Dynamic Purchasing Behavior in Healthcare Consumption

Lin Xu*

See Latest Draft.

This Draft May 19, 2018

Abstract

Existing empirical evidence from Medicare Part D documents significantly lower spending and suboptimal behavior at the "donut hole", a region of low insurance coverage sandwiched between two regions of higher coverage. Current efforts to explain the drop in drug adherence rely heavily on time-discounting models with strong assumptions that recover far lower rates of discounting than the broader literature would predict. This paper first develops an alternative simple heuristic model with intuitions on how enrollees ought to behave by formulating a perceived marginal out-of-pocket price based on objective probabilities that updates as enrollees accumulate spending through their plans and the year. This approach predicts that with continuously updating expectations, a beneficiary's perceived marginal price and spending should be smooth and occur far in advance of region boundaries or "kinks" early in the year. Further, the paper provides empirical evidence that beneficiaries do anticipate and respond more optimally to the Medicare Part D pricing schedule than previously suggested. It measures actual behavior throughout the entire insurance schedule using dynamic panel regressions that control for individual heterogeneity. Overall, prescription claim frequencies do respond with significant foresight to anticipated changes in coverage generosity.

*Please do not cite or circulate without permission. E-mail: lx82@cornell.edu. I am grateful to my graduate committee – Ted O'Donoghue, Levon Barseghyan, Colleen Carey, and Francesca Molinari – for their direction and support. Thanks to Gregory Besharov, Douglas Miller, Daniel Reeves, and numerous seminar participants at Cornell. For more information about my research and teaching, please visit my website: linxu.info.

1 Introduction

With the rapid growth of the US health care sector and the accompanying financial burden it poses, government and insurers have responded by experimenting with insurance plan characteristics and cost-sharing aimed at reducing their costs. Policy makers and academics have long had an interest in understanding how people's healthcare spending and health outcomes respond under these plans. However, it is only recently that economic research has begun to emphasize the extent of the dynamics and complexity of beneficiaries' decisions when facing multiple levels of cost-sharing within health insurance plans. This paper studies beneficiary behavior in the context of the Medicare Part D prescription drug insurance market in 2009-2012, where the government has had significant regulatory oversight on imposing plans with these features.

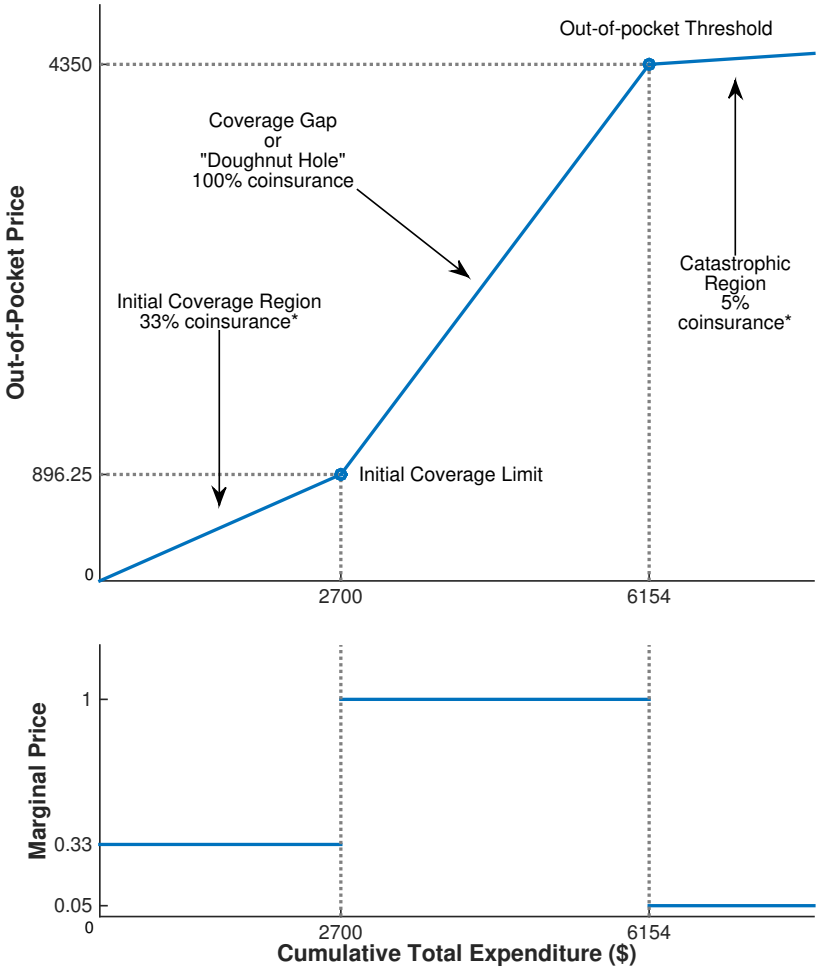
I focus on beneficiary responses to plans where the marginal out-of-pocket (OOP) price that the beneficiary is required to pay out of the total cost of a prescription is not constant (or non-linear) between coverage regions. Figure 1 illustrates an example Medicare Part D 2009 plan contract. The generosity of the plan explicitly depends on the cumulative amounts that patients, the insurance company, and Medicare have spent on prescriptions within the plan-year. In this example, there is a region of initial coverage (ICR) where the beneficiary is responsible for 33.2% of the total prescription costs (33.2% coinsurance), followed by a coverage gap or "donut hole" where beneficiaries are fully responsible for costs (100% coinsurance), after which beneficiaries reach the catastrophic region and have a minimal 5% coinsurance. See Section 3 for more information on the types of plans.[1] After the plan-year elapses, plans reset or beneficiaries switch to new plans, and the cumulative total spending for the new plan-year is reset to zero. Within a plan-year, beneficiaries face a truly dynamic problem—consuming healthcare today can impact the marginal price of future healthcare consumption.

It is important then to study the actual beneficiary response to these types of nonlinear pricing structures even before the actual nonlinearities. The coverage gap was initially included as a cost-saving measure for the government similar to a deductible but positioned in the middle of the patient benefit schedule (Baker, 2006). In addition to directly saving money in the donut hole, the presence of the donut hole mechanism was meant as an incentive to beneficiaries to restrain their spending even prior to the donut hole and avoid the catastrophic spending regions altogether.

Much of the literature has focused on beneficiary behavior at or around the controversial coverage gap, which has been shown to have a negative effect on drug adherence. Joyce et al.

[1] In general plans are structured similar to this example with either three or four regions.

Figure 1: Example Medicare Part D 2009 Contract Design



Note: The figure depicts the nonlinear structure of an example Medicare Part D benefit contract with no deductible from 2009. The plan depicted here is actuarially equivalent (with no deductible) to the government-defined contract depicted in Appendix Figure A.1. The premium or the amount the patient pays out-of-pocket for the benefit package is not displayed. The Cumulative Total Expenditure is the year-to-date cumulative sum of the beneficiary's total expenditures, which include the drug expenditure between the patient, insurance company, and Medicare. The Out-of-Pocket (OOP) Cost only includes the patient's drug expenditure in 2009. The initial coverage region's (ICR's) 33.2% coinsurance coverage is approximately the actuarially equivalent to the 2009 deductible plus an ICR coinsurance level of 25%. The 5% coinsurance coverage in the catastrophic region is also simplified for the figure. The actual 2009 coverage benefit requires beneficiaries to pay the maximum of either 5% the cost of the drug or $2.40 and $6.00 for a one-month supply of generic and branded drugs respectively. This means that patients may pay either the copay dollar amount or a percentage share of the drug price where the remainder is covered by insurance or the government. The bottom panel displays the marginal cost in each of the coverage regions, or the proportion that the beneficiary is responsible to pay of the total expenditure cost.

(2013) and Zhang et al. (2009) find that having a coverage gap disrupts the use of prescription drugs, with a higher decline on more expensive medications as compared to cheaper ones. And while Joyce et al. (2013) fail to detect a corresponding substitution from drugs to medical treatment in concurrent Medicare claims, if one believes that adherence to drug treatments is

good for patient outcomes, discontinuing the use of these drugs would have a negative welfare effect on patients.

In order to fully understand the ways in which changes to nonlinear plans can impact beneficiary behavior and welfare, researchers must also determine whether beneficiaries respond sub-optimally to them. The current consensus in the literature is somewhat mixed. Some of the literature cites the initial lack of knowledge of the donut hole where in 2007, only 40% of individuals knew about the coverage gap (Polinski et al., 2010). There are papers where beneficiaries appear to be somewhat optimal and fully forward-looking (Aron-Dine et al., 2015; Einav et al., 2015), and there are papers that support the conclusion that beneficiaries overly respond to the "spot" or current price (Dalton et al., 2015; Abaluck et al., 2015). As evidence of suboptimal behavior, Dalton et al. (2015) also present the stylized fact that even among individuals in their sample who were very likely to end the year in the coverage gap or beyond, there is a sharp drop in average spending on prescription purchases at the coverage gap.

Ignoring liquidity constraints, fully forward-looking optimal behavior suggests that beneficiaries should use their expected end-of-year marginal price for each purchase decision throughout the year. For example, a Medicare Part D beneficiary who fully expects to end the year in the catastrophic phase of their insurance coverage should not respond to temporary changes in their plan coverage and spot prices as they spend through earlier benefit phases. Under uncertainty about the end-of-year region and price, however, the beneficiaries may adjust their expected marginal prices as risks are realized. Assuming standard geometric discounting ($\delta >> 0$), these transitions should be smooth, especially early in the year.

In Section 2, because of the beneficiary's complicated optimization problem under uncertainty, this paper first approximates optimal behavior by developing a heuristic for constructing a beneficiary's perceived marginal out-of-pocket coinsurance rate. This rate is generated from the entire population's probabilities of ending the year in each region and is meant to help visualize how the average beneficiary's expected coinsurance rate may evolve across the different weeks of the year and cumulative spending levels. It differs from the optimal beneficiary's expected prices because it uses the ex-post population outcomes, which may not be representative of rational agents. Further, because beneficiaries may themselves have difficulties anticipating their cumulative total end-of-year spending and prices, the heuristic approach may provide intuitions on behavior as well. The heuristic approach predicts that, because probability distributions are quite smooth outside of the last weeks of the year, the expected marginal prices and thus spending should also be smooth with any pricing updating occurring prior to the kink.

Further, intuitions from this paper's heuristic approach indicate that errors in time discounting such as present bias may not be an appropriate model to represent beneficiary behavior that resemble a sharp spending drop at discrete changes in spot prices. A key takeaway from the heuristic model is that, outside of the last weeks of the year, only sharp discontinuities in a beneficiary's perceived marginal price or coinsurance rate should result in sharp discontinuities in beneficiary spending. As long as beneficiaries update their expectations weekly, only zero discounting would predict sharp changes at the coverage gap.

Section 3 provides background information on the structure of the Medicare health insurance program and details on the prescription insurance program, Medicare Part D. The section also discusses the data and sample. The data studied comes from the Center for Medicare and Medicaid Services' (CMS's) 5% sample of all beneficiaries in Medicare, with administrative information on enrollee selected plans and claims. The sample selected include enrollees, who are continuously in Medicare Part D plans without deductibles over the 2009-2012 period and who meet criteria to make sure they do not have unique pricing or non-optional choices. Then, the heuristic from Section 2 is applied to the empirical data in Section 4, supporting the intuitions developed.

In Section 5, this paper uses a flexible regression approach to provide a graphic representation of the empirically observed shape of spending patterns through the entirety of a beneficiary's plan year and across a wide range of cumulative total spending levels. In part, the goal of this empirical exercise is to generate a low-assumption view of the extent of a beneficiary's anticipatory response to the different pricing regions as a function of both the time of the year and the beneficiary's cumulative total spending. The regression approach uses an individual fixed effect to control for individual heterogeneity of spending frequencies in a dynamic panel that includes four years of claims data.[2]

Unlike the majority of the literature, this paper mainly focuses on the frequency of beneficiary claims. Suppose a beneficiary has a prescription to be filled, they have two broad decision options: wait to fill the prescription, or switch the prescription either from branded to generic or with more effort acquire an alternative prescription. The decision to stop taking a course of chronic treatment entirely likely has more of a negative effect on beneficiary health than a decision to switch medication. Thus, this paper measures changes in claims frequency, which should capture the beneficiary's first decision to postpone or stop (postpone indefinitely) taking a course of treatment.

[2]See Section 5.3 for a discussion regarding dynamic panel bias and why it is less of a concern in this setting.

This paper's estimates show that across four quarters of the year and across cumulative total spending amounts, beneficiaries have a statistically significant anticipatory response far in advance of the gap. As expected, the reduction in spending in advance of the coverage gap occurs at higher cumulative total expenditure values (closer to the ICL) in later parts of the year. Unlike the prior literature, this paper does not detect a large discontinuity in beneficiary spending frequencies directly at the coverage gap except in the last quarter of the year, as expected.

Dalton et al. (2015) also employ a fixed-effects-regression. They analyzed data from a proprietary source on a 2008 subset of employer-sponsored Medicare Part D individuals, who were likely to end the year in the coverage gap. Their analysis differs from this paper in that they only had indicators to measure the level of spending response (and other dependent variables including prescription occurrence) in four cumulative total spending zones ($310 before, between $310-$110 before, $110 before, and after the coverage gap) rather than a continuous response to cumulative total and cumulative out-of-pocket spending. They found no economic or statistically significant evidence of spending or claims frequency decreases in $310-$110 leading up to the coverage gap, but a sharp decrease in spending and claims frequency in the $110 right before the coverage gap.

This paper's reduced-form estimates also provide a simpler alternative to studying beneficiary behavior under nonlinear contracts as compared to the structural estimates of Einav et al. (2015) and Dalton et al. (2015). The reduced-form estimates provide a graphical explanation of why these papers found both evidence of forward-looking behavior and over-response to spot prices respectively. Einav et al. (2015) estimate a model with standard geometric discounting that allows for five types of individuals with different risk levels and sensitivities to the coverage gap, but they recover a weekly discount factor[3] δ equal to 0.96, which roughly translates to a yearly discount factor of only 11%. In order to explain observed drops in spending just prior to the coverage gap, Dalton et al. (2015) expand their model to allow for beta-delta time-inconsistency or present bias, the tendency to overweigh the "present" and have self-control problems. The discounting rates they estimate are indistinguishable between $\beta = \delta = 0$ indicating that beneficiaries only consider the spot price.

[3]Einav et al. (2015) refer to the estimate as a "'behavioral' parameter that also reflects individual's understanding of the insurance coverage contract, in particular the salience of the (future) nonlinearities of the contract", as part of the reason why δ is so low.

2 Intuition

In the Medicare Part D setting, patients make a combination of periodic and unexpected purchases of prescription drugs within a year. Within the decision to make each prescription purchase, as with all economic decision-making, a beneficiary employs some cost-benefit analysis comparing her costs with the perceived benefits of purchasing and then consuming drugs. While a beneficiary's beliefs on the medical benefits of consuming certain drugs are important, this paper focuses on the beneficiary's beliefs on his or her monetary costs with insurance. This analysis makes the assumption that the benefits of drug purchase and consumption do not depend directly on the arbitrary contract design or the time of the year.

Suppose, in week w of the year, a beneficiary with observables X faces a decision whether to fill a prescription that has a total cost of s, where the total cost is to be paid between the insurer and beneficiary. The beneficiary chooses to purchase if the perceived benefit of the drug exceeds her perceived cost of the drug under the insurance plan. Certainly a beneficiary's out-of-pocket costs (OOP), or the amount they are responsible to pay, could contribute to her perceived cost, and if she is forward looking, the impact that spending today has on her future costs should also enter into her decision.

Let us generalize

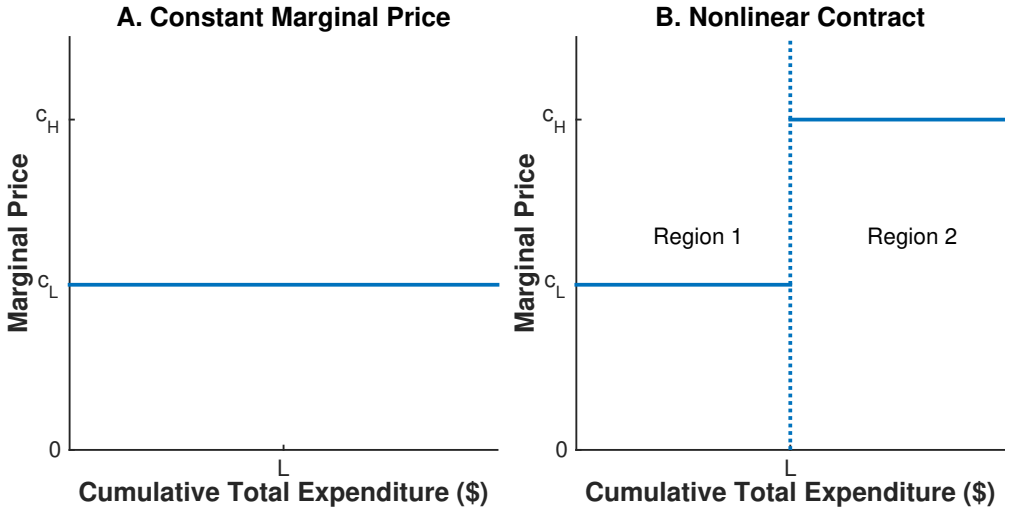
$$B(s|X) > sPMC(Z_w, w|X)$$

where $B(s|X)$ is the benefit of a drug for a unit cost of the drug, $Z_w = \sum_{u=1}^{w-1} s_u$ the cumulative total amount spent up until week w, and $PMC(Z_w, w|X)$ the perceived marginal cost of the drug to the beneficiary who has spent Z_w by week w conditional on their observables X.

To build intuition, consider how the beneficiary's perceived marginal cost may differ under two simple pricing contracts in Panel A and B of Figure 2. If the beneficiary were in a plan with contract A, she should always expect that her marginal cost of purchasing a drug would be equal to c_L, as it is the coinsurance rate applied to all purchases. It doesn't matter if she has to make a purchasing decision in the beginning of the year or the end, her $PMC(Z_w, w|X) = c_L$. Then her out-of-pocket (OOP) payment or the amount that she is responsible for is $c_L s$.

In contrast, Plan B is a nonlinear cost structure, and thus a beneficiary's perceived marginal price for purchasing a drug is less clear. Under this plan, her spot price and actual out-of-pocket payment in a given week would be $MC(Z_w) = c_L$ and $c_L s$ if $Z_w < L$ and $MC(Z_w) = c_H$ and

Figure 2: Marginal Prices



$c_H s$ if $Z_w \geq L$. At the two extremes of behavior, a beneficiary may only respond to the spot marginal price and her out-of-pocket payment for care determined by the current insurance region, or a beneficiary may be a "fully forward-looking", perfectly rational economic agent. If she only responds to the spot price, then her $PMC(Z_w, w|X) = MC(Z_w)$. Her PMC is just her coinsurance rate in the region she is in in week w, and she does not take into consideration how her spending can impact her future marginal costs. If a beneficiary is "fully forward-looking", in each purchasing period decision, she optimizes her decision by discounting the future stream of expected benefits and costs that result from her current decision, including any changes to her expected marginal prices due to the non-linear pricing. Ultimately if a beneficiary is fully forward-looking, she should anticipate her full stream of expected payments, and an additional prescription should incur her expected year-end marginal price. Thus, her expectation of ending in Region 1 or 2 matters. If she is entirely confident in week w that her end-of-year (week W) cumulative total expenditure will be below the limit L, $Pr(Z_W < L|Z_w) = 1$, then her perceived marginal price should always be $PMP_w = c_L$. Similarly if she is confident that her end-of-year cumulative total expenditure is greater than L, $P(Z_W \geq L|Z_w) = 1$, then her end-of-year marginal cost should be the cost in Region 2, or $PMP_w = c_H$. Depending on the beneficiary's perceived uncertainty about spending past L, and conditional on not passing L when evaluating her problem in week w, her perceived marginal price in w may be somewhere between c_L and c_H.

Within the "in-between" response, Dalton et al. (2015) has tested a model of inconsistent time-discounting and salience of the "donut hole" to explain beneficiaries' behaviors in their

data on 2008 Medicare Part D claims. While their structural model that allowed for present bias was a better fit than a standard discounting model, their estimation result of discount factors that were indistinguishable from $\beta = \delta = 0$ is indicative that present bias may not be an appropriate model. Abaluck et al. (2015) also allow for an "in-between" response to inter-year changes in the coinsurance rates by estimating the weights beneficiaries in Medicare Part D place on changes in coinsurance rates in the initial coverage region or coverage gap. While Abaluck et al. (2015)'s heuristic is somewhat similar to mine, in part due to the nature of their empirical approach, they limit their study to individuals who the researchers were confident to end the year in either the initial coverage region or coverage gap.

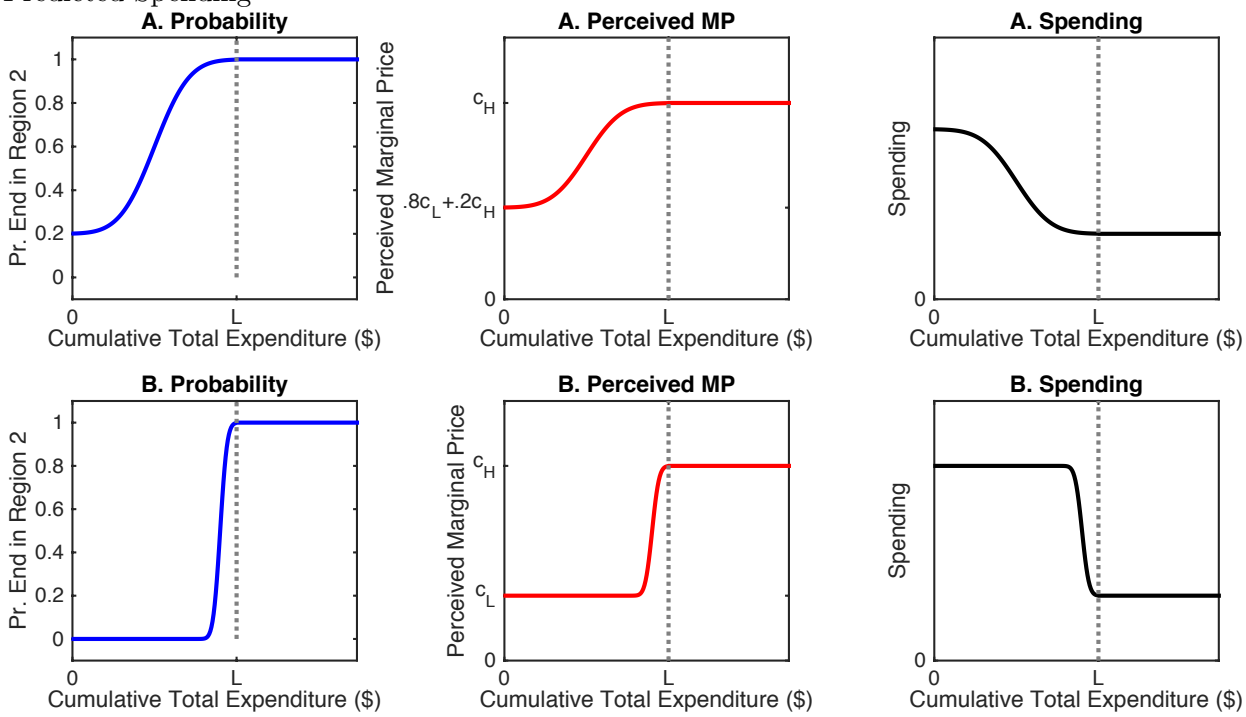
This paper proposes an alternative "in-between" response that a beneficiary may use to make the prescription purchasing decision. She may use a heuristic mental shortcut to calculate her objective expected year-end marginal price. Rather than comparing the net present value of the costs and benefits to conceive of an optimal expected year-end price, she may estimate a perceived marginal price based on her beliefs of the population objective probability of ending the year in any contract region. Using the objective population probability of being in any contract region, she can then infer the price in that region. Note that in the two region case, $Pr_w(\text{in Region 2 in W}|Z_w) = 1 - Pr_w(\text{in Region 1 in W}|Z_w)$.

If a beneficiary were to make a decision to spend on prescriptions in the last period of the year, no-matter her method of evaluating her "in-between" response, her perceived marginal price is just her marginal spot price. There is no uncertainty as to what her marginal cost is. However, at earlier parts of the year $w < W$, her beliefs on her probability of ending the year in any particular region and the associated expected marginal price can impact her spending patterns.

Consider the example probabilities of ending the year in Region 2 presented in Figure 3. In the series of Panel A graphs, suppose it is early in the year, the beneficiary who has \$0 cumulative total spending believes that there is only a 20% chance of ending the year in Region 2 (and necessarily a $1 - 0.2 = 0.8$ chance of ending the year in Region 1). As her cumulative total expenditure approaches the limit $Z_w \to L$, the probability of ending the year in Region 2 approaches 1, $Pr(\text{in Region 2 in W}|Z_w) \to 1$. Even before she reaches $Z_w = L$, because there are still many weeks left in the year, she anticipates an almost 100% chance of ending the year in Region 2.

Suppose she employs the heuristic with no discounting to generate her marginal cost, then

Figure 3: Example Probabilities of Ending the Year in Region 2, Perceived Marginal Price, and Predicted Spending



Note: The example assumes an individual faces marginal costs conditional on the depicted in the top graphs.

her objective expected marginal price (HMP) is below.

$$HMP_w = c_L Pr_w(\text{in Region 1 in W } |Z_w) + c_H Pr_w(\text{in Region 2 in W}|Z_w)$$

Her overall expected marginal price would then be the function depicted in in Figure 3 A. Her perceived marginal price transitions from $0.8c_L+0.2c_H$ to c_H as the cumulative total expenditure increases. If demand is a monotonically decreasing and continuous function of price, this implies that her spending should be a monotonically decreasing function of her perceived marginal price and should also adjust far prior to the limit L.

Suppose Panel B represents a week at the end of the year. Towards the end of the year, the beneficiary's uncertainty around ending the year in Region 2 decreases, because there is less time left in the year for health shocks as compared to Panel A. Any uncertainty only remains if she is in a narrow range of cumulative total expenditure values just prior to L. As she approaches the end of the year $w \to W$, her probability of ending the year in Region 2 approaches a piecewise formula, where this probability is 0 if $Z_w < L$ and 1 otherwise. Similarly, if demand is a monotonically decreasing function of price, then her spending should also approach a piecewise formula. If she has spent very little, she may be certain to end the year in Region 1 and

subsequently uses the c_L rate, while if she has crossed the limit L, she should spend according to the c_H rate, i.e. $s_w = s(c_L)$ if $Z_w < L$ and $s_w = s(c_H)$ otherwise. In aggregate data, it is natural to see how these expectations could generate spending that may appear to be discontinuous at the limit L, especially if there are few individuals observed in the transition region.

2.1 Discounting

While the examples in Figure 3 did not include any explicit discontinuities in the probabilities or expected marginal prices, an alternative scenario could exist. For example, suppose that in a period w early in the year beneficiaries believe that there is a 20% chance of ending the year in Region 2 for all $Z_w < L$ and by necessity the probability is 1 if $Z_w \geq L$. The explanation for why a beneficiary near the limit may not upwardly revise her 20% probability of ending the year in Region 2 may include her inattention to being so close to Region 2, or completely ignoring her potential future health shocks (having a zero discount factor).

This paper argues that non-zero geometric discounting and present bias models should not generate such discontinuities. Consider the adaptation to the heuristic model that allows for present bias. Assume the decision in the heuristic approach only considers today's medical impact (or benefit of the drugs) compared to the total cost today multiplied by the future perceived marginal price. The beneficiary would necessarily be making her spending decisions by discounting her entire heuristic marginal price $PMP_w = \beta\delta^{W-w} * HMP_w$. So she would purchase her prescriptions if

$$B(s|X) > \beta\delta^{W-w} HMP(Z_w, w|X)s$$

where β is the "present bias" discounting factor that represents the difference between the present t and all future outcomes and δ^{W-w} is the geometric or standard discounting factor that is the product of discounting in every week from the current week w to the end of the year W.

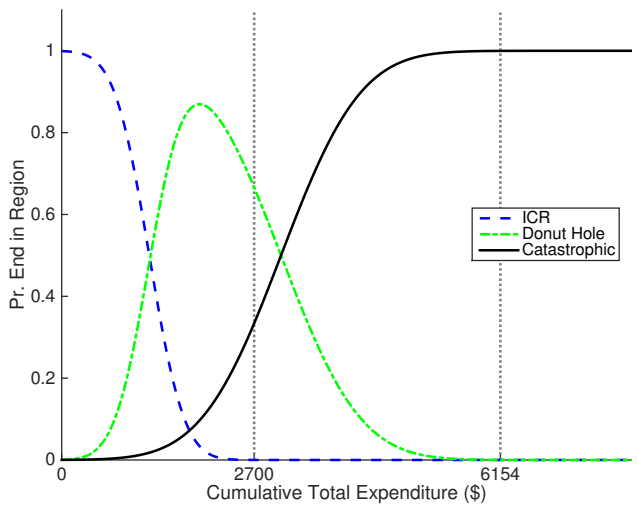
Assume a beneficiary has Panel A probability beliefs that generate a continuous heuristic marginal cost. Denote $\beta'_w = \beta\delta^{W-w}$ for a given week w. Then, introducing a $\beta > 0$ and $\delta > 0$ implies $\beta'_w > 0$ and will not generate a significant discontinuity in her perceived marginal prices. Instead her perceived marginal price would fall somewhere between $\beta'_w(0.8c_L + 0.2c_H)$ and $\beta'_w c_H$ in week w.

Further, introducing present bias as a explanation of suboptimal behavior observed in the

empirical data is not necessarily appropriate. Present bias models of behavior only generate suboptimal behavior when decisions are made between "the present" and "the future" and not when considering different time points within the future. The assumption that the benefits of prescription coverage are incurred "today" and in the present, while prevalent in the literature is not necessarily accurate. When beneficiaries fill prescriptions, they are often not for immediate consumption with 30 and 90 day supplies. Further, even if consumption was immediate, as Baicker et al. (2015) explain, the effects of many prescriptions such as statins to treat high cholesterol have far delayed benefits rather than any immediate symptomatic changes.

2.2 Intuition with three regions

Figure 4: Probabilities

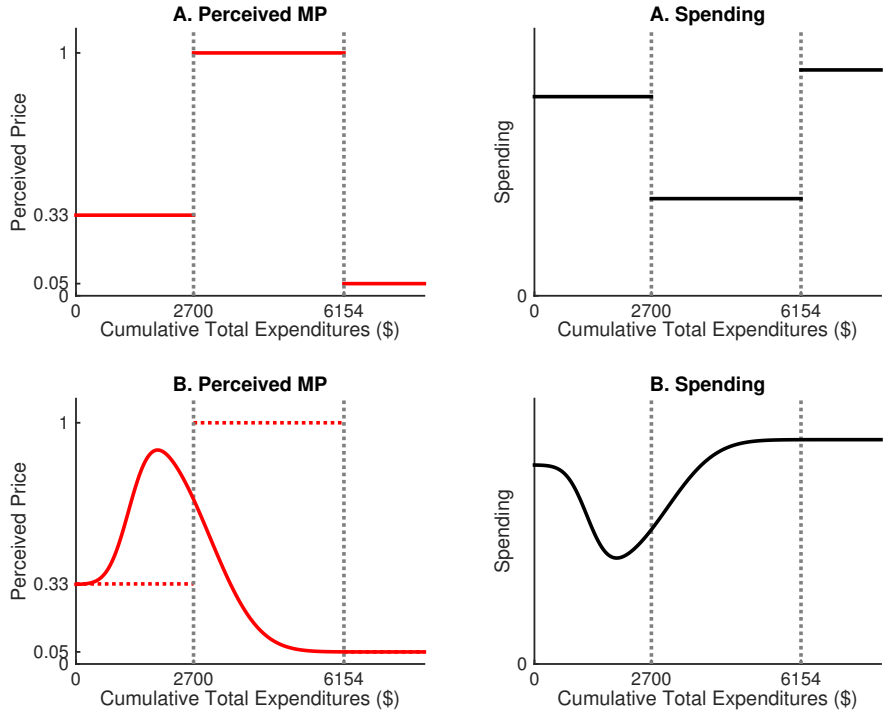


Note: An example graph of the probability of ending the year in any of three regions: initial coverage region (ICR), donut hole, or a catastrophic region. The region limits are based off of the 2009 example Medicare Part D contract from Figure 1. The probabilities are a stylized example.

The three-region setting is slightly more complex than the two-region setting, but the intuitions are similar. Assume a beneficiary has the example Medicare Part D contract and region limits displayed in Figure 1. Figure 4 displays hypothetical probabilities of ending the year in the three regions: the initial coverage region (ICR), the coverage gap (aka the donut hole), and the catastrophic region.

Similar to the probability of ending the year in Region 1 in the 2-region example, the probability of ending the year in the ICR is high at low cumulative spending values and low at high spending values. If it is earlier in the year, the probability of ending the year in the ICR can approach zero far prior to the subject crossing the $2,700 spending threshold. The catastrophic

Figure 5: Three Regions: Perceived Price and Spending Response



Note: Conditional on the probabilities displayed in Figure 4, the A. panels reflect the marginal price and resulting spending from a hypothetical agent who only responds to the spot price. the B. panels reflect the heuristic marginal price of an agent with the above probabilities and the 2009 example Medicare Part D contract prices and a stylized example of their spending.

region is the terminal state and the probability of ending in that state is similar to the probability of ending the year in Region 2 displayed in Figure 3. It is the intermediate state of the donut hole which differs, and the probability of ending the year in the donut hole peaks when the probability of ending the year in the other two regions are at their lowest. As shown in this example, the peak of the probability of ending the year in the donut hole can occur prior to an individual even entering the donut hole depending on their end-of-year expectations.

Figure 5 presents the difference between the response of an agent A who responds only to the spot price in panel A with the response of an agent B who responds using the Heuristic Marginal Price presented in panel B. The subject in panel A has a perceived marginal price that is equal to the Medicare Part D contract, while the panel B agent has a perceived marginal price that is equal to contract prices weighted by the probability of ending the year in each region. Assuming the agent's spending is still a continuous function of her perceived marginal price, then her spending would be discontinuous at the boundaries of the Medicare Part D contract regions, while agent B's spending would change continuously. As the agents approach the end

of the year, the probabilities should change and approach 1 if they are in the region and 0 otherwise, and agent B's perceived marginal price and spending should approach agent A's.

Section 4 applies this intuition to the data, and specifically presents the objective probabilities of ending the year in each of the spending regions and the implied perceived marginal price based on the observed data from Medicare Part D.

3 Background on Medicare Part D and Data

Before discussing the data that is used in this study, this section covers the institutional details about the Medicare Part D program and the specific plan types that are part of the program.

3.1 Background

In the United States, Medicare is a health insurance program for the elderly that covered approximately 46 to 51 million individuals from 2009 to 2012.[4] It is structured in four parts: A, B, C, and D. Parts A and B include hospital and medical insurance for in- and outpatient care. Patients who are enrolled or eligible for Parts A and B, are also eligible to enroll in Medicare Part D, which provides insurance for the prescription drug purchases, covering mostly self-administered drugs. Unlike Parts A and B, which are administered by the government, the Part D plans are administered by private insurers who are subject to the rules and regulations laid out by the government. Part C, also known as Medicare Advantage, is also administered by private insurers and is an all-inclusive alternative to Parts A, B, and D. While enrollment in Medicare is voluntary, individuals face significant penalties within the program if they choose not to sign up when first eligible (usually at 65) or have creditable (similar) health insurance coverage.

This paper focuses on beneficiary purchasing behavior of enrollees in Medicare Part D with stand-alone Prescription Drug Plans (PDP). In 2009, there were almost 18 million enrollees, and there were almost 20 million by 2012. The program began with the Medicare Modernization Act of 2003, which was enacted in 2006. The government regulates a "standard" plan with a baseline minimum amount of coverage, and private insurers can offer a variety of plans that on an actuarially equivalent basis meet or exceed the generosity of the standard plan. This results in a significant variety in prescription coverage plans with 1,689 stand-alone PDP plans in 2009

[4] Program statistics from the Centers for Medicare & Medicaid Services' Statistical Supplement https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Archives/MMSS/index.html

("The Medicare Part D", 2016).

In the context of studying beneficiary behavior when faced with nonlinear prices, the main advantage of studying Medicare Part D is the highly nonlinear structure of the coverage regions in the stand-alone Prescription Drug Plans (PDPs). This paper will focus on plans that do not have a deductible because the majority of beneficiaries do not choose plans with deductibles, but this section will discuss the full variety of plan structures. The 2009 standard Part D Plan included a deductible of $295, an initial coverage limit (ICL) of $2,700, and an out-of-pocket threshold (OOPT) of $4,350. The specific deductibles, ICLs, and OOPTs differ every year, but the structure of the standard plans and thus plans in the market are similar. The example plan in Figure 1 is actuarially equivalent to the standard plan displayed in the Appendix Figure A.1.

In the standard plan in all years, the patient is responsible for 100% of the cost of prescriptions until their cumulative spending reaches the deductible amount, after which they are in the initial coverage region (ICR). Note that both patient out-of-pocket and "total" spending (the total amount spent through a combination of patient, insurance company, government through Medicare, and drug companies) are the same up to the deductible. Patients in the Initial Coverage Region (ICR) are then responsible for 25% or less of the total price of prescription purchases. Once the patient's cumulative total spending amount reaches the ICL amount, patients enter the phase often referred to as the coverage gap or "donut hole" where they are again responsible for 100% of the total spending amount in 2009. Once the patient's cumulative out-of-pocket costs reaches the OOPT, they reach the "catastrophic" phase and are responsible for paying a greatly reduced share of the total costs of drugs. Specifically, they pay either the maximum of 5% of the total price of prescriptions or a $2.40 and $6.00 copay for a one-month supply of generic and branded drugs respectively. The pricing schedule resets at the end of the calendar year, and at the beginning of the next year beneficiaries begin anew with a total cumulative spending of 0 and the associated marginal costs.

The standard plan up until the catastrophic region has an exact mapping of out-of-pocket and total payments—an ICL of $2,700 corresponds exactly to cumulative OOP costs of $896.25, and an OOPT of $4,350 corresponds to $6,153.75 in total expenditures. In practice, an exact mapping between the two cumulative spending measures is difficult to ascertain in all plans. Plans only have to meet (or exceed) the coinsurance generosity of the standard plan on an actuarially equivalent basis through coinsurance, copays, or a combination of the two. Because of the actuarially equivalent clause, insurance companies have tremendous flexibility in structuring plans. Often, the cost sharing can be specific to drug tier or whether it is branded or generic.

Regulation just requires that on average, the plan is expected to be similar to or more generous than the standard plan.

The inclusion of the Part D coverage gap or "donut hole" (and a main component of the plan nonlinear structure) has been widely criticized and analyzed in the healthcare literature. The coverage gap was initially included as a cost-saving measure for the government similar to a deductible but positioned in the middle of the patient benefit schedule (Baker, 2006). The health policy literature has both criticized the arbitrary location of the donut hole and the impact it has on drug adherence.

Every year, the standard plan's spending limits are updated to adjust for rising costs. And under the Patient Protection and Affordable Care Act (ACA), the government began to fill in the coverage gap and will continue to increase the plan benefits in this region through 2020. The phase out began in 2010 when the standard plan included an automatic $250 rebate for beneficiaries who reached the coverage gap. Further coverage in the donut hole increased in 2011 and 2012, when instead of a rebate, standard plans included a 50% discount on brand name prescriptions that was paid by the drug manufacturer. This means that in 2011 and 2012, patients in the donut hole only paid 50% of the cost of branded drugs. In most cases, while not paid by the beneficiary, the 50% covered by drug manufacturers did contribute towards a beneficiary's cumulative out-of-pocket expenditures. Because of this set up, the discount did not significantly change the cumulative total expenditure amount it took for patients to get out of the donut hole. Table 1 shows the changes in the standard plan from 2009-2012.

Table 1: Medicare Part D Benefit Parameters for Defined Standard Benefit 2009-2012

Plan Characteristics	2009	2010	2011	2012
Deductible	295	310	310	320
Initial Coverage Limit (ICL)	2700	2830	2840	2930
Out-of-Pocket Threshold (OOPT)	4350	4550	4550	4700
Total Expenditure equivalent OOPT	6153.8	6440	6447.5	6657.5
Rebate (1)		250		
Brand discount(2)			50%	50%
Generic copay (3)	2.40	2.50	2.50	2.60
Branded copay (3)	6.00	6.30	6.30	6.50

(1) The rebate begins when patients reach their out-of-pocket threshold (OOPT).
(2) The brand discount only applies when patients are in the coverage gap, i.e. when their cumulative total spending is above the ICL, and their cumulative non-insurer spending (patient payments, any subsidies, brand discounts paid by the drug manufacturers) is below the OOPT.
(3) In the catastrophic region, beneficiaries pay the maximum of the copay or 5% the total cost of the prescription.

This paper will focus on plans that have the government-defined ICL and OOPT, but not

plans with the government-defined deductible. Enrollees have choices over a wide variety of Part D plans, with a majority of patients opting for plans with more generous plan benefits than the standard plans including no deductible plans. In 2006, fewer than 10% of beneficiaries were in plans with the standard design (Abaluck and Gruber, 2016), and our data support this finding as well. Even prior to 2010 when the ACA started phasing out the coverage gap, many plans offered some type of gap coverage, though these were typically on generic prescriptions. See Section 3.2 for a deeper discussion of plan types.

For the purpose of understand beneficiary behavior in the face of nonlinear contracts, there are other advantages to studying the Part D plans instead of other nonlinear employer-sponsored health insurance plans. These advantages include the widespread frequency of claims and large percentages of beneficiaries experiencing different coverage phases year-over-year. Hoadley et al. (2011) indicate that 16% of Medicare beneficiaries ended the year in the coverage gap, with 3% of beneficiaries reaching the gap and passing it to end the year in the catastrophic region. Across 2008-2009, almost 30% of patients experienced the gap. Further, they document that reaching the coverage gap is consistent, as 71 percent of enrollees who reached the gap in 2008 did so again in 2009. This recurrence of reaching the coverage gap is due to the fact that many patients take medications for chronic conditions rather than for acute, short-term medical needs.

Another advantage of studying Part D is that many of the medications patients take in Part D are for chronic conditions, so enrollees in the nonlinear Part D setting may have a better ability to forecast their yearly spending on prescription drugs than in other types of health insurance. In a MedPac report on Medicare Part D, they state that the list of top 15 therapeutic classes of drugs by spending and volume has remained relatively consistent since 2007. The values from 2013 indicate that drugs in the diabetic, asthma/COPD antihyperlipidemics,[5] antipsychotics, antihypertensive,[6] and peptic ulcer therapeutic classes are responsible for approximately 40% of drug spending (MedPAC, 2016). These drugs are all used to treat ongoing chronic conditions.

Further, in studying beneficiary behavior, it is also an advantage that Part D claims only cover self-administered drugs and do not cover drugs administered at the doctor's office or in the hospital. Unlike hospital claims, there is an increased likelihood that the prescription purchases are the beneficiary's decision rather than decisions made by a medical professional. However, it is still a concern for older patients that drug purchases could be done by a proxy.

One of the challenges of the Medicare Part D data for studying spending in the non-linear pricing schedule across years is that the pricing schedules change over years. As mentioned,

[5] Used to treat high cholesterol

[6] Used to treat high blood pressure.

each year the standard plan adjusts, and it is highly likely that the individual private plans adjust. Patients also have choices to switch between plans with different insurers across years and to switch across into Medicare Part C. However, while plans change every year, the schedule remains similar with mostly minor increases in the exact limits of each coverage region. The literature also document a significant amount of inertia and inattention in patient choice of plans, indicating that approximately 10% of Part D patients switching their plans between every two years (Abaluck and Gruber, 2016; Abaluck et al., 2015; Ho et al., 2015).

3.2 Data Description

The primary dataset includes the prescription drug and medical claims from a random 5% subsample of Medicare beneficiaries enrolled in Part D in the years between 2009 and 2012, with approximately 2 million individuals per year.[7] The data comes from the Center for Medicare and Medicaid Services. Broadly, the data cover the beneficiary demographics, their Part D prescription claims, and the plan characteristics of the specific Part D plan that each beneficiary chose. The prescription drug claims include the exact drug purchased, days supply, purchase date, the proportion paid by both the patients and insurance companies, and the benefit phase each claim occurs in. The plan characteristics supplement this information on the contracts that patients face with full details on the plan premiums and the exact cost-sharing characteristics: deductibles, coinsurance, copays for specific drug types and tiers.

Basic demographic information of the beneficiaries (gender, age, race) along with hospitalization and doctor claims information from Medicare Part A and B, which are used to determine patient health conditions, are also observed. With these data, I use the CMS-provided risk model to calculate a "risk score" or summary estimate of the expected average drug spending implied by patients' demographics and health conditions.[8] In order to normalize the risk score comparison, I use the risk and demographics score translation from 2011. Individuals who have a combined risk and demographic weight of 1 have an average prescription reimbursement liability of a Medicare Part D enrollee in 2011.

The analysis sample for this paper is constructed by keeping only individuals in the Medicare system who had Medicare Part D from 2009 through 2012 with specific beneficiary and cho-

[7] This analysis does not consider patients enrolled in Medicare Advantage (Part C), because while the Medicare Part D claims data include Part C prescription claims, the dataset does not include the doctor and hospitalization claims, which are used to control for heterogeneity.

[8] CMS use Hierarchical Conditional Codes (HCC) and RxHCC (for prescriptions) to adjust the reimbursement payments to insurance companies that offer plans in Medicare Advantage and other programs. The HCC and RxHCC scores are scaled reimburse the plans for managing patients with illnesses with expected increased medical and prescription medication costs respectively. http://setma.com/EPM-Tools/tutorial-hcc-rxhcc-risk

sen plan characteristics. To list the beneficiary characteristic restrictions, the sample contains beneficiaries who are 65 or older and are enrolled in Medicare PDPs from 2009-2012 through the Old Age and Survivors Insurance (OASI) and not for disability insurance or other qualifiers for Medicare. The sample also excludes individuals who are dual eligible for Medicaid financial assistance or receive other types of low-income subsidies (LIS) for premiums or cost-sharing. These individuals are excluded because they face very low cost-sharing and minor changes in their marginal costs. Even individuals who only receive premium subsidies are omitted, because they are more likely to be lower income and are more likely to be influenced by budget constraints. Further, the analysis of the paper also excludes individuals whose Medicare Part B claims indicate they were in long-term care institutions (LTI) such as nursing homes in the prior year. Including also a restriction for individuals who were in Medicare Part D for all twelve months of each year, this leaves approximately 300,000 beneficiaries in each of the 2009-2012 years respectively. This is the "full" sample and is used as a comparison group to the analysis sample.[9]

Further restrictions are made on the plan choices, because the variety of plans and a beneficiary's choices to switch plans between years pose a challenge for studying beneficiary behavior. There may be an endogenous relationship between a beneficiary's choice of plans and her spending patterns. That plan variety can be seen in Table 2. Very few enrollees over the years have chosen plans with deductibles; 65-75% of enrollees have plans without deductibles. Overall only 15-20% of enrollees have plans with the standard government-defined plan limits in the deductible, Initial Coverage Region, and Out-of-Pocket Thresholds. However, a majority of patients still have the same ICL and OOPT limits with 76% without a deductible. But despite the pervasiveness of the ICL and OOPT spending limits, less than 2% of beneficiaries have plans that use the exact 25% coinsurance rate suggested by the government for the initial coverage region. Because many beneficiaries do have plans with the Medicare-Defined spending limits for their ICL and OOPT, this paper will focus on the beneficiary's response to approaching these limits rather than her response to the specific coinsurance rates.

The analysis sample or baseline "No Deductible" sample then focuses on the balanced panel of individuals who qualify under the prior restrictions discussed, had the government prescribed limits for the ICL and OOPT, and who did not have deductibles. Individuals were also omitted if they were observed to have either zero spending in any year or had claims in every week of the

[9]The sample exclusion restrictions are further discussed in Appendix Section A and listed in Appendix Table A.1 with the percentage of the entire Medicare/Medicaid 5% sample they encompass. Note that not all individuals have Medicare Part D. See Appendix Table A.3 for summary statistics before plan choice restrictions.

Table 2: Summary of Full Sample of Medicare Beneficiary Plans 2009-2012

	2009 mean	2010 mean	2011 mean	2012 mean
Deductible: None	0.7627	0.6828	0.6258	0.6675
Deductible: Other	0.0391	0.1690	0.1901	0.1483
hasStandardDed	0.1970	0.1481	0.1841	0.1838
ICR: Standard Coinsurance	0.0101	0.0095	0.0119	0.0041
ICR: Cost Share Tiers	0.9899	0.9905	0.9881	0.9959
ICL: Standard	0.9897	0.9262	1	1
OOPT: Standard	1	1	1	1
Standard Plan Limits	0.1970	0.1481	0.1841	0.1838
No Ded, Standard ICL & OOPT	0.7528	0.6090	0.6258	0.6675
Main 4 Year Sample	0.3358	0.3216	0.3082	0.2938
Observations	291,550	304,477	317,670	333,309

*Note:*The full sample includes individuals in each year who satisfy the criteria for the "Full 12 Month Sample" from Table A.1.

year. Keeping only individuals with these plans limits, the sample contains 89,354 beneficiaries or about 30% of the full sample. Part of the reason for this restriction is to help standardize the spending limits for the analysis in later sections. This restriction has the negative effect of decreasing the sample size and reducing the generalizability of these results. Further, because these individuals choose plans without deductibles, the selected sample may be simultaneously more risky, more risk averse, and richer than an average Medicare Part D enrollee. Also because the sample requires the beneficiary have the same plan structure in all four years, they have higher inertia and may have higher costs of switching.[10]

Notice that this dataset of 2009-2012 claims differs from the data used in the papers previously mentioned. In all cases, this sample draws from a later sample of individuals with Medicare Part D than individuals who are in Joyce et al. (2013)'s 2006, Dalton et al. (2015)'s 2008, Einav et al. (2015)'s 2007-2009, and Abaluck et al. (2015)'s 2006-2009 sample. It is also a longer sample than most other papers covering four years. Because this sample involved individuals who retained similar plan structures through all four years, they are mechanically more likely to have experience with Medicare Part D[11], their plan structure, and their prescription needs than individuals described in these other papers. These differences possibly translate to different findings in the empirical section.

Table 3 displays the demographics of the baseline sample of individuals without deductibles. The average age of the population in 2009 is about 75, which is slightly older than the average

[10]An additional "Standard" sample is also created from the 9,178 individuals who signed up for plans with the government-defined deductible, ICL, and OOPT, for whom summary statistics are presented in Appendix Section A.

[11]Remember also Medicare Part D began in 2006.

Table 3: Demographics of Baseline No Deductible Sample of Medicare Beneficiaries in 2009-2012

	mean	sd
Age at End of 2009	74.9412	6.58
Start Medicare	1999.11	6.55
Female	0.6461	0.48
Race: White	0.9519	0.21
Race: Black	0.0240	0.15
Race: Other	0.0121	0.11
Race: Asian	0.0071	0.08
Race: Hispanic	0.0035	0.06
Observations	89,354	

	2009		2010		2011		2012	
	mean	sd	mean	sd	mean	sd	mean	sd
2011 RxHCC weight	0.4753	0.29	0.5063	0.29	0.5265	0.30	0.5487	0.31
2011 RxHCC demographic weight	0.4196	0.01	0.4196	0.01	0.4196	0.01	0.4197	0.01
Diabetes	0.2441	0.43	0.2591	0.44	0.2708	0.44	0.2800	0.45
Hypertension	0.6664	0.47	0.6918	0.46	0.7009	0.46	0.7072	0.46
Has Cancer	0.1016	0.30	0.1075	0.31	0.1107	0.31	0.1153	0.32
High Cholesterol	0.7325	0.44	0.7571	0.43	0.7650	0.42	0.7670	0.42

Note: The Baseline No Deductible sample includes individuals in each year who satisfy the criteria from Table A.1 and also were in a plan with standard ICL and OOPT limits with no deductible from 2009-2012. Risk scores are normalized to 2011 RXHCC scores for consistency across years.

Medicaid 5% sample population, but is consistent with the ages of beneficiaries enrolled in Medicare Part D. The vast majority (approximately 95%) of beneficiaries are white. The 2011 RxHCC weight and demographic weight indicate that the average person in this sample is healthier than the average Medicare Part D participant since their sum is lower than 1, which reflects the fact that disability, LIS, and LTI individuals were not included. However, they are sicker than the average individual in the full sample, which may reflect both the higher age and the optional nature of joining Medicare Part D for prescription purchases.

A majority of individuals have chronic conditions with 66% and 73% of individuals having a prior year condition code of hypertension and high cholesterol in 2009. Also, almost a quarter of the sample has diabetes and 10% has had cancer treatment of some kind in 2009. These are all conditions that often require constant prescription refills and spending using the Medicare Part D benefit. The Kaiser foundation documents that patients who took drugs to treat some of these conditions are far more likely to reach the coverage gap hole and catastrophic regions (Hoadley et al., 2011). They observed that the average Part D enrollee's probability of reaching either of the two regions as 19% in 2009 but 56% for patients on breast cancer treatment drugs, 40% for those taking oral anti-diabetics, 32% for those on statins etc. Also notice that as the beneficiaries progress through the years, they get older and sicker. The increase in sickness is

reflected in the claims data with an increasing trend in spending amount and frequency through 2009-2012.

Even though the majority of beneficiary plans do not follow the standard Medicare-Defined coinsurance amounts, the coinsurance levels between the different coverage regions are on average still economically and significantly different from each other. Table 4 displays the average coinsurance amount that beneficiaries in the baseline No Deductible sample face. Because the plans patients choose do not have deductibles, their effective coinsurance rate in the initial coverage region (ICR) is higher than the standard plan at 39% to 41% of the total cost of care. The coinsurance rate in the donut hole in 2009 and 2010 prior to the ACA legislation to filling in the donut hole was not quite 100% but still significantly high at approximately 92%. The beneficiary responsible portion of the coinsurance rate during the coverage gap in 2011-2012 was significantly lowered to effectively 54-55% with the addition of the 50% discount on branded drugs. While the difference between the ICR and Coverage Gap coinsurance rates in the first two years of the sample is higher than the latter two years, there is still a difference in latter two years. This means that beneficiaries should qualitatively still respond to these plan characteristics in the way laid out in Section 2. It is expected that any beneficiary response to the coverage gap in 2009-2010 may be muted in 2011-2012, because the marginal price difference between the two regions is smaller. The average coinsurance levels for patients with standard plan limits follow closer to the government recommended plan and are shown in the Appendix Table A.4.

Table 4: Average (person-week) Coinsurance in Phases in Baseline Sample No-Deductible Plans 2009-2012

	2009		2010		2011		2012	
	mean	count	mean	count	mean	count	mean	count
ICR	0.3863	4,285,602	0.4069	4,300,522	0.3975	4,277,143	0.4071	4,293,291
Coverage Gap	0.9223	319,117	0.9155	306,331	0.5508	325,492	0.5428	309,185
Catastrophic	0.0587	32,100	0.0580	30,742	0.0595	36,657	0.0597	38,708

Note: Table is generated from the baseline sample individuals with no deductible with standard ICL and OOPT limits. The coinsurance rates are averaged over the amount the patient pays (doesn't include the drug manufacture discounts in 2011 and 2012) divided by the total expenditure cost in the person-week observation where spending occurs. This rate is effectively weighted by the time individuals spend in each phase. The count reflects the fact that there are more person-week observations in the ICR region than others. These sums do not reflect the counterfactual coinsurance rates that beneficiaries with low spending would have faced if they had reached higher spending. While the data contain the actual structure of the beneficiary plans with exact coinsurance and copay rates for drug tiers, it is difficult to summarize in a specific coinsurance rate without knowing the mix of drugs that patients may consume.

Since the baseline sample is used to study beneficiary behavior as they cross spending phases, the sample does include individuals who are likely to reach the coverage gap and beyond. Table 5

shows that while the majority of the beneficiaries without deductibles end the year in the initial coverage region, over 21% of enrollees end the year past the coverage gap with approximately 2-3% reaching the catastrophic coverage region. While heterogeneity among beneficiaries mean that many would not have realistic expectations of reaching the higher spending coverage regions, there is certainly a significant subset of enrollees who might expect to end the year in these regions.

Table 5: Proportion of beneficiaries in each phase at the end-of-the-year in the baseline sample 2009-2012

	2009	2010	2011	2012
ICR	76.47	77.70	76.93	78.81
Gap	21.03	19.97	20.28	18.46
Catastrophic	2.49	2.32	2.79	2.73
Observations	89,354	89,354	89,354	89,354

Note: Table is generated from the baseline sample individuals with no deductible with standard ICL and OOPT limits. The proportion of beneficiaries that end the year in each phase is averaged over the individual beneficiary.

The sample beneficiary's average raw claim occurrence probability (i.e. the probability of submitting a claim) in each of the Medicare Part D insurance coverage regions is depicted in Table 6. Across the four years, while the probability of ever making a prescription claim in a week is 32-34%, the raw probabilities do differ significantly within the insurance regions. Of the beneficiaries who are in the initial coverage region, their average probability of spending is in the 31-33% range. The weekly claim probability for the observations in the coverage gap is higher at 43-47% and highest in the catastrophic region at 56-57%. The overall average is very similar to the average in the ICR region, since the majority of the individual-week observations occur in the ICR. This pattern can also be seen in Figure 6, which displays both the frequency of claims and the mean claim occurrence within $50 bins in 2009 as a function of the cumulative total expenditures. This image illustrate the raw probabilities of beneficiary's spending.

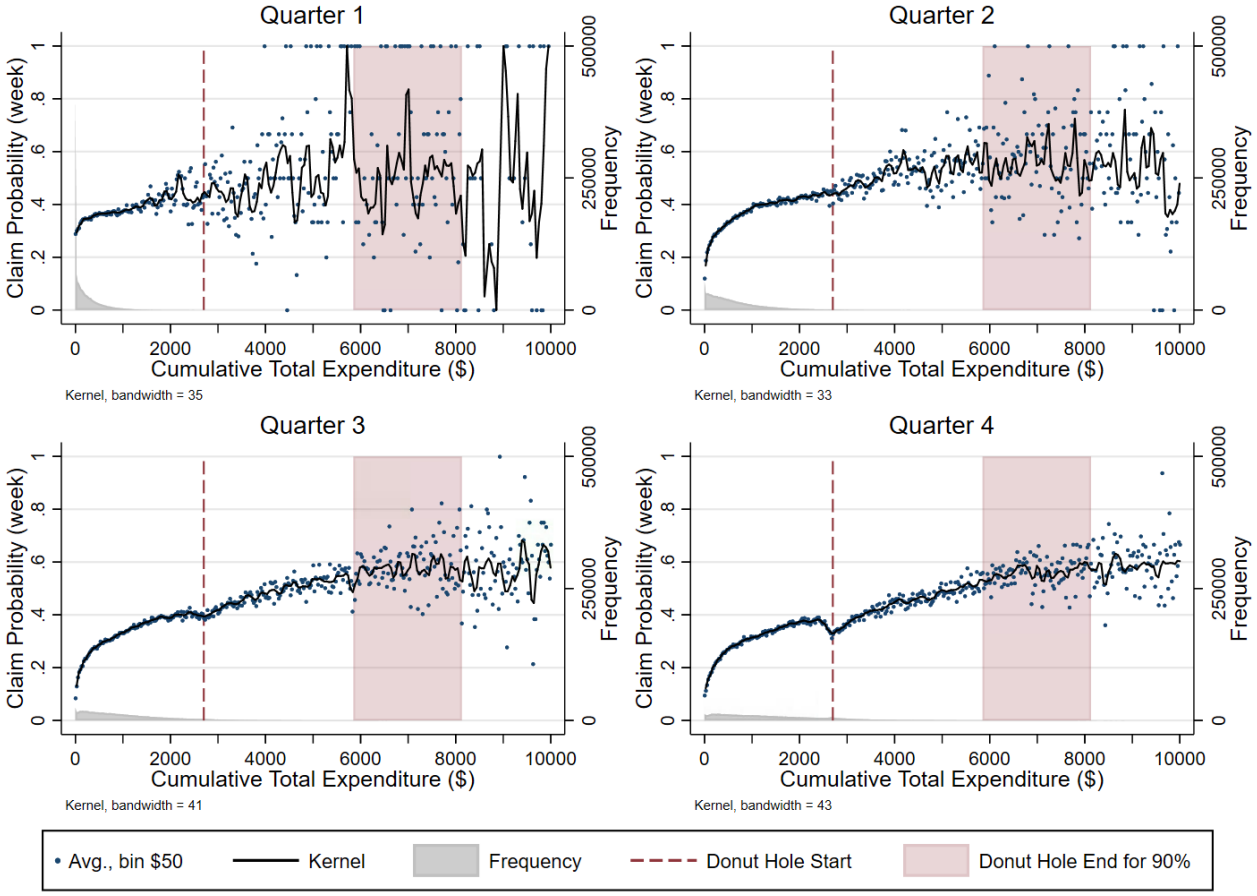
Given the number of individuals who end the benefit year in each region and the large number of person-week observations in the ICR relative to the catastrophic region, it makes sense that the ICR probability is closer to the total average probability. The differences in the probabilities across regions is a possible indication of heterogeneity in the probability of spending, where individuals with higher probabilities of claims, who may also have higher average spending amounts, are more likely to have observations in the coverage gap and catastrophic regions. Einav et al. (2015) uses a similar graph to Figure 6 of the monthly probability of spending to help illustrate the empirical patterns that beneficiaries engage in at the coverage gap. However,

Table 6: Average probability of weekly spending in coverage regions in no deductible plans 2009-2012

	2009	2010	2011	2012
ICR	31.48	31.90	32.15	32.71
Coverage Gap	42.65	43.63	45.01	46.93
Catastrophic	56.22	56.84	55.87	56.84
All weeks	32.44	32.86	33.26	33.87

Note: Table is generated from the baseline sample individuals with no deductible but standard limits for the ICL and OOPT. Table displays the raw probability of spending in a week in each coverage region and year. The coverage regions are the initial coverage region (ICR), the coverage gap (aka the donut hole), and the catastrophic region. The average is of the probability of a person-week observation having a claim and is low because there are more observations in the ICR per as seen in Table 4.

Figure 6: Weekly Claim Occurrence Conditional on Cumulative Total Expenditure and Spending



Note: Using a $50 dollar bin, the points are the average probability of spending on prescription purchases in a week conditional on the quarter of the year and the cumulative total expenditure. Einav et al. (2015) produced very similar graphs of the probability of a prescription purchase in a month rather than the week level. This image is only of 2009 claims.

in order to fully understand beneficiary behavior at and before the kink, researchers must take into consideration the significant amounts of heterogeneity in prescription needs that exist in the Medicare Part D population. Beneficiaries who are more likely to spend (and spend more), are also more likely to be observed in higher cumulative spending bins, while those who spend less are observed at the lower cumulative total expenditure levels. The approach that is taken in this paper to handle this heterogeneity is through the use of fixed effects in Section 5.

The limitations of these data include the fact that they do not capture prescription purchases outside of Part D such as large retailer generics since those purchases are outside the scope of the Medicare system. Further, this paper has limited data on beneficiary incomes. One potential explanation for non-standard behavior could be patients reaching budget constraints, and this paper is not able to directly measure individual liquid wealth.

4 Heuristic Approach Applied to Medicare Part D

This section applies the heuristic approach proposed in Section 2 to the empirical data. The goal is to recover the average objective expected end-of-year prices beneficiaries should expect as a function of week and current cumulative spending. One drawback of the heuristic method is that it requires a strong assumption that a beneficiary has access to the data to know her objective probabilities; however, this heuristic is still useful as a tool to understand how optimal agents should behave. Further, if the stylized facts about beneficiaries sharply reducing their spending when approaching the coverage gap earlier in the year are generally true, then under this heuristic this behavior would only be explained by discontinuities in beneficiary's subjective end-of-year probabilities, i.e. if they fail to update their beliefs or receive health shocks that lead to surprises.

There are more regions and more noise in these estimates than the simulated example, but basic interpretations hold. First, the probability distributions of ending in each of the contract regions based on the time of the year and beneficiary year-to-date spending are constructed. While these graphs are constructed at the cross-individual level, they should still provide insight for individuals to construct their internal beliefs.

For the baseline sample of beneficiaries with no deductible, Figure 7 displays the raw probability in weeks 13, 23, 33, and 43 of ending the year in each coverage region conditional on their cumulative total expenditures (through week 12, 22, 32, and 42 respectively). The heuristic expected marginal price (HMP) is constructed using these probabilities as discussed in Section 2, and it is displayed in Figure 8. Both probabilities and marginal costs are calculated from

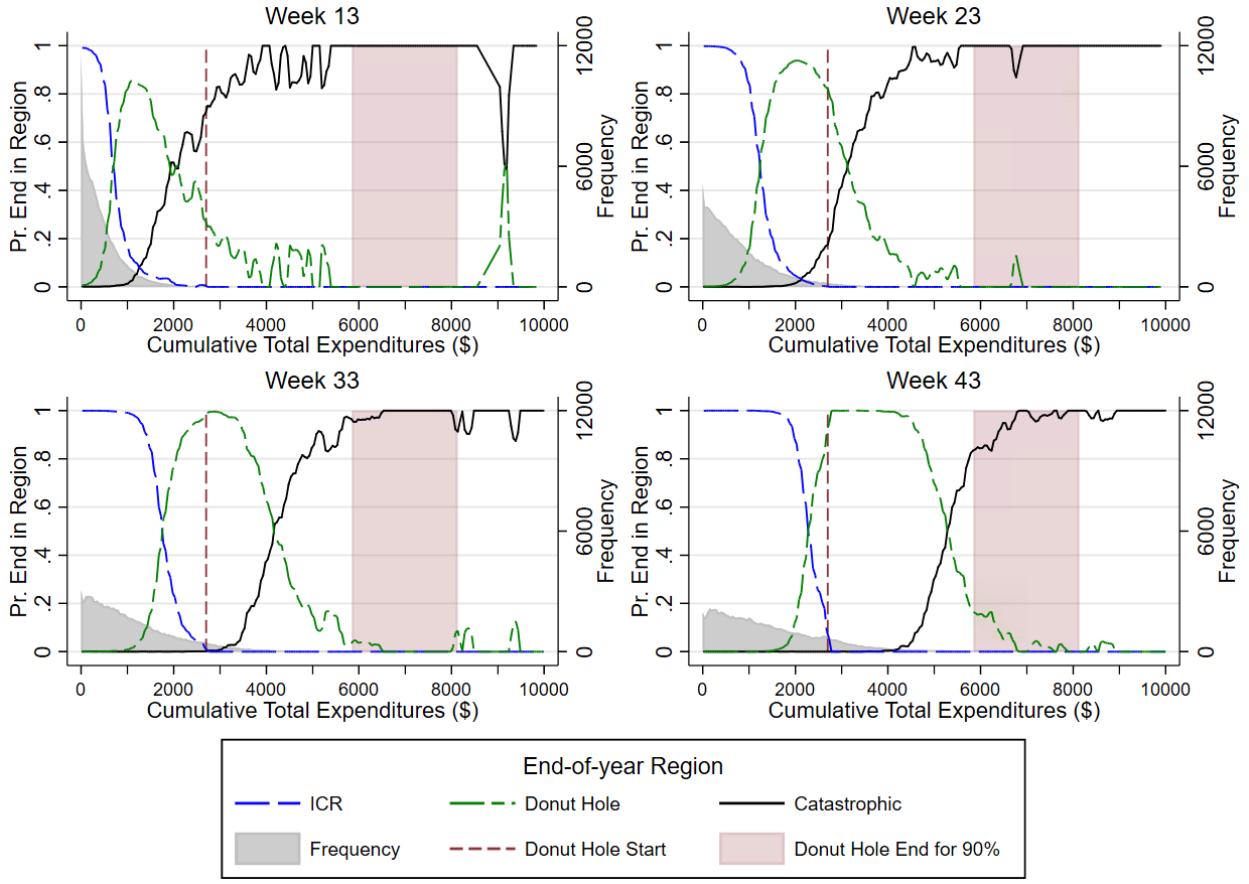
beneficiaries whose weekly cumulative total expenditure amounts fall in \$50 bins in the x-axis. The red line represents the initial coverage limit and the boundary between the initial coverage region of low coinsurance and the coverage gap. There are significantly more differences in the bin means for the probability ending the year in the catastrophic and initial coverage gap conditional on the cumulative total spending. This is due in large part because the out-of-pocket threshold (i.e. the boundary between the coverage gap and catastrophic region) translates to different cumulative total expenditures for different insurance plans and drug consumption patterns. The noise in probabilities is exacerbated in the week 13 panel because there are few individuals who have accumulated high spending levels that early in the year.

While much of the prior literature has focused on the behavior directly at the initial coverage limit, the effect of the non-linear pricing structure could be evident far prior to the coverage gap earlier in the year. In week 13, if a beneficiary's spending is between \$500 and \$1,500, she is most likely to end the year in the coverage gap, and this translates to her highest HMP being in that region. If a beneficiary's spending is already over \$1,500, she is most likely to end the year in the catastrophic zone, even though she is still far away from even the transition from the ICR to the coverage gap. If a beneficiary uses the HMP as her perceived marginal price, she should increase her spending as the HMP decreases, which begins well before the 2009 ICL of \$2,700.

As time progresses, the probability of ending the year in any of the regions also shift to higher cumulative totals and tend closer to 0 and 1, and the beneficiary's highest heuristic expected marginal price more closely resembles the plan spot prices. Over time the cumulative total spending level at which she experiences her highest HMP moves closer to the discontinuity between the initial coverage and the coverage gap regions. While the distribution of the probability of ending the year in the coverage gap was a maximum of 80% in week 13, there is an increasing group of individuals who have spent around \$2,000 by week 20 and \$2,700 in week 33 who are certain to end the year in the coverage gap. Thus the highest HMP, which should correlate with a beneficiary's lowest levels of spending move to about \$2000 in week 23 and to the ICL or \$2,700 in week 33. In week 43, or approximately 2 months before the end of the year, there already exists a discontinuity in the probability of ending the year in the coverage gap at the initial coverage limit. These graphs are not generated based on the behavior of necessarily standard forward-looking agents, so the HMP may approach the spot price earlier than for standard rational agents.

In translating the beneficiary's perceived marginal price to their spending patterns, if the

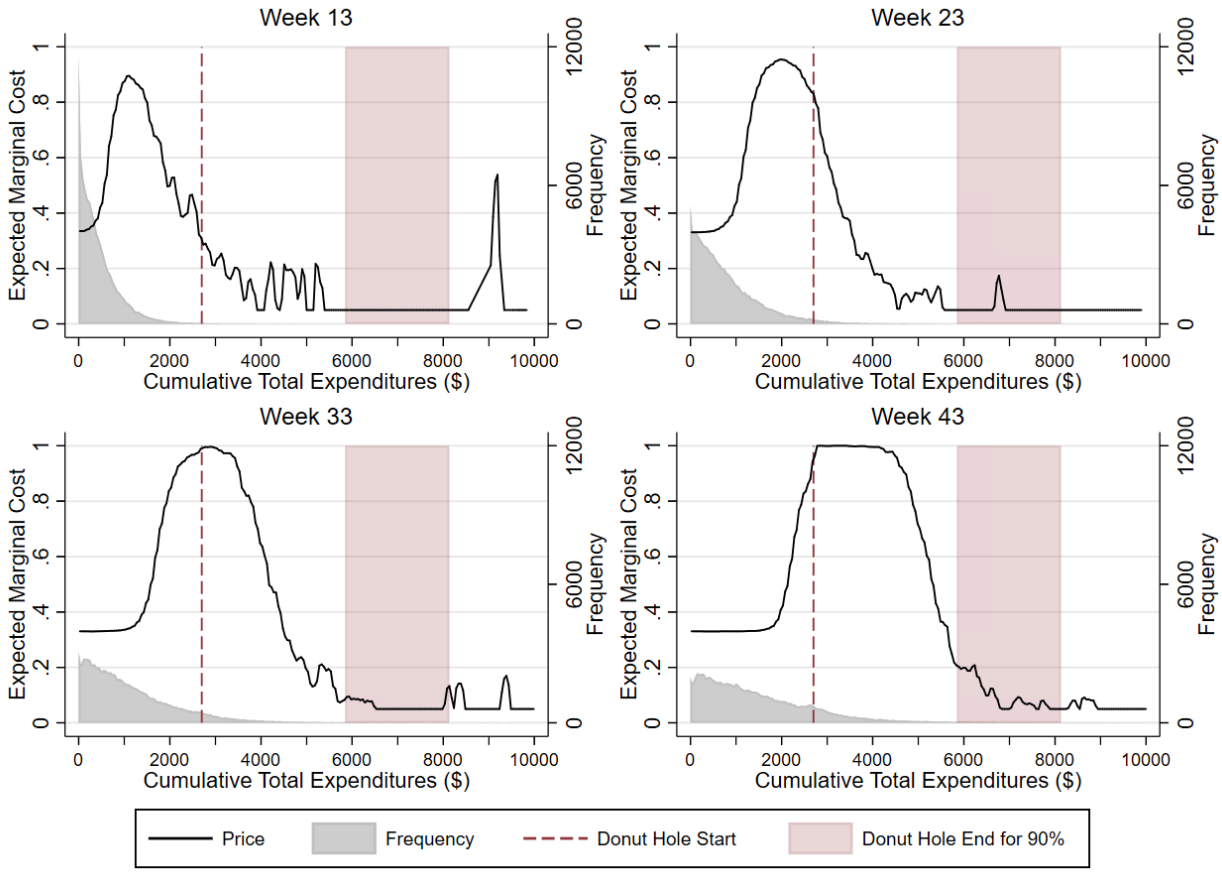
Figure 7: Distribution of the Probability of Reaching each Coverage Region in 2009 Conditional on Cumulative Total Spending



Note: The graph depicts an Epanechnikov kernel-weighted polynomial of the probability of a beneficiary being in coverage region r at the end of the year given week w and within a \$50 bin of the cumulative total spending Z_w. This makes up the distribution $F_r(Z_w, w)$. This figure is generated based on non-LIS beneficiaries who had PDP plans in 2009 that had the no deductible plans with the government-defined initial coverage limit at \$2,700 (shown by the red vertical line) and out-of-pocket threshold (not shown). The bin size = 50 and was chosen for illustrative purposes. Similar images for 2010-2012 are included in the Appendix. Because the out-of-pocket threshold limit for entering the catastrophic coverage region does not on aggregate map to a specific cumulative total amount, the average probability of ending in those phases as a function of the cumulative total amount has a wide dispersion of points.

demand function is smooth and quantity demanded is decreasing with the marginal price of prescription purchases, broad predictions can be made. The expectation is that within any time period the cumulative total expenditure with the lowest heuristic expected marginal prices in a time period should correspond with the cumulative total expenditure amount that has the highest level of spending. Similarly, within a time period, the cumulative total expenditure with the highest heuristic expected marginal price should correspond with the lowest levels of spending. If the demand function is smooth and the transitions in the HMP are smooth (as

Figure 8: Heuristic Expected Marginal Price in 2009



Note: Depicts the expected marginal price based off of the objective distributions of the probability of ending the year in each coverage region depicted in Figure 7. The heuristic expected marginal price is $HMP(Z_w, w)) = F_{ICR}(Z_w|w) * MC(ICR) + F_{Gap}(Z_w|w) * MC(Gap) + F_{Cat}(Z_w|w) * MC(Cat)$, where MC(ICR) = .33, MC(Gap)=1, and MC(C=.05) the government standard plan amounts.

they are for the majority of the periods), then the spending should also be smooth.

The relationship between the HMP and spending is not necessarily expected to be one-to-one and would not be as dramatic as a simple "flip" of the HMP curve, but the expectations is that the location of spending changes should however correspond with the peaks and troughs of the HMP. In fact, because prescription purchases often have immediate and significant health benefits, these drugs may be relatively inelastic goods and respond little if at all to the marginal price changes. Using the end-of-year purchases, Einav et al. (2016) measures the elasticities of different drug classes and find an overall elasticity of -0.037 so that a one percent increase in out-of-pocket cost leads to a 0.037 percent decrease in the probability of filling a claim. Thus, the expectation is that the HMP would result in small changes in the probability of filling a

claim as well.

The next section will discuss the empirical approach to estimate a graphical representation of beneficiary spending patterns. It will also discuss how consistent behavior is with the simple heuristic model of spending.

5 Estimation Model and Results

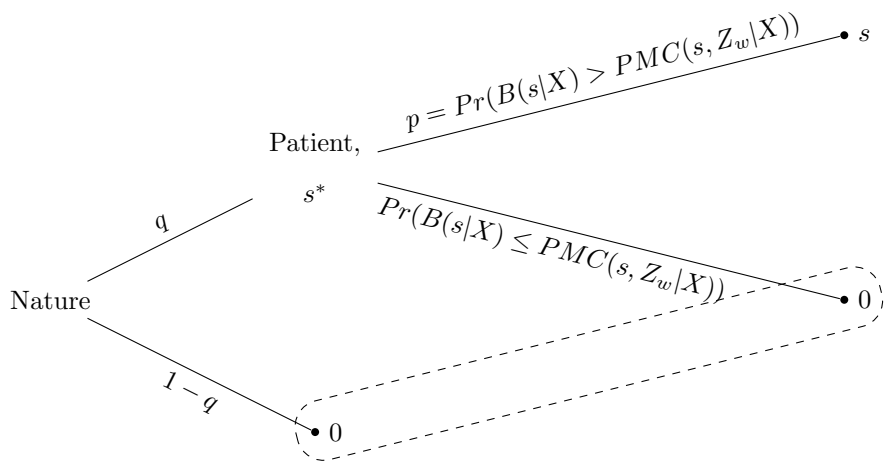
This section details a simplified model of prescription purchasing behavior under Medicare Part D that ascribes prescription spending to both opportunities to spend and then beneficiary decisions to spend. The model emphasizes the effects of past cumulative spending on weekly prescription purchases. Ultimately the estimation shown in this paper lumps both the opportunities to spend and the decision to spend as one process, but the model is laid out to provide intuition on the beneficiary's underlying motivations. The intent of the estimation is to document the patterns of consumer behavior that emerge as patients progress through their insurance plans and not to capture the full dynamics of the consumer prescription choice problem. The beneficiary's prescription spending decision is aggregated and examined on the weekly level to reduce the size of the problem.

The initial model below assumes beneficiaries have a choice to spend on prescriptions and a separate limited choice on the amount to spend. The only observables that are available in Medicare Part D data are prescription purchases and not the beneficiary's direct consumption of drugs. Because of the nature of chronic prescriptions, many are offered in 30 to 90 day refill amounts, which means that there is a periodicity to beneficiary's spending patterns that are not necessarily driven by a choice to spend or not. In order to capture some of this, the model assumes that the personal utility from medications should only apply if a patient has a medical event or shock that requires treatment. Thus, the decision to use Medicare Part D to purchase prescriptions should only occur when those events arrive and a doctor has written the patient a prescription. These events can be temporary health shocks that require treatment such as antibiotics for pneumonia, or continuations of existing conditions that require a prescription refill.

The probability of a medical event occurring q depends upon observables X such as demographics (age, risk scores, historic Medicare Part D usage, etc.) and observable environment characteristics (e.g. time of the year). It may also depend upon unobservable patient characteristics and health shocks. However, essential to the model is the idea that the actual probability of a true medical event occurring should not change due to the arbitrary insurance coverage

region (defined by the cumulative total spending Z_w up until week w) imposed by the patient's insurance plan coverage—that is the model assumes $E(q|X, Z_w) = E(q|X)\forall r$.

Once a patient receives a medical event, they have a choice to spend on prescriptions and a choice on the exact amount of out-of-pocket and total prescription spending. Within the second stage prescription filling decision, the patient has some flexibility in the total prescription costs and thus their out-of-pocket costs. They can choose branded or generic drugs, or they can potentially ask their doctor to prescribe alternative drugs in a class of drugs that treat their medical shock. I assume that beneficiary spending, conditional on receiving an event, falls in some truncated distribution. Willingness to consume given a medical shock is s, and the dollar amount of prescriptions patients actually purchase is censored at 0, and the probability $Pr(s > 0) = qPr(B(s|X) > PMC(s, Z_w|X))$. The total payment amount s depends on the patient's observable and unobservable characteristics X, but also depends on the region r of the beneficiary's insurance plan and the distance in spending to them. Changes in a patient's total spending between insurance regions is meant to capture changes in the patient's expectations of or response to out-of-pocket marginal price changes.



From what has been laid out so far, it is clear why a censored regression spending model such as a Tobit would be inappropriate. There are legitimate zeros when beneficiaries do not have prescriptions to refill or health events, where the benefit of any drug spending is minuscule and the cost would require a doctor's visit for a new prescription.

Rather than estimating the more complicated model of both the choice to spend and spending amount, this paper focuses on the empirical probability of observing non-zero spending in a week. This probability is represented by $1 - p' = 1 - q + qP(s \leq 0)$ the probability of observing zero spending either because the beneficiary did not have a claim, or because she had a claim and she chose not to fill the claim.

This paper takes a reduced-form fixed-effects approach that differs from the prior literature in order to control for the heterogeneity in enrollees spending patterns shown in Figure 6. Einav et al. (2015) and other papers take a structural approach that imposes substantial assumptions on beneficiary behavior. Along with a dynamic optimization model, Dalton et al. (2015) run simple linear regressions with fixed effects to examine the flat effect of being within $110 of and in the donut hole on individual's average weekly spending (and other measures of spending). This paper improves upon their reduced-form approach by applying it to a four-year panel of observed weekly spending patterns over a much larger number of beneficiaries to reduce the negative dynamic panel bias with fixed effects (see Section 5.3). Also, rather than using a single linear indicator of being near the coverage gap, this paper's analysis allows for a flexible cubic spline to characterize beneficiary's spending patterns. This is described in detail in Section 5.1.

Among the reduced-form literature, other papers have handled the heterogeneity in other ways. Kowalski (2014) uses a quantile regression with an instrumental variable to analyze the change in people's healthcare spending when their year-end marginal prices change due to accidental (and assumed exogenous) injuries to family members. Since Medicare consists of only individual plans, this paper does not use a similar instrumental variable approach. Abaluck et al. (2015) effectively net out individual fixed effects by levering their panel data and observing the difference in individual spending due to plan changes between different years. However, their analysis is purposefully focused on spending for individuals who are unlikely to cross coverage regions and is thus limited at spending kinks. Joyce et al. (2013) compares the difference in the spending patterns of patients who have standard Medicare Part D Plans with plan non-linearities with the patients who receive low-income subsidies (LIS) and thus do not have significant coverage gaps. However, using the spending patterns of the LIS as a baseline comparison group for non-LIS patients may ascribe inherent differences between the groups to the plan coverage structure.

5.1 Estimation

In order to identify how the probability of claims occurring responds to the nonlinear marginal prices of the Medicare Part D nonlinear contract, this analysis takes a fixed-effects-regression approach with a dynamic panel.

Model 1 is represented by Equation 1, a linear probability model and the main estimation approach for this paper. Suppose Medicare Part D individual claims are grouped on a weekly level with spending s_{iyw} in year y and week w. Let the occurrence of spending $o_{iyw} = I(s_{iyw} > 0)$

be a binary variable. Let

$$o_{iyw} = \alpha_i + \boldsymbol{\gamma}\mathbf{X}_{iy} + f(Q_{iyw}, Z_{iyw}, \tilde{Z}_{iyw}) + \tau_y + \varepsilon_{it}. \tag{1}$$

The variable $Z_{iyw} = \sum_{u=1}^{w-1} s_{iyw}$ is a measure that represents the cumulative total expenditures within year y up until week w, and the variable $\tilde{Z}_{iyw} = \sum_{u=1}^{w-1} OOP_{iyw}$ is a measure that represents the cumulative total out-of-pocket expenditures in the same time frame. The variable $Q_w \in \{1, 2, 3, 4\}$ indicates whether a week is in the first, second, third, or fourth set of 13 consecutive weeks in a year (simplified to be called quarter variables). While the data is observed on a weekly level and the heuristic in Section 4 was also presented on a weekly level, this estimation aggregates the time fixed-effects on the quarter level. The variable α_i is an individual-fixed-effect constant across years and nests any gender, race, and age in 2009 information about the beneficiary. $\mathbf{X}_{iy}$ is the set of year-varying individual demographics that include beneficiary RxHCC 2011 demographic and risk scores for that year that are based off of their known health conditions from the prior year.[12] The purpose of the risk scores are to reimburse prescription spending, and as such they are an important measure to capture any between year changes in beneficiary's probabilities of spending. The variable τ_y is the year fixed effect y that is the same for all individuals.

The variables Z_{iyw},$\tilde{Z}_{iyw}$, and Q_{iyw} enter the estimation through a flexible functional form:

$$f(Q_{iyw}, Z_{iyw}, \tilde{Z}_{iyw}) = \sum_{q=1}^{4} \sum_{r \in \mathcal{R}} I(Q_{iyw} = q) \left(\eta_{qr} + g(dZ^r_{iyw}) + \tilde{g}(d\tilde{Z}^r_{iyw}) \right) \tag{2}$$

This function f is parameterized as a piecewise function of 12 restricted cubic splines for each of the two cumulative spending measures over the interaction of four quarters of the year and three separate spending regions: ICR, donut hole, and catastrophic region. Each cubic spline is a natural spline and has 3 knots located at the 10, 50, and 90 percentiles suggested by Harrell (2001).[13] The percentiles are presented in Table 7. The cumulative total and cumulative out-of-pocket expenditure measures both define the spending regions and allow us to understand how beneficiaries change their spending patterns as they approach the regions across the year.

[12] While the demographics score is a function of an individual's fixed characteristics, it is a non-linear function and thus still included in this analysis. Within a year, the demographics score plus the risk score is a relative measure of the riskiness of an individual compared to others in Medicare Part D. People who have a demographics plus risk score equal to 1 are considered to have an average reimbursement liability.

[13] Section 5.4.1, discusses how this specification was selected for over alternative cubic splines with 3 knots and one with 4 knots.

The limits that define the spending regions, $\overline{ICL}_y$ and $\overline{OOPT}_y$, change every year. Hence, in order to study the response as beneficiaries reach these limits, define $dZ_{iyw} = Z_{iyw} - \overline{ICL}_y$ and $d\tilde{Z}_{iyw} = \tilde{Z}_{iyw} - \overline{OOPT}_y$. The cumulative spending measures are denoted within each region $r \in \mathcal{R}$ as dZ^r_{iyw} and $d\tilde{Z}^r_{iyw}$, where the regions are defined as functions of both the cumulative total spending and the cumulative out-of-pocket spending.

$$r = \begin{cases} ICR, & \text{if} \quad dZ < 0 \quad \& \quad d\tilde{Z} < 0 \\ DonutHole, & \text{if} \quad dZ \geq 0 \quad \& \quad d\tilde{Z} < 0 \\ Catastrophic, & \text{if} \quad dZ \geq 0 \quad \& \quad d\tilde{Z} \geq 0 \end{cases} \tag{3}$$

Further the functions g and $\tilde{g}$ denote the cubic splines for Z and $\tilde{Z}$ respectively.[14]

This functional form allows the slope (and form) of the relationship between the cumulative spending measures and the beneficiary weekly claim probability to vary separately in each of the 12 region-quarter grids. It assumes that the relationship in f is the same across all four years, with the year effects only altering the level of spending across all beneficiaries through τ_y in Equation 1. Alternative models were considered such as ones that included an individual-year fixed effect for more flexibility. However, such models were rejected in order to mitigate the potential bias with dynamic panels and fixed effects. See the discussion in Section 5.3.

Table 7: Percentile values

	dZ			$d\tilde{Z}$		
Percentile	ICR	Coverage Gap	Catastrophic	ICR	Coverage Gap	Catastrophic
5	-2,849	48	3,777	-4,674	2,976	11
10	-2,827	100	4,034	-4,593	-3,695	20
35	-2,565	432	5,258	-4,443	-3,848	84
50	-2,340	717	6,373	-4,342	-2,933	145
65	-2,565	1114	8,362	-4,233	-2,561	252
90	-981	2426	24,435	-3,882	-1,349	1,074
95	-577	2977	38,488	-3,720	-835	1,768

Note: Shows select percentile values of dZ and $d\tilde{Z}$ in each coverage region. These percentile values are used in determining the location of the knots in the estimation results.

[14] The function differs for Z and $\tilde{Z}$ because of the exact placement of the knots and thus the spline function notation differs slightly. Let k_i, $i = 1, 2, 3$ be the knot values, then the $g(\mathcal{V})$ function is a linear regression of $V_1 = \mathcal{V}$ and $V_2 = \frac{(\mathcal{V}-k_1)^3_+ - (k_3-k_2)^{-1}\{(\mathcal{V}-k_2)^3_+(k_3-k_1) - (\mathcal{V}-k_3)^3_+(k_2-k_1)\}}{(k_3-k_1)^2}$

5.2 Baseline Estimation Results

Figure 9 and Table 8 present the results of Model 1's linear probability fixed-effects estimation using a cubic spline. Errors are clustered at the individual level. Rather than displaying the coefficient estimates of the piecewise f function, the figure displays the predicted values from those estimates that describe the effect the cumulative spending values on the probability of filling a prescription claim in a week. The remaining coefficient estimates for the beneficiary risk scores and year (γ and τ_y) are included in Table 8.

The main takeaway from this analysis is that patients do anticipate and respond to the pricing incentives in the Medicare Part D contract. This analysis indicates that on average patients decrease their spending patterns far in advance of entering the donut hole and begin to increase it prior to exiting; however, these changes are smooth except in the end of the year. This analysis did not find the behavioral break at the entrance to the donut hole early in the year that previous literature has suggested.

Table 8: Model 1 estimates: Impact on the probability (%) of a claim in a week

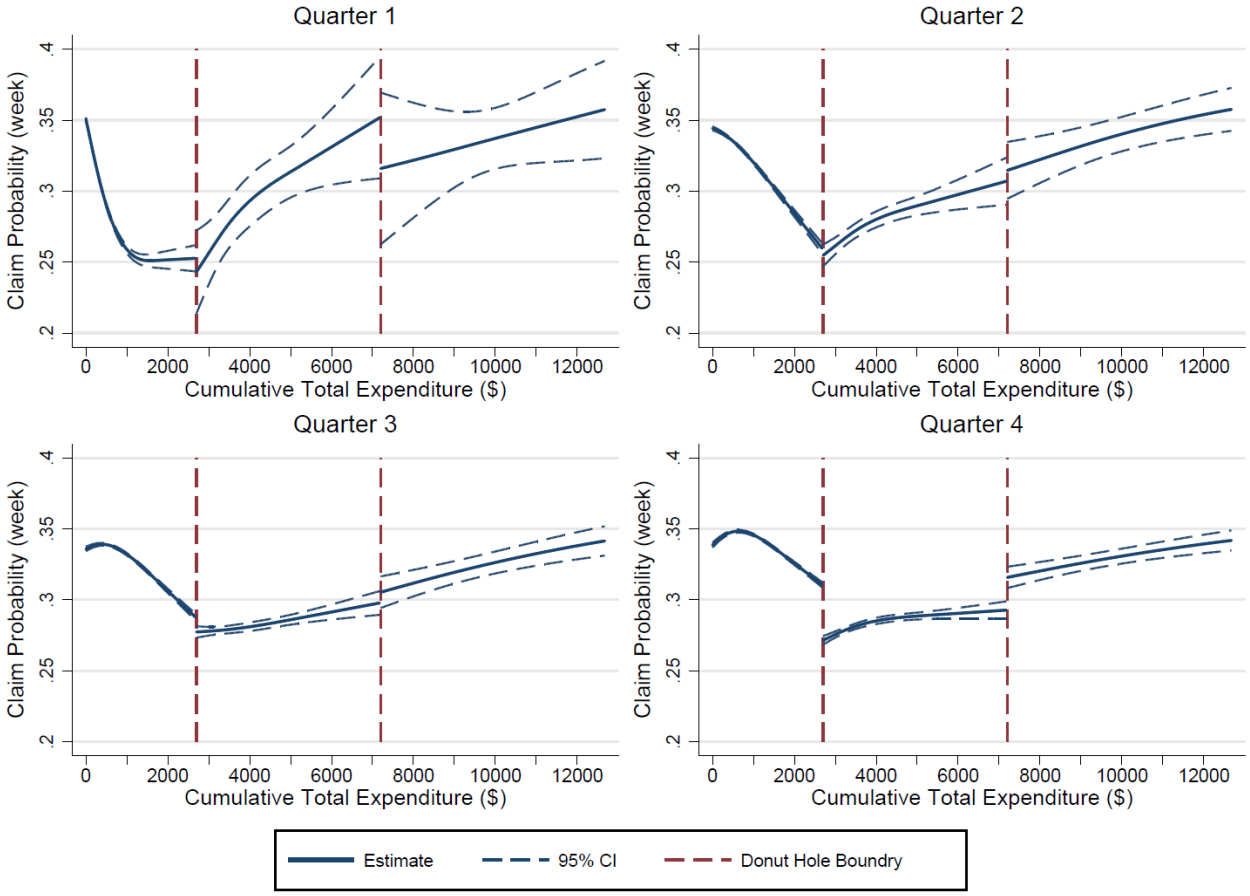
Coefficient	Estimate	Std. Error
RxHCC Risk Weight	3.57	0.11
RxHCC Demographic Weight	-26.71	6.12
2010	0.30	0.05
2011	0.61	0.05
2012	1.42	0.08
N	18,585,632	

Note: The estimated coefficients on the risk scores $\mathbf{X}_{iy}$ and year-time dummies τ_y from the fixed-effects panel linear probability regression of Equation 1 on the beneficiaries in the "No Deductible" sample. All estimates in the table are significant at less than the .1% level.

The coefficients of individual risk and demographics scores are highly significant, but the scale of the effects may be economically small. On average for the predicted sample, the point estimates of the probability of a claim occurring in a week all fall in the range of around 25-35%. If these predicted values of weekly claim probabilities are extrapolated multiple weeks, this translates into an economic meaning of visiting the pharmacy once every 4 weeks versus once every 2.9 weeks. These differences would be higher for individuals with lower risk scores and the differences would be smaller for high-risk types.

Holding individuals and other characteristics constant, if a patient's risk score increases by one standard deviation (approximately 0.3), the probability of claims being observed in a week is expected to increase by approximately 1%. Riskier patients result in higher prescriptions as expected. Also, if a patient's demographic score increases by one standard deviation (approx-

Figure 9: Linear: Probability of Claims Occurring in a Week, Conditional Year-Cumulative Total Spending



Note: Displays the predicted values of the claims occurrence probability $\hat{o}_{iyw}$ from the fixed-effects panel regression of Equation 1 on the beneficiaries in the "No Deductible" sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values.
The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure, which increases with the cumulative total expenditure at the average coinsurance rates presented in Table 4. The probabilities are predicted assuming that beneficiary has the sample average risk and demographic scores from 2009, 0.4753 and 0.4196 respectively. Each line segment represents a prediction made assuming the year is 2009 with 2009 spending limits and across time average individual risk and demographics. Since the estimates are from a linear fixed-effects-regression, the predicted values are plotted assuming the mean fixed-effects coefficient, which is zero.

imately 0.01), the probability of observing at least one claim in the week is expected to drop by 0.25%. At first glance the negative coefficient on the demographics score appears counter intuitive; however, this can be explained since the demographics score is a nonlinear function of age, race, and other fixed characteristics. Holding all else constant, the demographics score actually decreases as age increases. The positive and significant coefficients on the year fixed effects also capture both the effect of aging on the sample population and any year trends that

lead to increases in medication purchasing frequency. Individuals who are older and in 2012 are 1.4% more likely to have claim in a week than in 2009. Because this is a fixed-effects-regression on a balanced panel in all four years, this regression does not differentiate the effects of aging versus the year fixed effects.

Figure 9 depicts in four panels the predicted probability of a claim occurring in a week in each of four quarters (13 week periods) in the year as the beneficiary's cumulative total spending (Z_{iyw}) changes.[15] Within each panel, the predicted probabilities are displayed as a piecewise function with the first segment representing the initial coverage region, the second the coverage gap, and the third the catastrophic region. Further these estimates are predicted assuming other coefficients are held constant with year 2009 fixed effects, average demographics and risk scores for the sample, 2009 ICL and OOPT limits, and average coinsurance rates in each spending region. Because the prediction assumes fixed coinsurance rates, the cumulative total is a one-to-one mapping of the cumulative out-of-pocket expenditures. The example out-of-pocket total of $4,350 translates to a cumulative total expenditure amount of $7,200.

Focusing first on the Quarter 1 panel, the probability of a claim occurring in a week are tightly estimated for low cumulative expenditures and decreases from 35% to 25% prior to spending even $1,000, indicating that high spenders may anticipate entering the donut hole. The point estimates as individuals cross from the ICR to the coverage gap do not indicate a discontinuity and the probability of claim occurrences increase prior to entering the catastrophic region. The point estimate at the entrance to the catastrophic region indicates that individuals decrease their spending. The confidence interval around these estimates do become progressively wider as the cumulative total expenditure values increase, because there are few individual-week observations at such high spending levels early in the year. Even with the wide confidence intervals, the claim occurrence probability is lower for individuals who just entered the donut hole than for individuals who have spent a thousand more. The predicted values show a drop in claim occurrence frequency when exiting the donut hole into the catastrophic region, however, the confidence intervals around these predictions are wide at approximately 10 percentage points and are not significant.

The behavior observed in Quarter 2 and 3 are similar to Quarter 1. There is a significant decrease and then an increase in the claims occurrence probability in the ICR and donut hole, respectively, as the cumulative total expenditure increases. These changes are smooth, and there is no evidence for a discontinuity in the spending measure at either the entrance to the donut hole

[15] The cumulative out-of-pocket expenditures also change here but this is not displayed.

nor the start of the catastrophic region. The predicted values are generated off of more precise estimates in the later quarters with a maximum confidence interval around the entrance to the catastrophic region of 5% and 2.5% in the second and third panels respectively. The predicted values in Quarter 2 and 3 differ from Quarter 1 in that the decrease in the cumulative total spending in the ICR is not as steep, and the lowest probability occurs at a higher cumulative total expenditure in the later periods.

In Quarter 4, the probability of spending is still in the 25% to 35% range for cumulative total expenditures below $12,000, but the predicted spending values differ at specific cumulative total spending and are estimated with very tight confidence intervals. They indicate sharp discontinuities in spending at the entrance to the donut hole with a 4% drop in the probability of claims occurring in a week, and they indicate a 3% increase in claims occurrences upon entering the catastrophic region. The slope of the claim occurrence probabilities in the ICR and donut hole regions are a much lower magnitude and less economically significant even as they are estimated with more precision as time progresses. In addition, the slope within the ICR region appears to be convex.

This pattern supports the theory and simulations from Section 2, which predict that discontinuities in the probability of spending are most likely in the last time period because there is less ambiguity as to the beneficiary's end-of-year marginal costs. The lower magnitudes of the changes in the claim occurrence probability within each coinsurance region also reflect the theory that the perceived marginal cost should approach the actual coinsurance rates which are flat within a region. The convexity in the ICR that begins in Quarter 3 and is more evident in Quarter 4 was not predicted from the heuristic marginal price applied to data, but it is consistent with theory. At the end of the year, beneficiaries would have more information on their end-of-year region, and in the last months of the year, enrollees in the low values of the ICR may decrease their perceived marginal price as the probability of entering the donut hole decreases (and then increase their spending).

5.3 Bias in a Dynamic Panel with Fixed Effects

The estimates of fixed-effects models applied to a dynamic panel are known to be biased if the number of time periods (T) is small and the cross sectional size of the panel (N) is large (Nickell, 1981).[16] This model is affected by Nickell bias because the cumulative total and out-of-pocket

[16] Nickell (1981) highlighted that the time demeaning operation of fixed effects in a dynamic panel data model $y_{it} = \alpha_i + \beta y_{it-1} + \epsilon_{it}$ leads to a transformed regression model $y_{it} - \bar{y}_i = \beta(y_{it-1} - \bar{y}_{it-1}) + (\epsilon_{it} - \bar{\epsilon}_i)$ where the $\bar{y}_i, \bar{y}_{it-1}, \bar{\epsilon}_i$ indicate time averages. The error terms $(\epsilon_{it} - \bar{\epsilon}_i)$ and regressors $(y_{it-1} - \bar{y}_{it-1})$ are correlated even as

expenditures Z_{iyw} and $\tilde{Z}_{iyw}$ are both functions of the lagged dependent variable, the claim observation o_{iyw}. The paper does take precautions to reduce the influence of this bias on the analysis. First, the analysis spans beneficiary behavior over four years, or $T = 208$ weeks, a longer time span than normally cited in the literature. As Nickell (1981) demonstrates with a simple lag, as $N \to \infty$, the inconsistency of the estimated lagged parameter is of the order $1/T$. So while the number of beneficiaries in the "No Deductible" sample $N = 89,354$ is large, the potential bias with a larger T is greatly reduced.

The classic Nickell bias as it applies to the coefficient on the lag of the dependent variable is negative, and the coefficient on the cumulative spending measures in each region would be similarly negatively biased. To verify, the direction of the bias as it applies to the specific types of lags in this estimation are simulated with noise and presented in the Appendix Section D. Any bias in the estimates are more likely to be observed at low cumulative spending levels and leads to a more negative slope. The bias alone does not lead to discontinuities in the effect of the cumulative spending measures on the claim occurrence probability.

Other estimation approaches that would have circumvented the bias problem of the fixed-effects approach such as that presented by Anderson and Hsiao (1982) or Holtz-Eakin et al. (1988) (popularized and more commonly known as Arellano and Bond (1991)) have their own problems. These methodologies use a first-differences approach to net out the fixed-effect α_i and then use further lags of the lagged variable (in this case the Z_{iyw} and $\tilde{Z}_{iyw}$) as instruments in a 2SLS and GMM style estimation respectively. While not biased, Anderson and Hsiao (1982) does not use all available data and can result in imprecise estimates. Further, because of the large size of the data, the number of interactions of Z_{iyw} and $\tilde{Z}_{iyw}$ terms, and the long length of the panel, the Arellano-Bond methods become computationally intractable.

5.4 Robustness

I consider additional models to address some of the concerns of model specification and find results that support the initial findings. Alternative splines of the cumulative total and cumulative out-of-pocket spending measures are included as a falsification test to verify that the functional form definition is not driving the result. I also consider an alternative spending measure (the count of the number of claims in a week) and present both the results of a linear and Poisson model as an alternative to the linear probability model.

$N \to \infty$.

5.4.1 Falsification Test

In order to more fully verify the results of the estimation, this paper also conducted falsification tests with different spline functions and discontinuities allowed at different points.

Model 1 used a cubic spline with 3 knots located at the 10, 50, and 90 percentile values for dZ the cumulative total expenditures (centered on the ICL) and for $d\tilde{Z}$ the cumulative OOP expenditure (centered on the OOPT). One potential concern is that the cubic spline with 3 knots may over-smooth the beneficiary's response especially close to the region boundaries. Two alternative splines were tested in models called False 1 and False 2. One included three knots located at the 5, 50, 95 percentile values for both regressors and another with four knots at located at the 5, 35, 65, and 95 percentile values for both regressors. The knot values are displayed in Table 7. Model 1 was selected over these other two models by both the AIC and BIC criteria.

Another approach to verifying the accuracy of Model 1 was taken by introducing additional potential discontinuities in the domain. In False 3, the following linear probability model was conducted where the initial coverage region is broken down into two separate regions $ICR1$ and $ICR2$. Then the cumulative measures that enter Equation 1 and 2 are dZ^r_{iyw} and $d\tilde{Z}^r_{iyw}$ where $r \in \mathcal{R}'$ such that

$$r = \begin{cases} ICR1, & \text{if} \quad dZ < -1000 \quad \& \quad d\tilde{Z} < 0 \\ ICR2, & \text{if} \quad -1000 \geq dZ < 0 \quad \& \quad d\tilde{Z} < 0 \\ DonutHole, & \text{if} \quad dZ \geq 0 \quad \& \quad d\tilde{Z} < 0 \\ Catastrophic, & \text{if} \quad dZ >= 0 \quad \& \quad d\tilde{Z} \geq 0 \end{cases} \tag{4}$$

Equation 1 is estimated with the regions in $\mathcal{R}'$ and standard errors clustered at the individual level.

Table 9 displays the estimated coefficient values for the year fixed effects and risk scores for False 1-3. The predicted values of the claims occurrence as the cumulative total and out-of-pocket expenditures change are displayed in Figure 10 for False 3. The figures for False 1-2 are displayed in the Appendix Tables B.4 and B.5 respectively.

The results from these falsification tests are all very similar to the results from Model 1 and did not have a large impact on the coefficient estimates or the predicted claims occurrence values as the cumulative expenditures changed. In False 3, while we allowed for a discontinuity

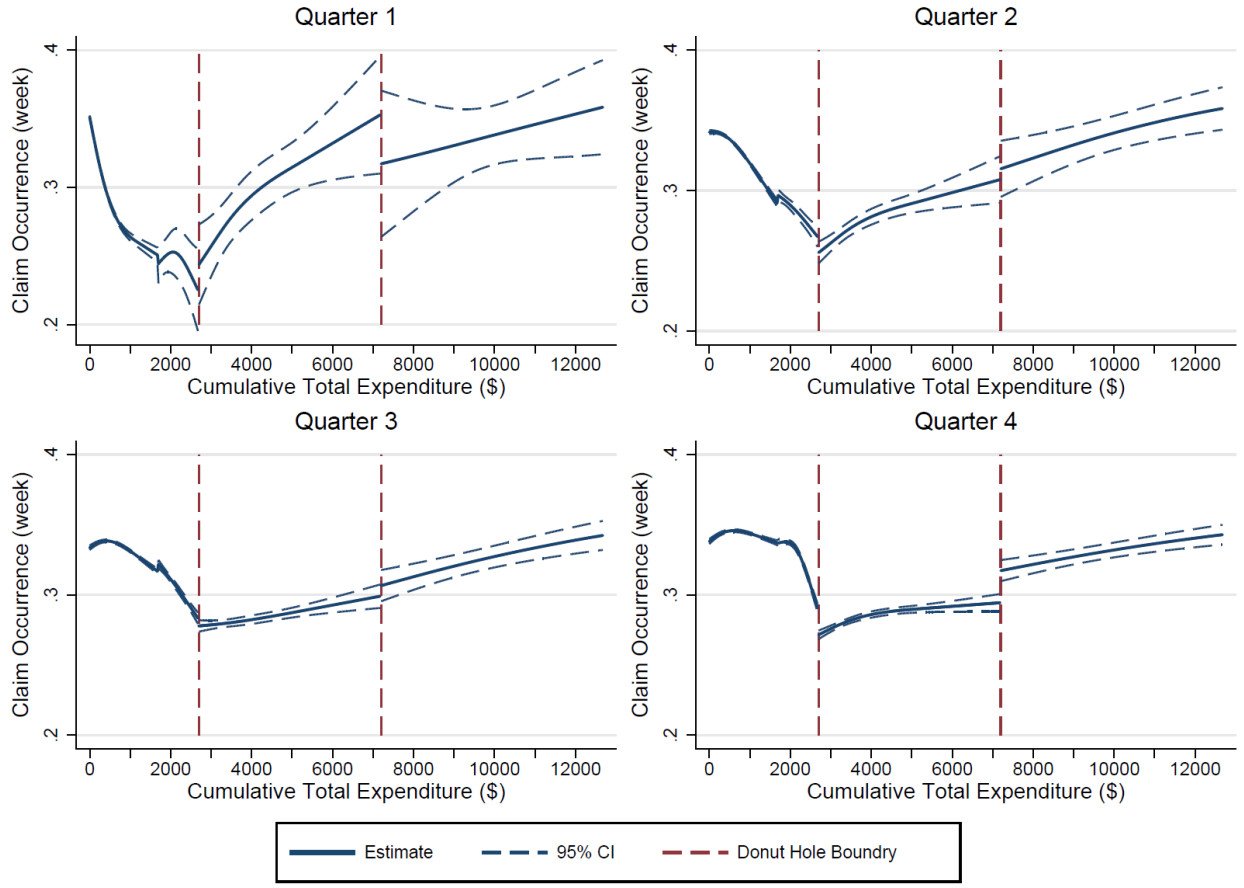
Table 9: Falsification 1, 2, 3: Impact on the claims occurrence probability (%)

	False 1		False 2		False 3	
Coefficient	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
RxHCC Risk Weight	3.57	0.11	3.57	0.11	3.58	0.11
RxHCC Demographic Weight	-26.63	6.12	-26.73	6.12	-26.58	6.12
2010	0.32	0.05	0.31	0.05	0.23	0.05
2011	0.63	0.05	0.61	0.05	0.53	0.05
2012	1.41	0.08	1.43	0.08	1.27	0.08
N			18,585,632			

Note: The estimated coefficients on the risk scores $\mathbf{X}_{iy}$ and year-time dummies τ_y from the fixed-effects panel linear probability regression of Equation 5 under different spline and region assumptions. In False 1&2, the spline of the cumulative total and cumulative out-of-pocket expenditures are defined by a cubic spline with four and three knots at the 5, 35, 65, 95 and 5, 50,95 percentiles respectively. The percentile values are shown in Table 7. In False 3, the spline is defined as in Model 1, but it introduced an additional potential discontinuity at $1000 less than the initial coverage limit (ICL) in the estimation. All estimates in the table are significant at less than the .1% level.

at ICL-$1000, neither a statistically significant nor economically significant discontinuity was estimated at this point. One minor but noticeable difference between these models and Model 1 is that in Quarter 1, the point estimates of the predicted values of the claims occurrence at the high values of the ICR are slightly higher than in Model 1, but they still fall within the confidence interval of the original estimates. Also from False 3, we observe that the predicted values in $ICR2$ do have a larger confidence interval than the predictions over the same domain in Model 1. This pattern does highlight the low number of observations of such high cumulative spending in the first quarter of the year.

Figure 10: Linear Falsification: Probability of Claims Occurring in a Week



Note: Displays the predicted values of the claim occurrence probability $\hat{o}_{iyw}$ in a week from the fixed-effects panel regression of Equation 5 under the assumptions of False 2 on the beneficiaries in the "No Deductible" sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values.
The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. See Figure 9 for the exact values used to generate the prediction.

5.4.2 Fill Count

The original dependent variable in Model 1 is the binary variable of the occurrence of any claim in a week and was chosen to highlight the extensive margin of the beneficiary's prescription purchasing decision. However, beneficiaries often have multiple health conditions and multiple prescription claims that they have the choice to fill in a week, so I also study the results using an alternative consumption measure of the number of prescriptions filled in a week.

Alternative specifications use the number of claims observed in a week n_{iyw} as the dependent variable in both a linear regression in Model 2 and one that assumes a Poisson regression in

Model 3. The fixed-effects linear and Poisson regressions face the same critique due to dynamic panel bias as the prior Model 1 linear regression.

The linear model is described in Equation 5 while the Poisson model in described in Equation 6 with a log-linear regression model where λ represents the mean number of claims.

$$n_{iyw} = \alpha_i + \boldsymbol{\gamma}\mathbf{X}_{iy} + f(Q_{iyw}, Z_{iyw}, \tilde{Z}_{iyw}) + \tau_y + \varepsilon_{it} \tag{5}$$

$$\ln \mathbb{E}(n_{iyw}) = \ln \lambda_{iyw} = \alpha_i + \gamma\mathbf{X}_{iy} + f(Q_{iyw}, R_{iyw}, Z_{iyw}) + \tau_y \tag{6}$$

For both models, the f retains the same functional form as in Equation 2 and the $\mathbf{X}$ variables are the same as in Model 1.

Table 10: Model 2 and 3 estimates: Impact on the number of claims in a week

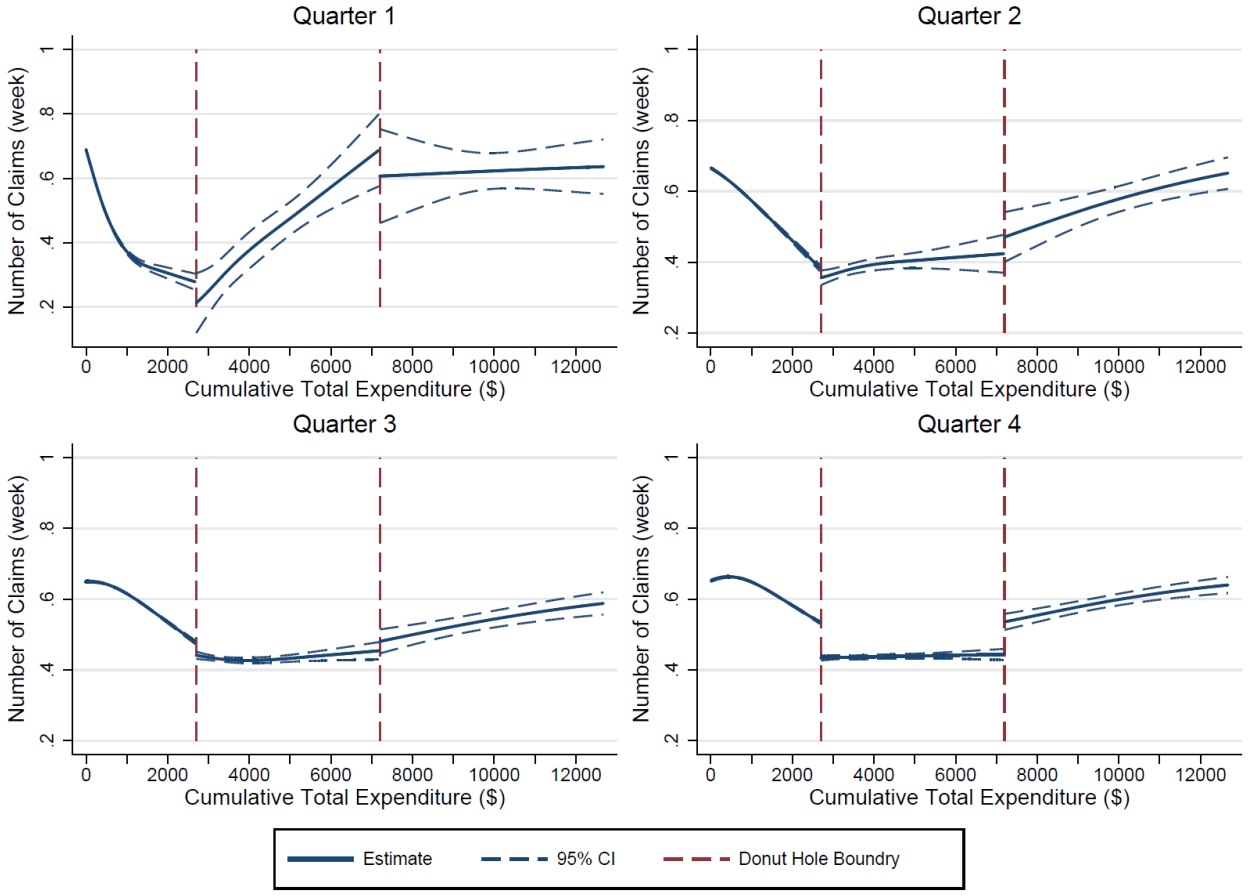
	Linear Estimates (x100)		Poisson Estimates	
Coefficient	Estimate	Std. Error	Estimate	Std. Error
RxHCC Risk Weight	8.37	0.26	0.12	0.004
RxHCC Demographic Weight	-34.31	13.6	-0.70	0.222
2010	-0.41	0.1	-0.0017	0.001
2011	0.25	0.12	0.0103	0.002
2012	1.11	0.02	0.0221	0.003
N		18,585,632		

Note: The estimated coefficients on the risk scores $\mathbf{X}_{iy}$ and year-time dummies τ_y from the fixed-effects panel linear probability regression of Equation 5 and the Poisson regression of Equation 6 on the "No Deductible" sample. All estimates in the table are significant at less than the 5% level. While the estimates from both models are presented in the same table, the interpretation for the Poisson estimates is multiplicative of the exponential of the estimate.

Qualitatively, the estimates are very similar to those estimated from Model 1 where the dependent variable was the occurrence of claims. In Model 2-3, the level of claims incidence is much higher, which reflect the fact that many beneficiaries file multiple claims in a week. The point estimates in both Model 2 and 3 are very similar and range from approximately 0.3 to 0.7, translating to a claim occurring every three weeks to every 10 days. The estimates from Model 3 are slightly less extreme than Model 2, particularly in Quarter 1. The confidence intervals in Model 1 and 2 are similarly scaled and precise (except in the high values of Quarter 1), but the confidence intervals from Model 3's Poisson regression are much wider than those from the linear case averaging about a 0.2 band around the estimates.

The results from Model 2 and 3 support the findings from Model 1, even though a different spending measure is used. While there are changes in the average number of claims in each week as the cumulative spending measures increase, these changes are smooth except potentially in

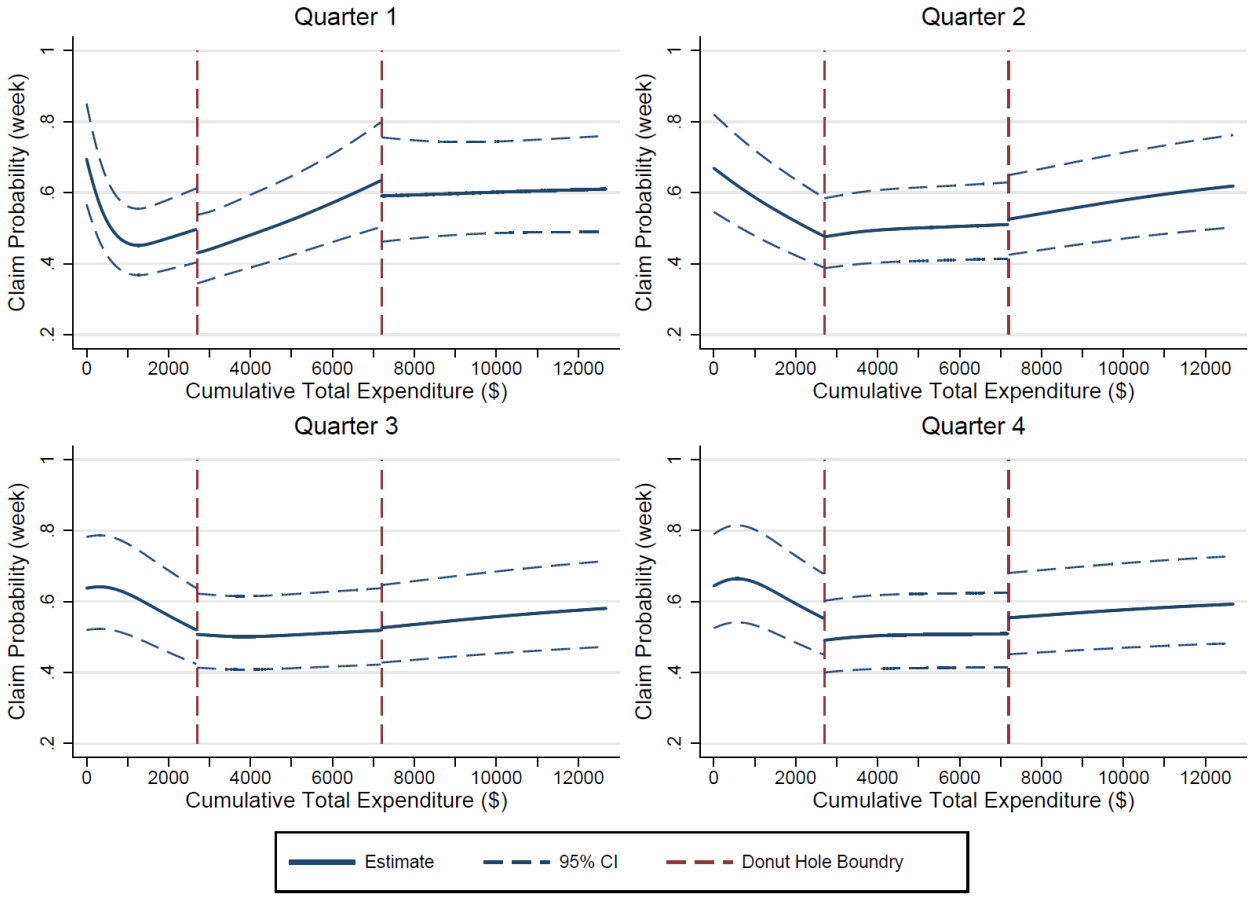
Figure 11: Linear: Incidence of Claims Occurring in a Week



Note: Displays the predicted values of the average number of claims in a week $\hat{n}_{iyw}$ from the fixed-effects panel regression of Equation 5 on the beneficiaries in the "No Deductible" sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values.
The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. See Figure 9 for the exact values used to generate the prediction.

Quarter 4. Claims decrease prior to entering the donut hole and increase prior to entering the catastrophic spending region, though both changes have lower magnitudes in the later quarters of the year.

Figure 12: Poisson: Incidence of Claims Occurring in a Week



Note: Displays the predicted values of the average number of claims in a week $\hat{n}_{iyw}$ from the fixed-effects panel regression of Equation 6 on the beneficiaries in the "No Deductible" sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values.
The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. The values used are similar to those from Figure 9; however, it differs in that the fixed effects used to generate these values is not zero. Unlike the linear model, the mean fixed-effects coefficient in the Poisson model is not zero, so the predicted values are scaled by the exponential of the mean fixed effect estimated $\alpha_i = -0.145$.

5.5 Discussion

The conclusion from the estimation of Model 1-3 is on average, individuals behave more optimally than some of the literature has found, because on average enrollees anticipated and responded to the pricing structure in Medicare Part D. This analysis is aggregated over many heterogeneous actors and while the fixed-effects approach helps to control for individual unobserved heterogeneity, it may still complicate the analysis. That is to say, not that any one individual behaves optimally, but there does not appear to be a systematic decrease in spending in this sample prior to entering the coverage gap. Further, it is also important to understand

why these results differ from some of the previous findings.

The results are aggregated over various heterogeneous variables. While the analysis included year fixed effects to control for differences in the levels of spending across year, slopes were not allowed to vary by year. It is possible that the slopes of the beneficiary claims response within the coverage gap differ across years, because the beneficiary plans differ across the years. Their marginal costs in 2011 and 2012 coverage gap are significantly closer to the ICR coinsurance levels, and thus this should result in fewer changes in spending. The expectation then by the estimation presented in Equation 1 and 2 across all four years is that the predicted f is "too flat" to describe 2009-2010 and not flat enough to describe 2011-12.

Further, in analyzing the f function estimates, it is important to remember that while the fixed effects may "net-out" the heterogeneity in the probability of spending between individuals, it does not control for the heterogeneity in health shocks or expectations that may exist at any point in time and cumulative expenditure level. This means that there are potentially heterogeneous response to the nonlinear contract. For example, in any quarter, but particularly earlier in the year the estimates for the change in the claims occurrence probability include both individuals who do not expect to end the year in the donut hole and decrease their spending as the probability of that outcome increases, and individuals who expect to end the year in the catastrophic region and may increase their spending as this outcome becomes more certain. This papers conclusions on not finding a discontinuity at the region boundaries is still reasonable even in light of this potentially heterogeneity. There is not a reasonable prior to think that individuals would have a discontinuous increase in their claims occurrence that perfectly offset another set of individuals discontinuous decrease.

Another source that could increase the heterogeneity in the response to the cumulative spending measures is the experience and knowledge individuals have with nonlinear pricing structures in general and each year's Medicare Part D plan specifically. Individuals with more experience (and no surprises) would be expected to have more constant spending patterns that result in a flatter overall f response curve. Further work should be done to understand how experience impacts the beneficiary response to the contract features.

Experience, sample selection, and estimation methods could be why this paper's estimates on finding a discontinuity at the donut hole differed from Dalton et al. (2015). Individuals in their dataset come from one year of observation in 2008 and are not on traditional Medicare and are known to be richer and have higher spending than Medicare patients. The sample in this paper joined prior to 2008, and we observe four years of their choices. The fact that this papers

sample is from later years also increases the probability that they either had more personal experience with or opportunities to learn about Medicare Part D which started in 2006.

6 Conclusion

This paper builds on the existing literature on beneficiaries' dynamic response as they approach the many discontinuities in the Medicare Part D pricing structure. Throughout all health insurance, the government and insurers have significant control over the cost-sharing features that are responsible for non-linear pricing, and with the rise in health-care costs, these institutions are more likely to use them as cost-control measures. Unfortunately, the effect of these cost-control measures on beneficiary behavior and health is not fully explained. While the literature has identified sharp drops in spending particularly at the Medicare Part D coverage gap, trying to explain this behavior using time-discounting models have resulted in discounting estimates far lower than the broader economics literature.

The first main contribution of this paper is its discussion of an expected price model that uses the objective probability distributions of beneficiaries' end-of-year prices given their spending probabilities. The key takeaway from this heuristic model is that if beneficiaries know the objective probabilities of ending the year in each region for the population, the expected marginal price that a beneficiary responds to can be constructed and be used to make purchasing decisions. Because the end-of-year region probabilities have relatively smooth transitions (in all but approximately the last 10 weeks of the year), beneficiaries' marginal price and then spending should also be smooth (except for the last weeks). Even if beneficiaries have inaccurate beliefs on their objective end-of-year probabilities or are present-biased, as long as they update those beliefs in each time period, a heuristic marginal price would not generate sharp spending changes unless there were sharp changes in probabilities.

A second significant contribution of this paper is the graphical representation of beneficiaries' claims rates conditioning on the cumulative sums of their total spending. Using separate linear probability and Poisson regressions with individual fixed effects to control for heterogeneity in base levels of spending, this paper illustrates beneficiary claims rates in a way that allows direct visual comparison with their heuristic marginal spending.

The estimation finds that there are significant changes to beneficiaries' claims rates, some of which are consistent with the predictions of the heuristic marginal price, but that those changes may not be economically significant. The lowest amount of claim rates in each quarter of the year broadly matches the cumulative total expenditures values that produced the lowest

expected marginal prices. The changes in claims rates for a predicted beneficiary with average risk and demographic scores indicate that the economic magnitude of the claims occurrences are on the scale of filing claims every 3 weeks versus filing claims every four weeks. Further work to understand the welfare consequences of these reductions would be to analyze whether the claims changes were by discontinuing drugs entirely or just small delays in going to the pharmacy for refills.

This finding in the paper differ empirical results in the prior literature. This difference may be due in part to the sample selection. Because this sample involved individuals who retained similar plan structures through all four years, they are mechanically more likely to have experience with Medicare Part D, their plan structure, and their prescription needs than individuals in the papers mentioned in the literature. Further, individuals who do not switch between plans with different limits may be different onto themselves as people who have high inertia and do not switch plans or just happen to have have expected spending amounts that align well with their chosen plans. Further work could explore whether experience with Medicare Part D promotes individuals to exhibit more optimal behavior, because that would imply that informational and educational programs could promote that behavior. A more detailed subsample analysis may be warranted.

The role of heterogeneous types of responses to the coverage gap should also be considered in future work. While this paper controlled for heterogeneous levels of claim rates for individual beneficiaries, it is likely that due to random health shocks or experience with the Medicare Part D pricing schedules, individuals may separately increase or decrease their claims in a predictable way that adds noise to this paper's estimates.

Another challenge to this paper's research question was the large variety of plans and coinsurance rates offered from 2009-2012 and the policy changes introduced by the Affordable Care Act to fill in the donut hole. The lower coinsurance rates within the donut hole in 2011 and 2012 could be partially responsible for the more stable spending estimates that are found throughout this paper. An obvious related research project would be to study the actual impact of the Affordable Care Act's policy of filling in the coverage gap and subsequent health outcomes to determine whether it truly impacted beneficiary spending rates. The plan variety that was a challenge for the analysis in this paper, would be a boon for follow-up research projects.

References

(2016). The Medicare Part D Prescription Drug Benefit. Technical report, The Henry J. Kaiser Family Foundation.

Abaluck, J. and Gruber, J. (2016). Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance. *American Economic Review*, 106(8):2145–2184.

Abaluck, J., Gruber, J., and Swanson, A. T. (2015). Prescription Drug Use Under Medicare Part D: A Linear Model of Nonlinear Budget Sets. SSRN Scholarly Paper ID 2572135, Social Science Research Network, Rochester, NY.

Anderson, T. W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297.

Aron-Dine, A., Einav, L., Finkelstein, A., and Cullen, M. (2015). Moral Hazard in Health Insurance: Do Dynamic Incentives Matter? *Review of Economics and Statistics*, 97(4):725–741.

Baicker, K., Mullainathan, S., and Schwartzstein, J. (2015). Behavioral Hazard in Health Insurance*. *The Quarterly Journal of Economics*, page qjv029.

Baker, D. (2006). The Origins of the Doughnut Hole: Excess Profits on Prescription Drugs. Technical Report 2006-19, Center for Economic and Policy Research (CEPR).

Dalton, C., Gowrisankaran, G., and Town, R. (2015). Myopia and Complex Dynamic Incentives: Evidence from Medicare Part D. *Unpublished*.

Einav, L., Finkelstein, A., and Polyakova, M. (2016). Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D.

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.

Harrell, F. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Series in Statistics. Springer-Verlag, New York.

Ho, K., Hogan, J., and Morton, F. S. (2015). The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program. *Working Paper*.

Hoadley, J., Summer, L., Hargrave, E., and Cubanski, J. (2011). Understanding The Effects of The Medicare Part D Coverage Gap in 2008 and 2009. Technical Report 8221, The Henry J. Kaiser Family Foundation.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating Vector Autoregressions with Panel Data. *Econometrica*, 56(6):1371–1395.

Joyce, G. F., Zissimopoulos, J., and Goldman, D. P. (2013). Digesting the doughnut hole. *Journal of Health Economics*, 32(6):1345–1355.

Kowalski, A. E. (2014). Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care. *Journal of Business and Economic Statistics*.

MedPAC (2016). Prescription Drugs. In *Health Care Spending and the Medicare Program*, pages 170–171.

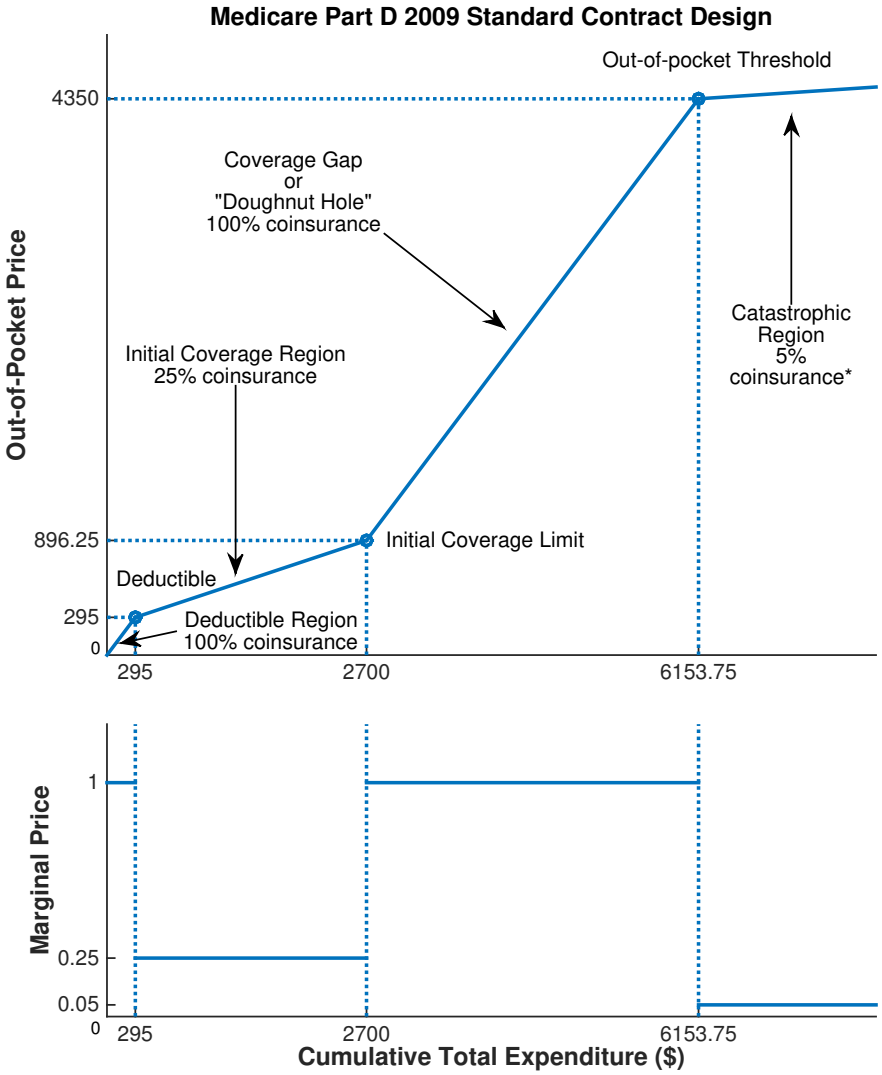
Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417–1426.

Polinski, J. M., Bhandari, A., Saya, U. Y., Schneeweiss, S., and Shrank, W. H. (2010). Medicare beneficiaries knowledge of and choices regarding Part D, 2005 present. *Journal of the American Geriatrics Society*, 58(5):950–966.

Zhang, Y., Donohue, J. M., Newhouse, J. P., and Lave, J. R. (2009). The Effects Of The Coverage Gap On Drug Spending: A Closer Look At Medicare Part D. *Health Affairs*, 28(2):w317–w325.

A Data Appendix

Figure A.1: Part D 2009 Coverage Regions



The figure depicts the nonlinear structure of the standard Medicare Part D benefit contract. Actual plans offered in 2009 were either actuarially equivalent or better. The premium or the amount the patient pays out-of-pocket for the benefit package is not displayed. The Total Expenditure includes the drug expenditure between the patient, insurance company, and Medicare, while the Out-of-Pocket (OOP) Cost only includes the patient's drug expenditure. The 5% coinsurance coverage in the catastrophic region is simplified for the figure. The actual 2009 coverage benefit requires beneficiaries to pay the maximum of either 5% the cost of the drug or $2.40 and $6.00 for a one-month supply of generic and branded drugs respectively. This means that patients may pay either the copay dollar amount or a percentage share of the drug price and the remainder is covered by insurance or the government. The bottom panel displays the marginal costs of the plan.

The reasons why beneficiaries are excluded from the full sample are listed in Appendix Table A.1 with the percentage of the 5% sample they encompass. Beneficiaries are limited to those who are 65 or older, are enrolled in Medicare PDPs in 2009-2012, and who originally and

are currently enrolled in Medicare through the Old Age and Survivors Insurance (OASI) and not for disability insurance or other qualifiers for Medicare. Similar to Einav et al. (2015), the sample also excludes individuals who are dual eligible for Medicaid financial assistance or receive other types of low-income subsidies (LIS) for premiums or cost-sharing. These individuals are excluded because they face very low cost-sharing and minor changes in their marginal costs. Even individuals who only receive premium subsidies are omitted, because they are more likely lower income and are more likely to be influenced by budget constraints. Joyce et al. (2013) used these LIS individuals as a control group for their analysis of the donut hole, but this paper takes a different approach. Further, the analysis of the paper also excludes individuals whose Medicare Part B claims indicate beneficiaries were in long-term care institutions (LTI) such as nursing homes in the prior year. The LTI beneficiaries are excluded on the chance that their Part D prescription purchases are managed by the nursing homes. In addition, the full sample excludes individuals who did not have a Medicare Part D plan for the full year ("Not Same Plan in 12 Months"), because those individuals would have varying time-frames over which their contracts run. See the Appendix for more detailed information on the subsample restrictions.

Further, beneficiaries may not have the same plan through the entire year if it is their initial year of enrollment, since patients are eligible to enroll for 7 months around their birthdays (three months before and after and including their birthday month).[17] Beneficiaries who died may also pass away mid-year. Beneficiaries who switch from Medicare Advantage (prescription coverage) to Medicare Part D, may also not be in plans for the full 12 months since the Medicare Advantage disenrollment period occurs within a year January 1-February 14, taking effect the first month after disenrollment (either February or March). Enrollees may also have switched plans mid-year under the special enrollment periods: after moving; losing current Medicaid/Employer/etc coverage; new creditable coverage options, specific plan changes, and other special circumstances.[18]

[17] If enrolled in the initial enrollment period, patients would be enrolled on the first day of their birthday month if enrolled prior. If sign-up occurs in the birthday month or in the three months after, coverage start is delayed 1-3 months after enrollment.

[18] https://www.medicare.gov/sign-up-change-plans/when-can-i-join-a-health-or-drug-plan/special-circumstances/join-plan-special-circumstances.html

Table A.1: Summary of 5% Sample of Medicare Data and Exclusion Reason 2009-2012

	(1) 2009		(2) 2010		(3) 2011		(4) 2012	
	mean	sd	mean	sd	mean	sd	mean	sd
Dual Eligible	0.2001	0.40	0.2009	0.40	0.2294	0.42	0.2271	0.42
Disabled or end-stage renal disease	0.1605	0.37	0.1700	0.38	0.1713	0.38	0.1704	0.38
Originally not OASI	0.2273	0.42	0.2386	0.43	0.2415	0.43	0.2426	0.43
Currently not OASI	0.1653	0.37	0.1789	0.38	0.1826	0.39	0.1825	0.39
Receive State Subsidy	0.1752	0.38	0.1792	0.38	0.1815	0.39	0.1816	0.39
No Prescription Coverage	0.4064	0.49	0.4046	0.49	0.3903	0.49	0.3691	0.48
Low-Income Subsidy	0.2232	0.42	0.2262	0.42	0.2272	0.42	0.2255	0.42
Long Term Care (or undef)	0.0438	0.20	0.0465	0.21	0.0469	0.21	0.0458	0.21
Cost Sharing Other (or undef)	0.4348	0.50	0.4346	0.50	0.4169	0.49	0.3940	0.49
Employer Subsidy	0.1494	0.36	0.1465	0.35	0.1638	0.37	0.1493	0.36
Undefined Creditable Coverage	0.0001	0.01	0.0001	0.01	0.0005	0.02	0.0006	0.03
Has Creditable Coverage	0.1848	0.39	0.1881	0.39	0.1629	0.37	0.1544	0.36
Died in plan year	0.1497	0.36	0.0430	0.20	0.0397	0.20	0.0390	0.19
Not Same Plan in 12 months	0.4376	0.50	0.4446	0.50	0.4196	0.49	0.4003	0.49
Not PDP	0.6389	0.48	0.6438	0.48	0.6362	0.48	0.6297	0.48
Employer Group Waiver (or undef)	0.4862	0.50	0.4842	0.50	0.4757	0.50	0.4705	0.50
Full 12 Month Sample	0.1089	0.31	0.1199	0.32	0.1213	0.33	0.1227	0.33
Main 4 Year Sample	0.0368	0.19	0.0388	0.19	0.0376	0.19	0.0363	0.19
Observations	2677143		2539492		2619222		2716094	

An additional exclusion were for individuals whose demographics or plan detail data were undefined or unavailable.

Table A.2: Summary of 5% Sample of Medicare Beneficiary Demographics 2009-2012

	(1) 2009		(2) 2010		(3) 2011		(4) 2012	
	mean	sd	mean	sd	mean	sd	mean	sd
2011 RxHCC weight	0.3826	0.43	0.3924	0.44	0.3958	0.44	0.3947	0.45
2011 RxHCC demo. weight	0.4863	0.19	0.4859	0.19	0.4873	0.19	0.4868	0.19
hasDiabetes	0.2095	0.41	0.2083	0.41	0.2108	0.41	0.2104	0.41
hasHypertension	0.4153	0.49	0.4108	0.49	0.4113	0.49	0.4069	0.49
hasCancer	0.0642	0.25	0.0630	0.24	0.0622	0.24	0.0618	0.24
highCholesterol	0.4290	0.49	0.4267	0.49	0.4293	0.49	0.4266	0.49
Long Term Care	0.0506	0.22	0.0496	0.22	0.0506	0.22	0.0497	0.22
Observations	2318357		2381965		2430325		2507147	

The total number of observations in the first panel differ from Table A.1 due to missing values in gender and race fields. The number of observations differ between the two panels since the information on risk factors and conditions are generated from the Medicare Part A and B claims from the prior year. Patients who did not have relevant claims to be scored from the prior year did not have values for these conditions.

Table A.3: Summary of Full Sample of Medicare Beneficiary Demographics 2009-2012

	(1) 2009		(2) 2010		(3) 2011		(4) 2012	
	mean	sd	mean	sd	mean	sd	mean	sd
Age at End of Reference Year	75.7853	7.18	75.9651	7.32	75.9497	7.30	75.8113	7.30
Start Medicare Year	1998.3	7.2	1999.0	7.3	2000.1	7.3	2001.2	7.3
Years in Medicare	10.7	7.2	10.9	7.3	10.9	7.3	10.8	7.3
Female	0.6546	0.48	0.6467	0.48	0.6407	0.48	0.6375	0.48
Race: White	0.9478	0.22	0.9490	0.22	0.9483	0.22	0.9464	0.23
Race: Black	0.0293	0.17	0.0279	0.16	0.0286	0.17	0.0294	0.17
Race: Other	0.0105	0.10	0.0109	0.10	0.0111	0.10	0.0118	0.11
Race: Asian	0.0064	0.08	0.0065	0.08	0.0066	0.08	0.0071	0.08
Race: Hispanic	0.0043	0.07	0.0039	0.06	0.0037	0.06	0.0037	0.06
Race: North American Native	0.0017	0.04	0.0017	0.04	0.0017	0.04	0.0017	0.04
2011 RxHCC weight	0.4546	0.32	0.4697	0.33	0.4767	0.33	0.4795	0.33
2011 RxHCC demographic weight	0.4202	0.01	0.4203	0.01	0.4201	0.01	0.4200	0.01
hasDiabetes	0.2300	0.42	0.2364	0.42	0.2429	0.43	0.2467	0.43
hasHypertension	0.6128	0.49	0.6172	0.49	0.6236	0.48	0.6249	0.48
hasCancer	0.0921	0.29	0.0962	0.29	0.0977	0.30	0.0986	0.30
highCholesterol	0.6568	0.47	0.6660	0.47	0.6776	0.47	0.6800	0.47
Main 4 Year Sample	0.3358	0.47	0.3216	0.47	0.3082	0.46	0.2938	0.46
Observations	291550		304477		317670		333309	

Table A.4: Average Coinsurance in Phases in Baseline Sample Standard Plans 2009-2012 (person-week)

	(1) 2009		(2) 2010		(3) 2011		(4) 2012	
	mean/sd	count	mean/sd	count	mean/sd	count	mean/sd	count
Deductible	0.7824	152348	0.8200	158028	0.7976	159524	0.8522	174615
	0.30		0.26		0.28		0.24	
ICR	0.2218	253123	0.1996	248750	0.2101	243824	0.2260	231479
	0.13		0.13		0.12		0.13	
Gap	0.6834	34249	0.5445	33136	0.3606	35389	0.3763	33197
	0.40		0.42		0.24		0.25	
Catastrophic	0.0619	4336	0.0546	4397	0.0502	5428	0.0533	4967
	0.08		0.04		0.03		0.03	
Total	0.4482	444056	0.4445	444311	0.4311	444165	0.4814	444258
	0.36		0.36		0.35		0.36	

Table is generated from the baseline sample individuals with standard deductible, ICL, and OOPT limits. The coinsurance rates are averaged over the amount the patient pays (doesn't include the drug manufacture discounts in 2011 and 2012) divided by the total expenditure cost in the person-week observation where spending occurs. The count reflects the fact that there are more person-week observations in the ICR region than others. These sums do not reflect the counterfactual coinsurance rates that beneficiaries with low spending would have faced if they had reached higher spending.

Table A.5: Proportion of Beneficiaries in Each Phase at the End of the Year in Baseline Sample Standard Plans 2009-2012

	Standard			
	2009	2010	2011	2012
Deductible	11.26	12.97	13.41	16.55
ICR	61.46	60.68	59.26	58.72
Gap	22.76	22.23	22.28	20.41
Catastrophic	4.52	4.13	5.04	4.33
Observations	9,178	9,178	9,178	9,178

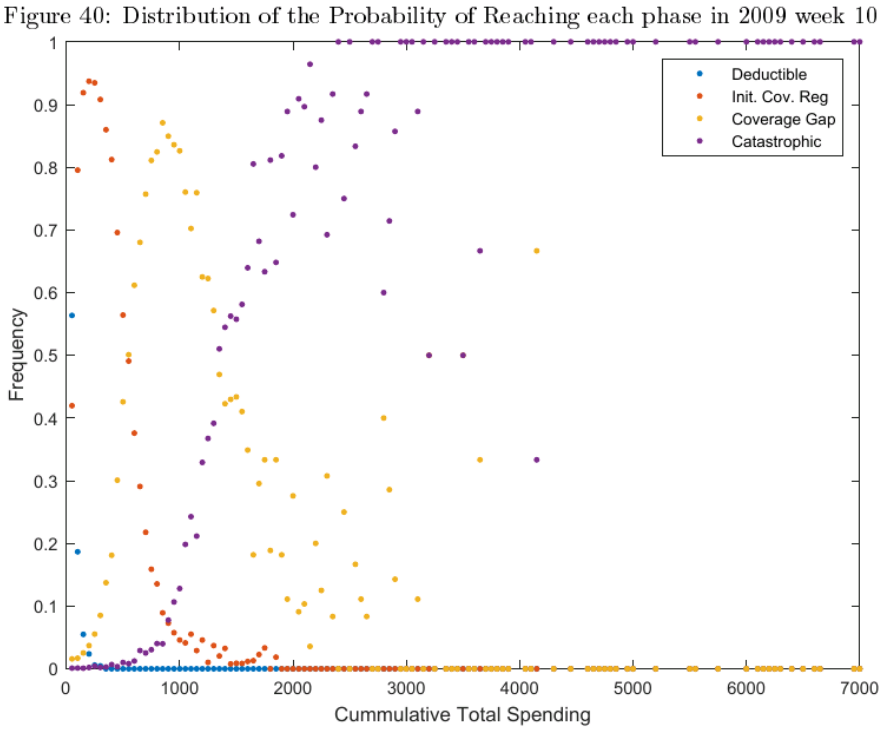
Table is generated from the baseline sample individuals with standard limits. The proportion of beneficiaries that end the year in each phase is averaged over the individual beneficiary.

Table A.6: Average Probability of Weekly Spending in Coverage Regions in Standard Plans 2009-2012 (person-week)

	(1)	(2)	(3)	(4)
	2009	2010	2011	2012
	mean/sd	mean/sd	mean/sd	mean/sd
Deductible	29.0821	29.8314	30.1629	30.1755
	45.41	45.75	45.90	45.90
ICR	40.0371	40.9394	41.1316	42.2222
	49.00	49.17	49.21	49.39
Gap	47.8506	48.5035	48.7668	51.6035
	49.95	49.98	49.99	49.98
Catastrophic	55.7429	56.0108	56.8488	57.3371
	49.67	49.64	49.53	49.46
Total	37.0524	37.7167	38.0082	38.3719
	48.29	48.47	48.54	48.63

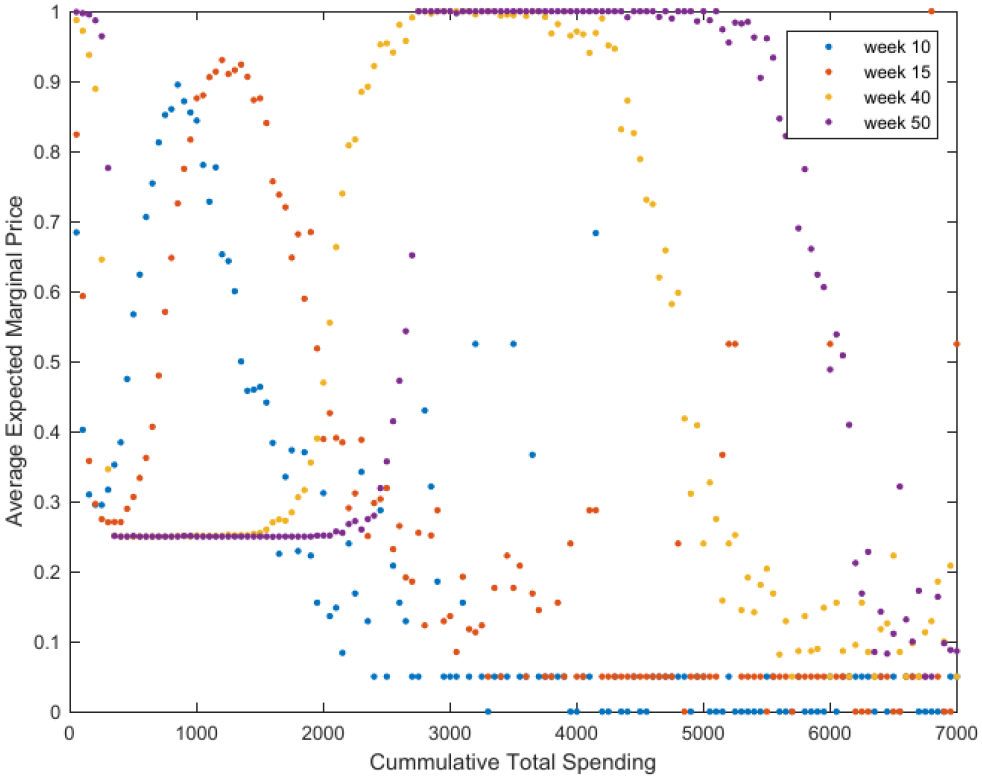
Table is generated from the baseline sample individuals with standard deductible, ICL, and OOPT limits. Table displays the raw probability of spending in a week in each coverage region and year.

Figure A.2: Distribution of the Probability of Reaching each Coverage Region in 2009 Conditional on Cumulative Total Spending in week 10



Each point is the conditional probability of a beneficiary reaching coverage region r at the end of the year given week w and their total cumulative spending $F_R(Z_w|w = 10)$. This figure is generated based on non-LIS beneficiaries who had PDP plans in 2009 that had the 2009 standard coverage region limits. Beneficiaries were also not in long term care institutions and received Medicare benefits for old-age. Points depict the actual observed probability of these beneficiaries ending the year in each of the coverage regions conditional on being in a specific cumulative total spending bin in week 10 of 2009. Bin size = 50 and was chosen for illustrative purposes (Using a Gaussian kernel regressions here, the STATA generated bin size had ranged from around 20-70. But to graph them all on the same figure, they needed to be consistent throughout so a bin size of 50 was chosen.)

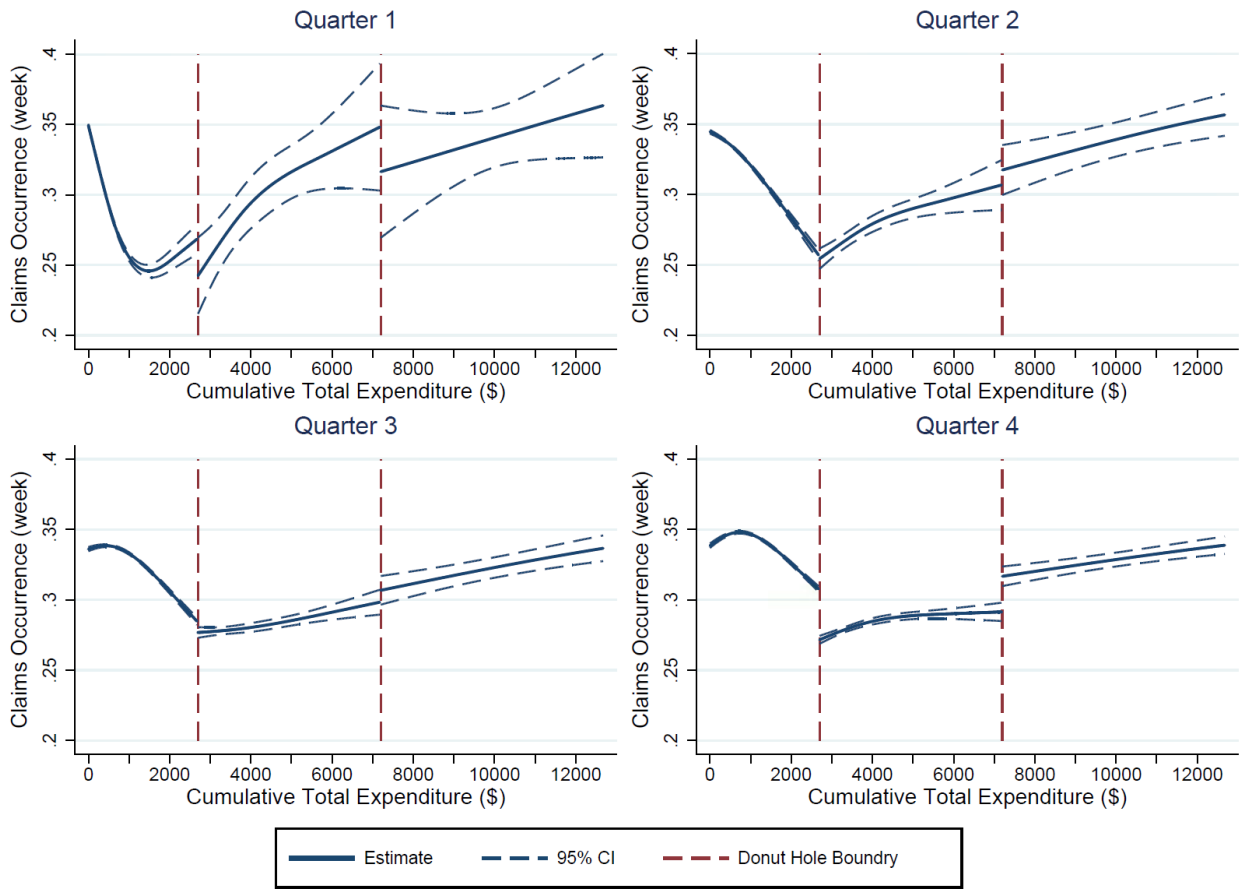
Figure A.3: Average Expected Marginal Price in 2009



Depicts the expected average marginal price $E(MP(\sum|w))$ in a given week in 2009 with the distributions depicted in Figure A.2. Where $E(MP(\sum|w)) = F_D(\sum|w) * MP(D) + F_P(\sum|w) * MP(P) + F_I(\sum|w) * MP(I) + F_C(\sum|w) * MP(C)$ and MP(D) = 1, MP(P) = .25, MP(I)=1, and MP(C=.05) the government standard plan amounts.

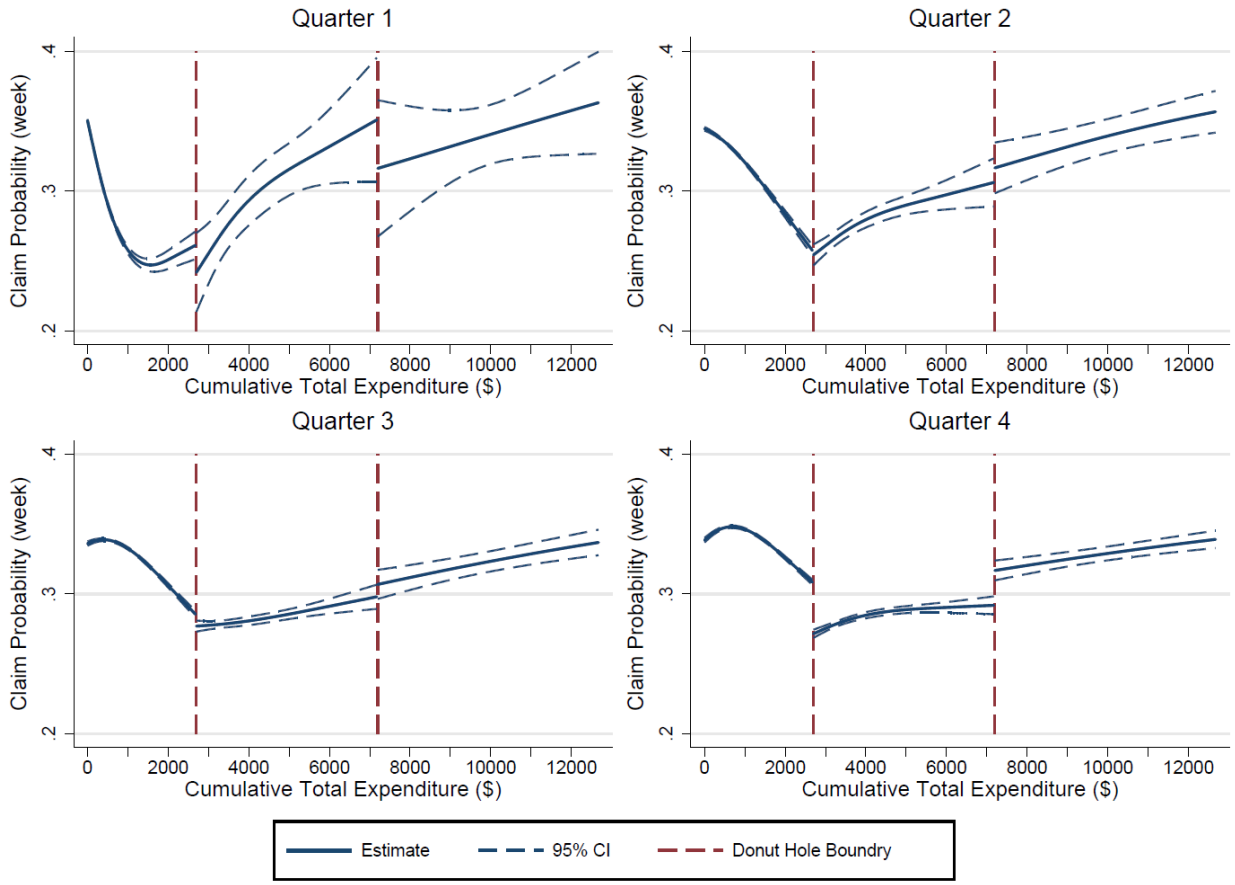
B Alternative Model Results

Figure B.4: False 1: Probability of Claims Occurring in a Week



Note: Displays the predicted values of the claim occurrence probability $\hat{o}_{iyw}$ in a week from the fixed-effects panel regression of Equation 5 under the assumptions of False 1 on the beneficiaries in the "No Deductible" sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values. The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. See Figure 9 for the exact values used to generate the prediction.

Figure B.5: False 2: Probability of Claims Occurring in a Week



Note: Displays the predicted values of the claim occurrence probability $\hat{o}_{iyw}$ in a week from the fixed-effects panel regression of Equation 5 under the assumptions of False 2 on the beneficiaries in the "No Deductible" sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values.
The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. See Figure 9 for the exact values used to generate the prediction.

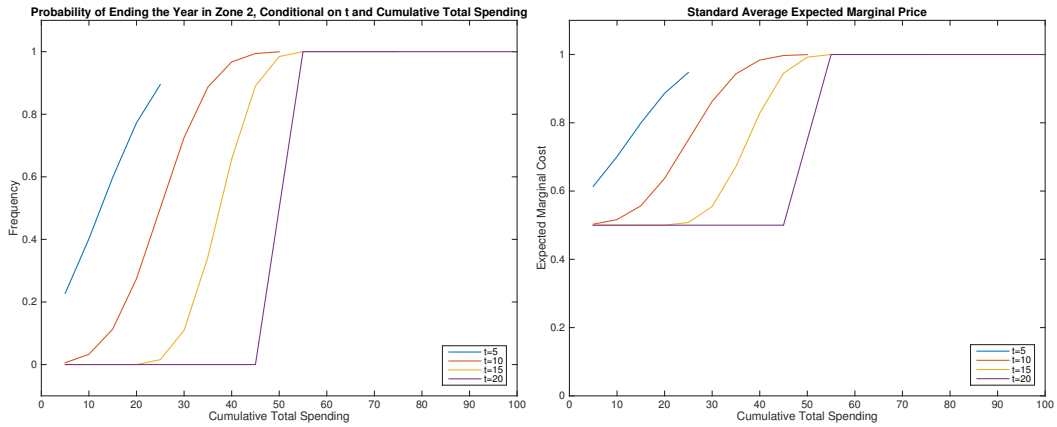
C Simulations of a Standard Agent

Now consider a simulated patient with accurate beliefs on a more complicated benefit cost calculation. The simulated patient behaves as a fully forward-looking standard agent with uncertainty in whether her health shock might necessitate health spending and no discounting. She faces the marginal costs of purchasing indicted by the contract presented in Figure 2 where $X = 50$. Her contract year is divided into 20 time periods in which she might get a health shock and need to spend a fixed amount \$5.

As the patient realizes his or her risks throughout the plan year and $\sum_t$ grows, the probability

of ending in Region 1 $Pr(\text{in Region 1 in T}|\sum_t)$ evolves. Certainly in making decisions in time T, the beneficiary should have very little ambiguity as to whether her cumulative total spending will exceed X. If the patient has spent more than X, then $Pr(\text{in Region 1 in T}|\sum_t > X) = 0$; if they have spent significantly less, then whether they will cross X depends on the beneficiaries' evaluation of their potential based on the costs and benefits of purchases their prescriptions.

Figure C.6: Simulated Frequency of Ending the Year in Region 2 and Expected Marginal Price

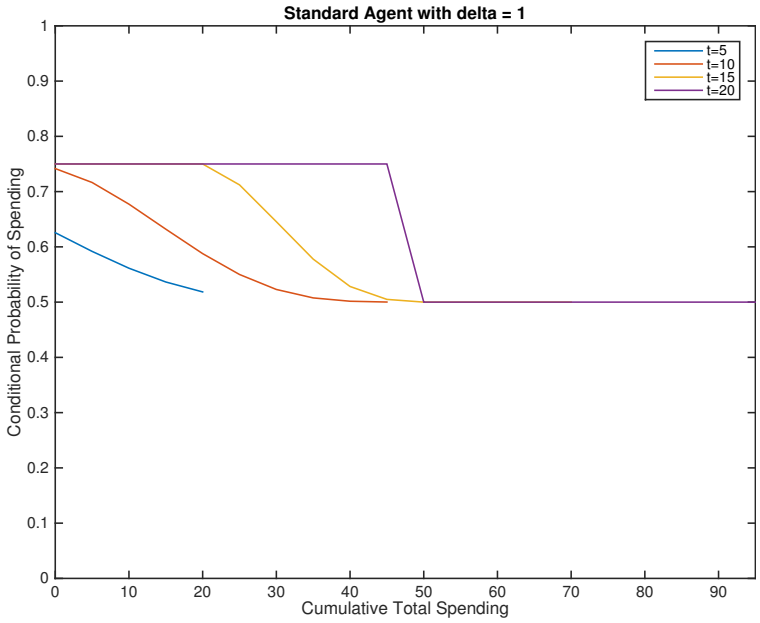


Note: The simulation assumes an individual who faces the example marginal prices from Figure 2. She has 20 opportunities to spend. In each period she receive a negative health shock that is distributed uniformly from 1 to 10. The cost of care in this scenario without subsidies in each period is \$5. And the spending limit $X = 50$ in this example. The conditional probability of spending is calculated at each \$5 increment. The x-axis plots the upper bound of each spending bin.

Figure C.6 Panel 1 presents the simulated beneficiary's period t probability of ending the year in Region 2 conditional on her cumulative total expenditures. Each line represents the conditional probability within each time period, and one can see that the beneficiary has more certainty about her year-end phase in later time periods. Figure C.6 Panel 2 depicts the heuristic version of the simulated beneficiary's expected marginal price, which is the weighted sum of her coinsurance rates times her probability of ending in each region. $PMP_t(\sum_t) = .5P_t(\text{in Region 1 in T}|\sum_t) + P_t(\text{in Region 2 in T})$. Notice that the PMP graph would indicate that beneficiaries who are likely to cross the $X = 50$ threshold should recognize their end-of-year region much earlier than the limit. They should be making their cost-benefit analysis using the PMP curve that is smooth at all but the last t=20 time period.

Ultimately, any uncertainty the beneficiary has about her end-of-year region is reflected in her perceived marginal price, and then in her probability of spending. Figure C.7 depicts the simulated beneficiary's actual expected spending patterns. Notice that there is an inverse relationship between the perceived marginal price of spending and the probability of making a

Figure C.7: Simulated Conditional Probability of Spending



*Note:*See Figure C.6 for simulation details.

purchase.

D Dynamic Panel with Fixed Effects Bias Simulated

Simulated data was created to have a claims occurrence value o_{iyw} and a cumulative total spending value Z_{iyw} with 208 weeks. The data generating process allows for individual heterogeneity in a base probability of spending. For the simulated case, the probability of spending and the amount spent is independent of the Z_{iyw} and region effects R_{iyw}.

The following 3 cases are graphed in Figure D.8. They show the shape the bias would have on the cumulative total spending measure.

T=208

$$o_{iyw} = \alpha_i + R_{iyw} + Z_{iyw} + \beta_r * Z_{iyw} + q_{iyw} + \tau_y + \varepsilon_{it}$$

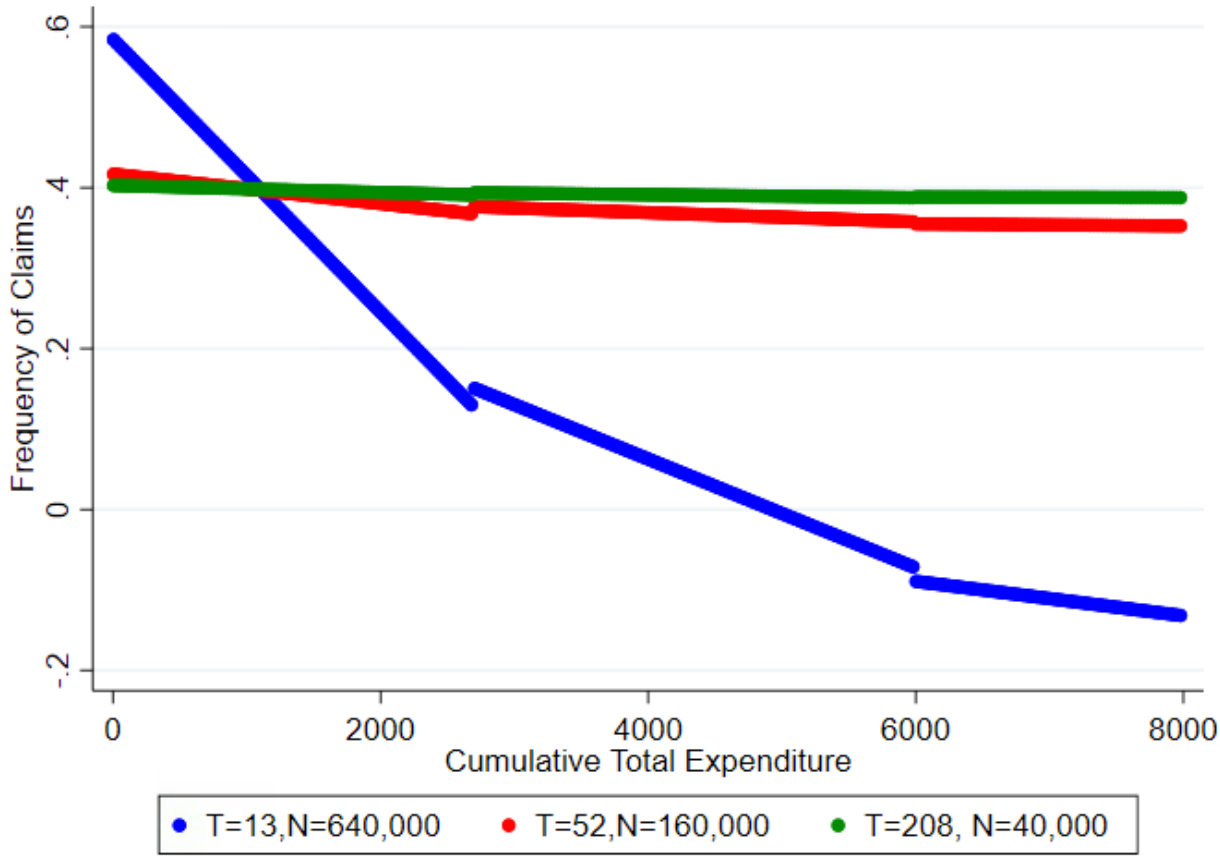
T=52

$$o_{iyw} = \alpha_{iy} + R_{iyw} + Z_{iyw} + \beta_r * Z_{iyw} + q_{iyw} + \varepsilon_{it}$$

T=13

$$o_{iyw} = \alpha_{iyw} + R_{iyw} + Z_{iyw} + \beta_r * Z_{iyw} + \varepsilon_{it}$$

Figure D.8: Estimation of a linear model of simulated data with different time frames



*Note:*The green- T=13, red - T=52, blue - T=208

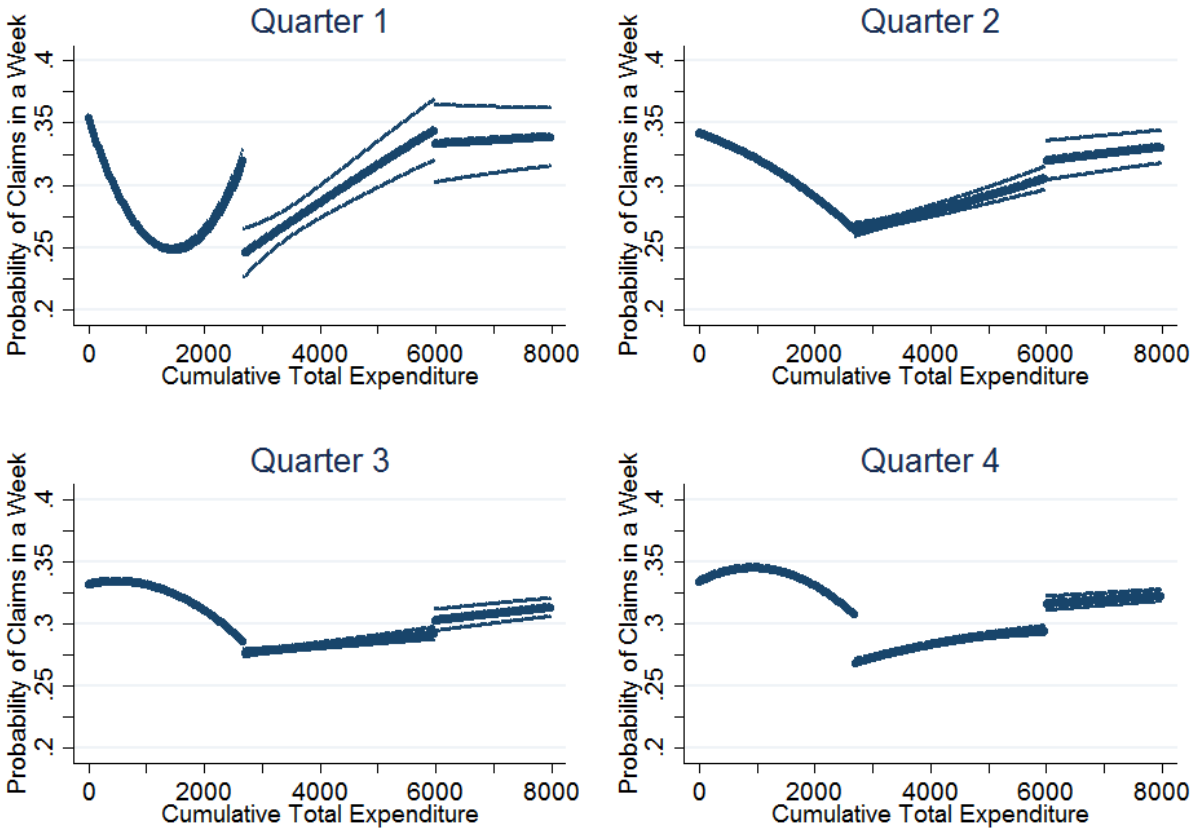
E Analysis with a quadratic polynomial functional form

Table E.7: Effect of Cumulative Total Expenditures on the Probability (%) of a Claim in a Week

Coefficient	Estimate	Std. Error
RxHCC Risk Weight	3.60	0.11
RxHCC Demographic Weight	-25.42	6.13
2010	0.32	0.036
2011	0.66	0.044
2012	1.16	0.05
N	18,585,632	

*Note:*The estimated coefficients on the risk scores $\mathbf{X}_{iy}$ and year-time dummies τ_y from the fixed-effects panel linear probability regression of Equation 1 on the beneficiaries in the "No Deductible" sample. All estimates in the table are significant at the 1% level.

Figure E.9: Linear: Probability of Claims Occurring in a Week, Conditional Year-Cumulative Total Spending



Note: The predicted values of the f function from a fixed-effects panel regression of Equation 1 on the beneficiaries in the "No Deductible" sample. Each panel represents a quarter of the year and displays the estimated probability of having at least one prescription claim in a week, conditional on the cumulative total spending in a year. Each quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. The probabilities are predicted assuming that the fixed effect is zero and the beneficiary has the sample average risk and demographic scores from 2009 .4753 and .4196 respectively. Images display a 95% confidence interval around the predicted values. Each line segment represents a prediction made assuming the year is 2009 with 2009 spending limits and across time average individual risk and demographics. The discontinuity for the OOPT into the catastrophic region was chosen at $6,153.8 the total expenditure equivalent OOPT from the government-defined 2009 plan (Table 1).

F Analysis with the sample of individuals with the "Standard" Plan Limits

Figure F.10: Probability of Spending in Each Quarter of the Year Conditional on Cumulative Total Spending