SALIENCE, MYOPIA, AND COMPLEX DYNAMIC INCENTIVES:
EVIDENCE FROM MEDICARE PART D

Christina M. Dalton
Gautam Gowrisankaran
Robert Town

Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D
Christina M. Dalton, Gautam Gowrisankaran, and Robert Town
NBER Working Paper No. 21104
April 2015, Revised March 2019
JEL No. D03,I13,I18,L88

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 a benchmark model with neoclassical preferences, 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 benchmark 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.

Christina M. Dalton
Department of Economics
Wake Forest University
tina.marsh.dalton@gmail.com

Gautam Gowrisankaran
Department of Economics
University of Arizona
P.O. Box 210108
Tucson, AZ 85721-0108
and HEC Montreal
and also NBER
gowrisankaran@eller.arizona.edu

Robert Town
Department of Economics
The University of Texas at Austin
2225 Speedway
BRB 1.116, C3100
Austin, TX 78712
and NBER
robert.town@austin.utexas.edu

1 Introduction

In 2006, the U.S. government added a new entitlement to the Medicare program, Part D, which offers 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 benefit structure. 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 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 (Heiss et al., 2010; Abaluck and Gruber, 2011, 2013; Schroeder et al., 2014; Ho et al.,

---

[1]The 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).

[2]The 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 spending of $627.50. Thus, the coverage gap ends after the initial coverage region total spending of $2,510 plus a mean of $3,422.50 (=$4,050 − $627.50) in further out-of-pocket/total spending, for a combined total of $5,932.50 in mean spending.

Figure 1: Coverage by region in 2008 with standard Medicare Part D plans
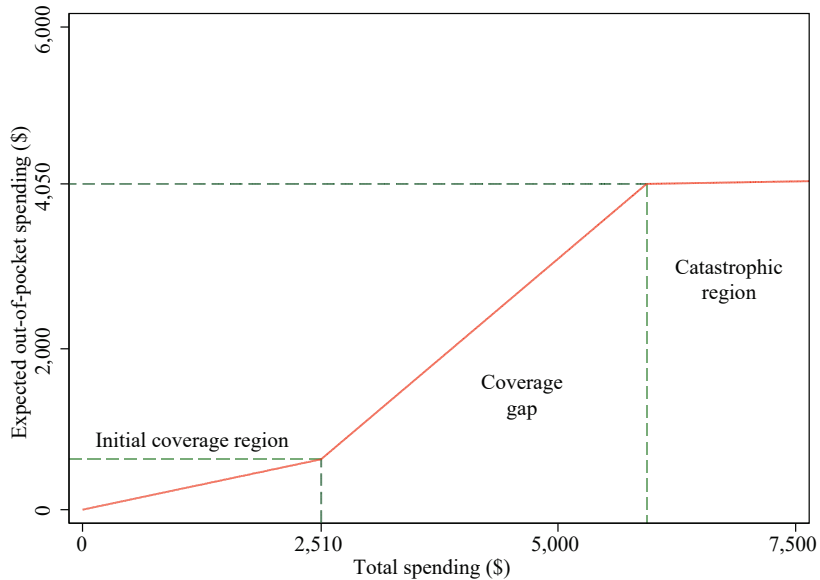
2017),[3] while other studies find that enrollees are, at least in part, rational in their Part D plan choice (Ketcham et al., 2012). Moreover, although the doughnut hole is specific to Part D, most health insurance plans have nonlinear aspects, such as out-of-pocket maxima and deductibles, implying 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 health care costs.

This paper has two goals. The first is to test whether the behavior of Part D enrollees in their prescription drug purchases meaningfully deviates from the predictions of a *benchmark model* defined by neoclassical preferences and a discount factor close to 1 at the annual level. 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.

We proceed by constructing two behavioral dynamic models of drug purchases: quasi-

---

[3]Also 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).

[4]This 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).

hyperbolic discounting (Laibson, 1997; Phelps and Pollak, 1968; Strotz, 1956) and price salience (Chetty et al., 2009; Bordalo et al., 2012). The benchmark 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 benchmark 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 benchmark 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 benchmark model.[5] We contribute to this literature by developing new tests of the benchmark 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 benchmark 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, quasi-hyperbolic discounting, the enrollee or her physician discounts future health expenditures in 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

---

[5]Brot-Goldberg et al. (2017) find that employees who were forced into a high deductible health insurance plan significantly reduced healthcare expenditures even when this would not reduce out-of-pocket expenditures. Ito (2014) shows that enrollees respond to average electricity prices, even though the benchmark model implies that people should respond to marginal prices. Grubb and Osborne (2014) find that consumers exhibit a range of biases in nonlinear cellular phone contracts. In contrast, Nevo et al. (2016) study forward-looking consumers faced with nonlinear broadband internet contracts using the benchmark model.

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 β or $\sigma < 1$), imply different consumption dynamics as enrollees approach and enter the doughnut hole. For β or $\sigma = 1$, both behavioral models are equivalent to each other and to geometric discounting with full salience.

These two behavioral models have very different counterfactual policy implications, highlighting the importance of distinguishing between them. For instance, the literature on quasi-hyperbolic discounting has argued that with "sophisticates," it might be useful to give people future commitment contracts (Laibson, 1997). However, if the deviation from the benchmark model is due to a lack of salience about the doughnut hole, then policies that provided information to help enrollees view future prices as more salient might be useful.

In the benchmark model, where β or $\sigma = 1$ and δ is close to 1 at the annual level, 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 benchmark 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

---

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

with the simple assumption that the enrollee will end the year in the gap with certainty. Moreover, since these enrollees will use all their insurance in the initial coverage region with very high probability, their Part D subsidy is very close to constant. We show that this implies that there should be little or no drop in prescription drug purchases upon entering the doughnut hole under the benchmark model. 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 consider the predictions of the benchmark model by examining whether there are drops in spending upon reaching the doughnut hole for the set of enrollees noted above. We further examine whether our data are consistent with 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 benchmark 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, 2015), 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 or σ. We show that we can identify the discount factor and behavioral parameter given a rank condition and sufficient variation in drug attributes.

Our structural estimation splits our sample into three subsamples based on an ex ante measure of expected pharmacy expenditures. For each subsamples, we can reject β or $\sigma > 0$. The price salience model fits the data best, 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 at the threshold, 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.[7] To isolate the importance of price salience, we examine how prescription purchase behavior changes under the benchmark model, using an annual discount factor of 0.95. Optimization under the benchmark model 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%.

Our paper is most closely related to the works of Einav et al. (2015) and Abaluck et al. (2018) who both also consider the implications of benefit design for Medicare Part D. We

---

[7]The 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.

develop complementary tests to Einav et al. (2015): we test for violations of the benchmark model by evaluating whether there are changes in behavior upon crossing into the doughnut hole when the benchmark 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 benchmark 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. (2018) 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 benchmark model. Finally, our structural model of quasi-hyperbolic discounting builds on Fang and Wang (2015) 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.[8] 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., 2018). 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 their underlying causes of the deviations from benchmark behavior implying different effective solutions to remedy these deviations. 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 benchmark model.[9]

A period in our model is a week, starting with Sunday.[10] Enrollees discount future weeks with a weekly (geometric) discount factor δ. Each week, the enrollee is faced with a number, zero or more, of health shocks. Each health shock is defined by its type. Each health shock type has a unique set of drugs that can be used as treatments. An example of a health shock type is "conditions treated with calcium channel blockers," which is treated exclusively with calcium channel blockers.[11] An example of a calcium channel blocker is Cardizem (diltiazem hydrochloride) in tablet form; our uniqueness assumption implies that this drug is not in a treatment for any other health shock type. Upon receiving a health shock, the enrollee makes

---

[8]Section 5 discusses estimation of the model which involves aggregation across enrollees.

[9]A 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.

[10]Our 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.

[11]For brevity, when unambiguous, we refer to this health shock type simply as "calcium channel blockers."

a discrete choice of one of the treatment drugs for the health shock type, 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.

Each week, the enrollee receives between 0 and N health shocks.[12] She receives the shocks sequentially, implying that upon receiving one shock, she does not know how many more she will receive, 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 distribution over health shock types.[13] Because the distribution of health shock types is specific to an enrollee, our model is consistent with within-enrollee correlations of health shock types, as would occur with a chronic disease. For instance, some enrollees might have type II diabetes, and those enrollees would draw from a health shock type distribution with type II diabetes while other individuals would not have type II diabetes and hence would not draw from this health shock type.[14]

The enrollee's decision problem is dynamic because each drug purchase brings her closer to the next phase of her nonlinear insurance contract (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, 2015). Under the "sophisticates" variant, the enrollee knows that in the future she will continue to act as a quasi-hyperbolic discounter. Under the "naïfs" variant, 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—makes her drug purchase decision prior to the point of sale, e.g., in the physician's office or at

---

[12]Hence, the realized number of health shocks received is distributed $i.i.d.$ categorical, or equivalently, multinomial with one trial.

[13]We 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.

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

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 remains some 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 distribution of the number of health shocks via conditional probabilities: let Q_n, for $n = 0, \ldots N$, 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 nth 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 $Q_n, \ldots, Q_N$.

Let H denote the number of health shock types. We assume that health shock type $h \in \{1, \ldots, H\}$ occurs with probability P_h. For each h, index the prescription drugs that can be used for treatment by $j = 1, \ldots, J_h$. 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 including the outside option.

The expected perceived flow utility from drug j for health shock type h is additive in: (1) the fixed utility from treatment, ϕ_{hj}, which is a parameter to estimate; (2) the disutility from the current expected perceived price of the drug, which we detail below;[16] and (3) an unobservable component ε_{hj}, which is distributed type 1 extreme value, *i.i.d.* across health

---

[15]This 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.

[16]Our 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.

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 each h, denote the outside option as good 0. We assume that $p_{h0} = oop_{h0} = \phi_{h0} = 0$ and that good 0's flow utility is ε_{h0}.

Denote a typical state by $(t, m, n, h, \vec{\varepsilon})$, where $t \geq 0$ is the number of weeks remaining in the year after the current week; $m \geq 0$ is the monetary distance to the doughnut hole at the start of a given purchase decision;[17] $n \in \{0, \ldots, N-1\}$ is the number of previous health shocks during the week; $h \in 1, \ldots, H$ is the health shock type; and $\vec{\varepsilon} \equiv (\varepsilon_{h0}, \ldots, \varepsilon_{hJ_h})$ is the vector of unobservables. At each state, the enrollee maximizes the expected discounted value of her expected 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.

Our estimation focuses on enrollees whose drug spending in the early part of the year have brought them near the start of the coverage gap. Given this, our tests of the benchmark 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.

The first part of Assumption 1 states that consumers will always expect to reach the doughnut hole, and that a one-time change in behavior will not affect this outcome. The second part of Assumption 1 states that consumers believe that they will never go beyond the doughnut hole.

We now show two related preliminary results that simplify the dynamic decision problem. First, we show that c_h does not affect choice probabilities when $\delta < 1$. Second, we show that Assumption 1 allows us to exposit the problem as an infinite horizon Bellman equation, and not account for the week of the year.

Lemma 1. *Consider two vectors of baseline health costs, $\vec{c} \equiv (c_1, \ldots, c_H)$ and $\vec{c}' \equiv (c'_1, \ldots, c'_H)$. Let $s^t(m, n, h, j | \vec{c})$ denote the optimizing probability of purchase of drug $j = 0, \ldots, J_h$ for*

---

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

a given set of state variables (t, m, n, h)—*integrating over* $\vec{\varepsilon}$ *and conditioning on* c. *Let* $s(m, n, h, j | \vec{c})$ *denote the analogous probability for an infinite horizon model where the dough-nut hole will always be reached (and hence where* t *is not a state variable). Then,*

(a) *For* $\delta < 1$, $s^t(m, n, h, j | \vec{c}) = s^t(m, n, h, j | \vec{c}\,')$ *and* $s(m, n, h, j | \vec{c}) = s(m, n, h, j | \vec{c}\,')$, $\forall m, n, h, j, \vec{c}, \vec{c}\,'$.

(b) *Given Assumption 1,* $s^t(m, n, h, j | \vec{c}) = s(m, n, h, j | \vec{c}\,'), \forall m, n, h, j$, *if* $\delta < 1$ *or* $c_h = c_h' = \gamma + \log\left(1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \overline{\alpha} \times \overline{p})\right)$ *for each* h, *so that the value function is* 0 *inside the doughnut hole given optimizing behavior.*

(Proofs of lemmas and propositions are in Appendix C.)

Lemma 1 (a) allows us to specify arbitrary values of c_h in computations and demonstrations regarding market shares. It occurs because c_h affects all options equally.[18] Lemma 1 (b) implies that the state space at the time of a drug purchase consists of four elements, with a typical state written as $(m, n, h, \vec{\varepsilon})$.[19] Appendix B formally exposits the behavioral dynamic optimization problems for the infinite horizon model.

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

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

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

---

[18] We impose $\delta < 1$ since for $\delta = 1$, the values can be infinite depending on the choice of c_h.

[19] With $\delta = 1$, we impose $c_h = \gamma + \log\left(1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \overline{\alpha} \times \overline{p})\right)$ as this gives a finite value function, making the problem conceptually much simpler.

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:

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

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 makes her decision using the prices from (1).

Finally, let the function $\alpha(p)$ denote the disutility from current expected 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. Applying (1) or (2), the current disutility from expected perceived price is $\alpha(p^{eff}(m, p_{hj}, oop_{hj}))$.

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 models 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 the benchmark model 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 enrollees will act approximately the same before and after the doughnut hole under the benchmark model (which is the geometric discounting model with fully salient prices and an annual discount factor close to 1). This is not true for the behavioral models with β or $\sigma < 1$. Intuitively, under the benchmark model, if 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 hence will not change their behavior upon crossing into the doughnut hole. 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 both full and out-of-pocket prices are the same across all drugs and $\delta = 1$:

Proposition 1. *Consider a dynamically-optimizing Part D enrollee for whom Assumption 1 holds and for whom β or $\sigma = 1$ and $\delta = 1$. As in Lemma 1 (b), fix $c_h = \gamma + \log\left(1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \overline{\alpha} \times \overline{p})\right)$ for each h. Suppose further that there is a common full price $\overline{p}$ and a common out-of-pocket price $\overline{oop}$ that are charged for every (inside-good) drug and that price disutility is linear so that $\alpha(p) \equiv \overline{\alpha}p$. Then, the purchase probability of each hj is the same across ex ante states, i.e., for all m, n, m', n', h, j, $s(m, n, h, j) = s(m', n', h, j)$.*

We note three 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,

---

[20]It is also be 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 sophisticates model, except for the relabeling of the parameter β to σ.

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. Finally, by Lemma 1, the normalization on the c_h values would not matter if $\delta < 1$ or if we considered the finite horizon model instead of the infinite horizon model.

Figure 2: Simulated drug spending for the geometric model across discount factors
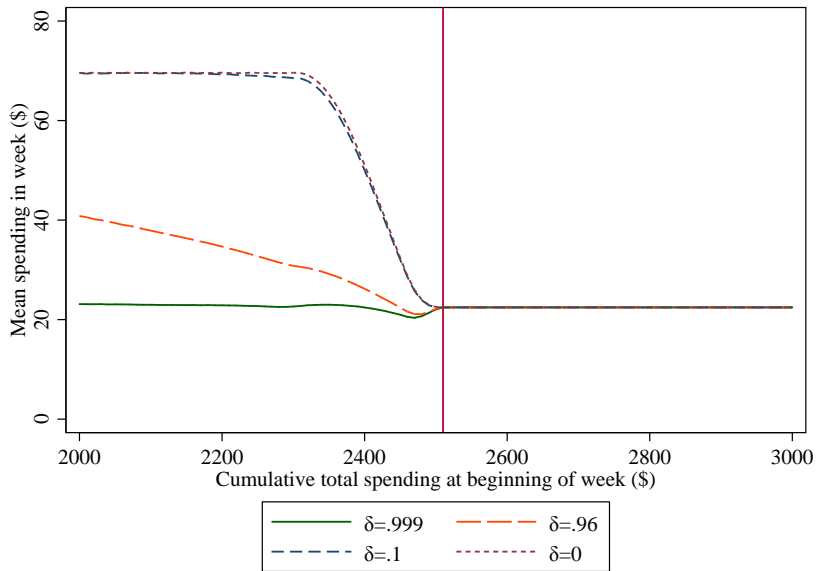
To provide further insight as to the role of δ in the geometric discounting model in affecting dynamic drug consumption patterns, we simulate the model for different values of δ. Figure 2 reports simulated mean total spending per week across discount factors as a function of the cumulative total spending at the beginning of the week. We report simulations for four 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; 0.1, to understand the impact close to 0; and 0, the case of perfect myopia.

We calculate dynamically-optimizing decision-making for enrollees and then simulate weekly spending in the figure. Enrollees in the simulation all have one health shock each week and each health shock is drawn with equal probability from one of 20 health shock types, 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, suggesting that Proposition 1 is approximately true more generally.[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 25% coinsurance, a foregone $100 purchase with $2,300 in total spending would result in $25 in immediate savings and $75 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 $75 in savings discounted more because the time until the expected crossing is longer. With $\delta = 0$, spending is flat in the pre-doughnut-hole region before $2,310 since discounted savings are worth nothing. Finally, the $\delta = 0.1$ line is only slightly downward sloped in this region, showing that the slope is continuous in δ.

Now we consider spending under 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* $\delta = 1$. *Fix* c_h *for each* h *as in Proposition 1. Suppose further that there is a common full*

---

[21]Drug 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. Price disutility is $\alpha(p) = p$. These ranges of prices are roughly similar to the sample.

[22]The 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.

price $\overline{p}$ and out-of-pocket price $\overline{oop}$ that is charged for every (inside-good) drug and that price disutility is linear so that $\alpha(p) \equiv \overline{\alpha}p$. 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 benchmark 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 $m, m' \geq \overline{p}$ 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 $m \geq \overline{p}$ and $0 < \beta = \sigma < 1$ and for all m, n, h and for $j = 1, \ldots, J_h$.

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 benchmark 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 or 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 shows simulation evidence for the same set of flow utility parameters as in

17

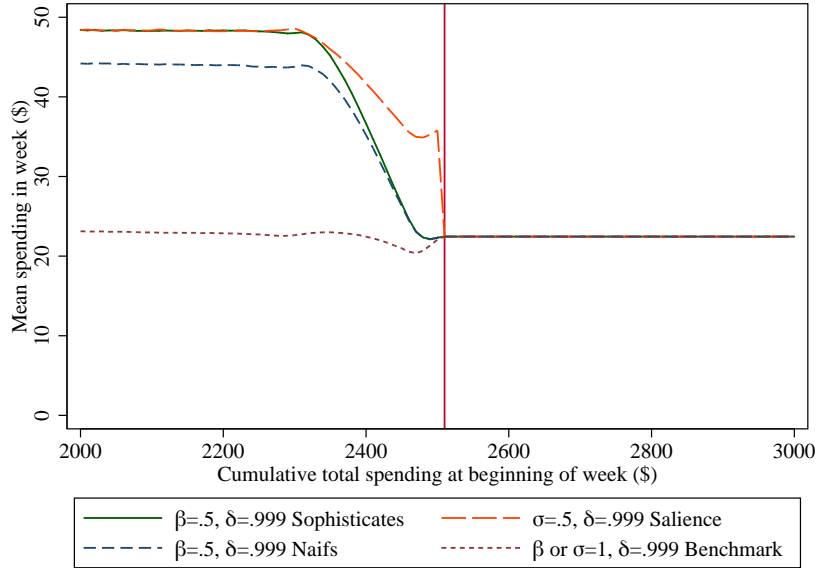Figure 3: Simulated drug spending for different behavioral models

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 benchmark model 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

18

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 weekly 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. The benchmark model predicts that there should be no drop in spending at the dough-nut hole while the other models that we consider predict a 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 pur-chases 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 manger 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 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. By construction, the coverage gap start is the same across the

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 provides generic coverage in doughnut hole.

base plans and the coverage gap end out-of-pocket 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.

The sixth plan in our data, plan F, is only used for falsification tests. Plan F has its coverage gap start at $4,000 in total spending, a much higher threshold than for the other plans. Its enrollees are older and disproportionately female relative to the plans in our base analysis sample.

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 benchmark 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.

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 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., Gowrisankaran et al., 2013; Handel, 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 health shock type meant to capture the treatment of the drug. A clinically trained research assistant performed the coding using the pill name, drug type (e.g., tablet or cream), most common indication, and National

---

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

[24]There 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.

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, they form separate health shock types in our analysis because their mechanisms are separate.

Table 2 lists the health shock types 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]

Table 2: Most common health shock types in base estimation sample

Health shock type	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 benchmark 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

---

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

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.

Figure 4: Spending near coverage gap for base estimation sample
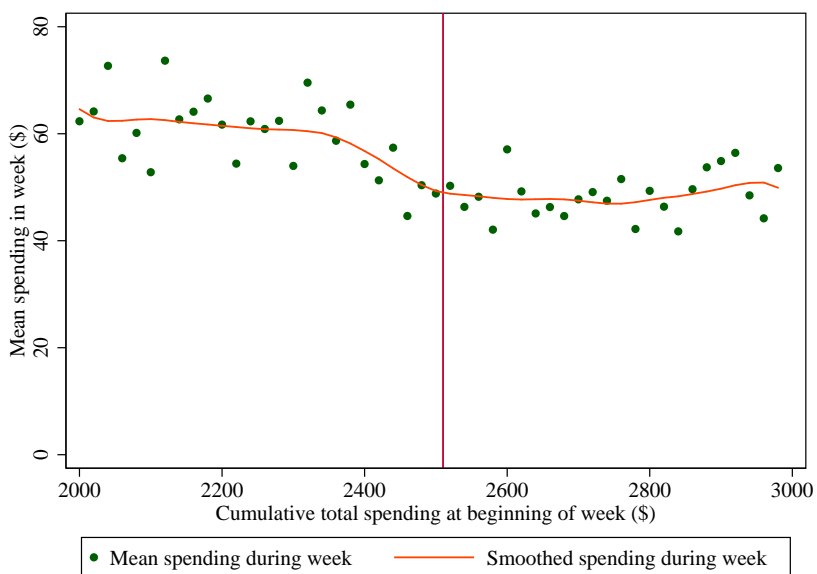
We start by graphing mean weekly spending levels and non-parametric regressions of these levels. Figure 4 plots mean total drug spending by $20 increments of beginning-of-week cumulative spending and a kernel smoothed "lowess" regression of mean total drug spending on beginning-of-week cumulative spending.[26] The mean total drug spending shows little 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

---

[26]We use a bandwidth of 0.3 for these regressions.

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 at the much higher level of $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 $2,510 in total spending. This finding 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:

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

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, if the enrollee starts the week near the doughnut hole, her spending during the week may move her into the doughnut hole.

By selecting a small region around the doughnut hole, we are comparing the same indi-

---

[27]Figure 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.

vidual at similar points in the year but faced with different contemporaneous 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: 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.

Table A2 shows the same analysis from Table 3, but on plan F, the falsification plan, which had no price change at $2,510 in spending. The coefficients with this sample are much smaller in magnitude than for the base sample, e.g., we find a $3.35 increase in weekly spending at the $2,510 point for this plan, compared with a $17.46 decrease for the base sample. They also do not show a consistent pattern, with three of the coefficients being positive and five being negative.

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 they have either a β or σ or a δ that is substantially less than one: the dynamics of their drug purchase decisions do not reflect the predictions of the benchmark model.

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

---

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

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	N
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.

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, the coefficient on total spending in the $2,200 to $2,399 range is not significant and close to 0. 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. This is consistent with the predictions of the behavioral models with β or σ much lower than δ. It is, however, inconsistent with the geometric discounting model with a sufficiently high discount factor. For instance, the analogous coefficient for $\delta = 0.96$ in Figure 2 (which uses simulated data) would be well above the confidence interval for our estimates here.[29]

Table A3 in Appendix A provides evidence on the five health shock types 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 for drugs that treat a health shock type as the dependent variable. We then report the health shock types with the biggest and smallest coefficients on the spending drop in the doughnut hole region. The five health shock types with the biggest drops in prescriptions are also among the ten most common health shock types, as reported in Table 2. Indeed, the only one of the top five health shock types that does not have a drop

---

[29]We perform formal tests on β and σ in the context of our structural estimation results in Section 6.1.

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: $2,510 - 2,999	$2,200 - 2,399	N
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.

that is also in the top five is opioids. The five health shock types with the biggest increases in prescriptions upon entering the doughnut hole are all health shock types with very few prescriptions (and the coefficients are all insignificant). Overall, this table shows that the percentage drops in prescriptions are similar across most health shock types. 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 the identification of our results rejecting the benchmark 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 $g = 1, \ldots, G$ 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.[30]

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 estimation window, with the logic being that the doughnut hole is sufficiently far away. Thus, we estimate P_h and Q_n for each ACG group from that group's enrollees' weekly drug purchases measured from their first week of purchases 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, we allow the other parameters to vary in three sets: the lowest, highest, and middle six ACG scores. For each estimation, we lump together health shock types with fewer than 100 prescriptions filled for the estimation sample over the entire year in a type called "Other." We also lump together drugs within a health shock type as "Other" until such point as every drug has at least 50 prescriptions filled over the entire year.[31]

Our basic approach to estimation is maximum likelihood 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 probabilities 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

---

[30]Table A4 in Appendix A provides details on the enrollees by group.

[31]We make these simplifications for computational tractability, since our estimation has fixed effects for each drug and requires an accurate estimation of the probability of each health shock type.

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.[32] 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 or σ and δ and then using a derivative-based search for all other parameters, given each value of β or σ and δ.[33] Not reported in the paper, we also performed Monte Carlo simulations to verify the accuracy of the code and power of the estimator.

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 or σ. We begin with an intuitive description of identification and then provide a proposition.

In dynamic discrete choice models, exclusion restrictions can be used to identify δ (Magnac and Thesmar, 2002; Fang and Wang, 2015; Abbring and Daljord, 2018). In our setting, the variability of drug prices near the doughnut hole provides such exclusion restrictions. To see this, consider the geometric discounting case with a cheap drug k —with $p_k = 40$ and $oop_k = 10$—and an expensive drug l—with $p_l = 100$ and $oop_l = 25$. From (1), at a state that is $m' = \$20$ dollars from the doughnut hole, there is no insurance subsidy for drug l but there is $\$10$ in insurance subsidy for drug k. Hence, our exclusion restriction is that the expected

---

[32]We 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.

[33]We 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.

discounted value from purchasing drug l at m' is the same as inside the doughnut hole.[34]

Focusing on the geometric discounting model, this exclusion restriction allows us to identify δ based on the change in the relative purchase probabilities of k to l compared to inside the doughnut hole.[35] We then identify the parameters of $\alpha(\cdot)$ using price variation across drugs inside the doughnut hole, and identify the ϕ parameters from choice probabilities net of the price disutility.

Finally, we can identify β or σ by considering the change in purchase probabilities as we move further back from the doughnut hole start. Intuitively, having identified the other parameters as in Figures 2 and 3, identification of β or σ follows from the difference in purchase probabilities at the second to last purchase occasion compared to the final purchase occasion before the doughnut hole. These probabilities will be similar if β is low and δ is close to 1, while the earlier purchase occasion will have a higher purchase probability if β is 1 and δ is low.

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; that $0 < \delta < 1$ and β or $\sigma > 0$; and that there is one health shock type, so that there are J drugs. Assume also that there is sufficient price variation across drugs such that for some drugs k and l and state variables m' and m'' that are reached by the data, (i) $oop_k < p_k$ and (ii) $oop_k < m' < m'' < \min\{p_1, \ldots, p_J, oop_l\}$. Finally, assume that the price disutility is linear so that $\alpha(p) \equiv \overline{\alpha}p$. Then:*

(a) For a given β or σ, $\overline{\alpha}$, $\phi_1, \ldots \phi_J$, and δ are identified given any of the three models—of quasi-hyperbolic discounting naïfs and sophisticates and price salience.

(b) Provided a rank condition holds, β or σ is also identified given any of the three models.

---

[34]Abbring and Daljord (2018) consider identification of the discount factor with restrictions on flow utility but not the value function. In our case, the theory directly imposes a restriction on the choice-specific value functions, as in Magnac and Thesmar (2002), and hence we do not apply the Abbring and Daljord results.

[35]Magnac and Thesmar (2002) also require a rank condition, which is satisfied in our case because there is insurance value for drug k at state m', unlike in the doughnut hole, implying that this state is different from the corresponding inside-the-doughnut-hole state.

The rank condition is:

$$\log\left(\frac{s(m'',1,1,0)s(0,1,1,l)}{s(m'',1,1,l)s(0,1,1,0)}\right)\left[\sum_{j=1}^{J}s(m',1,1,j)\log\left(\frac{s(m',1,1,j)s(0,1,1,l)}{s(m',1,1,l)}\right)\right.$$

$$\left. -\sum_{j=0}^{J}s(m',1,1,j)\log(s(m',1,1,j))+s(m',1,1,0)\log(s(0,1,1,0))\right]$$

$$-\log\left(\frac{s(m',1,1,0)s(0,1,1,l)}{s(m',1,1,l)s(0,1,1,0)}\right)\left[\sum_{j=1}^{J}s(m'',1,1,j)\log\left(\frac{s(m'',1,1,j)s(0,1,1,l)}{s(m'',1,1,l)}\right)\right.$$

$$\left. -\sum_{j=0}^{J}s(m'',1,1,j)\log(s(m'',1,1,j))+s(m'',1,1,0)\log(s(0,1,1,0))\right]\neq 0.\quad(4)$$

Several implications of the proposition deserve further comment. First, while we cannot verify that the rank condition in Proposition 3 (b) would always hold, we simulated choice probability data from the equilibrium of the model for a number of parameter values. We verified that the condition always holds for these parameter values and that we are always able to recover β and δ.[36]

Second, Proposition 3 focuses on states near the doughnut hole for tractability: for these states, any drug purchase leads to an exclusion restriction of the value function, as noted above.[37] However, in practice, states with multiple purchases necessary to reach the doughnut hole will help identify the parameters and add identifying variation.

Third, also for tractability, Proposition 3 imposes a number of assumptions—such as the presence of only one health shock type and only one shock per week—but more complex environments should yield more identifying variation.

Fourth, while we did not formally consider the identification of the different behavioral models, our proof shows that β and σ are identified from any pair of states m' and m'' within some range. We have many such pairs of states and our identifying equations are not collinear across pairs of states for the simulated parameters that we tried. Thus, intuitively, these additional pairs of states will also help identify the best model. Moreover, it is easy

---

[36]We include the simulation code, output, and instructions in the supplementary materials to the paper.

[37]Thus, unlike the above intuition, our proof does not use multiple drug purchase occasions, but the effect of observing the purchase probabilities at different amounts to the doughnut hole is mathematically similar.

to see how model specification between the quasi-hyperbolic discounting and salience models is identified: under quasi-hyperbolic discounting, the purchase probabilities at $0 < m < \min\{oop_1, \ldots, oop_J\}$ are identical to inside the doughnut hole, while under the salience model, they are strictly higher.

Finally, we note that our identification leverages the heterogeneity of prices across drugs and health shock types and responses to this heterogeneity. Our overall takeaway is that to identify discount factors from administrative data such as ours, it is useful 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 model with naïfs and sophisticates and the price salience model.

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 variants are identical. Since δ is not identified when β or $\sigma = 0$, we do not report δ.

We cannot compute a standard error for β or 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 or σ (Newey and McFadden, 1994), over a grid of these values. We reject all values of β or $\sigma > 0$ 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. It is difficult to formally test the two models, because our estimated β and

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^{**} (0.001)	-0.013^{**} (0.001)	-0.018^{**} (0.001)
Price spline $\geq$ \$150,	-0.006^{**} (0.001)	-0.006^{**} (0.001)	-0.003^{*} (0.001)
Behavioral parameter: β or σ	0 (–)	0 (–)	0 (–)
Discount factor: δ	–	–	–
log L	−95,594.6	−95,594.6	−95,456.7
log L β or $\sigma = 0.1, \delta = 0.1$	−95,602.6	−95,602.4	−95,460.4
P-value for LM test	0.00	0.00	0.00
log L β or $\sigma = 0.1, \delta = 0.4$	−95,604.8	−95,604.5	−95,462.0
P-value for LM test	0.00	0.00	0.00
log L at β or $\sigma = 0.1, \delta = 0.995$	− 95,619.6	−95,615.7	−95,471.6
P-value for LM test	0.00	0.00	0.00
log L at β or $\sigma = 0.3, \delta = 0.995$	−95,672.5	−95,663.8	−95,532.9
P-value for LM test	0.00	0.00	0.00
Number of health shock types H		60	
Number of drug fixed effects ϕ		245	
N		18,897	

Note: Standard errors reported in parentheses are calculated using standard outer product approximations, treating β, σ and δ as fixed. '**' 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 or σ and δ.

σ parameters are at the boundary. Thus, we perform a non-nested test of model selection that conditions on $\beta = \sigma = 0$. Using the Vuong (1989) test, we reject the quasi-hyperbolic discounting model in favor of the salience model (test statistic = 7.36, p<0.001).

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 the data reflect a more gradual

Figure 5: Fit of quasi-hyperbolic discounting and price salience models near coverage gap
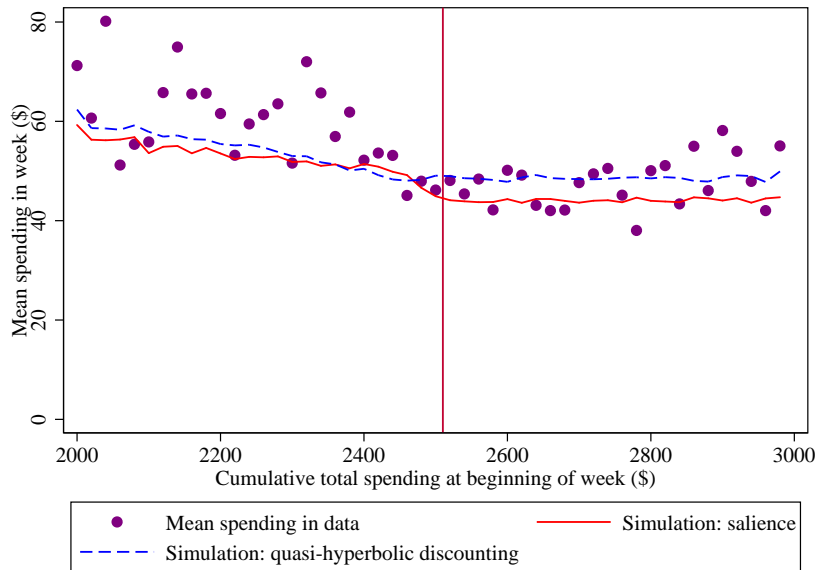
drop in spending before the doughnut hole than predicted by the estimated salience model.

In contrast, the quasi-hyperbolic discounting model predicts too early a spending decline from the doughnut hole. Given the estimated value of $\beta = 0$, this early decline is exclusively caused by health shocks with expensive prescriptions and weeks with multiple health shocks, rather than by forward-looking behavior by enrollees. Finally, the salience model also matches the drop in spending at the doughnut hole more closely than the quasi-hyperbolic discounting model.[38]

Table A5 in Appendix A presents the model fit for all the moments reported in the reduced-form analysis in Table 3. The salience model fits better on many of these moments, such as the drop in the number of expensive prescriptions and in the number of branded prescriptions.

Table A6 in Appendix A reports results for other ACG scores, which are very similar to

---

[38]Mean weekly spending in the data drops by 23.7% in the [\$2,510, \$3,000) region relative to the [\$2,000, \$2,400) region in the data, compared with 18.2% for the estimated salience model and only 12.9% for the estimated quasi-hyperbolic discounting model. Thus, while both models underpredict the drop, the salience model fits much better, at 77% of the realized drop.

the base results.[39] Overall, our estimates of $\sigma = 0$ 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,[40] Abaluck et al. (2018) 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. Since overly small estimated price coefficients may lead to overly low estimated β or σ in order to explain the drop in spending at the doughnut hole, this also adds to credibility that our estimated values of β and σ are not too low.

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

---

[39]Not reported in the paper, we also estimated the model with stratification both by ACG score 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.

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

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.[41] 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.[42]

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.[43] To quantify behavioral hazard, we compare the baseline Part D program to the benchmark 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 enrollee welfare using $\sigma = 1$ and an annualized 95% discount factor.[44]

The results, in Table 6, show that the benchmark 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 benchmark model. Interestingly, there is a small decrease in insurer expenditures in moving to geometric discounting,

---

[41]Our 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 having a health shocks (Q_n) and on health shock type probabilities (P_h) for the estimation and non-estimation samples.

[42]We compute welfare measures using $c_h = \gamma$. A different c_h would rescale the welfare of each counterfactual by the same amount.

[43]This is a relatively simple view of the potential behavioral hazard present here, focusing only on over-consumption of drugs, and not on substitution to non-pharmaceutical spending.

[44]Note 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 6: Relative impact of behavioral hazard and drug insurance

Statistic (per week)	Baseline: $\sigma = 0$ future not salient Case 1	Benchmark 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%.

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.

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 would 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. (2018) estimate that figure to be 6%.

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.

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.[45]

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 would 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

---

[45]Chandra 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.

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 benchmark model, with a geometric discount factor close to 1 and no quasi-hyperbolic discounting or price salience. 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 benchmark model. Enrollees lower their prescription drug purchases upon reaching the doughnut hole, with a disproportionate drop for branded drugs and drugs that cost over $150. 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 a health shock of a random type. Each health shock type requires treatment by a one of a particular set of drugs. The enrollee thus chooses to purchase one of the drugs in that set 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. We also reject the quasi-hyperbolic

discounting in favor of the price salience model. Finally, we find that future prices are not at all salient. This suggests that policies to make doughnut hole prices more salient in consumers' drug purchase decisions in the initial coverage region would add value.

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 that our estimated model matches data moments reasonably well.

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. (2018). Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets. *Journal of Public Economics*, 164:106–138.

Abbring, J. H. and Daljord, Ø. (2018). Identifying the discount factor in dynamic discrete choice models. Mimeo, Tilburg University.

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. C., Chandra, A., Handel, B. R., and Kolstad, J. T. (2017). What does a deductible do? the impact of cost-sharing on health care prices, quantities, and spending dynamics. *The Quarterly Journal of Economics*, 132(3):1261–1318.

Cabral, M. (2016). Claim timing and ex post adverse selection. *The Review of Economic Studies*, 84(1):1–44.

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 expenditure to nonlinear contract design: evidence from Medicare Part D. *The Quarterly Journal of Economics*, 130(2):841–899.

Fang, H. and Wang, Y. (2015). Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *International Economic Review*, 56(2):565–596.

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 system's 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 Scott Morton, F. (2017). The impact of consumer inattention on insurer pricing in the Medicare Part D program. *The RAND Journal of Economics*, 48(4):877–905.

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. Mimeo, University of Texas.

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.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333.

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	Health shock type	Bran-red	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
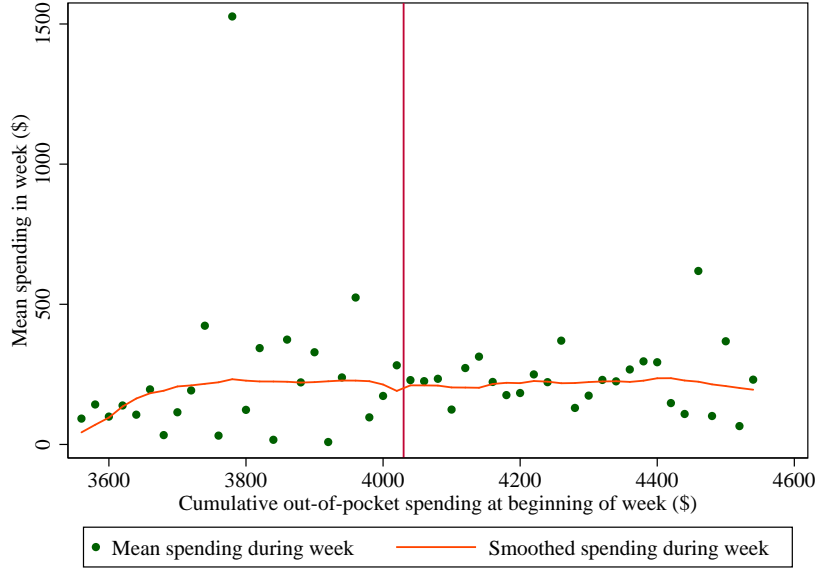
Note: Figure is based on plans in estimation sample, using enrollees who start a week with between $3,550 and $4,050 in out-of-pocket spending between Mar. 30 and Jul. 20, 2008.

Table A2: Behavior for falsification plan sample arriving near coverage gap

Dependent variable:	Mean value before $2,400	Beginning of week spending in: $2,510 - 2,999	N
Mean spending in week	72.18	3.35* (1.52)	42,687
Mean price per Rx	83.02	3.02* (1.47)	17,528
Number of Rxs	0.94	−0.02* (0.01)	42,687
Number of branded Rxs	0.30	−0.01 (0.01)	42,687
Number of generic Rxs	0.63	−0.02* (0.01)	42,687
Expensive Rxs	0.15	0.01** (0.00)	42,687
Medium Rxs	0.21	−0.01 (0.01)	42,687
Inexpensive Rxs	1.17	−0.01 (0.01)	42,687

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.

Figure A2: Spending near $2,510 for falsification Plan F
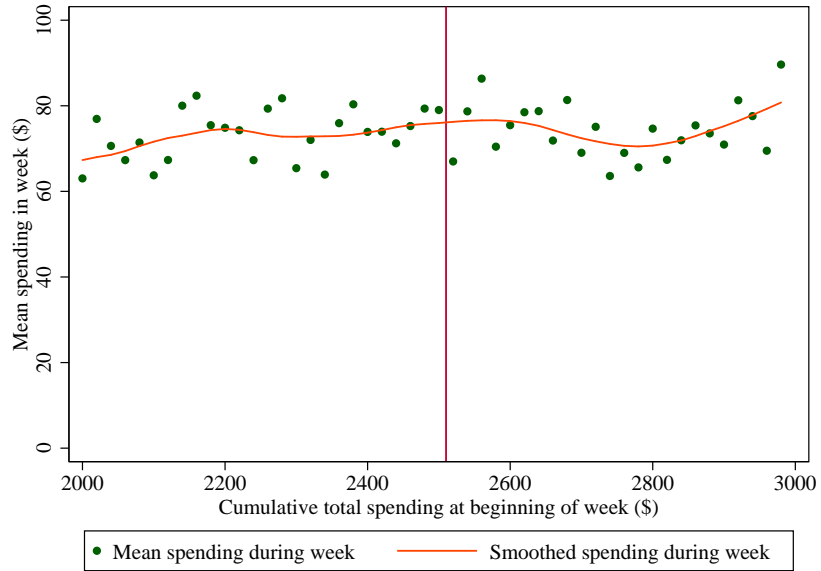
Note: Figure is based on analog to base estimation sample for plan F.

Figure A3: Histogram of total year drug spending for base estimation and full samples
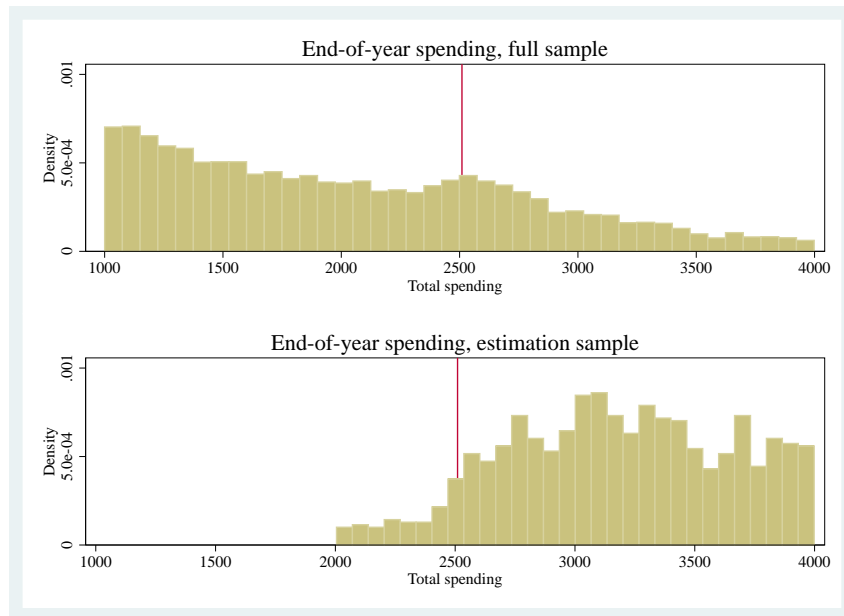
Figure A4: Information provided to Part D enrollees on distance to doughnut hole

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).

Table A3: Health shock types 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	N
Cholesterol-Lowering	0.081	−0.0177** (0.0034)	28,543
Beta-Blocker	0.046	−0.0135** (0.0023)	28,543
Gastroesophageal Reflux & Peptic 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 A4: ACG scores by base estimation sample status

ACG score	Enrollees in base sample	Enrollees not in base sample	Maximum number of health shocks, N	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 A5: Fit of estimated salience and quasi-hyperbolic discounting model to data

Statistic	Data	Estimated salience model	Estimated quasi-hyperbolic discounting model
Percent change in *Mean spending in week* from when beginning of week spending in [2000,2400) to when beginning of week spending in [2510,3000)	-23.7%	-18.2%	-12.9%
Percent change in *Mean price per Rx* from when beginning of week spending in [2000,2400) to when beginning of week spending in [2510,3000)	-18.7%	-5.9%	-3.8%
Percent change in *Number of Rxs* from when beginning of week spending in [2000,2400) to when beginning of week spending in [2510,3000)	-6.9%	-13.2%	-10.1%
Percent change in *Number of branded Rxs* from when beginning of week spending in [2000,2400) to when beginning of week spending in [2510,3000)	-14.5%	-15.6%	-10.3%
Percent change in *Number of generic Rxs* from when beginning of week spending in [2000,2400) to when beginning of week spending in [2510,3000)	-2.5%	-13.2%	-11.3%
Percent change in *Number of Expensive Rxs* from when beginning of week spending in [2000,2400) to when beginning of week spending in [2510,3000)	-30.8%	-24.3%	-17.1%
Percent change in *Number of Medium Rxs* from when beginning of week spending in [2000,2400) to when beginning of week spending in [2510,3000)	-9.2%	-12.5%	-8.3%
Percent change in *Number of Inexpensive Rxs* from when beginning of week spending in [2000,2400) to when beginning of week spending in [2510,3000)	2.1%	-1.2%	-1.3%

Table A6: 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**	−0.122**	−0.141**	−0.193**
	(0.011)	(0.012)	(0.014)	(0.018)
Price ∈ [$20, $50)	−0.015**	−0.022**	−0.025**	−0.030**
	(0.004)	(0.004)	(0.005)	(0.005)
Price ∈ [$50, $150)	−0.014**	−0.017**	−0.017**	−0.023**
	(0.002)	(0.002)	(0.002)	(0.002)
Price ≥ $150	−0.003	0.002	−0.002	0.002
	(0.004)	(0.004)	(0.003)	(0.002)
Behavioral β or σ	0	0	0	0
	–	–	–	–
δ	–	–	–	
log L	−27,244.4	−27,220.7	−19,558.9	−19,504.9
# health shock types	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 or 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 ≥ $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.

Appendix B: Details of Dynamic Optimization Problems and Likelihood Function

We now exposit the enrollee's dynamic optimization problem for the infinite horizon model where the doughnut hole is the absorbing state, 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 $\vec{\varepsilon}$ 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 $(m, n, h, \vec{\varepsilon})$, the enrollee's expected perceived utility from choice $j \in \{0, \dots, J_h\}$ can be written as:

$$\overline{u}_j(m, n, h) + \varepsilon_{hj} \equiv \phi_{hj} - \alpha(p^{eff}(m, p_{hj}, oop_{hj})) - c_h + \beta V\left(\max\{m - p_{hj}, 0\}, n + 1\right) + \varepsilon_{hj}. \tag{5}$$

Equation (5) 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 (5) into the mean utility, defined as $\overline{u}_j(m, n, h)$.

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 $\overline{u}_j(m, n, h)$, $s(m, n, h, j)$ takes on a standard logit functional form:

$$s(m, n, h, j) = \frac{\exp(\overline{u}_j(m, n, h))}{\sum_{k=0}^{J_h} \exp(\overline{u}_k(m, n, h))}. \tag{6}$$

Finally, we exposit the value function:

$$V(m, n) = (1 - Q_n)\delta V(m, 0) + Q_n \sum_{h=1}^{H} P_h \sum_{j=0}^{J_h} s(m, n, h, j) \times$$

$$\left[\phi_{hj} - \alpha(p^{eff}(m, p_{hj}, oop_{hj}) - c_h) + V(\max\{m - p_{hj}, 0\}, n + 1) - \log s(m, n, h, j) + \gamma\right], \tag{7}$$

where γ is Euler's constant. Equation (7) evaluates, in turn, the two possibilities ex ante to the health shock realization: first, that there are no more health shocks 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 health shock types. 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 (7) account for the expected future utility gross of ε_{hj}. The final

terms, $-\log s(m, n, h, j) + \gamma$, account for the expectation of ε_{hj} 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:

$$
V^{geo}(m, n) = (1 - Q_n)\delta V(m, 0) + Q_n \sum_{h=1}^{H} P_h \times
$$

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

(8)

The naïf enrollee will make choices with a utility function analogous to $\overline{u}_j(m, n, h)$ in (5), 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 $\overline{u}_j(m, n, h)$, $s(m, n, h, j)$ 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 expected 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 $n = 1, \ldots, N_{it}$, let m_{itn} denote the value of m, the dollars till the doughnut hole; $h_{itn} \in \{1, \ldots, H\}$ denote the realization of the health shock; and $j_{itn} \in \{0, \ldots, J_h\}$ 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:

$$
\log L_{it} = \sum_{n=0}^{N_{it}}
$$

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

(9)

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 9) 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 possible configurations. If the last drug purchase is at shock 3, the total number of shocks could be 3 or 4, yielding four possible configurations. Finally, if the last drug purchase is at shock 2, then the total number of shocks could be 2, 3, or 4, yielding three possible 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 $l_{itn}, n = 1, \ldots \hat{N}_{it}$ denote the places of each health shock, with $1 \leq l_{it1} < \cdots < l_{it\hat{N}_{it}} \leq N_{g(i)}$. 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:

$$
\log L_{it} = \log \left(\sum_{l_1,\ldots,l_{\hat{N}_{it}} \in \mathcal{L}(\hat{N}_{it},N)} \sum_{N_{it}=l_{\hat{N}_{it}}}^{N_{g(i)}} \left(\prod_{n=0}^{N_{it}-1} Q_{g(i)n} \right) (1 - Q_{g(i)N_{it}}) \right.
$$
$$
\prod_{n=1}^{\hat{N}_{it}} P_{g(i)h_{itn}} s\left(g(i), m_{itn}, l_n, h_{itn}, j_{itn} \right)
$$
$$
\left. \prod_{n=1,n\neq l_1,\ldots,n\neq l_{\hat{N}_{it}}}^{N_{it}} \left(\sum_{h=1}^{H} P_{g(i)h} s\left(g(i), \min_{\tilde{n} \text{ s.t. } l_{\tilde{n}}<n} m_{it\tilde{n}}, n, h, 0 \right) \right) \right). \quad (10)
$$

In words, the first line of (10) 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 (10) is similar to the earlier likelihood in equation (9) but with two main differences: first, it integrates over the

places of each observed shock, the total number of health shocks, and the health shock type 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 (10) 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 suppress 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 $(1, \ldots, N)$, which gives $o(\mathcal{L}(1, N)) = 1$.

Inductive case: Assume by induction that the theorem holds 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 Nth 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 $\overline{N} - 1$ 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. □

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 Lemmas and Propositions from Paper

Proof of Lemma 1 We now prove the lemma for the quasi-hyperbolic discounting model for sophisticates. We first prove part (a) of the lemma. Starting with the infinite horizon models, we show that $s(0, n, h, j|\vec{c}) = s(0, n, h, j|\vec{c}')$, i.e., that choice probabilities are the same across parameter values inside the doughnut hole. For $j = 0, \ldots, J_h$,

$$\overline{u}_j(0, n, h|\vec{c}) = \phi_{hj} - \alpha p_{hj} - c_h + \beta V(0, n + 1| \vec{c})$$

and

$$\overline{u}_j(0, n, h|\vec{c}') = \phi_{hj} - \alpha p_{hj} - c'_h + \beta V(0, n + 1| \vec{c}').$$

Since $\delta < 1$, $V(0, n + 1| \vec{c})$ and $V(0, n + 1| \vec{c}')$ are finite. Thus, $\overline{u}_j(0, n, h|\vec{c})$ and $\overline{u}_j(0, n, h|\vec{c}')$ differ by the same finite constant across all options, implying that $s(0, n, h, j|\vec{c}) = s(0, n, h, j|\vec{c}')$.

We now consider the $m > 0$ case. Suppose, by contradiction, that there is some m', n, h, j, where $m' > 0$ and $s(m', n, h, j|\vec{c}) \neq s(m', n, h, j|\vec{c}')$. Let $m' > 0$ be the smallest value of m such that this inequality holds; let the expected number of health shocks remaining, given that n shocks have already occurred, be $EHSR_n$ for $n = 0, \ldots, N$; and let the difference in the expected discounted future baseline health costs, given that n shocks have already occurred, be

$$DFBC_n = (EHSR_n + \frac{\delta}{1 - \delta}EHSR_0) \sum_{h=1}^{H} P_h(c'_h - c_h).$$

Because the choice probabilities are the same for $m < m'$, $V(m, n|\vec{c}) = V(m, n|\vec{c}') + DFBC_n$ for $m < m'$. It then follows that, for $j > 0$,

$$\overline{u}_j(m', n, h|\vec{c}) = \overline{u}_j(m', n, h|\vec{c}') - c_h + c'_h + \beta DFBC_{n+1}. \tag{11}$$

Given (11), (5) and (7) for $m = m'$ and $\forall h, n$ fully define $\overline{u}_0(m', n, h|\vec{c})$ and $V(m', n|\vec{c})$ respectively. We now propose candidate values for these terms:

$$\overline{u}_0(m', n, h|\vec{c}) = \overline{u}_0(m', n, h|\vec{c}') - c_h + c'_h + \beta DFBC_{n+1}, \forall h, n, \tag{12}$$

and

$$V(m', n|\vec{c}) = V(m', n|\vec{c}') + DFBC_n, \forall n. \tag{13}$$

At these candidate values, $\overline{u}_j(m', n, h|\vec{c})$ and $\overline{u}_j(m', n, h|\vec{c}')$ differ by the same constant across all options, implying that $s(m', n, h, j|\vec{c}) = s(m', n, h, j|\vec{c}')$. We now verify that, with these candidate values, (12) and (13) are jointly satisfied at $m = m'$, for all h, n when we substitute

in from (5) and (7). Starting with (12), for any n, h,

$$
\begin{aligned}
RHS_{(12)} &= \bar{u}_0(m', n, h | \vec{c}') - c_h + c_h' + \beta DFBC_{n+1} \\
&= -c_h' + \beta V(m', n+1 | \vec{c}') - c_h + c_h' + \beta DFBC_{n+1} \\
&= \beta V(m', n+1 | \vec{c}') - c_h + \beta DFBC_{n+1} \\
&= -c_h + \beta(V(m', n+1 | \vec{c}') + DFBC_{n+1}) \\
&= -c_h + \beta V(m', n+1 | \vec{c}) \\
&= \bar{u}_0(m', n, h | \vec{c}) = LHS_{(12)}.
\end{aligned}
$$

Continuing to (13), for any n,

$$
RHS_{(13)} = V(m', n | \vec{c}') + DFBC_n
$$

$$
= (1 - Q_n)\delta V(m', 0 | \vec{c}') + Q_n \sum_{h=1}^{H} P_h \sum_{j=0}^{J_h} s(m', n, h, j | \vec{c}') \times \left[\phi_{hj} - \alpha(p^{eff}(m', p_{hj}, oop_{hj}) - c_h') + \right.
$$

$$
\left. V(\max\{m' - p_{hj}, 0\}, n+1 | \vec{c}') - \log s(m', n, h, j | \vec{c}') + \gamma \right] + DFBC_n
$$

$$
= (1 - Q_n)(\delta V(m', 0 | \vec{c}) - \delta DFBC_0) + Q_n \sum_{h=1}^{H} P_h \sum_{j=0}^{J_h} s(m', n, h, j | \vec{c}) \times
$$

$$
\left[\phi_{hj} - \alpha(p^{eff}(m', p_{hj}, oop_{hj}) - c_h') + V(\max\{m' - p_{hj}, 0\}, n+1 | \vec{c}) \right.
$$

$$
\left. - DFBC_{n+1} - \log s(m', n, h, j | \vec{c}) + \gamma \right] + DFBC_n
$$

$$
= (1 - Q_n)\delta V(m', 0 | \vec{c}) + Q_n \sum_{h=1}^{H} P_h \sum_{j=0}^{J_h} s(m', n, h, j | \vec{c}) \times
$$

$$
\left[\phi_{hj} - \alpha(p^{eff}(m', p_{hj}, oop_{hj}) - c_h) + V(\max\{m' - p_{hj}, 0\}, n+1 | \vec{c}) - \log s(m', n, h, j | \vec{c}) + \gamma \right]
$$

$$
- \left((1 - Q_n)\delta DFBC_0 + Q_n \left(DFBC_{n+1} + \sum_{h=1}^{H} P_h(c_h' - c_h) \right) \right) + DFBC_n
$$

$$
= V(m', n | \vec{c}) - DFBC_n + DFBC_n = V(m', n | \vec{c}) = LHS_{(13)}.
$$

Since the candidate values at m' satisfy all the above conditions, they are the actual values, implying that $s(m', n, h, j | \vec{c}) = s(m', n, h, j | \vec{c}')$. This provides a contradiction.

Considering now the finite horizon model, the logic is very similar. The same base case holds, implying that $s^t(0, n, h, j | \vec{c}) = s^t(0, n, h, j | \vec{c}')$ for all n, h, j, t. By contradiction, we can then again examine the smallest m' such that $s^t(m', n, h, j | \vec{c}) \neq s^t(m', n, h, j | \vec{c}')$. We then consider the lowest t (latest week) among all the states at m' where this inequality holds. We can then redefine

$$
DFBC_n = \left(EHSR_n + \frac{\delta - \delta^{t+1}}{1 - \delta} EHSR_0 \right) \sum_{h=1}^{H} P_h(c_h' - c_h),
$$

which accounts for the finite nature of the problem. The rest of the proof is then analogous to the infinite horizon proof.

We now prove part (b) of the lemma. Here, we use an underlying cost of illness of $c_h = \gamma + \log\left(1 + \sum_{j=1}^{J_h} \exp(\phi_{hj} - \overline{\alpha} \times \overline{p})\right)$ for each h, as in the statement of the lemma. By part (a), the result holds across values of c_h if $\delta < 1$. Analogously to the infinite horizon problem, let $V^t(m, n)$ and $\overline{u}_j^t(m, n, h)$ be the value function and mean utility respectively, for the state where there are t weeks remaining in the year after the current week. Thus, the year, and the decision problem, is over at the end of the $t = 0$ period.

We first show that the choice probabilities are the same inside the doughnut hole for the finite and infinite horizon problems. Formally, we show that for all t, n, h, j, $s^t(0, n, h, j) = s(0, n, h, j)$. For the doughnut hole states, applying the analogs of (5) for the finite- and infinite-time models respectively, implies that

$$\overline{u}_j^t(0, n, h) = \phi_{hj} - \alpha p_{hj} - c_h + \beta V^t(0, n+1), \tag{14}$$

and

$$\overline{u}_j(0, n, h) = \phi_{hj} - \alpha p_{hj} - c_h + \beta V(0, n+1). \tag{15}$$

For both models, the future state is unaffected by the current choice and hence the mean utility functions are identical up to a constant, implying that $s^t(0, n, h, j) = s(0, n, h, j)$. By Assumption 1 (a), individuals inside the doughnut hole will never leave the doughnut hole for the catastrophic region. Moreover, once inside the doughnut hole, after subtracting the baseline health cost c_h, optimizing behavior from (14) or (15) produces expected mean flow utility of 0. This renders the value functions $V^t(0, n) = V(0, n) = 0, \forall n, t$ inside the doughnut hole for both models.

Consider now the mean utility function at states t, m, n:

$$\overline{u}_j^t(m, n, h) = \phi_{hj} - \alpha(p^{eff}(m, p_{hj}, oop_{hj})) - c_h + \beta V^t(\max\{m - p_{hj}, 0\}, n+1) \tag{16}$$

and accompanying value function:

$$V^t(m, n) = \mathbb{1}\{t > 0\}(1 - Q_n)\delta V^{t-1}(m, 0) + Q_n \sum_{h=1}^{H} P_h \sum_{j=0}^{J_h} s^t(m, n, h, j) \times$$
$$\left[\phi_{hj} - \alpha(p^{eff}(m, p_{hj}, oop_{hj}) - c_h) + V^t(\max\{m - p_{hj}, 0\}, n+1) - \log s^t(m, n, h, j) + \gamma\right]. \tag{17}$$

From (16), for it to be the case that $s^t(m, n, h, j) \neq s(m, n, h, j)$, it would have to be the case that $V^t(m, n) \neq V(m, n)$. Suppose then, by contradiction, that there is some t, m, n, h, j, where $m > 0$ and $V^t(m, n) \neq V(m, n)$. Consider the smallest $m' > 0$ such that the equality does not hold. It follows that $V^t(m, n) = V(m, n)$ for all $m < m'$ and for all t, n. Let t' be a period such that the equality does not hold at m'. Because m' is the lowest value where the future value functions of each model are not equal, the finite horizon equations (16) and (17) are similar to their infinite horizon analogs (5) and (7) respectively,

except for the term $\mathbb{1}\{t' > 0\}(1 - Q_n)\delta V^{t'-1}(m', 0)$ in (17). This term occurs because, when $t' = 0$ in the finite horizon model, the decision process ends. But, Assumption 1 (b) implies that the expected probability of ending the year without reaching the doughnut hole is 0 with any single change in behavior. Substituting (17) into (16), if we consider the case of $t' = 0$, then the only way that the assumption is met is when $m' = 0$. But at $m' = 0$, $V(m', 0) = 0$, eliminating the term which was different between the two models. Hence, the system of equations at m' is identical to its infinite horizon analog for all n. This then implies that the solution to the infinite horizon system is also a solution to this system, which generates a contradiction. Furthermore, since the mean utility and value functions are the same, $s^t(m, n, h, j) = s(m, n, h, j)$.

We omit the proofs for the salience model and the quasi-hyperbolic discounting model for naïfs as they are similar. $\square$

Proof of Proposition 1 Our proof uses Lemma 1 (b) and hence uses the infinite horizon problem modeled in Section 2.3. For further tractability, we analyze a model with exactly one health shock per week. A simple manipulation of (8) 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 also 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. Note also that our assumption on the c_h values implies that the expected flow utility from optimizing behavior when inside the doughnut hole is equal to the illness cost, which renders the value function equal to 0 inside the doughnut hole.

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

$$V^{geo}(m, 1) = \overline{\alpha}\left(\left\lfloor \frac{m}{\overline{p}} \right\rfloor (\overline{p} - \overline{oop}) + \max\left\{0, m\%\overline{p} - \overline{oop}\right\}\right), \tag{18}$$

where "$\lfloor \cdot \rfloor$" is the floor function and "$\%$" is the remainder function. In (18), $V^{geo}(m, 1)$ 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 benchmark agents away from their optimal decisions. On the first $\left\lfloor \frac{m}{\overline{p}} \right\rfloor$ drugs, the enrollee receives a subsidy of $(\overline{p} - \overline{oop})$ 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, $V^{geo}(0, 1) = 0$.

We now verify our claim that V^{geo} satisfies (18) by showing that the Bellman operator,

$\mathcal{T}(V^{geo})$, defined here by:

$$\mathcal{T}(V^{geo})(m,1) = \sum_{h=1}^{H} P_h \Big(\gamma +$$

$$\log \left[\exp(V^{geo}(m,1)) + \sum_{j=1}^{J_h} \exp\left(\phi_{hj} - \overline{\alpha}p^{eff}(m,\overline{p},\overline{oop}) + V^{geo}(\max\{m-\overline{p},0\},1))\right)\right] - c_h\Big),$$

$$(19)$$

will have as its value a function equal to (18) when its argument is the same.

We divide our analysis of (19) into three sets of ex ante states. First, we consider all states $V^{geo}(m,1)$, $0 \leq m \leq \overline{oop}$, 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, $p^{eff}(m,p_{hj},oop_{hj}) = \overline{p}$. Substituting these values into (19), we obtain:

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

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

Second, we consider all states $V^{geo}(m,1)$ with $\overline{oop} < m \leq \overline{p}$. Here, the remaining insurance amount is $m - \overline{oop}$, $V^{geo}(m,1) = \overline{\alpha}(m-\overline{oop})$, and the current out-of-pocket payment is $\overline{p} - m + \overline{oop}$. 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 (19), we obtain:

$$\mathcal{T}(V^{geo})(m,1) =$$

$$\sum_{h=1}^{H} P_h \left(\gamma + \log\left[\exp(\overline{\alpha}(m-\overline{oop})) + \sum_{j=1}^{J_h} \exp\left(\phi_{hj} - \overline{\alpha}(\overline{p} - m + \overline{oop})\right)\right] - c_h\right)$$

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

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

Finally, we consider all states $V^{geo}(m,1)$ with $m > \overline{p}$. The remaining insurance amount is $\iota \equiv \overline{\alpha} \left\lfloor \frac{m}{\overline{p}} \right\rfloor (\overline{p} - \overline{oop}) + \max\{0, m\%\overline{p} - \overline{oop}\}$, $V^{geo}(m,1) = \overline{\alpha}\iota$, and the current out-of-pocket payment is $\overline{oop}$. Thus, the future value is $\overline{\alpha}(\iota - \overline{p} + \overline{oop})$ upon choosing an inside option; it

remains $\overline{\alpha}\iota$ with the outside option choice. Substituting these values into (19), we obtain:

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

if $m > \overline{p}$.

Thus, for all cases, $\mathcal{T}(V^{geo}) = V^{geo}$, where V^{geo} is defined using (18). Applying the standard contraction mapping approach to dynamic programming theory (Stokey et al., 1989), (18) is accurate. Note that the mean utility function from (21) specializes to:

$$\overline{u}_j(m,1,h) = \phi_{hj} - \overline{\alpha}p^{eff}(m,\overline{p},\overline{oop}) - c_h + V^{geo}\left(\max\{m - \overline{p}, 0\}, 1\right), \tag{20}$$

for all $j = 1, \ldots, J_h$ and all h. Similarly, $\overline{u}_0(m,1,h) = -c_h + V^{geo}(m,1)$ for all h. Substituting from (18), $\overline{u}_j(m,1,h) - \overline{u}_0(m,1,h) = \phi_{hj} - \overline{\alpha} \times \overline{p}$ for all $j = 1, \ldots, J_h$ and each of the three cases considered above. Thus, $s(m,1,h,j) = s(m',1,h,j)$ for all m, m', h, j. $\qquad \square$

Proof of Proposition 2 Our proof again uses Lemma 1 (b). We also again 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 (7) 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 one shock assumption. Thus, our results are again valid without this additional assumption.

We first prove part (a) of the proposition. Across the three models, $\overline{u}_j$ deviates from the benchmark 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 (8), naïfs believe that they will act as benchmark optimizers from next period on. Thus, specializing to our case, the naïf enrollee will have:

$$\overline{u}_j(m,1,h) = \phi_{hj} - \overline{\alpha}p^{eff}(m,\overline{p},\overline{oop}) + \beta V^{geo}\left(\max\{m - \overline{p}, 0\}, 1\right) - c_h, \tag{21}$$

for all $j = 0, \ldots, J_h$ and all h and where V^{geo} is defined in (18). Applying (18), $\overline{u}_j(m,1,h) - \overline{u}_0(m,1,h) = \phi_{hj} - \overline{\alpha}(\overline{oop} + \beta(\overline{p} - \overline{oop}))$ for all $m, m' \geq \overline{p}$, h, and $j = 1, \ldots, J_h$. This implies that $s(m,1,h,j) = s(m',1,h,j)$ for all h, j and $m, m' \geq \overline{p}$. Since $\overline{oop} + \beta(\overline{p} - \overline{oop}) < \overline{p}$ for $\beta < 1$, $s(m,1,h,j) > s(0,1,h,j)$ for $m \geq \overline{p}$, h, and $j = 1, \ldots, J_h$.

We now prove part (b) for the quasi-hyperbolic discounting sophisticates and price salience model. For ease of notation, let $\tilde{p}(m) = p^{eff}(m, \bar{p}, \overline{oop})$, $\tilde{m}(m) = \max\{m - \bar{p}, 0\}$, $\mathcal{V}(m, 1) = V(m, 1) - V(\tilde{m}(m), 1)$, and $x_{hj}(m) = \exp(\phi_{hj} - \bar{\alpha}\tilde{p}(m))$, for all h and $j = 1, \ldots, J_h$. 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 $\bar{u}_j(m, 1, h) = \phi_{hj} - \bar{\alpha}\tilde{p}(m) - c_h + \beta V(\tilde{m}(m), 1)$, for all h and $j = 1, \ldots J_h$, and $\bar{u}_0(m, 1, h) = -c_h + \beta V(m, 1)$, for all h. Thus, for all h and $j = 1, \ldots J_h$,

$$s(m, 1, h, j) = \frac{\exp(\phi_{hj} - \bar{\alpha}\tilde{p}(m))}{\exp(\beta(V(m, 1) - V(\tilde{m}(m), 1))) + \sum_{k=1}^{J_h} \exp(\phi_{hk} - \bar{\alpha}\tilde{p}(m))}$$
$$= \frac{x_{hj}(m)}{\exp(\beta\mathcal{V}(m, 1)) + \sum_{k=1}^{J_h} x_{hk}(m)}. \quad (22)$$

Similarly, for all h,

$$s(m, 1, h, 0) = \frac{\exp(\beta\mathcal{V}(m, 1))}{\exp(\beta\mathcal{V}(m, 1)) + \sum_{k=1}^{J_h} x_{hk}(m)}. \quad (23)$$

Now, specializing the value function (7) to the case of the proposition and separating out the

outside option, we can write:

$$V(m,1) = \sum_{h=1}^{H} P_h \Big(\gamma - c_h + s(m,1,h,0) \left[V(m,1) - \log\left(s(m,1,h,0)\right) \right]$$

$$+ \sum_{j=1}^{J_h} s(m,1,h,j) \left[\phi_{hj} - \overline{\alpha}\tilde{p}(m) + V(\tilde{m}(m),1) - \log\left(s(m,1,h,j)\right) \right] \Big)$$

$$\Rightarrow V(m,1) - V(\tilde{m}(m),1) = \sum_{h=1}^{H} P_h \Big(\gamma - c_h + s(m,1,h,0) \left[V(m,1) - V(\tilde{m}(m),1) - \log\left(s(m,1,h,0)\right) \right]$$

$$+ \sum_{j=1}^{J_h} s(m,1,h,j) \left[\phi_{hj} - \overline{\alpha}\tilde{p}(m) - \log\left(s(m,1,h,j)\right) \right] \Big),$$

$$\Rightarrow \mathcal{V}(m,1) =$$

$$\sum_{h=1}^{H} P_h \Big(\gamma - c_h + \frac{\exp(\beta\mathcal{V}(m,1))}{\exp(\beta\mathcal{V}(m,1)) + \sum_{k=1}^{J_h} x_{hk}(m)} \left[\mathcal{V}(m,1) - \log\left(\frac{\exp(\beta\mathcal{V}(m,1))}{\exp(\beta\mathcal{V}(m,1)) + \sum_{k=1}^{J_h} x_{hk}(m)} \right) \right]$$

$$+ \sum_{j=1}^{J_h} \frac{x_{hj}(m)}{\exp(\beta\mathcal{V}(m,1)) + \sum_{k=1}^{J_h} x_{hk}(m)} \left[\phi_{hj} - \overline{\alpha}\tilde{p}(m) - \log\left(\frac{x_{hj}(m)}{\exp(\beta\mathcal{V}(m,1)) + \sum_{k=1}^{J_h} x_{hk}(m)} \right) \right] \Big),$$

$$(24)$$

where the second expression subtracts $V(\tilde{m}(m),1)$ from both sides, and the third expression substitutes from (22) and (23).

Importantly, the last expression in (23) implicitly defines the function $\mathcal{V}(m,1)$ for $m > 0$.[46] Note that, for $m, m' \geq \overline{p}$ and all h, j, $\tilde{p}(m) = \tilde{p}(m')$, $\tilde{m}(m) = \tilde{m}(m')$, and $x_{hj}(m) = x_{hj}(m')$ so $\mathcal{V}(m,1) = \mathcal{V}(m',1)$. Applying (22) and (23), $s(m,1,h,j) = s(m',1,h,j)$ for these cases.

To sign the change in purchase probabilities between the two regions, let $\mathcal{V}^{geo}(m,1) = V^{geo}(m,1) - V^{geo}(\tilde{m}(m),1)$, analogously to $\mathcal{V}(m,1)$. Then, $\mathcal{V}(m,1) < \mathcal{V}^{geo}(m,1)$ for $m \geq \overline{p}$ since V^{geo} represents the value with optimal behavior while V represents the value with suboptimal behavior, both from the point of view of the same (benchmark) agent. From (22), choice probabilities under the benchmark model are equivalent to probabilities under the two behavioral models with the substitution of $\mathcal{V}^{geo}(m,1)$ for $\beta\mathcal{V}(m,1)$. Combining this with (i) the fact that a decrease in $\beta\mathcal{V}$ increases all inside good choice probabilities, (ii) the result from Proposition 1 that the benchmark 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 benchmark model inside the doughnut hole, we find that $s(m,1,h,j) > s(0,1,h,j)$ for $m \geq \overline{p}$ and for all h, j.

We now prove part (c). Because the last expression in (23), which implicitly defines $\mathcal{V}$, is

---

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

identical for the sophisticates quasi-hyperbolic discounting model and for the price salience model if $m \geq \bar{p}$, $s(m, 1, h, j)$ 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) choice probabilities for the naïfs model use an expression identical to (22) except for the substitution of $\mathcal{V}^{geo}$ for $\mathcal{V}$, (ii) $\mathcal{V}(m,1) < \mathcal{V}^{geo}(m,1)$ for $m \geq \bar{p}$, and (iii) a decrease in $\beta\mathcal{V}$ increases all inside good choice probabilities, $s(m, 1, h, j)$ will be lower for the naïfs model than for the other models if $m \geq \bar{p}$ and for all h and $j = 1, \ldots, J_h$. $\qquad\square$

Proof of Proposition 3 We prove this proposition by first normalizing the baseline health cost (for the unique health shock type that we consider) to be $c_1 = \gamma - \log s(0, 1, 1, 0)$ and then expressing the structural parameters as a function of the choice probabilities given this normalization. Since $\delta < 1$, by Lemma 1 (a), the choice probabilities are invariant to c_1. Thus, the same formulas hold for any value of c_1, implying that the model is identified without imposing this normalization.

As in Propositions 1 and 2, with this normalization, c_1 is the expected value of optimizing behavior for one purchase occasion at $m = 0$, which results in $V(0,1) = 0$ and $\bar{u}_0(0,1,1) = -\gamma + \log s(0,1,1,0)$. This in turn implies that, for $j = 1, \ldots, J$,

$$\log\left(\frac{s(0,1,1,j)}{s(0,1,1,0)}\right) = \bar{u}_j(0,1,1) - \bar{u}_0(0,1,1) = \phi_{1,j} - \bar{\alpha}p_{1,j}, \tag{25}$$

and

$$\bar{u}_j(0,1,1) = \log\left(\frac{s(0,1,1,j)}{s(0,1,1,0)}\right) + \log s(0,1,1,0) - \gamma = \log s(0,1,1,j) - \gamma. \tag{26}$$

Note also that for the sophisticate quasi-hyperbolic discounting model, we can write the mean utility for the outside option as:

$$\bar{u}_0(m,1,1) = -c_1 + \beta\delta V(m,1) \Rightarrow V(m,1) = \frac{\bar{u}_0(m,1,1) + c_1}{\beta\delta}. \tag{27}$$

The other models have analogous expressions to (27): for the price salience model, σ substitutes for β and for the quasi-hyperbolic discounting naïfs model, V^{geo} substitutes for V.

We now prove identification for the sophisticates model for a given $\beta > 0$ (which includes the geometric discounting model). Consider state m' as defined in the statement of the proposition. At this state, there is no insurance value from buying drug l and hence $\bar{u}_l(m',1,1) = \bar{u}_l(0,1,1)$. Using this exclusion restriction and substituting from (26),

$$\log\left(\frac{s(m',1,1,j)}{s(m',1,1,l)}\right) = \bar{u}_j(m',1,1) - \bar{u}_l(m',1,1), \forall j = 0, \ldots, J$$
$$\Rightarrow \bar{u}_j(m',1,1) = \log\left(\frac{s(m',1,1,j)}{s(m',1,1,l)}\right) + \bar{u}_l(0,1,1) = \log\left(\frac{s(m',1,1,j)s(0,1,1,l)}{s(m',1,1,l)}\right) - \gamma, \tag{28}$$

implying that $\bar{u}_j(m',1,1)$ is known for each $j = 0, \ldots, J$.

This then allows us to identify δ given some $\beta > 0$. Specifically:

$$V(m', 1) = s(m', 1, 1, 0)[-c_1 + \delta V(m', 1)] + \sum_{j=1}^{J} s(m', 1, 1, j)\overline{u}_j(m', 1, 1)$$

$$- \sum_{j=0}^{J} s(m', 1, 1, j) \log(s(m', 1, 1, j)) + \gamma$$

$$\Rightarrow V(m', 1) =$$

$$\frac{\sum_{j=1}^{J} s(m', 1, 1, j)\overline{u}_j(m', 1, 1) - \sum_{j=0}^{J} s(m', 1, 1, j) \log(s(m', 1, 1, j)) - c_1 s(m', 1, 1, 0) + \gamma}{1 - \delta s(m', 1, 1, 0)}$$

$$\Rightarrow \frac{\overline{u}_0(m', 1, 1) + c_1}{\beta \delta} =$$

$$\frac{\sum_{j=1}^{J} s(m', 1, 1, j)\overline{u}_j(m', 1, 1) - \sum_{j=0}^{J} s(m', 1, 1, j) \log(s(m', 1, 1, j)) - c_1 s(m', 1, 1, 0) + \gamma}{1 - \delta s(m', 1, 1, 0)}$$

$$\Rightarrow \delta = (\overline{u}_0(m', 1, 1) + c_1) \times$$

$$\left(\beta \left[\sum_{j=1}^{J} s(m', 1, 1, j)\overline{u}_j(m', 1, 1) - \sum_{j=0}^{J} s(m', 1, 1, j) \log(s(m', 1, 1, j)) - c_1 s(m', 1, 1, 0) + \gamma \right] \right.$$

$$\left. + s(m', 1, 1, 0)(\overline{u}_0(m', 1, 1) + c_1) \right)^{-1}$$

$$\Rightarrow \delta = \log \left(\frac{s(m', 1, 1, 0)s(0, 1, 1, l)}{s(m', 1, 1, l)s(0, 1, 1, 0)} \right) \times$$

$$\left(\beta \left[\sum_{j=1}^{J} s(m', 1, 1, j) \log \left(\frac{s(m', 1, 1, j)s(0, 1, 1, l)}{s(m', 1, 1, l)} \right) - \sum_{j=0}^{J} s(m', 1, 1, j) \log(s(m', 1, 1, j)) \right.\right.$$

$$\left.\left. + s(m', 1, 1, 0) \log(s(0, 1, 1, 0)) \right] + s(m', 1, 1, 0) \log \left(\frac{s(m', 1, 1, 0)s(0, 1, 1, l)}{s(m', 1, 1, l)s(0, 1, 1, 0)} \right) \right)^{-1}. \quad (29)$$

The first line of (29) follows because there is no remaining insurance value for the inside choices after the current choice. The second line follows by rearranging terms. The third line follows by substituting for $V(m', 1)$ from (27), using the fact that $\beta, \delta > 0$. The fourth line follows by rearranging terms, and requires that the denominator be non-zero. With $\beta = 1$, this denominator equals $V(m', 1)$. For $0 < \beta < 1$ it lies somewhere between $\beta V(m', 1)$ and $V(m', 1)$. Since $V(0, 1) = 0$, this term would be equal to zero if m' were equal to 0. However, since m' is sufficiently high that there is insurance value from drug k, $V(m', 1) > 0$ implying that the denominator is positive. Finally, the last line substitutes choice probabilities for the $\overline{u}$ terms (from equation 28) and for c_1 from its normalized value. Thus, (29) defines δ as a

function of choice probabilities conditional on β.

Similarly, we can define $\overline{\alpha}$ as a function of choice probabilities. Specifically,

$$\overline{u}_k(m', 1, 1) = \overline{u}_k(0, 1, 1) + \overline{\alpha}(m' - oop_k)$$
$$\Rightarrow \overline{\alpha} = \frac{\overline{u}_k(m', 1, 1) - \overline{u}_k(0, 1, 1)}{m' - oop_k}$$
$$\Rightarrow \overline{\alpha} = \log\left(\frac{s(m', 1, 1, k)s(0, 1, 1, l)}{s(m', 1, 1, l)s(0, 1, 1, k)}\right)/(m' - oop_k). \quad (30)$$

The second line rearranges terms and the third line substitutes in choice probabilities for the $\overline{u}$ terms. Since (30) expresses $\overline{\alpha}$ in terms of shares, it shows that $\overline{\alpha}$ is identified. This then allows us to identify ϕ_j using (25) for $j = 1, \ldots, J$, implying that the sophisticates model is identified, given $\beta > 0$.

We now prove identification of β for the quasi-hyperbolic discounting sophisticates model, under a rank condition. Consider the analog of (29) but using m'' (as defined in the statement of the proposition) instead of m'. Since both equations equal δ, we can set them equal to

each other, obtaining:

$$(\overline{u}_0(m',1,1) + c_1) \times$$

$$\left(\beta \left[\sum_{j=1}^{J} s(m',1,1,j)\overline{u}_j(m',1,1) - \sum_{j=0}^{J} s(m',1,1,j)\log(s(m',1,1,j)) - c_1 s(m'1,1,0) + \gamma \right] \right.$$

$$\left. + s(m',1,1,0)(\overline{u}_0(m',1,1) + c_1) \right)^{-1}$$

$$= (\overline{u}_0(m'',1,1) + c_1) \times$$

$$\left(\beta \left[\sum_{j=1}^{J} s(m'',1,1,j)\overline{u}_j(m'',1,1) - \sum_{j=0}^{J} s(m'',1,1,j)\log(s(m'',1,1,j)) - c_1 s(m'',1,1,0) + \gamma \right] \right.$$

$$\left. + s(m'',1,1,0)(\overline{u}_0(m'',1,1) + c_1) \right)^{-1}$$

$$\Rightarrow \beta = \left[(\overline{u}_0(m'1,1,) + c_1)(s(m'',1,1,0)(\overline{u}_0(m'',1,1) + c_1) \right.$$

$$\left. - (\overline{u}_0(m'',1,1) + c_1)(s(m',1,1,0)(\overline{u}_0(m',1,1) + c_1) \right] \times$$

$$\left((\overline{u}_0(m'',1,1)+c_1)\left[\sum_{j=1}^{J} s(m',1,1,j)\overline{u}_j(m',1,1) - \sum_{j=0}^{J} s(m',1,1,j)\log(s(m',1,1,j)) - c_1 s(m',1,1,0) + \gamma \right] \right.$$

$$\left. -(\overline{u}_0(m',1,1)+c_1)\left[\sum_{j=1}^{J} s(m'',1,1,j)\overline{u}_j(m'',1,1) - \sum_{j=0}^{J} s(m'',1,1,j)\log(s(m'',1,1,j)) - c_1 s(m'',1,1,0) + \gamma \right] \right)^{-1}$$

$$\Rightarrow \beta = \left[\log\left(\frac{s(m',1,1,0)s(0,1,1,l)}{s(m',1,1,l)s(0,1,1,0)} \right) s(m''1,1,0) \log\left(\frac{s(m'',1,1,0)s(0,1,1,l)}{s(m'',1,1,l)s(0,1,1,0)} \right) \right.$$

$$\left. - \log\left(\frac{s(m'',1,1,0)s(0,1,1,l)}{s(m'',1,1,l)s(0,1,1,0)} \right) s(m',1,1,0) \log\left(\frac{s(m',1,1,0)s(0,1,1,l)}{s(m',1,1,l)s(0,1,1,0)} \right) \right] \times$$

$$\left(\log\left(\frac{s(m'',1,1,0)s(0,1,1,l)}{s(m'',1,1,l)s(0,1,1,0)} \right) \left[\sum_{j=1}^{J} s(m',1,1,j)\log\left(\frac{s(m',1,1,j)s(0,1,1,l)}{s(m',1,1,l)} \right) \right. \right.$$

$$\left. - \sum_{j=0}^{J} s(m',1,1,j)\log(s(m',1,1,j)) + s(m',1,1,0)\log(s(0,1,1,0)) \right]$$

$$- \log\left(\frac{s(m',1,1,0)s(0,1,1,l)}{s(m',1,1,l)s(0,1,1,0)} \right) \left[\sum_{j=1}^{J} s(m'',1,1,j)\log\left(\frac{s(m'',1,1,j)s(0,1,1,l)}{s(m'',1,1,l)} \right) \right.$$

$$\left. \left. - \sum_{j=0}^{J} s(m'',1,1,j)\log(s(m'',1,1,j)) + s(m'',1,1,0)\log(s(0,1,1,0)) \right] \right)^{-1}.$$

$$\tag{31}$$

68

Thus, provided that the denominator term in (31) is not 0, β is identified, further implying that all the parameters are identified under this rank condition.

We briefly consider identification for the price salience case. This model is slightly different than the quasi-hyperbolic discounting model because σ enters into the inside option values at state m', while β did not. Because of this, we employ a different proof order: conditional on σ, we first show that $\overline{\alpha}$ is identified and then that δ is identified. Note that:

$$\log\left(\frac{s(m',1,1,l)}{s(m',1,1,k)}\frac{s(0,1,1,k)}{s(0,1,1,l)}\right) = \overline{\alpha}(1-\sigma)(p_l - oop_l) - \overline{\alpha}(1-\sigma)(p_k - oop_k) - \overline{\alpha}\sigma(m' - oop_k),$$
(32)

where the first term in (32) is the extra utility from the lower expected perceived price for drug l, the second term is the negative of the utility from the lower expected perceived price for drug k, and the final term is the negative of the utility from the lower actual price for drug k. We can solve for $\overline{\alpha}$ here, implying that $\overline{\alpha}$ is identified conditional on σ.

As in the quasi-hyperbolic discounting case, we can then identify $\phi_1, \ldots, \phi_J$ using (25). We can then evaluate $\overline{u}_j(m')$ for all $j = 0, \ldots, J$ using the analog of (28).

An equation analogous to (29) holds, which then allows us to identify δ conditional on σ. Finally, we can solve for σ with an expression analogous to (31), subject to the analogous rank condition.

We omit the proof for the quasi-hyperbolic discounting naïfs model, which is similar to the sophisticates model, but using the value function V^{geo}.

$\square$

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 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.[47] 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 benchmark model if enrollees curtail purchases upon reaching the doughnut hole. A rational reaction to the upcoming presence of the catastrophic region 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 benchmark 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 benchmark model, our finding is conservative.

Fourth, our results are unlikely to be due to cross-year substitution. Cabral (2016) 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.[48] Also, most of our enrollees are in plans with deductibles, and it would be

---

[47]This dataset provides a 10% sample of all Medicare eligibles.

[48]To 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

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, 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 types 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.

---

doughnut hole in 2009, while 19.1% did not (and we were unable to match the remaining 4.7%).