, j) + \gamma$, account for the expectation of ε_j conditional on choice j (Hotz and Miller, 1993).

The quasi-hyperbolic naïfs case is slightly different. Here, the enrollee perceives that she will act as a geometric discounter in the future. Hence, we can rewrite the Bellman equation, which is used to account for perceived future behavior, in its standard geometric (geo) form:

$$V^{geo}(m, n) = (1 - Q_n)\delta V(m, 0) + Q_n \sum_{h=1}^H P_h \times \\ \left(\gamma + \log \left[\sum_{j=0}^{J_h} \exp(\phi_{hj} + \alpha(p^{eff}(m, p_{hj}, oop_{hj}) - c_h) + V^{geo}(\max\{m - p_{hj}, 0\}, n + 1)) \right] \right). \quad (7)$$

The naïf will make choices with a utility function analogous to $\bar{u}_j(m, n, h)$ in (4), but using V^{geo} instead of V for future valuations.

The equations underlying behavior for the price salience model are analogous to those in the sophisticates quasi-hyperbolic discounting model, but with effective prices from (2) instead of (1). Using this p^{eff} and substituting σ for β , the same equations define $\bar{u}_j(m, n, h)$,

$s(m, n, h, j)$ and $V(m, n)$ for the salience model as for the sophisticates model. Note that σ takes the place of β because the enrollee assesses probability σ of there being a price change in the future, while with probability $1 - \sigma$, there are no perceived future price changes. This is quite similar to weighting the future with quasi-hyperbolic discount factor β .

We now define the likelihood. Let $g(i)$ denote the group of individual i and now index terms by group g , so that we have Q_{gn} , N_g , P_{gh} , and $s(g, m, n, h, j)$ respectively. For each person/week observation it , let N_{it} denote the number of health shocks. For $n = 1, \dots, N_{it}$, let m_{itn} denote the value of m , the dollars till the doughnut hole; $h_{itn} \in \{1, \dots, H\}$ denote the realization of the health shock; and $j_{itn} \in \{0, \dots, J_h\}$ denote the drug chosen.

We first explain what our likelihood would be if we observed outside option choices, and then explain how the likelihood is different based on not observing the outside option. If all health events were observable, then N_{it} , h_{itn} , j_{itn} , and m_{itn} would all be observable. We could then write the log likelihood for individual i at week t as:

$$\log L_{it} = \sum_{n=0}^{N_{it}} \log \left(1\{n = N_{it}\}(1 - Q_{g(i)n}) + 1\{n < N_{it}\}Q_{g(i)n}P_{g(i)h_{itn+1}}s(g(i), m_{itn+1}, n + 1, h_{itn+1}, j_{itn+1}) \right). \quad (8)$$

In words, the log likelihood for an observation can be broken down into a sum across health shocks n . For each n (starting at 0), there are two possibilities: an additional health shock occurrence or none. If there is an additional health shock what matters is the probability of seeing the additional shock multiplied by the conditional probability of the observed shock (given that one is observed) and the conditional probability of the drug chosen for that shock (given the observed shock). If there is no additional shock, then the likelihood is simply the probability of seeing no more shocks.

We now consider the likelihood accounting for the fact that we only observe health shocks when the individual purchases an inside good instead of the outside option. The likelihood is the sum of the likelihood conditional on a configuration of outside option purchases (which

is given by equation 8) times the probability of each outside option purchase configuration.

We illustrate with an example. Consider an enrollee/week observation with 2 purchased drugs, with A being purchased before B, where the enrollee has a maximum of 4 health shocks in a period. The drug purchases could occur at the following health shocks (with A being before B always): $((1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4))$. If the last drug purchase is at shock 4, the total number of health shocks must be 4, yielding three configurations. If the last drug purchase is at shock 3, the total number could be 3 or 4, yielding four configurations. Finally, if the last drug purchase is at shock 2, then the total number could be 2, 3, or 4, yielding three configurations. The likelihood sums the probability of the observed data conditional on each of these 10 outside good configurations times the probability of each outside good configuration.

Formally, let $\hat{N}_{it}$ denote the number of health shocks where the purchase included an inside good. Let $l_{itn}, n = 1, \dots, \hat{N}_{it}$ denote the places of each health shock, with $1 \leq l_{it1} < \dots < l_{it\hat{N}_{it}} \leq N_{g(i)}$. Let $\mathcal{L}(\hat{N}, N)$ denote the set of possible vectors of places when there are $\hat{N}$ health shocks with an inside good purchase and N possible purchase occasions; e.g. $\mathcal{L}(2, 4)$ has six elements as listed above. Then, the log likelihood is:

$$\log L_{it} = \log \left(\sum_{l_1, \dots, l_{\hat{N}_{it}} \in \mathcal{L}(\hat{N}_{it}, N)} \sum_{N_{it}=l_{\hat{N}_{it}}}^{N_{g(i)}} \left(\prod_{n=0}^{N_{it}-1} Q_{g(i)n} \right) (1 - Q_{g(i)N_{it}}) \right. \\ \left. \prod_{n=1}^{\hat{N}_{it}} P_{g(i)h_{itn}} s(g(i), m_{itn}, l_n, h_{itn}, j_{itn}) \right) \\ \prod_{n=1, n \neq l_1, \dots, n \neq l_{\hat{N}_{it}}}^{N_{it}} \left(\sum_{h=1}^H P_{g(i)h} s \left(g(i), \min_{\tilde{n} \text{ s.t. } l_{\tilde{n}} < n} m_{it\tilde{n}}, n, h, 0 \right) \right) \right). \quad (9)$$

In words, the first line of (9) represents the double sum over the possible places of each health shock (l) and the number of health shocks ($N_{g(i)}$), and, for each case, lists the probability of observing that many health shocks. The second line provides the probabilities of seeing the drugs chosen for the health shocks with observed drug choices, where the places of the drug shocks show up through l_n . The third line is the probability of seeing an outside option

chosen at each place without a drug purchase, where the dollar amount until the doughnut hole m is simply the dollar amount from the most recent drug purchase (which is also the minimum dollar amount across previous purchases). Note that equation (9) is similar to the earlier likelihood in equation (8) but with two main differences: first, it integrates over the places of each observed shock, the total number of health shocks, and the drug class for health shocks with the outside option chosen; and second, it combines all health shocks in a week because they are no longer separable given the unknown places and number of shocks.

The advantage of our formulation in (9) is that it derives the likelihood in closed form conditional on any set of health shock occurrences $\mathcal{L}(\hat{N}_{it}, N)$. By solving for the likelihood in closed form, we eliminate the need for simulation which improves the efficiency and computational time required to estimate our model.

The remaining challenge is in enumerating the elements of $\mathcal{L}(\hat{N}_{it}, N)$. We now describe our method in more detail, which follows Gowrisankaran (1999) closely. For brevity of notation, we now suppresses the dependence of variables on individual i , group g , or time t . Recall that each element in $\mathcal{L}(\hat{N}, N)$ corresponds to one vector of places for the health shocks with inside good purchases when there are $\hat{N}$ health shocks with inside good purchases and N is the maximum number of health shocks. For instance, if $N = 8$ and $\hat{N} = 3$, an element of $\mathcal{L}(\hat{N}, N)$ is $(1, 5, 8)$.

As in Gowrisankaran (1999), let $o(\cdot)$ denote the number of elements in a set. Using a similar proof structure to Gowrisankaran (1999) Theorem 1, we offer the following:

Proposition A1. *Using induction, the number of elements in $\mathcal{L}(\hat{N}, N)$ can be described as follows:*

Base case 1: $\hat{N} = 1$. $o(\mathcal{L}(1, N)) = N$.

Base case 2: $\hat{N} = N$. $o(\mathcal{L}(N, N)) = 1$.

Inductive case: $1 < \hat{N} < N$. $o(\mathcal{L}(\hat{N}, N)) = o(\mathcal{L}(\hat{N}, N - 1)) + o(\mathcal{L}(\hat{N} - 1, N - 1))$.

Proof We split the proof into assertions of the base cases and the inductive case.

Base case 1: $\mathcal{L}(1, N)$ enumerates all possible places for the single health shock with an inside good purchase. This single health shock can occur at any of the purchase occasions

between 1 and N . There are thus N possible places.

Base case 2: Here $\mathcal{L}(N, N)$ represents all possible place vectors for the inside good purchases when the number of inside good purchases is equal to the maximum number of purchase occasions. Here, each purchase occasion must be used for an inside good purchase. Thus, the unique place vector is $(1, \dots, N)$, which gives $o(\mathcal{L}(1, N)) = 1$.

Inductive case: Assume by induction that the theorem hold for all cases with maximum number of purchase occasions less than N and also for the (N, N) case. We now prove that it holds for the $(\hat{N}, N)$ case by induction, where $1 < \hat{N} < N$.

We divide the possible place vectors into two exhaustive and mutually exclusive cases. Either the N th health shock has no inside good purchase or it has one. Suppose first that it has none. Then, all the $\hat{N}$ inside good health shocks must occur at the first $\bar{N} - 1$ places. By the inductive assumption, there are $o(\mathcal{L}(\hat{N}, N - 1))$ possible place vectors that satisfy this criterion. Now suppose that the last place contains the last inside good purchase. Then the $\hat{N} - 1$ earlier inside good purchases must occur sometime during the first $N - 1$ places. Again by the inductive assumption, there are $o(\mathcal{L}(\hat{N} - 1, N - 1))$ possible place vectors that satisfy this vector. Adding up the number of elements in both cases, we have proven the inductive case. $\square$

Note that the inductive formula in Proposition A1 is the same as the inductive formula that defines binomial coefficients. Hence, we could also write $\mathcal{L}(\hat{N}, N) = \text{Binom}(N + 1, \hat{N}) \equiv \frac{(N+1)!}{(\hat{N})!(N+1-\hat{N})!}$. Finally, note that Gowrisankaran (1999) Theorem 2 provides a computationally efficient method for enumerating and accessing individual elements of $\mathcal{L}(\hat{N}, N)$. The analogous method works here and hence we use the method from that paper here also.

Appendix C: Proofs of Propositions

Proof of Proposition 1 Our proof imposes Assumption 1 and hence allows us to use the infinite horizon problem modeled in Section 2.3. For further tractability, we analyze a model with two additional assumptions (but show that our results are still valid without these additional assumptions). First, we assume an underlying cost of illness of $c_h = \gamma + \log \left(1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \bar{\alpha} \times \bar{p}) \right)$. With this assumption, the expected flow utility from optimizing behavior when inside the doughnut hole is exactly equal to the illness cost, which will render the value function 0 inside the doughnut hole. This then avoids the possibility of payoffs being infinite, which can occur since we consider $\delta = 1$. Since the value of c_h does not affect enrollee choices and Proposition 1 exclusively concerns enrollee choices, the proposition is valid for other choices of c_h . Second, we specify that there is exactly one health shock per week. A simple manipulation of (7) shows that, when $\delta = 1$, any V^{geo} that is a solution with this assumption is a solution without the assumption, thus showing that our results apply without this assumption. Intuitively, this result holds because with the infinite horizon and the absence of discounting, the number of health shocks in a week is irrelevant. Employing this second assumption, we ease notation by removing n from the state space.

We now claim that V^{geo} with these assumptions has the following functional form:

$$V^{geo}(m) = \bar{\alpha} \left(\left\lfloor \frac{m}{\bar{p}} \right\rfloor (\bar{p} - \overline{oop}) + \max \{0, m \% \bar{p} - \overline{oop}\} \right), \quad (10)$$

where “ $\lfloor \cdot \rfloor$ ” is the floor function and “ $\%$ ” is the remainder function. In (10), $V^{geo}(m)$ is equal to the marginal utility of money multiplied by the remaining maximum insurance amount, which essentially implies that the insurance coverage does not bias neoclassical enrollees away from their optimal decisions. On the first $\left\lfloor \frac{m}{\bar{p}} \right\rfloor$ drugs, the enrollee receives a subsidy of $(\bar{p} - \overline{oop})$ with a smaller, possibly zero subsidy on the next drug, and no subsidy thereafter. The complication in the expression, e.g., as reflected in the second term, is only due to the fact that the drug price does not necessarily divide by the initial coverage amount equally and, on the last insured drug, the enrollee pays her out-of-pocket cost before the insurance coverage starts. Finally, note that, as defined here, $V^{geo}(0) = 0$.

We now verify our claim that V^{geo} satisfies (10) by showing that the Bellman operator, $\mathcal{T}(V^{geo})$, defined here by:

$$\mathcal{T}(V^{geo})(m) = \sum_{h=1}^H P_h \left(\gamma + \log \left[\exp(V^{geo}(m)) + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \bar{\alpha} p^{eff}(m, \bar{p}, \overline{oop}) + V^{geo}(\max\{m - \bar{p}, 0\})) \right] - c_h \right), \quad (11)$$

will have as its value a function equal to (10) when its argument is the same.

We divide our analysis of (11) into three sets of ex ante states. First, we consider all states $V^{geo}(m)$, $0 \leq m \leq \overline{oop}$, i.e., all states with no future insurance value including the doughnut hole state. For these states, V^{geo} is zero for the state reached from m following any choice. Further, note that, in this case, $p^{eff}(m, p_{hj}, oop_{hj}) = \bar{p}$. Substituting these values into (11), we obtain:

$$\mathcal{T}(V^{geo})(m) = \sum_{h=1}^H P_h \left(\gamma + \log \left[1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \bar{\alpha} \times \bar{p}) \right] - c_h \right) = 0,$$

if $0 \leq m \leq \overline{oop}$.

Second, we consider all states $V^{geo}(m)$ with $\overline{oop} < m \leq \bar{p}$. Here, the remaining insurance amount is $m - \overline{oop}$, $V^{geo}(m) = \bar{\alpha}(m - \overline{oop})$, and the current out-of-pocket payment is $\bar{p} - m + \overline{oop}$. Thus, the future value is zero upon choosing an inside option; it remains $m - \overline{oop}$ with the outside option choice. Substituting these values into (11), we obtain:

$$\begin{aligned} \mathcal{T}(V^{geo})(m) &= \sum_{h=1}^H P_h \left(\gamma + \log \left[\exp(\bar{\alpha}(m - \overline{oop})) + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \bar{\alpha}(\bar{p} - m + \overline{oop})) \right] - c_h \right) \\ &= \bar{\alpha}(m - \overline{oop}) + \sum_{h=1}^H P_h \left(\gamma + \log \left[1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \bar{\alpha} \times \bar{p}) \right] - c_h \right) = \bar{\alpha}(m - \overline{oop}), \end{aligned}$$

if $\overline{oop} < m \leq \bar{p}$.

Finally, we consider all states $V^{geo}(m)$ with $m > \bar{p}$. The remaining insurance amount is $\iota \equiv \bar{\alpha} \left\lfloor \frac{m}{\bar{p}} \right\rfloor (\bar{p} - \overline{oop}) + \max \{0, m \% \bar{p} - \overline{oop}\}$, $V^{geo}(m) = \bar{\alpha}\iota$, and the current out-of-pocket payment is $\overline{oop}$. Thus, the future value is $\bar{\alpha}(\iota - \bar{p} + \overline{oop})$ upon choosing an inside option; it remains $\bar{\alpha}\iota$ with the outside option choice. Substituting these values into (11), we obtain:

$$\begin{aligned} \mathcal{T}(V^{geo})(m) &= \\ & \sum_{h=1}^H P_h \left(\gamma + \log \left[\exp(\bar{\alpha}\iota) + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \bar{\alpha}(\overline{oop} + \iota + \bar{p} - \overline{oop})) \right] - c_h \right) \\ &= \bar{\alpha}\iota + \sum_{h=1}^H P_h \left(\gamma + \log \left[1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \bar{\alpha} \times \bar{p}) \right] - c_h \right) = \bar{\alpha}\iota, \end{aligned}$$

if $m > \bar{p}$.

Thus, for all cases, $\mathcal{T}(V^{geo}) = V^{geo}$, where V^{geo} is defined using (10). Applying the standard contraction mapping approach to dynamic programming theory (Stokey et al., 1989), (10) is accurate. Note that the mean utility function from (13) specializes to:

$$\bar{u}_j(m, h) = \phi_{hj} - \bar{\alpha}p^{eff}(m, \bar{p}, \overline{oop}) - c_h + V^{geo}(\max\{m - \bar{p}, 0\}), \quad (12)$$

for all $j = 1, \dots, J_h$ and all h . Similarly, $\bar{u}_0(m, h) = -c_h + V^{geo}(m)$ for all h . Substituting from (10), $\bar{u}_j(m, h) - \bar{u}_0(m, h) = \phi_{hj} - \bar{\alpha} \times \bar{p}$ for all $j = 1, \dots, J_h$ and each of the three cases considered above. Thus, $s(m, h, j) = s(m', h, j)$ for all m, m', h, j . $\square$

Proof of Proposition 2 Our proof again imposes Assumption 1. We also again use $c_h = \gamma + \log \left(1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \bar{\alpha} \times \bar{p}) \right)$ and specify that there is exactly one health shock per week. Note that the sophisticates and price salience models use V and not V^{geo} . Similarly to Proposition 1, if $\delta = 1$, a simple manipulation of (6) shows that any solution to V with the one health shock assumption is a solution without the assumption. While uniqueness does not follow from standard dynamic programming theory, our assumption that there is a unique solution to V further ensures that this is the only solution to the model without the

assumption. Thus, our results are again valid without these additional assumptions.

We first prove part (a) of the proposition. Across the three models, $\bar{u}_j$ deviates from the neoclassical model only in inclusion of the V or V^{geo} term. But, the future state is always the same, $m = 0$, and hence current choices are unaffected by this term. Hence, the enrollee solves the statically optimal policy, exactly as in Proposition 1.

We now prove part (b) for the quasi-hyperbolic discounting naïfs model case. As in (7), naïfs believe that they will act as neoclassical optimizers from next period on. Thus, specializing to our case, the naïf enrollee will have:

$$\bar{u}_j(m, h) = \phi_{hj} - \bar{\alpha}p^{eff}(m, \bar{p}, \overline{oop}) + \beta V^{geo}(\max\{m - \bar{p}, 0\}) - c_h, \quad (13)$$

for all $j = 0, \dots, J_h$ and all h and where V^{geo} is defined in (10). Applying (10), $\bar{u}_j(m, h) - \bar{u}_0(m, h) = \phi_{hj} - \bar{\alpha}(\overline{oop} + \beta(\bar{p} - \overline{oop}))$ for all $m, m' \geq \bar{p}$, h , and $j = 1, \dots, J_h$. This implies that $s(m, h, j) = s(m', h, j)$ for all h, j and $m, m' \geq \bar{p}$. Since $\overline{oop} + \beta(\bar{p} - \overline{oop}) < \bar{p}$ for $\beta < 1$, $s(m, h, j) > s(0, h, j)$ for $m \geq \bar{p}$, h , and $j = 1, \dots, J_h$.

We now prove part (b) for the quasi-hyperbolic discounting sophisticates and price salience model. For ease of notation, let $\tilde{p}(m) = p^{eff}(m, \bar{p}, \overline{oop})$, $\tilde{m}(m) = \max\{m - \bar{p}, 0\}$, $\mathcal{V}(m) = V(m) - V(\tilde{m}(m))$, and $x_{hj}(m) = \exp(\phi_{hj} - \bar{\alpha}\tilde{p}(m))$, for all h and $j = 1, \dots, J_h$. Note that the price salience and sophisticates models are mathematically identical except for a different $\tilde{p}(m)$ (when $0 < m < \bar{p}$) and the use of β instead of σ . In our exposition below, we will use β but the proof would apply equally well to the price salience model with the substitution of σ .

First, note that $\bar{u}_j(m, h) = \phi_{hj} - \bar{\alpha}\tilde{p}(m) - c_h + \beta V(\tilde{m}(m))$, for all h and $j = 1, \dots, J_h$, and $\bar{u}_0(m, h) = -c_h + \beta V(m)$, for all h . Thus, for all h and $j = 1, \dots, J_h$,

$$\begin{aligned} s(m, h, j) &= \frac{\exp(\phi_{hj} - \bar{\alpha}\tilde{p}(m))}{\exp(\beta(V(m) - V(\tilde{m}(m)))) + \sum_{k=1}^{J_h} \exp(\phi_{hk} - \bar{\alpha}\tilde{p}(m))} \\ &= \frac{x_{hj}(m)}{\exp(\beta\mathcal{V}(m)) + \sum_{k=1}^{J_h} x_{hk}(m)}. \end{aligned} \quad (14)$$

Similarly, for all h ,

$$s(m, h, 0) = \frac{\exp(\beta \mathcal{V}(m))}{\exp(\beta \mathcal{V}(m)) + \sum_{k=1}^{J_h} x_{hk}(m)}. \quad (15)$$

Now, specializing the value function (6) to the case of the proposition and separating out the outside option, we can write:

$$\begin{aligned} V(m) &= \sum_{h=1}^H P_h \left(\gamma - c_h + s(m, h, 0) [V(m) - \log s(m, h, 0)] \right. \\ &\quad \left. + \sum_{j=1}^{J_h} s(m, h, j) [\phi_{hj} - \bar{\alpha} \tilde{p}(m) + V(\tilde{m}(m)) - \log s(m, h, j)] \right) \\ \Rightarrow V(m) - V(\tilde{m}(m)) &= \sum_{h=1}^H P_h \left(\gamma - c_h + s(m, h, 0) [V(m) - V(\tilde{m}(m)) - \log s(m, h, 0)] \right. \\ &\quad \left. + \sum_{j=1}^{J_h} s(m, h, j) [\phi_{hj} - \bar{\alpha} \tilde{p}(m) - \log s(m, h, j)] \right), \\ &\Rightarrow \mathcal{V}(m) = \\ &\sum_{h=1}^H P_h \left(\gamma - c_h + \frac{\exp(\beta \mathcal{V}(m))}{\exp(\beta \mathcal{V}(m)) + \sum_{k=1}^{J_h} x_{hk}(m)} \left[\mathcal{V}(m) - \log \frac{\exp(\beta \mathcal{V}(m))}{\exp(\beta \mathcal{V}(m)) + \sum_{k=1}^{J_h} x_{hk}(m)} \right] \right. \\ &\quad \left. + \sum_{j=1}^{J_h} \frac{x_{hj}(m)}{\exp(\beta \mathcal{V}(m)) + \sum_{k=1}^{J_h} x_{hk}(m)} \left[\phi_{hj} - \bar{\alpha} \tilde{p}(m) - \log \frac{x_{hj}(m)}{\exp(\beta \mathcal{V}(m)) + \sum_{k=1}^{J_h} x_{hk}(m)} \right] \right), \end{aligned} \quad (16)$$

where the second expression subtracts $V(\tilde{m}(m))$ from both sides, and the third expression substitutes from (14) and (15).

Importantly, the last expression in (15) implicitly defines the function $\mathcal{V}(m)$ for $m > 0$.³⁷ Note that, for $m, m' \geq \bar{p}$ and all h, j , $\tilde{p}(m) = \tilde{p}(m')$, $\tilde{m}(m) = \tilde{m}(m')$, and $x_{hj}(m) = x_{hj}(m')$ so $\mathcal{V}(m) = \mathcal{V}(m')$. Applying (14) and (15), $s(m, h, j) = s(m', h, j)$ for these cases.

To sign the change in purchase probabilities between the two regions, let $\mathcal{V}^{geo}(m) = V^{geo}(m) - V^{geo}(\tilde{m}(m))$, analogously to $\mathcal{V}(m)$. Then, $\mathcal{V}(m) < \mathcal{V}^{geo}(m)$ for $m \geq \bar{p}$ since V^{geo}

³⁷For $m = 0$, $\mathcal{V}(m) = V(0) - V(0)$, so there is nothing to define.

represents the value with optimal behavior while V represents the value with suboptimal behavior, both from the point of view of the same (neoclassical) agent. From (14), shares under the neoclassical model are equivalent to shares under the two behavioral models with the substitution of $\mathcal{V}^{geo}(m)$ for $\beta\mathcal{V}(m)$. Combining this with (i) the fact that a decrease in $\beta\mathcal{V}$ increases all inside good shares, (ii) the result from Proposition 1 that the neoclassical model has the same behavior inside and outside the doughnut hole, and (iii) the result from part (a) of the proposition that the behavioral models have the same behavior as the neoclassical model inside the doughnut hole, we find that $s(m, h, j) > s(0, h, j)$ for $m \geq \bar{p}$ and for all h, j .

We now prove part (c). Because the last expression in (15), which implicitly defines $\mathcal{V}$, is identical for the sophisticates quasi-hyperbolic discounting model and for the price salience model if $m \geq \bar{p}$, $s(m, h, j)$ will be identical for these two models if $m \geq \bar{p}$ and for all h and j . To sign the difference in purchase probabilities between the naïfs and other models, because (i) market shares for the naïf model use an expression identical to (14) except for the substitution of $\mathcal{V}^{geo}$ for $\mathcal{V}$, (ii) $\mathcal{V}(m) < \mathcal{V}^{geo}(m)$ for $m \geq \bar{p}$, and (iii) a decrease in $\beta\mathcal{V}$ increases all inside good shares, $s(m, h, j)$ will be lower for the naïfs model than for the other models if $m \geq \bar{p}$ and for all h and $j = 1, \dots, J_h$.

Proof of Proposition 3 Since the proposition concerns markets with only one drug class and one shock per week, to ease notation, we drop h —and when present n —from the terms ϕ_{hj} , J_h , c_h , $s(m, n, h, j)$, and $\bar{u}_j(m, n, h)$. Without loss of generality, assume that the states with data are $m \in [0, 1, \dots, 2510]$. Because we show identification, assume that $s(m, j)$ is observable for these values. As in Propositions 1 and 2, we normalize $c = \gamma + \log\left(\frac{1}{s(0,0)}\right)$. Since this value of c is the expected value of optimizing behavior for one purchase occasion at $m = 0$, this then results in $V(0) = \bar{u}_0(0) = 0$. This in turn implies that

$$\log \frac{s(j, 0)}{s(0, 0)} = \bar{u}_j(0) = \phi_j - \bar{\alpha}p_j, \forall j = 0, \dots, J. \quad (17)$$

Finally, note that for the sophisticate quasi-hyperbolic discounting models, we can write the

mean utility for the outside option as:

$$\bar{u}_0(m) = \beta\delta V(m) \Rightarrow V(m) = \frac{\bar{u}_0(m)}{\beta\delta}. \quad (18)$$

The other models have analogous expressions to (18): for the price salience model, σ substitutes for β ; for the quasi-hyperbolic discounting naïfs model, V^{geo} substitutes for V ; and for the the geometric model, $\beta = 1$.

We now prove identification for the sophisticate model with a fixed $\beta > 0$, which includes the geometric discounting model. Fix a cheap drug ‘ k ’ and an expensive drug ‘ l ’ as given in the statement of the proposition. For this model, it can be shown that $V(m) = 0$, if $m \leq oop_k$ —and not just for $m = 0$ —because the choices and insurance here are identical to inside the doughnut hole. Now consider any state m' for which $oop_k < m' \leq \min\{p_1, \dots, p_J, oop_l\}$. Such a state exists by the assumptions of the proposition. At m' , there is no insurance value from buying drug l and hence $\bar{u}_l(m') = \bar{u}_l(0)$ (as in equation 17). Using this exclusion restriction, for all $j = 0, \dots, J$,

$$\log \frac{s(m', j)}{s(m', l)} = \bar{u}_j(m') - \bar{u}_l(m') \Rightarrow \bar{u}_j(m') = \log \frac{s(m', j)}{s(m', l)} + \bar{u}_l(0),$$

which implies that $\bar{u}_j(m')$ is known for each $j = 0, \dots, J$. This then allows us to identify δ given that β is fixed. Specifically:

$$\begin{aligned} V(m') &= s(m', 0)[\delta V(m') - \log(s(m', 0))] + \sum_{j=1}^J s(m', j)[\bar{u}_j(m') - \log(s(m', j))] + \gamma - c \\ \Rightarrow V(m') &= \frac{\sum_{j=1}^J s(m', j)[\bar{u}_j(m') - \log(s(m', j))] - s(m', 0) \log(s(m', 0)) + \gamma}{1 - \delta s(m', 0)} - c \\ \Rightarrow \frac{\bar{u}_0(m')}{\beta\delta} &= \frac{\sum_{j=1}^J s(m', j)[\bar{u}_j(m') - \log(s(m', j))] - s(m', 0) \log(s(m', 0)) + \gamma}{1 - \delta s(m', 0)} - c, \quad (19) \end{aligned}$$

by substituting for $V(m')$ from (18). Using the fact that $V(m') > 0$ and $\beta > 0$, (19) defines δ as a linear equation which implies that δ and $V(m')$ are identified conditional on a fixed β .

We can then identify $\bar{\alpha}$. Specifically,

$$\bar{u}_k(m') = \bar{u}_k(0) + \bar{\alpha}(m' - oop_k) \Rightarrow \bar{\alpha} = \frac{\bar{u}_k(m') - \bar{u}_k(0)}{m' - oop_k}. \quad (20)$$

Since every term on the right side of (20) is known, $\bar{\alpha}$ is identified. This then allows us to identify ϕ_j using (17) for $j = 1, \dots, J$, implying that any sophisticate model with a fixed $\beta > 0$, including the geometric discounting model, is identified.

We now prove identification for the quasi-hyperbolic discounting sophisticates model. Consider m'' such that $oop_k + p_k < m'' \leq 2p_k$. This state has two purchase occasions with positive insurance value for drug k implying that a positive continuation value with choice k , i.e. $V(m'' - p_k) > 0$. We can write:

$$V(m'') = \delta V(m'')s(m'', 0) + \sum_{j=1}^J [\phi_j - p^{eff}(m'', p_j, oop_j) + \beta \delta V(\max\{m'' - p_j, 0\})] s(m'', j) + \gamma - \log(s(m'', 0)). \quad (21)$$

For any β , we have already shown that δ and $V(m'' - p_j), \forall j = 1, \dots, J$ are identified. Thus, (21) implicitly defines β . Provided a full rank condition holds so that (21) has a unique solution, β is then identified. While we do not verify that this full rank condition is satisfied, we expect that it will be satisfied because, given that there is a unique δ that fit the data at m' for any β , inside good market shares will be increasing in β for $m > m'$, as in Figure 2 and 3.

Thus, we have shown identification for the quasi-hyperbolic discounting sophisticate case. We omit the proof for the quasi-hyperbolic discounting naïf case, which is similar, though it uses the value function V^{geo} .

We now prove identification for the price salience case. Analogously to the quasi-hyperbolic discounting case, we condition on $\sigma > 0$, show identification at m' conditional on σ and then identify σ from data at m'' . However, this model is slightly different than the quasi-hyperbolic discounting model because σ enters into the purchase decision at state m' . Because of this, we employ a different proof order: conditional on σ , we first show that $\bar{\alpha}$ is identified and

then that δ is identified. Considering again state m' , note that:

$$\log \left(\frac{s(m', l)}{s(m', k)} \frac{s(0, k)}{s(0, l)} \right) = \bar{\alpha}(1 - \sigma)(p_l - oop_l) - \bar{\alpha}(1 - \sigma)(p_k - oop_k) - \bar{\alpha}\sigma(m' - oop_k), \quad (22)$$

where the first term in (22) is the extra utility from the lower perceived price for drug l , the second term is the negative of the utility from the lower perceived price for drug k , and the final term is the negative of the utility from the lower actual price for drug k . It is straightforward to solve for $\bar{\alpha}$ here implying that $\bar{\alpha}$ is identified conditional on σ . As in the quasi-hyperbolic discounting case, we can then identify all ϕ_j using doughnut hole data. We can then evaluate $\bar{u}_j(m')$ for all $j = 0, \dots, J$, since δ does not enter into the expression for $j = 1$ (or any inside good). In addition, an equation analogous to (19) holds with these $\bar{u}_j(m')$ values and the substitution of σ for β . This then allows us to identify δ and to recover $V(\cdot)$ conditional on σ . Finally, (21) holds with the substitution of σ for β , which allows us to recover σ provided that the analogous rank condition holds.

□

Appendix D: Analysis of Threats to Identification of Results in Section 4

We now consider, and eliminate, threats to the identification of our tests in Section 4. First, one might believe that a drop in spending at the doughnut hole reflects a simple alternate scenario, where the treatment value of a drug always lies somewhere between its out-of-pocket price and its full price, so that individuals would find it optimal to stop purchasing drugs in the doughnut hole. If this alternate model were to hold, Assumption 1 would be violated and many individuals would end the year right at the doughnut hole. This is very much unlike what we find, in Table 1. In addition, this would lead to total year drug spending “bunching” right after the doughnut hole start. Bunching has been observed in the broad sample of all Part D enrollees (Einav et al. (2015), Starc and Town (2016)). Figure A3 in Appendix A considers bunching for the full sample of enrollees in the plans we consider and

in our selected sample of enrollees who reach near the doughnut hole early in the year. While we observe bunching in the full sample, we do not observe bunching for our sample. This implies that enrollees continue spending well past the doughnut hole and that there is enough heterogeneity in drug values that some drugs are worth more than their full price.

Second, one might believe that our results are due to enrollees simply being misinformed regarding the benefit structure of Part D. Yet, because our data are from the third year of the program, it is unlikely that our results on myopia are driven by a lack of understanding about the presence of the doughnut hole and its implications. Enrollees are mailed detailed monthly information that lists their out-of-pocket and total costs for the month, the cost of their drugs to the plan as well as the out-of-pocket costs and explains how far they are from the doughnut hole. Figure A4 in Appendix A shows an example of the part of the mailing that pertains to the distance to the doughnut hole. In our view, the frequency and detail of the information provided suggest that rational enrollees have the opportunity to be informed about the coverage gap. Moreover, because Medicare enrollee drug consumption is principally tied to the treatment of chronic conditions, those who reach the doughnut hole in one year are likely to reach or approach the doughnut hole in the next year. While we lack data from 2007 for enrollees in our sample, we verified this proposition with national panel data, from the Medicare Part D Prescription Drug Event data.³⁸ Using this dataset, we examined the probability of reaching the doughnut hole in 2007 for enrollees who would have been in our sample (because they started a week early in 2008 near the doughnut hole). We find that 83.4% of these enrollees reached the doughnut hole in 2007, while 16.6% did not (and we were unable to match the remaining 0.8%). This suggests that, even at the start of 2008, most enrollees in our sample should have been directly informed about the presence and attributes of the doughnut hole from their previous experience.

Third, although our testable implications may be somewhat biased because they omit the presence of the catastrophic coverage region, we can sign the direction of this bias. Specifically, we will reject the neoclassical model if enrollees curtail purchases upon reaching the doughnut hole. A rational reaction to the upcoming presence of the catastrophic region

³⁸This dataset provides a 10% sample of all Medicare eligibles.

would be to increase weekly purchases as the cumulative out-of-pocket spending increases towards the doughnut hole start. Although we would expect this reaction to be small at the start of the doughnut hole (since there is some distance to the catastrophic coverage start), this effect will cause enrollees to increase weekly spending as their cumulative spending increases. Thus, this effect would work in the opposite direction from our test and thereby would bias us towards a null finding that the neoclassical model is accurate. Note that we find no evidence of an increase in spending for the range we consider (Figure 4). Moreover, to the extent that we reject the neoclassical model, our finding is conservative.

Fourth, our results are unlikely to be due to cross-year substitution. Cabral (2013) has found evidence that people move dental services to an earlier year when they have spare insurance benefits in the current year. But, individuals in our sample have little incentive to stockpile since they mostly end the year in the doughnut hole. Another possibility is that they strategically curtail spending during our sample in order to make up that spending in the following year during the initial coverage region (as shown by Einav et al., 2015)). But, enrollees in our sample are very likely to hit the doughnut hole in the year after our sample, which is 2009, implying that this strategy would not add substantial value for enrollees with δ close to 1.³⁹ Also, most of our enrollees are in plans with deductibles, and it would be medically costly for these enrollees to wait until they are past the deductible for treatment. Finally, we can understand the extent of cross-year substitution by comparing doughnut hole spending at the end of 2008 to earlier in 2008, using the fact that it would also be medically more costly for enrollees to defer expenditures to 2009 from relatively early in 2008 than from the end of 2008. We regress mean weekly spending on enrollee fixed effects and indicators for four week intervals, for the part of our base estimation sample that is inside the doughnut hole. We find that the four-week-indicators for September, October, and November are not significantly different from the December indicator. We also find qualitatively similar effects for our structural estimation when we limit our sample to end by November 1, 2008. Together,

³⁹To verify this, we again use the Medicare Part D Prescription Drug Event national panel data to examine the probability of reaching the doughnut hole in 2009 for enrollees who would have been in our sample (because they started a week early in 2008 near the doughnut hole). We find that 76.2% of these enrollees hit the doughnut hole in 2009, while 19.1% did not (and we were unable to match the remaining 4.7%).

these factors suggest that cross-year substitution is limited in our sample.

Finally, our results are unlikely to be due to a variety of other factors. Liquidity constraints cannot explain why people would spend more earlier on but less later. Indivisibility of drugs is not likely to explain the magnitude of our findings either, since for the classes with the largest drops, such as cholesterol lowering drugs, patients are better off taking a partial amount of the drug to none. Precautionary savings due to uncertainty about future medical shocks also is not likely to explain this pattern, since greater price uncertainty would exist prior to the doughnut hole to inside the doughnut hole.