NBER WORKING PAPER SERIES

CHOICE INCONSISTENCIES AMONG THE ELDERLY:
EVIDENCE FROM PLAN CHOICE IN THE MEDICARE PART D PROGRAM

Jason T. Abaluck
Jonathan Gruber

Working Paper 14759
http://www.nber.org/papers/w14759

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2009

Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part
D Program
Jason T. Abaluck and Jonathan Gruber
NBER Working Paper No. 14759
February 2009
JEL No. I11,I18

ABSTRACT

The Medicare Part D Prescription Drug Plan represents the most significant privatization of the delivery
of a public insurance benefit in recent history, with dozens of private insurers offering a wide range
of products with varying prices and product features; the typical elder had a choice of roughly 40 stand-alone
drug plans. In this paper we evaluate the choices of elders across this wide array of Part D options
using a unique data set of prescription drug claims matched to information on the characteristics of
choice sets. We first document that the vast majority of elders are choosing plans that are not on the
"efficient portfolio" of plan choice in the sense that an alternative plan offers better risk protection
at a lower cost. We then estimate several discrete choice models to document three dimensions along
which elders are making choices which are inconsistent with optimization under full information: elders
place much more weight on plan premiums than they do on expected out of pocket costs; they place
almost no value on variance reducing aspects of plans; and they value plan financial characteristics
beyond any impacts on their own financial expenses or risk.These findings are robust to a variety of
specifications and econometric approaches. We develop an "adjusted" revealed preference approach
that combines data from consumer choices with ex ante restrictions on preferences, and find that in
a partial equilibrium setting, restricting the choice set to the three lowest average cost options would
have likely raised welfare for elders under the program.

Jason T. Abaluck
MIT Department of Economics
50 Memorial Drive
Cambridge, MA 02142
abaluck@mit.edu

Jonathan Gruber
MIT Department of Economics
E52-355
50 Memorial Drive
Cambridge, MA 02142-1347
and NBER
gruberj@mit.edu

The Medicare Modernization Act of 2003, better known as the legislation that added the Part D prescription drug benefit to the Medicare program, represents the single most significant expansion of public insurance programs in the U.S. in the past 40 years. The most novel, and controversial, feature of this legislation was the use of multiple private insurance providers to deliver this new public insurance product. Unlike the traditional model of government mandated uniform insurance packages for all enrollees, under the Part D program dozens of private insurers were allowed to offer a wide range of products with varying prices and product features. Perhaps most well-known was the extent to which plans covered the "donut hole", a broad uncovered range of expenditures in the minimum mandated plan.

This unprecedented privatization of the delivery of a public insurance product raises a host of important policy questions. Primary among these is the impacts of allowing choice across so many private insurance options. The typical elder in our data (described below) faces a choice of over 40 stand-alone drugs plans, and our estimates suggest that the range of cost from the most to least expensive option facing an elder is comparable to the mean of those costs. Choice is clearly meaningful in this context. Yet, to date, we know almost nothing about how elders are making these crucial choices.

This paper investigates the choices of elders for the newly formed Part D program in 2006. We analyze data that provides information on the Part D plans chosen and prescription drug utilization for a large sample of elders in the U.S. These data were collected by Wolters Kluwer (WK), a "switch agent" that lies between pharmacies that fill prescriptions and the insurance companies and prescription benefit managers that pay for them. WK collects information on almost one-third of all third party prescription drug transactions, and we will use the universe of their data for those over age 65 during 2005-2006 to examine choice of Part D

plan. We match to this data set a comprehensive set of information from the Centers for Medicare and Medicaid Services (CMS) on the Part D plans available to each person in our data set.

Specifically, for each elder whose claims appear in our sample, we model the financial implications of each of the plans in their choice set, based on both 2005 and 2006 drug utilization and several different models of expectations. We begin by presenting the basic facts on choice, documenting that the vast majority of elders are choosing plans that are not on the "efficient portfolio" of plan choice for that elder. We then turn to more rigorous multinomial models of individual choice to incorporate non-financial characteristics, preference heterogeneity and unobserved plan characteristics into our analysis.

Our findings are striking: along three dimensions, elders are making choices which are inconsistent with optimization under full information. First, elders place much more weight on plan premiums than they do on the expected out of pocket costs that they will incur under the plan. Second, they substantially under-value variance reducing aspects of alternative plans. Finally, consumers appear to value plan financial characteristics far beyond any impacts on their own financial expenses or risk. These findings are robust to a variety of specifications and econometric approaches.

We attempt to interpret the magnitude of our results by analyzing the impact of restricting the option set offered to elders under Part D. We develop an "adjusted" revealed preference approach that combines data from consumer choices with ex ante restrictions on preferences. We find that in a partial equilibrium setting, restricting the choice set to the three lowest average cost options would have likely raised welfare for elders under the program.

Our paper proceeds as follows. Part I provides background on the Part D program and reviews the growing literature on its impacts. Part II discusses our data sources, and Part III presents initial results on choice set variation and choice behavior. Part IV describes our choice framework, and Part V presents results. Part VI shows the welfare analysis based on these models, and Part VII concludes.

Part I: Background

The Medicare Part D Program

Medicare, which provides universal health insurance coverage to those over age 65 and to those on the disability insurance program, was established in 1965. The original program covered most medical needs for the elderly and disabled, including hospital and doctor costs, but it excluded coverage for prescription drugs. This omission was not perceived as a major one in the early years of the Medicare program, but in the 1990s the advancement of prescription drug treatments for common illnesses among the elderly drew attention to this gap in Medicare coverage. Medicare recipients, for example, spent an average of $2,500 each on prescription drugs in 2003, more than twice what the average American spent on all health care in 1965.[1]

In 2003, the Bush administration and Congress reached agreement on a far-reaching prescription drug benefit package at a projected cost to the federal government of $40 billion per year for its first ten years. The most noticeable innovation of the Part D plan is that this new Medicare benefit is not delivered by the government, but rather by private insurers under contract with the government. Beneficiaries can choose from three types of private insurance plans coverage of their drug expenditures. The first is stand-alone plans called Medicare

---

[1] Data for prescription drug spending comes from the Congressional Budget Office (2002). Data for average Americans' health spending comes from the "National Health Expenditures" section of the Centers for Medicare and Medicaid Services' *National Health Accounts*.

Prescription Drug Plans (PDP) (a plan that just offers prescription drug benefits). In 2006, there were 1429 total PDPs offered throughout the nation, with most states offering about forty PDPs. The majority of PDPs are offered by a dozen national or near national companies.

The second alternative is Medicare Advantage (MA) plans, plans that provide all Medicare benefits, including prescription drugs, such as HMO, PPO, or Private FFS plans. There were 1314 total plans nationally in 2006. Finally, beneficiaries could retain their current employer/union plan, as long as coverage is "creditable" or at least as generous (i.e. actuarially equivalent) as the standard Part D plan, for which they would receive a subsidy from the government

Under Part D, recipients are entitled to basic coverage of prescription drugs by a plan with a structure actuarially equivalent to the following: none of the first $250 in drug costs each year; 75% of costs for the next $2,250 of drug spending (up to $2,500 total); 0% of costs for the next $3,600 of drug spending (up to $5,100 total, the "donut hole"); and 95% of costs above $5,100 of drug spending. Over 90% of beneficiaries in 2006, however, are not enrolled in the standard benefit design, but rather are in plans with low or no deductibles, flat payments for covered drugs following a tiered system, or some form of coverage in the donut hole. The main requirement for plans is that they must have equal or greater actuarial value than the standard benefit.[2] The government also placed restrictions on the structure of the formularies that plans could use to determine which prescription medications they would ensure. Overall, Part D sponsors have great flexibility in terms of plan design.

Enrollment in Part D plans was voluntary for Medicare eligible citizens, although Medicare recipients not signing up by May 15, 2006 were subject to a financial penalty if they eventually joined the program (to mitigate adverse selection in the choice of joining the

---

[2] Cover Memo for Medicare Part D Benefit Parameters: Annual Adjustments for Standard Benefit in 2007 (CMS)

program). One group, however, was automatically enrolled: low income elders who had been

receiving their prescription drug coverage through state Medicaid programs (the "dual

eligibles"). These dual eligibles were enrolled in Part D plans by default if they did not choose

one on their own. The Part D plans for dual eligibles could charge copayments of only $1 for

generics/$3 for name brand drugs for those below the poverty line, and only $2 for generics/$5

for name brand drugs for those above the poverty line, with free coverage above the out of

pocket threshold of $3600.[3]

Despite reluctance voiced before the legislation passed, there was enormous interest from

insurers in participating in the Part D program. By November 2006, 3,032 plans were being

offered to potential Part D enrollees. Every county in the nation had at least 27 plans available;

the typical county had 48 plans, while some counties featured more than 70 choices, primarily

due to high number of MA plans.[4]

Enrollment in the new Part D program was initially fraught with problems, but in the

following months the federal government was able to iron out many of the problems that had

arisen during the initial transition. As of June 2006, there were 10.4 million people enrolled in

stand alone PDP plans, 5.5 million people enrolled in MA plans and about 6 million dual

eligibles.[5] Yet 73% of people over 65 felt that the Medicare prescription drug benefit was too

---

[3] In addition, two other groups receive substantial subsidies – those found eligible for Low Income Subsidy (LIS) or for Partial Subsidy by the SSA. To qualify for LIS, beneficiaries must have income less than 135% of poverty and resources less than $7,500/individual or $12,000 couple. This group received benefits comparable to the dual eligibles with incomes above 100% of poverty. To qualify for Partial Subsidy, beneficiaries must have income at 135%-150% of poverty and resources less than $11,500/individual or $23,000/couple. This group can enroll in plans with a $50 deductible, a 15% copayment up to the out of pocket threshold, and $2/$5 copayments above that point.   In addition, premiums are fully paid by the government up to 135% of poverty, and then partially subsidized up to 150% of poverty.

[4] Details on number of plans in a median county obtained from Prescription Drug Plan Formulary and Pharmacy Network Files for 2006, provided by CMS.
[5] Enrollment data (rounded) taken from CMS, State Enrollment Data spreadsheet at http://www.cms.hhs.gov/PrescriptionDrugCovGenIn/02_EnrollmentData.asp#TopOfPage. Enrollment numbers also available at http://www.kff.org/medicare/upload/7453.pdf.

complicated, while 91% of pharmacists and 92% of doctors expressed this concern. When asked

if they agree with the statement "Medicare should select a handful of plans that meet certain

standards so seniors have an easier time choosing," 60% of seniors answered "Yes."[6]

Despite these reservations, there were no signs of diminished plan choice in subsequent

years. The number of PDPs increased by about 30% in 2007, from 1,429 to 1,875 and remained

at this level in 2008.[7]

Issues of Elder Choice in Part D

The use of this private delivery device, with such a multiplicity of choices, is a novel

feature of the Part D legislation. Standard economic theory would suggest that this is a

beneficial plan feature: allowing individuals to choose across a wide variety of plans that meet

their needs, rather than constraining them to a limited set of choices being made by the

government, can only increase welfare in the standard model in a partial equilibrium setting.

But there are reasons to believe that the standard model is insufficient, particularly for a

population of elders. There is growing interest in behavioral economics in models where agents

are better off with a more restricted choice set, as nicely reviewed in Iyengar and Kamenica

(2006). Recent theoretical work shows that the traditional "more is better" principle may be

reversed in choice set contexts, for example when the presence or absence of options conveys

information (Kamenica, 2006; Kuksov and Villas-Boas, 2005) or when agents have preferences

with regret (Irons and Hepburn, 2003; Sarver, 2005). And a growing body of empirical work

shows that individuals are less likely to participate in markets where they face more choice;

decisions to purchase a good (Iyengar and Lepper, 2000; Boatwright and Nunes, 2001), take a

---

[6] Kaiser Family Foundation and Harvard School of Public Health (2006).
[7] Hoadly et al. (2006). Data on 2008 plans taken from CMS 2008 PDP Landscape Source (v. 09.25.07) available at http://www.cms.hhs.gov/prescriptiondrugcovgenin/.

loan (Bertrand et al., 2005) or enroll in a 401(k) plan (Iyengar, Huberman and Jiang, 2004) are found to decrease when participation requires choosing from a larger set of alternatives.

Iyengar and Kamenica (2006) find that not only the decision to participate in a market, but also the nature of choice itself, is affected by the size of the option set. They investigate choice over asset allocation in both laboratory and real-world (pension plan choice) settings, and find that individuals opt for safer investments when faced with a larger range of risky choices. In particular, they find that the presence of more investment options in a 401(k) plan leads to more frequent choice of money market or bond options rather than equity investment. Iyengar and Lepper (2000) also find that satisfaction with choices made falls with the size of the choice set in several experimental settings.

These issues may be paramount within the context of the elderly, given that the potential for cognitive failures rises at older ages. Salthouse (2004) shows clear evidence that the performance on a series of memory and analytic tasks declines sharply after age 60. Part of the reason for this may be the rise in incidence of dementia with age; starting at age 60, dementia rates roughly double every five years (Fratiglioni et al., 1999). A recent study by Agarwal et al. (2006) shows that in ten different contexts, ranging from credit card interest payments to mortgages to small business loans, the elderly pay higher fees and face higher interest rates than middle-aged consumers. These types of findings raise particular concern about choice in the Part D context.

<u>Previous Studies of Part D Choice</u>

We are aware of only three previous studies of these issues in the context of Part D. The first is a set of studies by Dan McFadden and colleagues, as summarized in McFadden (2006). These researchers surveyed a set of elders about their plans for enrolling in Part D programs, and

evaluate whether enrollment intentions in the plan were "rational" given the penalties for delay. They find that 71% of potential enrollees were making the appropriate decision (under various assumptions about discount rates, etc.), while 10% of enrollees did not intend to enroll when it would be in their interests to do so, and 19% intended enroll when it would be in their interest to delay. Thus, for most potential enrollees, the decision over whether to enroll seems to be made rationally.

Their findings are less sanguine, however, for choice of Part D plan. This survey offered individuals a choice of the standard plan described above versus alternatives that provide different levels of insurance coverage (e.g. catastrophic only, complete coverage, etc.), with corresponding actuarially fair premiums. They find that only about 36% of enrollees choose the cost-minimizing plan, and they do not place much value on the insurance aspects of more comprehensive plans. They conclude that "consumers are likely to have difficulty choosing among plans to fine-tune their prescription drug coverage, and do not seem to be informed about or attuned to the insurance feature of Part D plans."

While this is an interesting set of findings, it provides only a preliminary look at the crucial issue of plan choice. These conclusions are based on data which do not contain precise detail about the prescription drugs used by individuals; assumptions about utilization are made using aggregate imputations from other sources. Moreover, this is based on hypothetical choices across a set of non-existing plans; individuals may become educated about the program when they are actually faced with plan choices. Thus, the failures of choice documented by McFadden (2006) may not hold when we use data on actual individual utilization and choices.

A recent paper by Lucarelli, Prince and Simon (2008) uses aggregate data on plan market shares to conduct a study of how plan features impact demand and to undertake a welfare

analysis of choice restrictions. They estimate sizeable welfare losses from limiting the option set facing seniors. But they do so in a framework which assumes that seniors are choosing optimally so that by definition restricting the choice set can only be harmful. Without individualized data on plan choices, they are unable to evaluate the underlying efficacy of plan choice.

Most closely related to our work is a recent field experiment by Kling et. al. (2008). They examine how providing people with information about the relative costs of each of the available plans in 2007 computed using their 2006 claims impacts their choices. They find that individuals who receive this intervention are more likely to switch plans, and more likely to end up with lower predicted and realized costs. Using our richer dataset on patient claims, we are able to model the individualized risk characteristics of plans in addition to looking just at average costs. Our model is also more general in terms of sample and implications. While they investigate the consequences of one particular intervention on a sample of patients at a single hospital, our model allows us to calculate the potential welfare gains from reforms which change the structure of the choice set, and to do so for a large fraction of Medicare Part D enrollees.

Part II: Data

Our primary data source is a longitudinal sample of prescription drug records from the Wolters Kluwer (WK) Company. They are the largest "switch" operator in the prescription drug market: they collect the electronic claims from pharmacies and pass them on to the Pharmacy Benefit Managers (PBMs) and insurance companies that will pay the claims. After adjudicating the claim, it is passed back through the switch to the pharmacy. WK performs this function for a large sample of pharmacies throughout the U.S. Once pharmacies are in their sample, there is a

93% chance that they remain enrolled, so this is effectively a longitudinal sample of pharmacies.

On average the claims captured by the WK system represent almost 31% of all 3rd party

prescription claims filled in the U.S.[8] The geographic distribution of these data is very closely

representative of the geographic distribution of 3rd party claims as well; the correlation between

the WK market share and the overall 3rd party market share across each of the states is 0.86.

WK keeps a longitudinal file that tracks prescription drug use for more than 100 million

persons in the U.S. They have made available to us for research purposes a longitudinal sample

of prescription claims for any individuals age 65 and over in 2005.

These data are crucial because they are the only available data (of which we are aware) that

contain information both on specific drug utilization by elders and on plan choice. Information

about specific drug utilization is key because plan costs vary tremendously based on drug

utilization, as we discuss below.

We begin with a sample of 2.7 million elders who (a) have a Part D claim, (b) are not

employer-insured, dual eligibles or eligible for low-income subsidies/partial subsidies, (c) have

claims for only one region of the country, (d) have no claims with missing payment information,

(e) are in the sample of consistently reporting pharmacies, and (f) have data for both 2005 and

2006. This data file has a rich set of information about every drug claim for individuals in the

longitudinal sample, including information on: month in which the prescription was filled;

county of location of the pharmacy; a de-personalized patient id which allows longitudinal

patient linkages; patient age; NDC code for the drug; quantity measures (days supply, dosage,

package size); patient and insurer payments; price of purchase; and insurer or PBM name.

WK has created a sample for us that links longitudinally all claims from elders that fill

prescriptions at a pharmacy in their sample. Thus, there are three types of attrition from the

---

[8] Figure based on data provided by WK for Q3:2006.

sample. First, elders may die (in which case we still observe all of their claims). Second, pharmacies may enter or leave the sample. This can be addressed by using only pharmacies that are continuously in their sample.[9] Finally, individuals may switch pharmacies. If the switch is to a pharmacy within the WK sample, then the company does a detailed statistical match to ensure that the patient is captured and matched to other prescriptions (based on the de-identified form of data fields such as first name, last name, date of birth, year of birth, gender, health insurance id and zip code). If the switch is outside of the WK sample, then the individuals will be lost to this sample.

Unfortunately, there is no way to capture such transitions. But we can assess their importance by taking advantage of the fact that Wolters Kluwer provided us with a coverage level variable which indicates the proportion of pharmacies in each county which are covered by Wolters Kluwer. We have rerun our models on the 10% of counties where WK covers at least 40% of all third-party prescriptions, and our results are very similar to what we report below. This suggests that attrition is not significantly biasing our results.

The CMS Plans Database

We obtain information on availability of Part D plans and specific plan features directly from four files provided by CMS: the plan information file, the beneficiary cost file, the formulary file and the geographic locator file. The plan information file lists plan names and identifiers, and regions/counties in which plans are offered. The beneficiary cost file contains copays and coinsurance rates for different tiers of each plan. The formulary file contains a list of all the drugs that are included on the formulary for each plan. The geographic locator file allows

---

[9] A store is flagged as continuously enrolled provided that the store does not miss more than 11 days (including weekends and holidays) of reporting in a month.

us to identify all the Social Security Administration (SSA) counties that correspond to different PDP and MA regions.

The major strength of the CMS data is that it allows us to fully parameterize any elder's plan choice set based on their location. We have used these data to build a "cost calculator" that mimics the calculator provided on Medicare's web site. This calculator uses a given set of prescriptions for a given elder to compute their projected out of pocket spending in each plan available in their county.

Matching patients to their Part D Plan

One challenging aspect of the WK data is that we know each patient's county and the name of the company that provides the Part D plan that is covering each prescription, but not specifically which Part D plan offered by that company is covering the prescription. For example, we know that an elder is covered by a Humana product, but not whether it is Humana Complete, Humana Enhanced, etc.

Fortunately, we can resolve this matching problem in most cases by using a combination of county code, company name, and copayment structure. For each claim and each of the plans within the same company offered in a particular county, we check if the copay that the patient paid for this claim matches any of the prescribed copays of the plan. We assign a person to a plan if most of their claims match to the same unique Part D plan. We carry out this exercise for each month. To confirm that a person has been matched to a correct Part D plan, we look at all the months together and insist that a person be consistently matched to the same plan in each month from June 2006 on, since enrollment into Part D plans was open until May 15th 2006.

Of the approximately 2.7 million individuals in our sample, 776,118 were matched to Part D plans. The remainder were excluded either because they had a large number of non-Part

D claims (implying that they have some other form of coverage), because they had too few

claims to reliably match, or because their copays were inconsistent with the copays listed for Part

D plans in their region.

Of the matched individuals, 57.1% were uniquely matched to a Part D plan, 42.9% were

multiply matched (meaning that more than one Part D plan was consistent with their copays).

While the unique matches are clear, excluding multiple matches leads us to misstate the

proportion of enrollment in some plans. This problem is especially severe among Humana plans

because Humana offers several plans which differ only in the deductible and donut hole coverage

and thus cannot generally be distinguished on the basis of copays. While comprising 20% of all

matches, Humana plans are only 10% of unique matches. To deal with this problem, we include

both unique and multiple matches, with multiple matches randomly assigned to one of the plans

to which they are matched with probability equal to the proportion of total national enrollment in

that plan in 2006.[10]   In fact, if we restrict only to unique matches, our results are considerably

stronger (e.g. the anomalies we document below are heightened as are the utility gains from

restricting the choice set).[11]

Construction of Out of Pocket Cost Variables

The total enrollee costs of Part D can be decomposed into premiums, which are known

for certain at the time of plan choice, and the distribution of out of pocket costs given the

information available at the time when plans are chosen.   Our focus is on estimating the

distribution of costs given all of the information potentially available to individuals at the time

---

[10] Regional enrollment figures are not available at the plan level in 2006 for most plans.

[11] One might still worry that if we include all matches our results are driven by misassignment of multiple matches, while if we include just unique matches our results are driven by selecting for plans which are easier to match (although it seems unlikely that both issues would coincidentally lead to the same estimates). To deal with this objection, we estimated an earlier version of our models using the full sample of unique and multiple matches and the correct likelihood function given the random assignment of multiple matches (following Hausman et. al. 1997). This correction appears to make little difference, in part due to the fact that most multiple matches could be assigned with a high level of confidence.

when they make their choice. There are three reasons that estimating this distribution is challenging: first, we only observe realized out of pocket costs for the plan in which an individual is enrolled; second, we observe only a single realization of out of pocket costs for each individual (making it impossible to compute a variance measure); and third, we do not observe all of the information available to individuals at the time when they make their choice.

To handle the first difficulty, we assume that the set of 2006 claims is fixed and would remain constant had the individual in question chosen a different plan; that is, we assume no moral hazard. This assumption allows us to use the calculator to determine what each individual's realized costs would be for each plan in their choice set. Given typical estimates of the elasticity of prescription drug utilization in the range of 0.2 to 0.5, and considering that this would only impact our results to the extent that individuals have sufficient foresight to take into account future utilization effects in their plan choices, this is a fairly innocuous assumption. A formal model and empirical results are provided to justify this claim in Appendix A.

To handle the second difficulty, we sample realized costs from 200 individuals who are "identical" to the individual in question at the time when the plan choice is made. In practice, we define "identical" as individuals with the same decile of 2005 drug expenditures, 2005 days supply of branded drugs and 2005 days supply of generic drugs; after extensive searching, we found that this combination provided the best prediction of 2006 prescription drug spending based on 2005 characteristics. We therefore assign each individual to one of 1000 cells demarcated by the interacted deciles of these measures. We restrict our sample to individuals for whom there are at least 200 other individuals in their cell, and we use these 200 individuals in each cell to compute both our rational expectations measure of utilization in 2006 (described below) and our variance measure.

The third difficulty is that individuals may actually know more than can be predicted given 2005 costs at the time when they make their plan choices. Intuitively, we can attempt to determine whether individuals know more than can be predicted given 2005 costs by analyzing whether their choices are sensitive to the component of the variation in realized costs across plans which cannot be predicted given 2005 characteristics. We develop a formal model of this approach in Part V.

Final Sample Creation

Under Part D individuals could enroll not only in a stand alone PDP plan, but also in a more comprehensive MA plan; we distinguish between individuals matched to MA and those matched to PDP plans based on copay and exclude the former. We focus just on PDP plans (and therefore, just on individuals who chose PDP plans) because MA plans involve broader health care decisions which are beyond the scope of our data (e.g. regarding HMOs and fee-for-service plans). This exclusion is justified by the "independence of irrelevant alternatives" assumption that underlies our logit modeling, as discussed (and tested) further below. We also exclude individuals who have fewer than 500 observations in their state or fewer than 100 observations in their brand/state cell to increase the speed of estimation of the model by reducing the required number of brand/state fixed effects; this restriction has no effect on our final results.

Our final sample consists of 477,393 individuals. The typical patient in this sample is almost 75 years old, three-fifths are female, and they have an average of 34 claims per year. Their total prescription drug spending averages $1711 per year. While some individuals were enrolled in Part D for the full year, others enrolled as late as May. The average total premiums paid after enrollment was $287 and the average OOP costs paid out over the same period was $666. This is the sample used in the efficient frontier analysis below. In our conditional logit

models, we randomly subsample 20% of these individuals for computational reasons. We

estimate the more computationally demanding random coefficients models on a randomly chosen

subsample of 15,000 patients.

The distribution of enrollees across Part D plans is highly correlated in this final sample

with the national facts on PDP enrollment provided by CMS. The correlation between the share

by brand in our sample and the CMS sample is 0.98, and the correlation between the share of our

sample in the top 10 plans is correlated with the CMS reported share in those plans at 0.89 (the

correlation for the top 100 plans is 0.91).

Part III: Facts on Plan Choice

To motivate our regression framework, Figure 1 shows the basic facts on the relationship

of plan choice to total plan costs. For each individual in the data, we estimate the total cost of

enrolling in each PDP plan in their county, adding both premiums and expected out of pocket

costs. We then estimate the difference in total costs between the plan chosen by that individual

and the lowest cost plan in their county. For this exercise, we use a perfect foresight model of

expectations, using actual 2006 expenditures to estimate the costs that individuals face in each

plan.

As Figure 1 shows, only 12.2% of individuals choose the lowest cost plan in their state.

Indeed, on average, individuals could save 30.9% of their total Part D spending by choosing the

lowest cost plan rather than the plan they chose. If we redo these calculations using actual 2005

expenditures, or predicted 2006 expenditures based on 2005 expenditures rather than actual 2006

expenditures, we find even stronger deviations from the lowest cost plan.[12]

---

[12] It appears that some plans may have offered low premiums in 2006 in order to entice consumers to choose their plan in the first year of the Part D program before raising their premiums in subsequent years. This behavior should

Of course, individuals are not simply choosing a fixed payment stream when choosing a Part D plan; individuals who are highly risk averse may explicitly be choosing plans with higher mean expenditure to protect themselves against variance in expenditure. Yet this does not seem to be the case. Even if we *only* include plan choices where the variance is non-increasing, over 70% of enrollees could have chosen a lower cost plan, and the typical enrollee could have saved 23.3% of their Part D expenditures without raising their variance.[13]

The explanation for these facts is shown in Figure 2, which shows the choice set for individuals in California. The X axis in this graph is the mean of total costs for each plan, and the Y axis is the average standard deviation in costs (where the standard deviation is computed using the 1000 cell method, and the average is taken across individuals). In this graph, there is a clear "efficient frontier" of plans which dominate others in terms of both cost and variance. This graph masks considerable heterogeneity across individuals: different plans lie on the efficient frontier for different individuals, so the fact that a plan lies off the efficient frontier in this graph does not imply that it is suboptimal for each individual. Nonetheless, most of the plans are well off the efficient frontier, meaning individuals could have either lowered their mean costs or their variance by picking a different plan.

As we will document below, one reason for the large amount of choice off the efficient frontier is that individuals consider plan characteristics in making their choices – but not how those plan characteristics matter for themselves. This is perhaps best illustrated by a simple examination of the decision to choose a plan with donut hole coverage. Figure 3 shows the

---

not impact our analysis except insofar as there are large switching costs because consumers have the option to switch plans after each year, but one might still wonder to what extent the above results are driven by such plans. To assess this issue, we repeated the above analysis using the 2007 premiums for all plans and found that the average potential cost savings fell slightly from 30.9% to 25%.

[13] The fact that this number is smaller than the 30.9% number is because we are searching for cost savings over a small set of plans, not because individuals are especially sensitive to risk, a point we document further below.

probability of choosing donut hole coverage, and the financial implications of doing so, sorted by 2006 spending percentiles; the results are once again similar for other measures such as 2005 actual spending or 2006 predicted spending. The bottom line shows the percent of the population at each percentile choosing donut hole coverage; the top line shows the average savings for individuals in that quantile from switching from the lowest cost plan in their region which offers donut hole coverage to the lowest cost plan that does not.

The plans which offer donut hole coverage actually have slightly inferior coinsurances relative to the lowest cost non-donut hole plans in the initial coverage range, and so the cost of donut hole coverage is rising with expenditures until the point when individuals become likely to enter the donut hole.

The results here are striking: the percentage choosing donut hole coverage is virtually flat throughout the spending distribution at around 10%. Even if individuals are willing to pay extra in mean costs for the protection provided by donut hole coverage, it is hard to rationalize the fact that the same proportion of individuals in the 10th and 85th percentile of the spending distribution choose donut hole coverage.

Part IV: Base Model of Part D Plan Choice

In this section, we extend the efficient frontier analysis presented above by considering several discrete choice models. These models serve three general purposes in our setting. First, they allow us to control for additional plan characteristics such as plan quality. Second, they allow us to understand more precisely how preferences combine with choice set characteristics so we can forecast how individuals might choose in counterfactual choice environments. Third, they allow us to quantify the welfare consequences of choices.

We begin by specifying a CARA utility model with normally distributed costs C and constant wealth W:

$$U(C) = -\exp\big(-\gamma(W - C)\big) \text{ where } C \sim N(\mu, \sigma^2) \tag{1}$$

In this case, indirect utility is given by:

$$u(\mu, \sigma^2) = EU(C) = -\alpha \exp\big(\gamma\mu + \frac{1}{2}\gamma^2\sigma^2\big) \tag{2}$$

where $\alpha = -\exp(\gamma W)$ is a constant. A first-order Taylor expansion about the point $(\mu', \sigma^{2'})$ yields:

$$u(\mu, \sigma^2) \approx u(\mu', \sigma^{2'}) - \alpha\gamma u(\mu', \sigma^{2'})(\mu - \mu')$$
$$-\frac{1}{2}\alpha\gamma^2 u(\mu', \sigma^{2'})(\sigma^2 - \sigma^{2'}) \tag{3}$$

Dropping constant terms (since these are irrelevant in the logit model), we obtain:

$$u(\mu^*, \sigma^2) \approx -\alpha\gamma u(\mu', \sigma^{2'})\mu - \frac{1}{2}\alpha\gamma^2 u(\mu^*, \sigma^{2'})\sigma^2 \tag{4}$$

We can write total costs as $C = \pi + OOP$ and since π is known for any given plan, $Var(C) = Var(OOP) = \sigma^2$ and $\mu = E(C) = \pi + E(OOP) = \pi + \mu^*$. Adding an error term, we can rewrite equation (4) as:

$$u = -\alpha\gamma u(\mu', \sigma^{2'})(\pi + \mu^*) - \frac{1}{2}\alpha\gamma^2 u(\mu^*, \sigma^{2'})\sigma^2 + \epsilon \tag{5}$$

This maps into a conditional logit model of plan choice where the utility of individual i from choosing plan j is given by:

$$u_{ij} = \pi_j\beta_0 + \mu_{ij}^*\beta_1 + \sigma_{ij}^2\beta_2 + \boldsymbol{x_j}\boldsymbol{\gamma} + \boldsymbol{q_{b(j)}}\boldsymbol{\delta} + \epsilon_{ij} \tag{6}$$

with $\beta_0 = \beta_1 = \alpha\gamma u(\mu^*, \sigma^{2'})$ and $\beta_2 = -\frac{1}{2}\alpha\gamma^2 u(\mu^*, \sigma^{2'})$. In this equation $\boldsymbol{x_j}$ represents any financial plan characteristics which impact choice, $\boldsymbol{q_{b(j)}}$ represents plan quality ratings and other

non-financial aspects of plans (which vary only across brands), and ϵ_{ij} are i.i.d. type I extreme value random variables.

In this expression, $\gamma = \frac{2\beta_2}{\beta_1}$ allows us to map the ratio of the coefficients on the variance of costs and the coefficient on the mean of costs into the coefficient of absolute risk aversion. This assumes that wealth is constant across all states of the world: the only risk facing individuals is uncertainty about the distribution of out of pocket costs. The same expression would hold if we added idiosyncratic risk that was uncorrelated with prescription drug expenditures, but it is not implausible that there would be correlated risks: in states of the world where prescription drug expenditures are higher, other medical expenditures are higher as well. Such correlated risks would tend to bias upwards our already low estimates of risk aversion.

We include in our model several financial plan characteristics beyond premiums, out of pocket costs, and the variance of out pocket costs. These are: the deductible of the plan; a dummy for whether the plan covers all donut hole expenditures; a dummy for whether the plan covers generic expenditures in the donut hole only; and a cost-sharing index. The cost sharing index is calculated for each plan as the average percentage of expenditures covered by the plan between the deductible and the donut hole. This variable differs from expected out of pocket costs in that it has the same value for everyone in the sample for each plan, and because it is not directly impacted by whether plans have deductibles or donut hole coverage. We also include two measures of plan quality: the share of the top 100 drugs used by elders that is included in the plan's formulary and a quality index. The quality index is computed by CMS on a 1-5 scale by aggregating consumer ratings at the brand level collected along 17 dimensions which are categorized as "Customer Service," "Drug Pricing Information" (availability / rate of price changes), and "Using Your Plan to Get Your Prescription Filled".

Identification is a natural concern in this context. All of the plan characteristics included in our model may be endogenous due to unobserved demand factors, and they may be biased by correlation with unobserved plan characteristics. To address this concern, we observe and include in our model all of the publicly available information that might be used by individuals to make their choices. We also consider models where we control for a full set of brand dummies, as well as a full set of interactions of state dummies with brand dummies. When we include brand dummies, the coefficient on the quality index (which is measured at the brand level) is no longer separately identified although it can be recovered by a GLS regression of these dummies on the quality variable. When brand-state dummies are included, coefficients on plan characteristics such as the premium, deductible and donut hole coverage are identified by the variation across plans offered by the same brands in a given state.[14]

Even with these fixed effects, it is possible that premiums are endogenous because they are set based on brand-state specific assessments of demand conditions. If premiums are higher in regions where insurers anticipate more demand for their particular plan (relative to other plans offered by the same insurer), our estimate of the coefficient on premiums will be biased towards zero since individuals will appear to be less averse to higher premiums. To the extent that these factors make high premiums appear less undesirable than they actually are, our conclusion that premiums are overweighted relative to out of pocket costs would be strengthened, as would our estimates of the welfare loss due to consumer mistakes.[15]

---

[14] For instance, in many states Humana offers a Standard plan with lower premiums but limited coverage, an Enhanced plan with higher premiums but no deductible, and a Complete Plan which offers superior cost sharing and full donut hole coverage at much higher premiums.

[15] We did attempt estimating the models reported below using two instruments using the control function approach: these were the average premium for a given plan in all states where the plan is offered (designed to avert local demand shocks) and a "marginal cost" instrument constructed using the average covered expenditures for individuals enrolled in the plan. In both cases, the magnitude of the coefficient on premiums increased in the IV models. We are not confident that the exclusion restriction is satisfied for either of these instruments, so we continue to estimate the model without an instrument below.

<u>Restrictions on Preferences</u>

The model laid out above suggests three natural restrictions on preferences which extend the efficient frontier concept to the discrete choice setting.

Restriction 1: $\beta_0 = \beta_1$

This restriction states that the coefficient on premiums should equal the coefficient on expected out of pocket costs. Controlling for the risk characteristics of plans, individuals should be willing to pay exactly one dollar in additional premiums for coverage which reduces expected out of pocket costs by one dollar. If this restriction fails to hold, individuals are not choosing on the efficient frontier: they could switch to alternative plans with comparable risk characteristics but lower total costs.

Restriction 2: $\gamma = 0$

This restriction states that financial plan characteristics other than premiums, expected out of pocket costs and the variance of out of pocket costs do not impact choices. Individuals should not care about deductibles, donut hole coverage or copays *per se*; they should only care about these factors to the extent that they impact the distribution of out of pocket costs. Once we control for this distribution, these factors should be redundant.

Restriction 3: $\beta_2 < 0$

This restriction states that individuals should be risk averse.

While these restrictions follow naturally from utility maximization with full information and standard preferences, the model from which they are derived makes several important functional form assumptions: we assume that the distribution of out of pocket costs can be summarized by its mean and variance, that indirect utility is a linear function of this mean and variance, and that the errors are i.i.d. type I extreme value. In Appendix B, we show that the

restrictions assumed in the previous section still hold even when these functional forms assumptions are weakened.[16] Of course, it is always possible to write down preferences that would violate the above restrictions, but these restrictions are generally compatible with commonly used expected utility functions given the observed cost distributions.

Choice Model Results

Table 1 reports the results from several conditional logit models. Model (1) includes only the premium, realized out of pocket costs and the variance of out of pocket costs. As noted in the discussion of the cost variables, expected out of pocket costs – meaning the individual's expectation of out of pocket costs at the time of plan choice – is not directly observed, so we use realized costs as a proxy for expected out of pocket costs. This proxy has noise (where "noise" includes the component of realized costs unknown to the individual at the time when the choice is made) and so its coefficient is biased downwards. We address this problem at length in the next section and show that it does not much impact our conclusions.

The cost variables – premiums and out of pocket costs – are measured in hundreds of dollars. Model (1) therefore shows that a $100 increase in premiums leads to a 32% reduction in the probability that a given plan is chosen, implying an average elasticity of --0.75.[17] There are two ways to interpret the remaining coefficients. First, we can divide by the premium coefficient in order to compute the willingness to pay in dollars for a one unit increase in the characteristic. Second, the coefficient itself can be interpreted as the percentage increase in the probability that

---

[16] In particular, we simulate choices using the actual distribution of costs and several commonly used utility functions (CRRA, CARA) with varying levels of risk aversion. In some cases, the restrictions do not hold exactly, but the violations are much smaller in magnitude than we observe when we estimate the model using actual choices.

[17] The implied elasticity varies across plans based on premium level and market share. The "32%" number given in the text is derived from the equation $\frac{\partial \log(p_{ij})}{\partial x_{ij}} = (1 - p_{ij})\beta$. Thus, for $p_{ij} \approx 0$ which holds for a large number of plans, we can interpret β as the percentage change in p_{ij} associated with a one unit change in x_{ij}.

a plan is chosen from a one unit increase in the characteristic provided that probability is small (as it is for most plans).

Two points about the model (1) results are noteworthy. First, the coefficient on out of pocket costs is only about ½ as large as the coefficient on premiums, violating Restriction 1. Second, the coefficient on the variance term is negative and significant, but extremely small, implying risk aversion substantially less than we obtained in the simulations with CRRA = 1.

Model (2) adds additional covariates to control for deductibles, donut hole coverage, average cost sharing, formulary coverage and plan quality. Many of these covariates enter the model with significant coefficients. When we add plan characteristics, the coefficient on premiums increases suggesting that it was initially biased downward due to omitted variable bias. The coefficient on the variance term drops even further once we add a control for the # of the most popular 100 drugs which are included in the plan's formulary. This suggests that while individuals prefer plans which cover more drugs, they do not have sufficient foresight to choose plans which cover drugs which they (or at least people in their cell) might need in the future but are not already taking.

Models (3) and (4) add brand dummies and brand-state dummies respectively. The coefficient on premiums actually shrinks once we include brand-dummies, but the effects of the premium remain large; a $100 increase in annual premiums leads to a 50% reduction in the probability that a plan is chosen, corresponding to an average elasticity of -1.17. The coefficient on out of pocket costs has similar magnitude across all of the models, which reflects the fact that it is identified based on individual variation. In columns (3) and (4) the coefficient on the premium is more than five times as large as the coefficient on out of pocket costs.

The coefficients on plan characteristics are also very large in all specifications. Controlling for the out of pocket cost consequences, model (4) – which has the smallest plan characteristics - suggests that individuals are willing to pay over $300 for full donut hole coverage, $50 for generic donut hole coverage, about $80 to go from a deductible of 250 to a deductible of 0, about $80 to go from the plan with the least cost sharing (25%) to the plan with the most cost sharing (65%), and $12 for each of the top 100 drugs which appear on the formulary. These numbers are not enormous, but they are an order of magnitude larger than the results in the simulations, and have non-trivial consequences for the welfare evaluation of plan choice as we investigate in the welfare analysis section.

Thus, this formal modeling of choice reveals a violation of all three of the preference restrictions we laid out above. The coefficient on premium is an order of magnitude larger than the coefficient on out of pocket expenditures; generalized plan characteristics enter the model highly significantly, even conditional on individual out of pocket risk; and individuals are not willing to pay more for plans with lower variance in expected spending.

Part V: Modeling the Information Set of Consumers

In the previous section, we presented results from a conditional logit model of plan choice and identified three apparent irregularities in choices. Our interpretation is that these results reflect consumer errors – plan characteristics are more salient than are their implications for the distribution of out of pocket costs, and individuals are unable to compute the individualized risk characteristics of the alternative plans. In this section we consider an alternative explanation: we have misspecified out of pocket costs because we have failed to

appropriately model the information available to individuals at the time when they make their plan choice.

Thus far we have measured out of pocket costs using the realized cost measure constructed from 2006 claims. An alternative measure that we consider in this section we label our "rational expectations" measure. Recall that to create our variance measure we classified all individuals into 1000 cells defined by deciles of 2005 total spending, generic prescriptions and branded prescriptions, and ran the 2006 claims of 200 persons in each cell through the cost calculator for that plan. This procedure generates a distribution of costs for each patient and plan. Our rational expectations measure is defined as the mean of this distribution. Under the strong assumptions discussed above (CARA utility and a normal distribution of costs), the mean and the variance would completely summarize the impact of the cost distribution on utility. Our simulations in Appendix B show that they summarize this distribution well anyway even if these assumptions are relaxed.

It is useful to compare this rational expectations measure to the perfect foresight/realized costs measure we have been using. The latter measure is "too broad" in the sense that it includes information not available to individuals at the time when they choose (provided that is, that they do not know exactly what their drug needs and drug prices will be for the coming year). The former measure is "too narrow" in the sense that individuals may have private information at the time they choose beyond what can be inferred from their 2005 costs. If a patient learns they have cancer just prior to choosing their 2006 plan, they would correctly forecast that their drug needs would likely exceed the average of those with similar 2005 spending.

We address these concerns by developing a model with which we can identify the information available to consumers at the time when they choose. The intuition behind this

model is that we can determine if individuals know more than we can predict given just their

2005 spending by evaluating whether their plan choices are responsive to the component of 2006

spending which is not known in 2005.

Suppose that utility is given by:

$$u_{ij} = u^*_{ij}\beta_1 + \epsilon_{ij} \tag{7}$$

where μ^*_{ij} represents expected costs, defined as the individual's expectations of out of pocket

costs at the time when they make their choice (for ease of exposition, we momentarily ignore the

premium and variance terms). μ^*_{ij} is not observed. However, we do observe realized costs,

which can be written as the sum of expected costs and a noise term, defined as the component of

realized costs unknown to the individual at the time of plan choice:

$$C_{ij} = \mu^*_{ij} + \eta_{ij} \tag{8}$$

where C_{ij} denote the realized costs of individual i upon enrolling in plan j, μ^*_{ij} denotes expected

costs, and η_{ij} denotes noise. We can further decompose expected costs into the component of

expected costs predictable from 2005 data, μ_{ij}, and the component which is private information,

e_{ij}. This yields:

$$C_{ij} = \mu_{ij} + e_{ij} + \eta_{ij} \tag{9}$$

We assume that μ_{ij} and e_{ij} are independent of η_{ij}.[18] This assumption implies that individuals

are aware at the time when they make their choices of the component of costs that can be

predicted based on their previous year's consumption. This "rational expectations" assumption

is substantive, but conforms with the baseline rational choice model that is implicitly tested by

our analysis. We discuss this issue further below.

---

[18] Combined with the additive structure assumed above, the assumption of independence also rules out the case in which the degree of uncertainty about costs varies with the level of expected costs. We relax this strong assumption below by assuming only that μ_{ij} and η_{ij} are conditionally independent given the measured variance of costs.

If this were a linear model, the assumption of independence would be sufficient to identify β_1. This assumption implies that we have a classical measurement error problem: C_{ij} is a noisy measure of μ_{ij}^*. As usual, this problem can be solved with instrumental variables and in this case, μ_{ij} is a valid instrument − it is correlated with μ_{ij}^* and uncorrelated with η_{ij}, so instrumenting for C_{ij} with μ_{ij} would consistently estimate β_1. Because the model is non-linear, we need to be more explicit about the form of the measurement error to obtain consistent estimation. First, we rewrite equation (7) substituting in equation (8):

$$u_{ij} = C_{ij}\beta_1 - \eta_{ij}\beta_1 + \epsilon_{ij} \tag{10}$$

We assume further that $e_{ij} \sim N(0, \tau_{ij}^2)$ and $\eta_{ij} \sim N(0, \sigma_{ij}^2)$. Combined with equation (9), the normal updating formula implies:

$$f(\eta_{ij}|C_{ij}, \mu_{ij}) \sim N\left(\frac{\sigma_{ij}^2}{\sigma_{ij}^2 + \tau_{ij}^2}(C_{ij} - \mu_{ij}), \frac{\tau_{ij}^2 \sigma_{ij}^2}{\sigma_{ij}^2 + \tau_{ij}^2}\right) \tag{11}$$

We do not observe σ_{ij}^2 or τ_{ij}^2. However, provided we assume that $Var(e_{ij} + \eta_{ij}|i,j) = Var(e_{ij} + \eta_{ij}|Z_i, j)$ where Z_i are the variables which define each cell – that is, we assume that there is no heterogeneity in the variance of costs within cells – we do observe $Var(e_{ij} + \eta_{ij}) \equiv \widetilde{\sigma_{ij}^2}$. This is the variance we construct from the 1000 cell exercise. This still leaves us with a separate parameter to identify for each (i,j) pair. We additionally assume that a constant fraction τ_{frac} of the variance of costs within cells is due to private information . That is, we assume that $\tau_{ij}^2 = \tau_{frac}\widetilde{\sigma_{ij}^2}$ and $\sigma_{ij}^2 = (1 - \tau_{frac})\widetilde{\sigma_{ij}}^2$. As written, this is a random coefficients model with one additional parameter beyond the β 's - τ_{frac}, the degree of private information.

How is τ_{frac} identified? Equation (11) suggests a simple intuition. We can think of the model as one with a fixed coefficient β_1 on C_{ij} and a random coefficient with mean $\frac{\beta_1 \sigma_{ij}^2}{\sigma_{ij}^2 + \tau_{ij}^2} =$

$$\frac{\beta_1(1-\tau_{frac})\widetilde{\sigma}_{ij}^2}{\widetilde{\sigma}_{ij}^2} = \beta_1\big(1-\tau_{frac}\big) \text{ on } C_{ij} - \mu_{ij}.$$ Thus, the degree of private information is identified

by the degree to which the coefficient on $C_{ij} - \mu_{ij}$ falls short of the coefficient on C_{ij}. If

individuals have no information beyond what can be predicted from 2005 costs, we will observe

$\tau_{frac} = 0$, and equation (10) will simplify to:

$$u_{ij} = C_{ij}\beta_1 - \big(C_{ij} - \mu_{ij}\big)\beta_1 + \epsilon_{ij} = \mu_{ij}\beta_1 + \epsilon_{ij} \tag{12}$$

If on the other hand individuals have perfect information about 2006 costs, we will observe

$\tau_{frac} = 1$, and equation (10) will simplify to:

$$u_{ij} = C_{ij}\beta_1 + \epsilon_{ij} \tag{13}$$

In the interim case, individual's choose based on a linear combination of C_{ij} and μ_{ij}, and the

random coefficient on $C_{ij} - \mu_{ij}$ captures the fact that different individuals with the same C_{ij} and

μ_{ij} can have varying amounts of private information.

    This model also has implications for the variance term and the measurement of risk

aversion. The measured variance from the 1000 cell exercise $\widetilde{\sigma}_{ij}^2$ overstates the true variance in

costs because some of this variation represents variation in realized costs which is unpredictable

based on 2005 costs but is known to the individual at the time when they choose. Thus, the

correct variance to use in the model is $(1 - \tau_{frac})\widetilde{\sigma}_{ij}^2$, the variance of the noise term. To the

extent that individuals are responsive to the variance term, omitting this correction will tend to

bias our estimates of risk aversion downward by a factor of $\frac{1}{1-\tau_{frac}}$.[19]

---

[19] Note that this model does not directly allow for private information about the variance of costs; in some cases, individuals may learn that they are at risk of developing a certain condition which would require treatment with prescription drugs. This knowledge would increase their expected out of pocket costs in the coming year and would also increase the variance in their forecast. The model above does not allow for this type of information; while the model allows individuals in the same cell to have different values of expected costs based on their realization of private information, we continue to assume that all individuals in the same cell in the 1000-cell model face the same

Reintroducing the variance and the premium term, we obtain:

$$u_{ij} = \pi_{ij}\beta_0 + C_{ij}\beta_1 - \eta_{ij}\beta_1 + (1 - \tau_{frac})\widetilde{\sigma_{ij}^2}\beta_2 + \epsilon_{ij} \tag{14}$$

where the distribution of η_{ij} is given by equation (11).

Table 2 reports the results from estimating equation (14). For computational reasons, we estimate this model on a much smaller sample by randomly selecting 15,000 patients from our earlier sample. Column (1) reports our earlier results, column (2) reports the earlier specification on the new sample, and column (3) the results from equation (14) which allows for private information. The model is estimated using the Laplace approximation developed in Hausman-Harding (2007) with bootstrapped standard errors, including controls for the various plan characteristics. This model is identical to the model in equation (6), adding the normally distributed noise term (which is a function of $C_{ij} - \mu_{ij}$) and the variance adjustment.

The results in Table 2 suggest that there is substantial private information: individual choices take into account about 60% of the variation in out of pocket costs which cannot be predicted given their cell. The results also imply that our earlier finding that the coefficient on realized costs is smaller than the coefficient on premiums is robust to any measurement error generated from the fact that individuals do not know realized costs at the time when they choose.[20] Moreover, financial plan characteristics such as the donut hole and deductible continue to enter highly significantly in this model. Therefore, two of the major choice inconsistencies persist even when we model private information.

Interpreting the risk-related results in Table 2 requires examining in more detail our assumptions about what individuals know at the time when they choose. If we take the model in

---

variance. To the extent that this assumption is false, our model could be viewed as substituting the predicted variance given the variables used to construct the 1000-cell model for the actual variance.

[20] In other words, in a linear context, instrumenting realized costs for 2006 by predicted costs based on 2005 characteristics does not much change the coefficient on out of pocket costs, indicating little bias from measurement error in our out of pocket cost coefficient.

this section literally, we would conclude that individuals know much of what their costs will be to each plan in their choice set in the coming year so there is little insurance motive. Under this interpretation, the variance in out of pocket costs is small for all plans because there is little uncertainty. This means that any measured response to the variance term would imply high levels of risk aversion, and that the standard errors in our estimates of risk aversion are much larger than we concluded in the model ignoring private information. The risk index in these models (obtained by dividing the variance coefficient by the premium coefficient and multiplying by 200) is comparable to what we obtained in our Appendix B simulations for CRRA = 3 with wealth = 17000.

The results reported in Table 2 are actually consistent with two models of choice with private information, however, each of which has very different consequences for our measurement of risk aversion. The first model – assumed in the preceding analysis - is that individuals are using all available information to make forward-looking choices of plans for 2006, but are simply mis-weighting premiums and out of pocket costs in making those choices. The alternative is that individuals are not using all available information, but rather are paying attention only to a part of their prescription drug expenditures. For that portion to which they are attentive, individuals are rationally weighting premiums and out of pocket costs in the same way in making their decision. Yet individuals do not respond to variation in out of pocket costs beyond that portion. For example, it may be that individuals sometimes check whether plans cover one particular drug they will need in the coming year out of several that they actually take. For this one drug, they have foresight – even if they do not currently take the drug, they know they will need it in the coming year and choose on that basis. Nonetheless, they are insensitive to variation across plans in the costs of covering the remainder of their drugs.

In Appendix C, we supplement the model developed in the previous section by integrating the fact that the coefficient on premiums exceeds the coefficient on out of pocket costs into our measurement of private information in order to capture the possibility that individuals might be insensitive to a component of out of pocket costs. Doing so, our estimate of the degree of private information now shrinks to 20% since we are interpreting the gap between premiums and out of pocket costs as evidence that some costs are unforeseen. This gap – the "Fraction Observed" – implies that consumers are inattentive to almost 80% of the variation in out of pocket costs when they make their choices. Because of this, there is now substantial variation in the risk facing individuals across plans which reduces the standard deviation in our measurement of the variance term. As in the base model, we measure the variance coefficient is close to zero since the choices we observe are interpreted as a small response to substantial uncertainty rather than extreme sensitivity to a small amount of uncertainty.

The bottom line from our models of private information is that our conclusions about the gap between the premium and out of pocket expenditure coefficients, and the powerful role for general plan financial characteristics in driving choice, are robust to a wide variety of specifications of out of pocket spending risk. Our conclusion about the low degree of estimated risk aversion, however, is more sensitive to the precise specification of the model.

Part VI: Heterogeneity

The independence of irrelevant alternatives assumption that underlies the conditional logit model places strong restrictions on how elasticities vary across plans and will lead to inconsistent estimates if preferences are heterogeneous across the population. To address this concern, we assess the robustness of our model to heterogeneity driven by both observed and

unobserved factors. We first note that our model already allows for a substantial amount of individual variation: we estimated the coefficients on individualized out of pocket cost parameters. Nonetheless, it may still be the case that preferences vary in ways not included in our model. In terms of observed heterogeneity, we have reestimated our model for a number of separate samples: by gender; by age; and by tercile of the 2005 prescription drug expenditure distribution. In every case, we find that our results are very similar across all samples. In particular, each of these samples illustrates the three choice inconsistencies documented thus far: the premium coefficient is many multiples of the out of pocket cost coefficient; financial plan characteristics enter significantly; and the estimated degree of risk aversion is very low.

We therefore turn to considering unobserved heterogeneity. We use the Laplace approximation developed by Hausman-Harding (2007) to estimate a model with normally distributed random coefficients on all included characteristics. Our goal here is primarily a robustness check: does accounting for heterogeneity change any of our qualitative conclusions?

Table 3 shows the results of this analysis. As before, column 1 is the original model on a small sample. Column 2 adds random coefficients on premium, perfect foresight OOP, variance and quality, while column 3 adds random coefficients on all variables. Again, we see that the choice inconsistencies are present even after accounting for unobserved heterogeneity. Further, the magnitude of the coefficients estimated in the model without heterogeneity (which correspond to the mean of the random coefficients in this model) is not much affected. We do estimate significant heterogeneity in the coefficients on premium, quality, the deductible and the generic donut hole term; allowing for this heterogeneity turns out not to have a significant impact on the welfare results we report below.

We can also interpret the results in Table 3 as a test of the IIA assumption. To the extent that any of the coefficients are significant, this suggests that the IIA assumption does not hold exactly. Nonetheless, the fact that the magnitude of the coefficients does not change substantially once we allow for random coefficients suggests that this assumption is not altering our conclusions.

Part VII: Implications

Thus far, we have provided evidence of choice irregularities and shown that these irregularities are robust to the consideration of private information and preference heterogeneity. In this section, we attempt to determine the implications of these irregularities. We do so in a strictly partial equilibrium framework, ignoring supply side considerations, computation costs and many other factors; these are discussed at the end of this section.

We proceed by defining a normative utility function that differs from the positive function we estimated above in that it satisfies the proposed restrictions on rationality. That is, if individuals were fully informed, their choices would be given by the model estimated above but satisfying three additional restrictions: the coefficient on premiums is equal to that on expected out of pocket costs; financial plan characteristics other than premiums are excluded from the utility function once we control for the individual's expected out of pocket costs; and individuals exhibit risk aversion in their plan choice. We assume that the coefficient on premiums represents the marginal utility of a dollar if individuals were fully informed (this in turn determines the dollar value of quality variables and risk characteristics).

We begin by extending the results reported in the previous sections to take into account risk and quality characteristics of plans. We define the normative utility function to include

premiums and out of pocket costs (equally weighted), variance and quality and value the latter

characteristics in terms of dollars of premiums. We then ask: if individuals had chosen the plan

which maximizes this normative utility function rather than the plan which they did in fact

choose, by how much would utility be improved? The answer in this model is about 27% of

total costs – this is comparable to the 30.9% we found when we looked only at cost savings. The

small difference is due to the fact that the lowest cost plans also have slightly lower quality

ratings on average.[21]

We can interpret the 27% number as telling us the scope of the potential partial

equilibrium utility gains. If there were some intervention that would make individuals fully

informed and fully rational, this is the amount by which their utility could be improved. An

alternative question is what reforms we could use to attempt to realize some of these gains.

Some possibilities include directly providing individualized information about costs (as in Kling

2008 et. al.) or appointing surrogates such as doctors or pharmacists to play some role in plan

choice. We pursue a different alternative here: can we improve welfare by removing some

options from the choice set to reduce the scope for mistakes, or will individual heterogeneity

mean that removing options inevitably reduces welfare?

To pursue this exercise, we need to forecast choices in a modified choice environment.

We assume that the positive utility function we estimate on the full choice set also guides

behavior when choosing from the smaller choice set. We then evaluate the utility difference

between the larger and smaller choice sets using a normative utility function which satisfies the

restrictions discussed in the previous section. When positive and normative utility functions

---

[21] The 27% number uses the measured coefficient on the variance of costs which is close to zero. We can alternatively impose a coefficient which corresponds to a coefficient of absolute risk aversion of .0003 (roughly CARA = 3 with wealth of 17,000). In that case, the number rises to 27.6%. The difference is small because the lowest cost plans offer comparable risk protection to the plans which are actually chosen.

coincide, this is guaranteed to (weakly) reduce utility. When positive choice behavior diverges from optimizing normative utility functions because individuals are confused or not fully informed, there may be gains from moving to a smaller choice set.

Appendix D develops the formal tools we use in this section to perform utility comparisons in a setting where positive and normative utility functions differ. We show that the expected consumer surplus (CS) from a given choice set is given by:

$$E(CS) = \sum_j (x_{ij}\beta - x_{ij}\hat{\beta}) \frac{e^{x_{ij}\hat{\beta}}}{\sum_k e^{x_{ik}\hat{\beta}}} + \ln \sum_k e^{x_{ik}\hat{\beta}} \tag{15}$$

where $\hat{\beta}$ denotes the coefficients in the positive model and β denotes the coefficients in the normative model. This equation has a natural interpretation. The second-term is the familiar log-sum term from logit consumer surplus evaluated on the positive utility function: this is what CS would be if the positive utility function were also the normative utility function. The first term corrects the CS calculation for the fact that the log-sum term misevaluates the observed component of CS. For each plan, this correction is equal to the difference in utility between the normative and positive utility functions times the probability that the plan is chosen. Throughout this section, the positive model we use is the one given in column (4) of Table 1, which includes all plan characteristics and brand-state fixed effects.

A major problem with carrying out this exercise is the role of omitted characteristics in the logit model. It has been observed repeatedly in the discrete choice literature that assumption of independently distributed errors leads to implausible predictions of the welfare impact of additional choices (Berry and Pakes, 1999; Ackerberg and Rysman, 2002; Petrin, 2002). For example, Ackerberg and Rysman (2002) point out that even in random coefficients models that relax the IIA assumption, the fact that errors are independent conditional on the value of the

random coefficients implies implausible welfare gains from additional products because there is no "congestion"; each additional product is assumed to add an additional desirable characteristic.

There are a number of ways of approaching this problem. One route suggested by the analysis of Berry and Pakes (1999) is to consider a random coefficients model with either no error term or with a one dimensional omitted characteristic. While we continue to use a positive model with the logit error terms included (in part because we think it plausible that there are more than one omitted characteristics which have some impact on choice), we begin by considering a normative model in this spirit which excludes the i.i.d. error terms. In this model, only observed characteristics (including brand-state dummies) enter the normative utility function.[22]

A middle ground alternative is suggested by Ackerberg and Rysman (2002). In their approach, an additional parameter is estimated to directly determine the impact of choice set size on utility. We adopt a version of their approach here in which we use the variation in the number of PDP plans across states to determine whether PDP plans become more desirable as a whole relative to MA plans in states where there are a greater number of PDP plans. In particular, we estimate a nested logit model where PDP plans are one nest and MA plans another, and include as "top level" regressors the number of PDP plans, the number of MA plans, and an MA plan dummy. Because our information on the characteristics of MA plans is incomplete, we aggregate all MA plans into a single "outside option" whose premium is given by the average of the three most popular MA plans in each state, and we normalize all other characteristics to zero and include interactions of the out of pocket cost variables with a dummy for the MA plan (to prevent the specification of the outside option from spuriously contributing to the identification

---

[22] The results we report here are for the model without random coefficients, although including them does not impact our results. It would be necessary to include these in a positive model which excludes the error term so that all choices have a non-zero probability of being chosen.

of these coefficients). We randomly select for inclusion in the sample enough individuals who chose MA plans so that the proportion matches the proportion in the CMS data in 2006. The results of this approach are given in Table 4.

The log-sum coefficient lies between 0 and 1 indicating that the model is consistent with utility maximization. The model implies that the logit errors overstate the utility gains from additional plans by a substantial margin. If we normalize the estimated coefficient by the coefficient on premiums, the model implies that the logit errors overstate the utility losses of moving from a choice set with 47 PDP plans (the number in CA) to one with 3 plans by almost $300. This value is difficult to assess in the abstract: the utility gains implied by the logit error terms from moving to 3 to 47 plans are highly dependent on which three plans we choose. We can however compare this value to the utility gain implied by the logit errors in the context of a specific restriction, which we now do.

In particular, we select the three plans which are predicted using the previous year's data to have the lowest average cost in each state for inclusion in the small choice set. Note that this does not imply that these plans are lowest cost for each individual in the state, let alone the optimal plans when other factors (such as quality and risk) are taken into account.

Figure 4 shows the "utility landscape" that results from restricting to the three lowest cost plans in each state if we omit both brand-state dummies and omitted characteristics from the normative model.[23] The utility landscape shows the distribution that results if we randomly select

---

[23] We still find welfare gains from the small choice set if we include brand-state fixed effects in the normative model, although they are reduced on average to 13.8%. We feel they should be omitted from the model for three reasons. Firstly, they generate implausible variation in the value of identical plans across states – the value of the Humana fixed effect has a range of almost $900 across states. Secondly, they generate implausibly large variation across plans. The range of brand-state fixed effects implies that individuals would be willing to pay $1700 (about 170% of total costs) to go from the least desirable to the most desirable plan *controlling for all financial characteristics*. The interquartile range is well over $600 (by way of comparison, the coefficient on the quality variable implies less than a $200 difference between the 25th and 75th percentile plans). Thirdly, there are no obvious characteristics of plans beyond financial characteristics and quality which we think would impact

a state in proportion to its population and determine the average percentage change in utility in that state from moving from the full choice set to the smaller choice set.[24] We see from the figure that restricting to the three lowest average cost plans yields increases in utility ranging from 0% to 30% in most states. The variation across states stems from the fact that the lowest cost plans are relatively more inexpensive in some states than others and vary in their relative quality ratings. The average percentage utility gain is 16.7%. If we instead include omitted characteristics in the normative model, but adjust utility based on the estimated "number of plans" term, we find an average utility gain of 14.1%. That is, whether we exclude omitted characteristics or include them and use a revealed preference approach to estimate the degree of "congestion", we find comparable gains.

Figure 5 examines how the utility increases are distributed across the population. We see from these figures that the potential gains from moving to a small choice set are not evenly distributed across people: welfare increases for about 50% of the population, decreases for about 32% and remains constant for about 18%.

A priori, one might expect that any reduction in utility from this policy would arise due to individuals purchasing an insufficient amount of insurance. In fact, however, we find that the plans chosen in the small choice set offer slightly better risk protection on average than the plans which are chosen in the full choice set. Because of this, we find that the average welfare gains are actually larger if we impose more substantial risk aversion than we estimated in our model. The fact that we exclude the lone plan which offers full coverage in the donut hole does

---

individuals if they were perfectly rational and fully informed. That is, while individuals may (even rationally) decide to choose the AARP plan because they believe AARP will treat them well and because they are unable to fully evaluate the implications of plan characteristics for their financial well-being, in light of the fact that we *can* make this evaluation, we should no longer count the AARP fixed effect as part of welfare if it was only a heuristic for factors that we do observe.

[24] We report the results in this form to highlight the fact that we are choosing a different three plans in each state.

adversely impact some of the highest cost individuals. But the fact that all individuals are insured against catastrophic expenditures with only a modest co-insurance rate (once the donut hole ends) mitigates any utility losses from this exclusion.

As noted at the start of this section, these results are subject to a number of caveats. Most importantly, we ignore general equilibrium considerations. We do so because of uncertainty about how to specify the institutional structure following a reform which reduced the size of the choice set. Of course, there are possible reforms which would preserve the competitive nature of the bidding process while reducing the number of plans ultimately offered to consumers, such as first stage bidding across plans to offer one of a limited set of plan structures. Our results are also driven in part by the fact that the three lowest cost plans in each choice set are also among the best plans in terms of average utility all things considered. If we select three random plans in each choice set, we find welfare losses from restricting choice.

On the demand side, our analysis assumes that the estimated choice process is fixed. We assume that individuals choose according to the same positive utility function regardless of the size of the choice set – any utility increases from smaller choice sets arise because there is less scope for error. If individuals are in fact better able to evaluate alternatives in a smaller choice set, then our analysis would understate the potential gains. Moreover, surveys indicate that elders spend an average of 3 hours selecting their Part D plan (Kling et. al. 2008), so the dollar value of the hours saved by dramatically simplifying the choice process may be non-trivial as well.

Part VIII: Conclusion

The new delivery mechanism for a public insurance benefit introduced by the Medicare Part D program is a radical departure from the traditional public insurance model – and an

exciting opportunity to understand the role of choice in the delivery of public insurance. Using a unique data set we have provided the first evidence on the efficacy of the choices made by individuals under Part D. While individual choices are consistent with maximizing behavior such as preferring plans with lower premiums, lower out of pocket exposure and higher quality, they are inconsistent with the standard model in three important respects: individuals underweight out of pocket spending relative to premiums; they overweight plan characteristics beyond their own circumstances; and they do not fully appreciate the risk-reducing aspects of plans for themselves.

Our analysis in the last section suggests that given the normative assumptions we have outlined and given the positive model we estimated in the previous sections, there is substantial scope for increases in utility if consumers made better choices, and some of these gains could be realized by restricting to the three lowest cost plans. Our models do not distinguish between the case of boundedly rational consumers choosing plans they trust as a heuristic given the time-costs of fully evaluating choices, and the case where consumers simply err in underweighting out of pocket costs due to a lack of cognitive ability. While this distinction is important for evaluating the potential efficacy of providing consumers with additional information, it is less relevant to considering the welfare impact of altering the choice set: in either case, our estimates imply that consumers would be better off if there were less scope for choosing the wrong plan.

This analysis is subject to a number of caveats and should only be the first step in a richer modeling of the implications of choice in Part D plans. In particular, this approach should be contrasted with the impacts of more benign approaches such as improving the information set available to seniors and their success in accessing that information set. One interesting question raised by our analysis is why private firms have not emerged which provide such information.

The analysis used here could form the basis of a general equilibrium model which takes into account the fact that consumer inattention to individualized characteristics may lessen the threat of adverse selection, but pose new problems as firms choose product characteristics in part to exploit consumer error. Our analysis underscores the importance of such richer modeling before the public sector follows this model further, either within Part D or in other public insurance programs.

References

Ackerberg, D. A. and M. Rysman. 2002. "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects" *NBER Working Paper*.

Agarwal, S., J. C. Driscoll, X. Gabaix, and D. Laibson. 2007. *The Age of Reason: Financial Decisions Over the Lifecycle*. Vol. Pension Research Council; Wharton School: National Bureau of Economic Research Cambridge, Mass., USA.

Berry, S. and A. Pakes. 2001. "Estimating the Pure Hedonic Discrete Choice Model" *manuscript, Department of Economics, Yale University*.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman. 2005. "What's Psychology Worth A Field Experiment in Consumer Credit Market" *Yale University Economic Growth Center Discussion Paper*, 918.

Boatwright, P. and J. C. Nunes. 2001. "Reducing Assortment: An Attribute-Based Approach" *Journal of Marketing*, 65(3): 50-63.

Fratiglioni, L., D. De Ronchi, and H. A. Torres. 1999. "Worldwide Prevalence and Incidence of Dementia." *Drugs & aging*, 15(5): 365.

Harding, M. C. and J. Hausman. 2007. "Using a Laplace Approximation to Estimate the Random Coefficients Logit Model by Nonlinear Least Squares" *International Economic Review*, 48(4): 1311-1328.

Hausman, Jerry, Jason Abrevaya, and Fiona Scott-Morton. 1998. "Misclassification of the Dependent Variable in a Discrete-Response Setting" *Journal of Econometrics*, 87(2): 239-269.

Heiss, F., D. McFadden, and J. Winter. 2006. "Who Failed to Enroll in Medicare Part D, and Why? Early Results" *Health affairs*, 25(5): 344-354.

Heiss, F., D. L. McFadden, and J. K. Winter. 2007. "Mind the Gap! Consumer Perceptions and Choices of Medicare Part D Prescription Drug Plans" *NBER Working Paper*.

Irons, Ben and Cameron Hepburn. 2007. "Regret Theory and the Tyranny of Choice*" *Economic Record*, 83(261): 191-203.

Iyengar, S. S., W. Jiang, and G. Huberman. 2004. "How Much Choice is Too Much? Contributions to 401 (k) Retirement Plans" *Pension Design and Structure: New Lessons from Behavioral Finance*: 83-96.

Iyengar, S. S. and E. Kamenica. 2006. "Choice Overload and Simplicity Seeking" *University of Chicago Graduate School of Business Working Paper*.

Iyengar, S. S. and M. R. Lepper. 2000. "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of personality and social psychology*, 79(6): 995-1006.

Kamenica, E. 2008. "Contextual Inference in Markets: On the Informational Content of Product Lines" *American Economic Review*, 98(5): 2127-2128-2149.

Kling, J. R., S. Mullainathan, E. Shafir, L. Vermeulen, and M. V. Wrobel. 2008. "Misperception in Choosing Medicare Drug Plans" *Unpublished paper, June*.

Kuksov, D. and J. M. Villas-Boas. 2006. "When More Alternatives Lead to Less Choice" *Haas School of Business, University of California, Berkley, Mimeo*.

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

Petrin, A. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan" *Journal of Political Economy*, 110(4): 705-729.

Salthouse, T. A. 1996. "The Processing-Speed Theory of Adult Age Differences in Cognition" *Psychological review*, 103(3): 403-428.

Figure 1:

Histogram of cost savings from switching to lowest cost plan
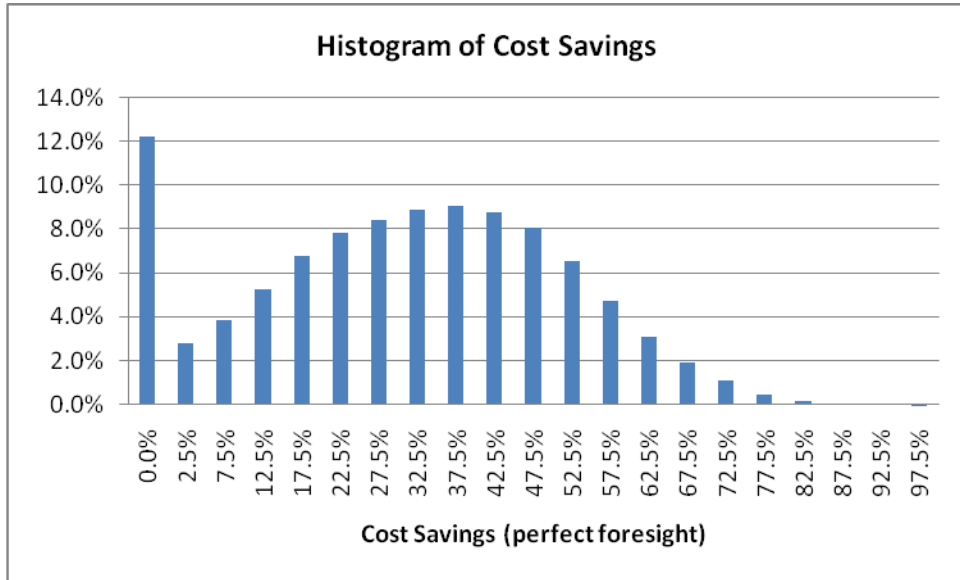
Figure 2:

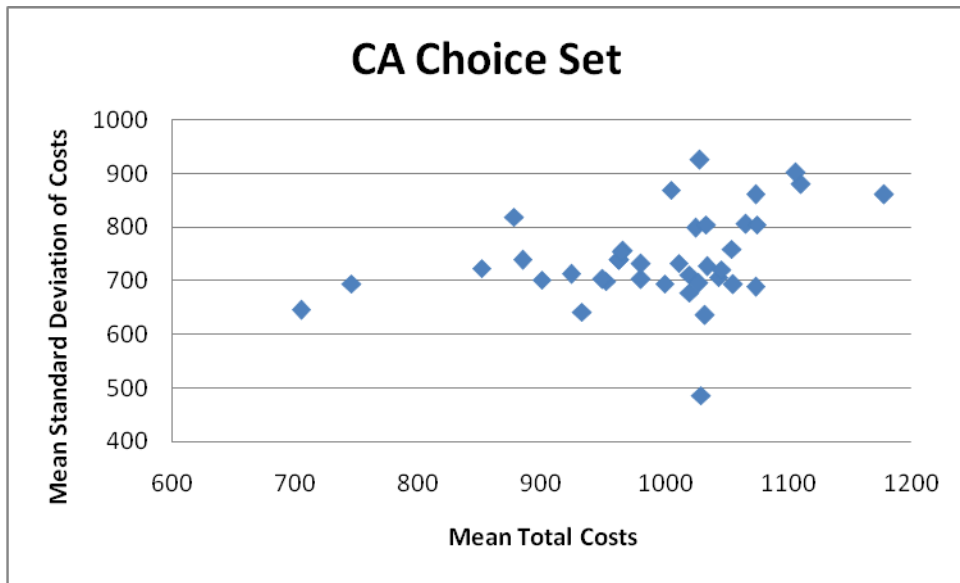**Average mean and standard deviation for each PDP plan in CA**
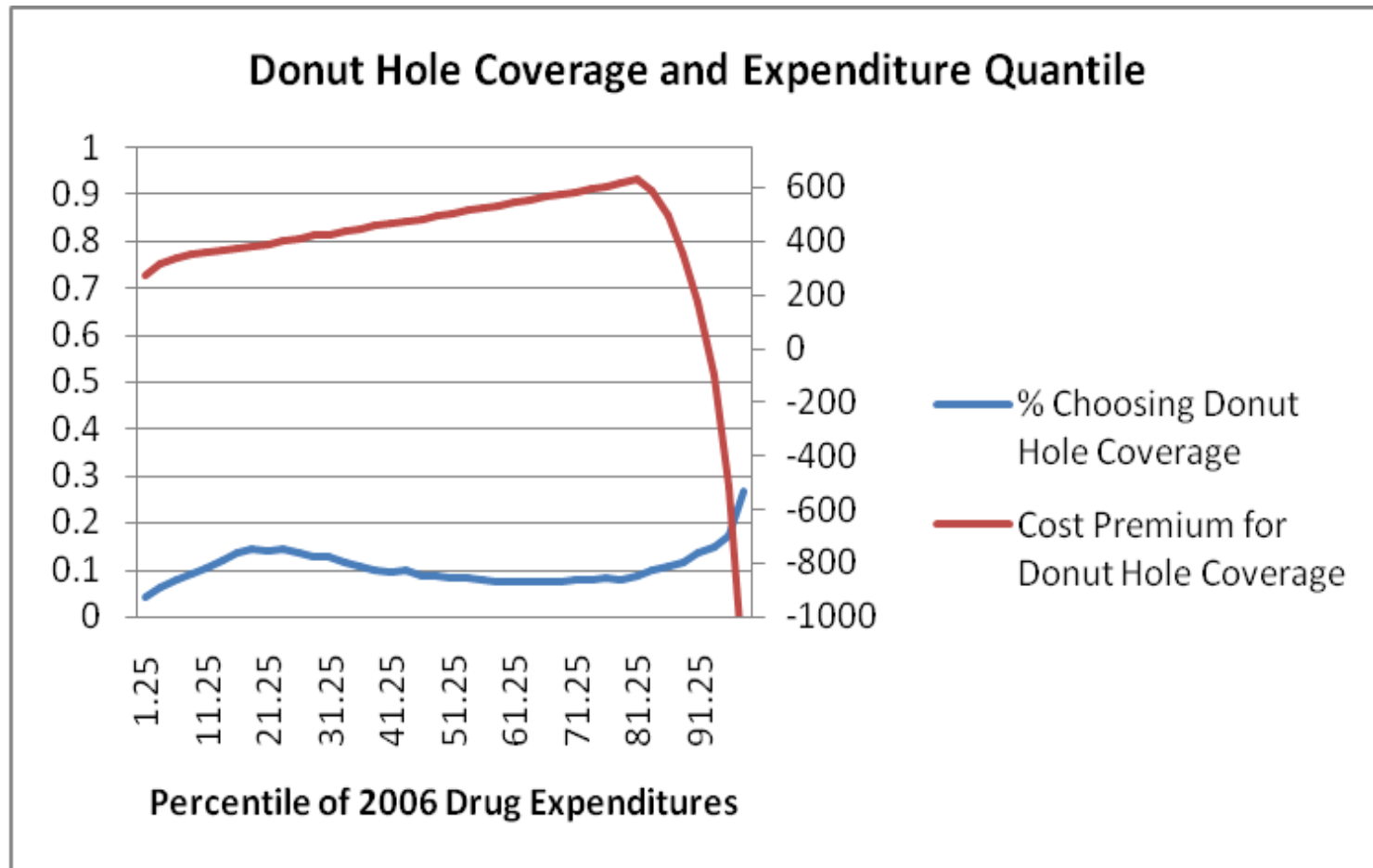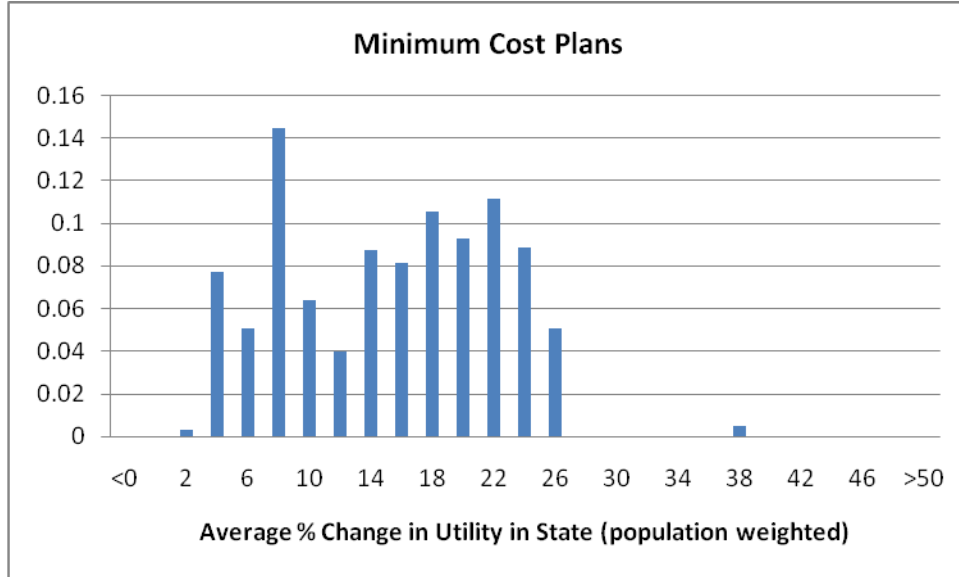
Figure 3:

Percent Choosing Donut Hole Coverage and Added Cost by Expenditure Quantile
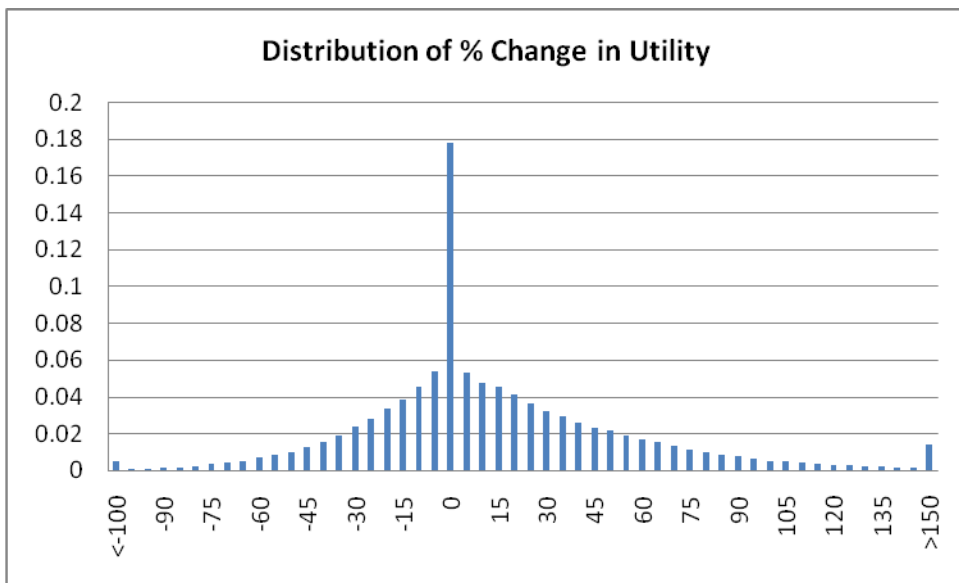
Figure 4:

Welfare Landscape for Three Lowest Cost Plans in Each State

Figure 5:

Distribution of the Percentage Change in Utility in the Minimum Cost Plans Treatment

Table 1: Conditional Logit Results

	(1)	(2)	(3)	(4)
Premium	-.3208**	-.7663**	-.4990**	-.5218**
(hundreds)	(.0024)	(.0038)	(.0061)	(.0069)
OOP Costs (realized)	-.1861**	-.1172**	-.0961**	-.0967**
(hundreds)	(.0014)	(.0015)	(.0015)	(.0016)
Variance	-.0163**	-.0004	-.0006	-.0005
(times 10^6)	(.0023)	(.0007)	(.0007)	(.0007)
Deductible	x	-.2899**	-.1628**	-.1674**
(hundreds)		(.0049)	(.0067)	(.0072)
Donut Hole	x	3.023**	1.762**	1.865**
		(.0181)	(.0277)	(.0303)
Generic Coverage	x	.4203**	.3004**	.2700**
		(.0140)	(.0175)	(.0177)
Cost Sharing	x	3.282**	1.189**	1.057**
		(.0538)	(.0741)	(.0778)
# of top 100 on Form	x	.0937**	.0587**	.0644**
		(.0007)	(.0017)	(.0018)
Avg. Quality	x	.7398**	x	x
		(.0039)		
Brand Dummies	NO	NO	YES	NO
Brand-State Dummies	NO	NO	NO	YES
Risk Index	10	0	0	0
# of patients	95742	95742	95742	95742
# of plans	702	702	702	702
# of states	47	47	47	47
# of brands	36	36	36	36

Notes: Table shows conditional logit results from estimating the model given in equation (6) by maximum likelihood. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. * indicates significance at the 5% level and ** indicates significance at the 1% level. The first column includes only premium, realized out of pocket cost and the variance measure. The second column adds controls for the indicated plan characteristics, the third column adds brand fixed effects and the fourth column adds brand-state fixed effects. Premiums, out of pocket cost and deductibles are in hundreds of dollars and the variance term is in millions. The cost sharing variable is computed as the average value of covered expenditures divided by total drug expenditures for individuals in the choice set. The average quality variable is a normalized version of the "average rating" index provided by CMS. The risk index is twice the coefficient on the variance divided by the coefficient on premiums scaled by 100. In the model in the text, this value equals one million times the coefficient of absolute risk aversion.

Table 2: Results with Private Information

	Final Sample	Restricted Sample	Private Info
% Private Information			.5818**
			(.0618)
Premium	-.7383**	-.7156**	-.7489**
(hundreds)	(.0038)	(.0094)	(.0132)
OOP Costs (realized)	-.1169**	-.1040**	-.1687**
(hundreds)	(.0016)	(.0039)	(.0094)
Variance	-0.0026	-.1103*	-.8574**
(times 10^6)	(.0014)	(.0517)	(.2947)
Deductible	-.2677**	-.3079**	-.2767**
(hundreds)	(.0014)	(.1257)	(.0137)
Donut Hole	2.823**	2.805**	2.870**
	(.0181)	(.0490)	(.0478)
Generic Coverage	.3066**	.4743**	.4784**
	(.0143)	(.0341)	(.0347)
Full Cost Sharing	2.990**	3.391**	2.829**
	(.0546)	(.1417)	(.1743)
# of top 100 on Form	.0939**	.0995**	.1005**
	(.0007)	(.0019)	(.0021)
Avg. Quality	.7167**	.7418**	.7512**
	(.0039)	(.0098)	(.0095)
Brand Dummies	NO	NO	NO
Brand-State Dummies	NO	NO	NO
Risk Index	1	31	229
# of patients	95742	15001	15001
# of plans	702	702	702
# of states	47	47	47
# of brands	36	36	36

Notes: Table compares conditional logit results with results from estimating the random coefficients model given in equations (11) and (14) using the Laplace approximation to the likelihood function developed by Hausman and Harding (2007) with bootstrapped standard errors. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. * indicates significance at the 5% level and ** indicates significance at the 1% level. The first column is identical to the second column of Table 1. The second column estimates the same model on a random subsample of 15,000 and the third column estimates the random coefficients model on this same subsample. Variable definitions are identical to Table 1. The "Percent Private Information" field corresponds to the variable τ_{frac} in the model in the text.

Table 3: Random Coefficients Results

	(1)	(2)	(3)
Premium	-.7156**	-.7677**	-.7354**
(hundreds)	(.0094)	(.0111)	(.0841)
Std. Deviation of Premium	x	.2940**	.2659**
		(.0560)	(.0330)
OOP Costs (realized)	-.1040**	-.1091**	-.1193**
(hundreds)	(.0039)	(.0052)	(.0171)
Std. Deviation of OOP Costs	x	.0001	.0226
		(.0003)	(.0506)
Variance	-.1103*	-.1486	-.0866
(times 10^6)	(.0517)	(.0926)	(.1237)
Std. Deviation of Variance	x	.2035**	.2603
		(.0685)	(.1769)
Deductible	-.3079**	-.2524**	-.2354
(hundreds)	(.1257)	(.0107)	(.1903)
Std. Deviation of Deductible	x	x	.2922*
			(.1452)
Donut Hole	2.805**	2.775**	2.523**
	(.0490)	(.0630)	(.3538)
Std. Deviation of Donut Hole	x	x	.8078
			(.5997)
Generic Coverage	.4743**	.4845**	-.0037
	(.0341)	(.0427)	(.2320)
Std. Deviation of Generic Coverage	x		1.106*
			(.5108)
Full Cost Sharing	3.391**	2.083**	1.829
	(.1417)	(.1852)	(1.304)
Std. Deviation of Cost Share	x		.1290
			(.4828)
# of top 100 on Form	.0995**	.0992**	.1405**
	(.0019)	(.0023)	(.0078)
Standard Deviation of top 100	x		.1071
			(.0790)
Avg. Quality	.7418**	.7729**	.7622**
	(.0098)	(.0090)	(.0358)
Standard Deviation of Quality	x	.3753**	.3503
		(.0146)	(.2182)
Brand Dummies	NO	NO	NO
Brand-State Dummies	NO	NO	NO
Risk Index	124	124	124

# of patients	15001	15001	15001
# of plans	702	702	702
# of states	47	47	47
# of brands	36	36	36

Notes: Table shows results from estimating the random coefficients model discussed in the heterogeneity section, estimated using the Laplace approximation developed in Hausman-Harding (2007) with bootstrapped standard errors. This model is identical to the model in equation (6), adding the normally distributed noise term (which is a function of $C_{ij} - \mu_{ij}$) and the variance adjustment. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. * indicates significance at the 5% level and ** indicates significance at the 1% level. In columns (2) and (3), each set of four rows reports the mean and standard deviation of random coefficient (and their standard errors). The first column estimates the conditional logit model on the subsample of 15000 (and is identical to the second column of Table 2). The second column adds random coefficients on financial characteristics and quality and the third column adds random coefficients on all variables. Variable definitions are otherwise identical to Table 1.

Table 4: Nested Logit Estimates of the Coefficient on Number of Plans

	Conditional Logit	Nested Logit
	Restricted Sample	Linear # of Plans
Premium	-.7156**	-.6090**
(hundreds)	(.0094)	(.0237)
OOP Costs (realized)	-.1040**	-.1150**
(hundreds)	(.0039)	(.0072)
Variance	-.1103*	.0233
(times 10^6)	(.0517)	(.0144)
Deductible	-.3079**	-.1202**
(hundreds)	(.1257)	(.0172)
Donut Hole	2.805**	2.585**
	(.0490)	(.1124)
Generic Coverage	.4743**	.1957**
	(.0341)	(.0527)
Full Cost Sharing	3.391**	2.248**
	(.1417)	(.2005)
# of top 100 on Form	.0995**	.1279**
	(.0019)	(.0065)
Avg. Quality	.7418**	.6829**
	(.0098)	(.0293)
# PDP Plans	x	-.0384**
(impact on PDP utility)		(.0055)
Log-sum Coefficient	x	.5943**
		(.0260)
MA Constant, # of MA Plans	NO	YES
OOP / Var * MA Dummies	NO	YES
Brand Dummies	NO	NO
Brand-State Dummies	NO	NO
Risk Index	31	-8
# of patients	15001	26385
# of plans	702	703
# of states	47	47
# of brands	36	37

Notes: Table compares conditional logit results with nested logit results from the model with the number of plans adjustment. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. * indicates significance at the 5% level and ** indicates significance at the 1% level. The first column estimates the conditional logit model on the subsample of 15000 (and is identical to the second column of Table 2). The second column estimates the

nested logit model on this subsample, plus an additional sample drawn from the set of individuals who chose MA plans so that the proportions match that in the CMS data from 2007. An additional choice is added to all choice sets and is indicated as chosen for the additional MA individuals. The nests include a nest for all PDP plans and a degenerate nest for MA plans. The value of premiums for this choice is set to the average of the three most popular MA plans in the choice set. The values of all other characteristics are set to zero for the MA option. The model includes an MA constant and interactions between quintiles of OOP costs and quintiles of variance with an indicator for the MA plan so that the estimation results are not sensitive to normalization that all OOP cost and variance variables are zero for the MA plan. In addition, the model controls for the number of MA plans in each choice set. Variable definitions are otherwise identical to Table 1.

Appendix A: Utilization Effects

To evaluate the impact of utilization effects, let us consider a simple model in which total drug spending S is a linear function of the coinsurance rate: $S = a - b \cdot CI$ and out of pocket costs are a linear function of total spending: $OOP = S \cdot CI$.

Let CI_0 denote the coinsurance rate of the plan in which the individual was actually enrolled and CI_1 denote the coinsurance rate of an alternative plan. Thus far, we have been operating under the assumption that $S_1^* = a - bCI_0$ and $OOP_1^* = CI_1(a - bCI_0)$. Since in fact $OOP_1 = CI_1(a - bCI_1)$, we have $OOP_1 - OOP_1^* = -bCI_1(CI_1 - CI_0)$. Since in fact $OOP_1 = CI_1(a - bCI_1)$, we have $OOP_1 - OOP_1^* = -bCI_1(CI_1 - CI_0)$. Normalizing the marginal utility of a dollar of out of pocket costs equal to 1, we have:

$$u_j = -OOP_j + \gamma CI_j + \epsilon_{ij} = -OOP_j^* + bCI_j(CI_j - CI_0) + \gamma CI_j + \epsilon_{ij} \qquad (16)$$

Our current model omits the $bCI_j(CI_j - CI_0)$ term. This equation suggests that our current model understates the value of plans with high coinsurances because it omits the fact that out of pocket costs would be smaller in these plans because consumers would consume fewer drugs. Of course, so far, this analysis ignores half the picture: the value of the foregone drugs. This value is given by the area under the demand curve: $CI_0(S_0 - S_1) + \frac{1}{2}(CI_1 - CI_0)(S_0 - S_1)$. Subtracting this term from the previous equation, we obtain:

$$u_j = -OOP_j^* + bCI_j(CI_j - CI_0) + bCI_0(CI_0 - CI_j)$$

$$+ \frac{1}{2}b(CI_j - CI_0)(CI_0 - CI_j) + \gamma CI_j + \epsilon_{ij} \tag{17}$$

$$= -OOP_j^* + \frac{1}{2}bCI_j(CI_j - CI_0)^2 + \gamma CI_j + \epsilon_{ij}$$

If coinsurances differ from the plan that consumers' actually chose, consumers would always be better off than we predicted assuming no utilization effects because they are choosing their utilization optimally.

What is the magnitude of this coefficient? The elasticity of spending with respect to the coinsurance rate is given by: $\frac{\partial \ln(S)}{\partial CI} = \frac{\frac{1}{S}\partial S}{\partial CI} = -\frac{b}{S}$ so $b = -S\epsilon$. For b approximately -0.2 and S approximately 1400 (the mean found in our data over the Part D enrollment period), this gives $b = 280$ if individuals fully internalize the impact of utilization effects on their plan choices. The magnitude of the omitted term is extremely small: even if an individual is considering switching from a plan with 30% coinsurances to one with 75% coinsurances (moving from the 1st to the 99th percentile of what is observed in the data), the welfare difference due to utilization effects would be only $21 ($\frac{1}{2} \cdot 280 \cdot .75 \cdot .45^2$). By contrast, the coefficient on cost sharing we estimate in the data implies that, controlling for the individual financial implications of plan choice, individuals would be willing to pay between $50 and $200 for an equivalent increase in coinsurances (dividing the estimated coefficient on cost sharing in our models by the coefficient on premiums and multiplying by 100 gives the dollar value of increasing cost sharing from 0% to 100%).

While it is difficult to analytically sign any bias from omitting the utilization effect term, we can directly include this term to determine whether it has any impact on

our estimates. This exercise shows the impact it would have on our estimates if we assumed that individuals did fully incorporate utilization effects, in contrast to our current analysis where we assume that they lack sufficient foresight to take them into account. Note that we cannot directly estimate the coefficient on this term because the value of the term itself depends on which outcome was chosen. Instead, we include the term

$CI_j (CI_j - CI_0)^2$ constraining the coefficient to be 140 times the coefficient on a dollar of

premiums (our estimate of $\frac{1}{2}b$ above) and evaluate whether this impacts the other

coefficients in our model.[25] Doing so, we find almost no impact on our results. The only noticeable change is an increase by 25% in the coefficient on cost-sharing, implying that individuals overvalue cost sharing slightly more than we had estimated in the absence of utilization effects.

---

[25] Note that even if the functional form assumptions made in this section are inexact, we could interpret this term as an attempt to capture parametrically the fact that with utilization effects, plans with different coinsurances become relatively more desirable than we give them credit for being because individuals can reoptimize.

Appendix B: A Simulation Exercise

In this section we demonstrate that the choice restrictions we test in our conditional logit model would be satisfied by a broad range of generalized utility functions. We start by considering CARA and CRRA utility functions with a range of values for risk aversion. We then take the cost distributions generated from the Part D data for each plan and simulate individuals' choices using the assumed utility function. Finally, we estimate the conditional logit model using these simulated choices and check whether the restrictions considered above hold. We add a small amount of noise to each observation so that the coefficients are identified at small levels of risk aversion.[26] The results of this exercise are reported in Appendix Table 1.

The CRRA utility function is evaluated at wealth $17,000, the median financial wealth of those age 65-74 in 2004 (EBRI, 2005). This is a conservative assumption which will tend to increase the curvature – and thus the degree of misspecification (it is especially conservative given that our analysis excludes individuals eligible for low-income subsidies). The absolute magnitude of the coefficients is determined by the amount of added noise (since this is the only omitted factor). A more informative measure is the size of each coefficient relative to the coefficient on premiums: this measure gives the dollar value of a one unit change in the included variable.

Regarding the first restriction, we see that, provided risk aversion is not too large (CRRA < 3, CARA < .0001), the coefficient on premiums equals the coefficient on OOP costs, and the two are very comparable in magnitude even at more extreme levels of risk aversion. The second restriction appears to hold roughly over the same range: the plan characteristics are insignificant controlling for the mean and variance of out of pocket

---

[26] The standard deviation of the noise to $1/20^{\text{th}}$ of the interquartile range of utility.

costs provided CRRA < 3 and CARA < .003. Even in the cases when they are significant, they are small in magnitude relative to premiums. Dividing the coefficient on each variable by the coefficient on premiums gives the dollar value of a 1 unit increase in the variable. In the CRRA = 10 case, the results would imply that individuals are willing to pay $9 for (full) donut hole coverage, would have to be paid $22 to go from a 0 deductible to a $250 deductible, and would have to be paid $8 to accept generic donut hole coverage (since these values are driven entirely by misspecification there is no reason the signs should be sensible). The third restriction is satisfied in the sense that we estimate risk aversion in all cases when the coefficient of risk aversion is greater than 0. The "risk index" is obtained by dividing two times the coefficient on the variance term by the coefficient on premiums. We showed above that with CARA utility and normal noise, this index should approximate to (10^6 times) the coefficient of absolute risk aversion. We see in Appendix Table 1 that this approximation seems to get things roughly correct in our sample (despite the fact that costs are non-normal), although it begins to break down when risk aversion is grows very large.

Appendix Table 1: Simulation Results

	CRRA (wealth = 17000)			CARA		
	1	**3**	**10**	**0.0001**	**0.0003**	**0.0005**
Premium	-5.385**	-5.057**	-3.407**	-5.288**	-4.774**	-3.877**
(hundreds)	(.0364)	(.0337)	(.0199)	(.0354)	(.0312)	(.0237)
OOP Cost	-5.355**	-4.911**	-2.767**	-5.284**	-4.517**	-3.379**
(hundreds)	(.0369)	(.0335)	(.0237)	(.0359)	(.0324)	(.0268)
Variance	-1.903**	-6.293**	-19.87**	-2.600**	-9.244**	-16.28**
(times 10^6)	(.1536)	(.1329)	(.1985)	(.1489)	(.1436)	(.1919)
Deductible	-.0409	-.1567**	-.4452**	-.0424	-.2177**	-.3728**
(hundreds)	(.0225)	(.0211)	(.0188)	(.0221)	(.0211)	(.0198)
Donut Hole	-.0390	-.2506**	.2974**	.0210	-.0254	-.2393**
	(.0863)	(.0781)	(.0549)	(.0657)	(.0703)	(.0626)
Generic Donut Hole	-.0470	-.1412*	-.2829**	-.0325	-.1326**	-.2782**
	(.0680)	(.0652)	(.0560)	(.0657)	(.0629)	(.0571)
Cost Sharing	.1905	1.207**	2.791**	.2915	1.672**	2.726**
	(.2546)	(.2425)	(.2114)	(.2505)	(.2393)	(.2230)
# in Top 100	.0115	-.0045	-.0056**	.0110	-.0121**	-.0098**
	(.0036)	(.0031)	(.0023)	(.0036)	(.0029)	(.0024)
Avg. Quality Rating	-.1281	-.0508	.0525**	-.1186	-.0082	.0349**
	(.0181)	(.0164)	(.0128)	(.0175)	(.0155)	(.0137)
Risk Index	71	249	1166	98	387	840
# of patients	94732	94732	94732	94732	94732	94732
# of plans	702	702	702	702	702	702
# of states	47	47	47	47	47	47
# of brands	36	36	36	36	36	36

Notes: Table shows conditional logit results from estimating the model given in equation (6) by maximum likelihood using simulated choices. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. * indicates significance at the 5% level and ** indicates significance at the 1% level. The sample differs slightly from that in Table 1 because individuals with greater than 17000 in total costs for any plan are dropped. All simulated choices are based on the cost distribution generated from the realized costs of 200 individuals in the same decile of 2005 total costs, decile of 2005 total days supply of branded drugs and decile of 2005 days supply of generic drugs. The first three columns compute expected utility using a CRRA utility function with wealth of 17000 and the indicated coefficient of relative risk aversion, assuming that individuals select the choice which maximizes expected utility. The final three columns compute expected utility using a CARA utility function with the indicated coefficient of absolute risk aversion. Variable definitions are otherwise identical to Table 1.

Appendix C: Modeling Unknown Component of Spending

As discussed in the text, the results of our private information model are consistent with two very different normative models of choice with private information. The first is that individuals are using all available information to make forward-looking choices of plans for 2006, but are simply mis-weighting premiums and out of pocket costs in making those choices. The alternative is that individuals are not using all available information, but rather are paying attention only to a part of their prescription drug expenditures. For that portion to which they are attentive, individuals are rationally weighting premiums and out of pocket costs in the same way in making their decision. Yet individuals do not respond to variation in out of pocket costs beyond that portion.

To address this point, we make two changes to the model. First, we allow for a component of predicted costs which − while it can be predicted given 2005 characteristics − is still unknown to individuals at the time when they make their choice. Since previously we allowed the coefficient on predicted costs to fall short of the coefficient on premiums, this modification just allows for heterogeneity across individuals in the portion of the variance in predicted out of pocket costs which is observed. Second, we estimate the degree of private information by assuming that if OOP costs were fully observable, it would be weighted identically with π_{ij} in patients' utility functions. Conceptually, the estimates in the previous section apply to the case where the small coefficient on OOP costs reflects misweighting even though out of pocket costs are known (so there is no uncertainty), whereas the estimates in this section apply to the case where a component of out of pocket costs remains unknown. We must distinguish these

two cases to determine the appropriate variance in out of pocket costs for each alternative.

We begin by decomposing realized costs into predicted costs and the component which cannot be predicted given 2005 characteristics:

$$C_{ij} = \mu_{ij} + \nu_{ij} \tag{18}$$

Formerly, we assumed $\nu_{ij} = e_{ij} + \eta_{ij}$ and attempted to determine the magnitude of the component of ν_{ij} that was observable relative to μ_{ij} which we assume was fully observable. In this model, we make an analogous decomposition, but we write $\nu_{ij} = \hat{\eta}_{ij} + \eta_{ij}$ to distinguish e_{ij}, which is the degree of private information assuming individuals are aware of predicted out of pocket costs from $\hat{\eta}_{ij}$, which is identified assuming individuals are fully aware only of premiums. We also decompose predicted costs so that $\mu_{ij} = \bar{\mu}_i + \hat{\mu}_{ij} + \tilde{\mu}_{ij}$ where $\hat{\mu}_{ij} \sim N(0, \alpha \mu_{ij}^2 \sigma_\mu^2)$ and $\tilde{\mu}_{ij} \sim N(0, (1-\alpha)\mu_{ij}^2 \sigma_\mu^2)$, where $\hat{\mu}_{ij}$ is the component of predicted costs known to the individual at the time of choice, and $\tilde{\mu}_{ij}$ is the residual component. The variance of μ_{ij} captures the degree of heterogeneity in the amount of information possessed by individuals. If this variance is large, then the same observed predicted costs might correspond to very different observed predicted costs across individuals and plans. We assume that this variance is proportional to $\mu_{ij} - \bar{\mu}_i$. This allows for the fact that information about drugs has a multiplicative flavor: the information that a plan covers certain drugs has a larger impact on your choices if you consume more of those drugs or if those drugs are more costly.[27] The

---

[27] The multiplicative model is also easier to estimate because it generates a random coefficients model where the random coefficient has fixed mean and variance across plans for a given individual.

constant σ_μ^2 which determines the degree of heterogeneity in information about predicted costs across individuals and plans is estimated.

Define (positive) utility as:

$$u_{ij} = \pi_j \beta_0 + \mu_{ij}^* \beta_0 + \epsilon_{ij} \tag{19}$$

where $\mu_{ij}^* = \hat{\mu}_{ij} + e_{ij}$ is the component of out of pocket costs known to the individual at the time of choice. Substituting in for μ_{ij}^*, we obtain:

$$u_{ij} = \pi_j \beta_0 + C_{ij} \beta_0 - \tilde{\mu}_{ij} \beta_0 - \eta_{ij} \beta_0 + \epsilon_{ij} \tag{20}$$

As before, we apply the normal updating formula and obtain:

$$f(\eta_{ij} | C_{ij}, \mu_{ij}) \sim N\left(\frac{\sigma_{ij}^2}{\sigma_{ij}^2 + \tau_{ij}^2}(C_{ij} - \mu_{ij}), \frac{\tau_{ij}^2 \sigma_{ij}^2}{\sigma_{ij}^2 + \tau_{ij}^2}\right) \tag{21}$$

And:

$$f(\tilde{\mu}_{ij} | C_{ij}, \mu_{ij}) \sim N\left((1 - \alpha)(\mu_{ij} - \bar{\mu}_i), \alpha(1 - \alpha)(\mu_{ij} - \bar{\mu}_i)^2 \sigma_\mu^2\right) \tag{22}$$

As before, we assume that $\tau_{ij}^2 = \tau_{frac} \widetilde{\sigma_{ij}^2}$ and $\sigma_{ij}^2 = (1 - \tau_{frac})\widetilde{\sigma_{ij}}^2$. As noted, we have now assumed that the coefficient on the observed portion of OOP costs is β_0, the same as the coefficient on premiums. The fact that the observed coefficient on realized costs is less than β_0 in the original conditional logit model is rationalized by two factors: first, part of predicted costs is unobserved (the $\tilde{\mu}_{ij} \beta_0$ term) and second, part of the difference between realized and predicted costs is unobserved (the $\eta_{ij} \beta_0$ term). Note that we can interpret the $\tilde{\mu}_{ij}$ term as a random coefficient on $\mu_{ij} - \bar{\mu}_i$ with mean $\beta_0(1 - \alpha)$ and variance $\beta_0 \alpha(1 - \alpha)\sigma_\mu^2$.

In both cases, we determine the portion of the variance explained by the unobserved factor by determining magnitude of the coefficient relative to the magnitude of the coefficient on premiums. τ_{frac} is now defined so that the coefficient on $C_{ij} - \mu_{ij}$ is $\beta_0(1 - \tau_{frac})$. Likewise, α is identified by the fact that the coefficient on μ_{ij} is $\beta_0(1 - \alpha)$.

Appendix Table 2 reports the results from estimation of this model. A few points are notable. First, our estimate of the degree of private information now shrinks to 20% since we are interpreting the gap between premiums and out of pocket costs as evidence that some costs are unforeseen. This gap – the "Fraction Observed" – implies that almost consumers are inattentive to almost 80% of the variation in out of pocket costs when they make their choices. Because of this, there is now substantial variation in the risk facing individuals across plans which reduces the standard deviation in our measure of the variance term, which we now measure to be close to zero (the point estimate is very small but would actually indicate risk-loving behavior). This model controls more flexibly than previous models for the full impact of out of pocket costs on choice (taking into account both heterogeneity in the degree of private information and in the fraction of predicted out of pocket costs which is observed). The fact that the magnitude of the premium coefficient increases by about .2 suggests that this coefficient was previously biased downwards by the failure to adequately account for this heterogeneity.

Appendix Table 2: Results w/ private information and unobserved component of costs

	Restricted Sample	Private Info.	Unobserved Costs
Tau (% Private Info)		.5818**	.1703**
		(.0618)	(.0042)
Fraction Observed (alpha)			.2235**
			(.0121)
Std. Deviation Observed			1.778**
			(.5466)
Premium	-.7156**	-.7489**	-.9207**
(hundreds)	(.0094)	(.0132)	(.0692)
OOP Costs (realized)	-.1040**	-.1687**	
(hundreds)	(.0039)	(.0094)	
Variance	-.1103*	-.8574**	.0538*
(times 10^6)	(.0517)	(.2947)	(.0237)
Deductible	-.3079**	-.2767**	-.2668**
(hundreds)	(.1257)	(.0137)	(.1300)
Donut Hole	2.805**	2.870**	3.670**
	(.0490)	(.0478)	(.0457)
Generic Coverage	.4743**	.4784**	.9791**
	(.0341)	(.0347)	(.0510)
Full Cost Sharing	3.391**	2.829**	2.909**
	(.1417)	(.1743)	(.1580)
# of top 100 on Form	.0995**	.1005**	.1006**
	(.0019)	(.0021)	(.0017)
Avg. Quality	.7418**	.7512**	.7791**
	(.0098)	(.0095)	(.0081)
Brand Dummies	NO	NO	NO
Brand-State Dummies	NO	NO	NO
Risk Index	31	229	-12
# of patients	15001	15001	15001
# of plans	702	702	702
# of states	47	47	47
# of brands	36	36	36

Notes: Table compares conditional logit results with results from estimating the random coefficients model given in equations (20), (21) and (22) using the Laplace approximation to the likelihood function developed by Hausman and Harding (2007) with bootstrapped standard errors. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. * indicates significance at the 5% level and ** indicates significance at the 1% level. The first column estimates the conditional logit model on the subsample of 15000 (and is identical to the second column of Table 2). The second column again reports the results from the model with no unknown component of spending (the third column of Table 2) and the third column

estimates the model with an unknown component of spending. The "Percent Private Information" field corresponds to the variable τ_{frac} in the model in the text. The fraction observed field corresponds to the variable α and the standard deviation observed field corresponds to $\beta_0\alpha(1-\alpha)\sigma_\mu^2$. Variable definitions are otherwise identical to Table 1.

Appendix D: Mathematical Details of Utility Analysis with Mistakes

In this section, we derive the formulae used to compute utility when positive and normative utility functions differ. We assume that choices are given by the positive utility function:

$$\hat{u}_{ij} = x_{ij}\hat{\beta} + \hat{\epsilon}_{ij} \tag{23}$$

Whereas they are evaluated using the normative utility function:

$$u_{ij} = x_{ij}\beta + \epsilon_{ij} \tag{24}$$

We assume in all of our models that the marginal utility of income is constant and given by the estimated coefficient on premiums (β_0). In the case where omitted characteristics are excluded from the normative utility function ($\epsilon_{ij} = 0$), the analysis is straightforward. Utility depends only on the characteristics of the plan which is chosen, and since we assume a constant marginal utility of income, this is given by $\frac{u_{ij}}{\beta_0}$. To evaluate the utility gains from moving to a small choice set, we compare the plan chosen from the full choice set with choices simulated using equation (23) in the small choice set.

The case where omitted characteristics do continue to enter normative utility is more involved. In this case, $\epsilon_{ij} = \hat{\epsilon}_{ij}$. For simplicity, we normalize $\beta_0 = 1$. As is standard in discrete choice models, we compute the expected value of consumer surplus:

$$E(CS) = E(\max u_{ij})$$

$$= \sum_{j}^{J} E(u_{ij} \,|\, \hat{u}_{ij} > \hat{u}_{ik} \text{ for all } k \neq j) P(\hat{u}_{ij} > \hat{u}_{ik} \text{ for all } k \neq j) \tag{25}$$

The second term in the summation is the ordinary logit probability computed from the positive utility. To simplify the first term, we use the fact that if $X_1, \dots, X_n$ are independent type I extreme value random variables with scale parameter 1 and location parameters $\mu_1, \dots, \mu_n$ respectively, then the conditional distribution of X_i given that X_i is the maximum of $X_1, \dots, X_n$ is also type I extreme value with location parameter $\ln \sum_{j}^{n} e^{\mu_j}$.

We can rearrange the first term so that:

$$E\left(u_{ij} \,\middle|\, \hat{u}_{ij} > \hat{u}_{ik} \text{ for all } k \neq j\right) =$$

$$E(x_{ij}\beta + \epsilon_{ij} \,|\, x_{ij}\hat{\beta} + \epsilon_{ij} > x_{ik}\hat{\beta} + \epsilon_{ik} \text{ for all } k \neq j) \tag{26}$$

$$E(x_{ij}\beta + \epsilon_{ij} \,|\, x_{ij}\beta + \epsilon_{ij} > x_{ik}\hat{\beta} + x_{ij}\beta - x_{ij}\hat{\beta} + \epsilon_{ik} \text{ for all } k \neq j)$$

Substituting in using the above result, we obtain:

$$E\left(x_{ij}\beta + \epsilon_{ij} \mid x_{ij}\beta + \epsilon_{ij} > x_{ik}\hat{\beta} + x_{ij}\beta - x_{ij}\hat{\beta} + \epsilon_{ik} \text{ for all } k \neq j\right) =$$

$$\ln\left(e^{x_{ij}\beta} + \sum_{k \neq j} e^{x_{ik}\hat{\beta} - x_{ij}\hat{\beta} + x_{ij}\beta}\right) =$$

$$\ln\left(e^{x_{ij}\beta}\left(1 + \sum_{k \neq j} e^{x_{ik}\hat{\beta} - x_{ij}\hat{\beta}}\right)\right) = \tag{27}$$

$$x_{ij}\beta + \ln\left(e^{-x_{ij}\hat{\beta}}\left(e^{x_{ij}\hat{\beta}} + \sum_{k \neq j} e^{x_{ik}\hat{\beta}}\right)\right) =$$

$$x_{ij}\beta - x_{ij}\hat{\beta} + \ln\sum_{k} e^{x_{ik}\hat{\beta}}$$

Substituting this back into the expression for expected consumer surplus, we obtain:

$$E(CS) = \sum_{j}\left(x_{ij}\beta - x_{ij}\hat{\beta}\right)\frac{e^{x_{ij}\hat{\beta}}}{\sum_{k} e^{x_{ik}\hat{\beta}}} + \ln\sum_{k} e^{x_{ik}\hat{\beta}} \tag{28}$$

This equation has a natural interpretation. The second-term is the familiar log-sum term from logit welfare evaluated on the positive utility function: this is what welfare would be if the positive utility function were also the normative utility function. The first term is the "mistake term". This term corrects the utility calculation for the fact that the log-sum term misevaluates the observed component of utility. For each plan, this correction is equal to the difference in utility between the normative and positive utility functions times the probability that the plan is chosen.