ONLINE APPENDIX

A Institutional Context and Data

Institutional Context Congress expanded Medicare to include prescription drug coverage via Medicare Part D in 2006. In 2016, approximately 41 million individuals benefited from the Medicare Part D program and the Congressional Budget Office estimates that the government currently spends over $94 billion on Part D annually.

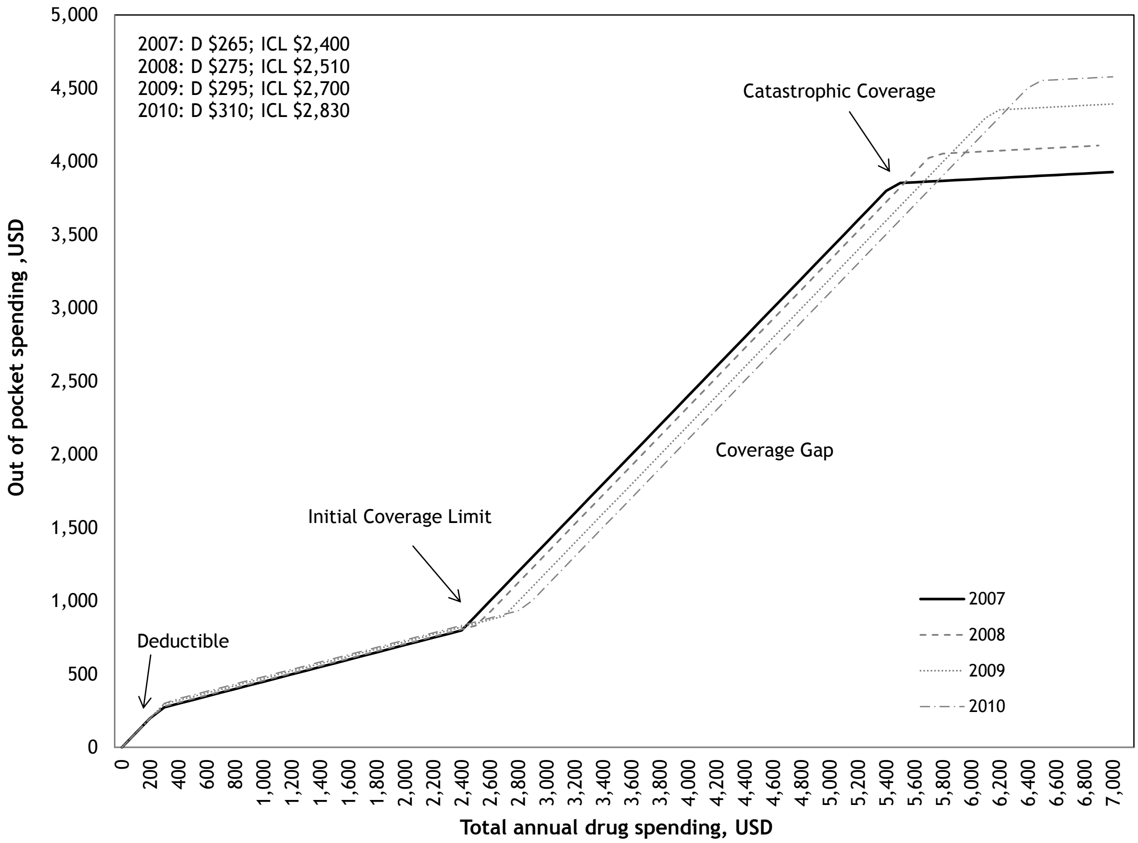
The supply-side of the Part D program has a unique, and controversial, design. Oliver, Lee and Lipton (2004) discuss the political origins of Part D and its mixed reception in the first years of the program, particularly among consumers. Unlike the rest of Medicare, the drug insurance benefit is administered exclusively by private insurance companies. At the same time, the setting differs from more conventional private insurance markets in two key ways. First, firms are highly regulated and product selection is restricted; CMS sets an annual Standard Defined Benefit (SBD), which defines the minimum actuarial level of insurance that the private plans are required to provide. The SDB has a non-linear structure illustrated in Figure A1; it includes a deductible, a 25 percent co-insurance rate and the infamous "donut hole," which is a gap in coverage at higher spending levels. As long as an actuarial minimum is satisfied, insurers are allowed to adjust and/or top up the SDB contract design, which generates variation in contracts' financial characteristics. In addition, contracts may be differentiated by the quality of insurer's pharmacy networks, which drugs are covered, and other non-pecuniary quality measures.

The second way in which Part D environment differs from more conventional insurance markets is that consumers bear only a fraction of the cost in the program. As much as 90 percent of insurer revenues come from the government's per capita subsidies.[26] For individuals, who are eligible for low-income-subsidies, these subsidies can go up to 100 percent.

Subsidies are determined through a complex system that depends on firm behavior. First, the government administers an annual "simultaneous bidding" mechanism. According to this mechanism, the insurers that want to participate in the program submit bids for each insurance plan in each region they want to offer. By statute, the bids are supposed to reflect how much revenue the insurer "needs," including a profit margin and fixed cost allowances, to be able to offer the plan to an average risk beneficiary. There are several nuances buried in the set-up of the bidding procedure that are important for insurers' incentives and will

[26] See Table IV.B11 of 2012 Trustees of Medicare Annual Report.

Figure A1: Standard Defined Benefit in Medicare Part D



Insurers in the Medicare Part D program are required to provide coverage that gives at least the same actuarial value as the Standard Defined Benefit (SDB) that is illustrated in this Figure for years 2007-2010. The SDB design features a deductible, a co-insurance rate of 25 percent up to the initial coverage limit (ICL) and the subsequent "donut hole" that has a 100 percent co-insurance until the individual reaches the catastrophic coverage arm of the contract. The graph illustrates these features of the SDB by mapping the total annual drug spending (on the x-axis) into the out-of-pocket expenditure (on the y-axis). The legend in the top left corner specifies the key parameter values of the contract and their evolution over time: deductible level ("D") and the start of the coverage gap phase ("ICL").

enter the insurers' profit function in our empirical model. First, Medicare sets a minimum required actuarial benefit level that plans have to offer. Plans are allowed to offer more coverage ("enhance" the coverage), but that enhanced portion is not subsidized. Thus, when submitting their bids plans are supposed to only include the costs they expect to incur for the baseline actuarial portion of their benefit. The incremental premium for enhanced coverage has to be directly passed on to consumers. CMS takes the bids submitted by insurers for each of their plans and channels them through a function that outputs which part of the bid is paid by consumers in way of premiums, and which part is paid by CMS through subsidies. This function takes the bids of all Part D PDP plans nationwide, adds in the bids submitted by MA-PD plans, weights them by lagged enrollment shares of the plans, and takes the average. 25 percent of the national bid average together with the difference between the plan's bid and the national average is set as the consumer's premium.

The second feature of the subsidy mechanism concerns the role of low income beneficiaries. CMS utilizes the same insurer bids to determine which insurance plans qualify to enroll randomly assigned LIS beneficiaries. For each geographic market, CMS calculates the enrollment-weighted average consumer premium. This average constitutes the subsidy amount that low-income beneficiaries receive, known as LIS benchmark or LIPSA. Plans that have premiums below the LIS benchmark qualify for random assignment of LIS enrollees (Decarolis, 2015).

Last but not least, subsidies vary depending on the health of individual enrollees. Insurers receive higher subsidies for sicker beneficiaries through the process known as risk adjustment. Each beneficiary is assigned a continuous risk score that is calculated such that the individual of average health within the Medicare program has a risk score of one. Sicker beneficiaries get assigned higher risk scores. CMS payments to insurers are scaled by this risk score to reflect higher expected expenditures that insurers would incur for sicker enrollees. For a consumer with risk score r_i, CMS pays the insurer the bid times the risk score, minus consumer premium.[27] While the premiums do not vary across consumers, CMS payment is higher for less healthy enrollees. For beneficiaries of especially poor health, insurers also receive a so-called "re-insurance" payment, by which the government pays about 80 percent of individual's spending after this individual has incurred relatively high pharmaceutical

[27] While risk adjustment decreases the expected differences in costs across enrollees, the adjustment is typically imperfect. Mathematically, multiplicative risk adjustment cannot equalize level differences in ex post costs. Further, as a large literature on the design of risk adjustment systems notes, perfect risk adjstment is difficult to achive for a variety of reasons, such as imperfect information and/or time lags about individuals' health status. Geruso and Layton (2017) provides a recent overview.

costs. This effectively further increases subsidies for sicker beneficiaries.

These mechanisms of risk adjustment and re-insurance are parts of a three-pillar risk equalization system within Medicare Part D. The third part of this system—risk corridors—directly decreases insurers' exposure to bottom line risk in profits by capping certain levels of losses (and symmetrically taxing unexpected gains). These three mechanisms play two roles in the market. First, they effectively result in higher subsidies for individuals of worse health. Second, they serve to mute insurers' incentives to cream-skim healthier enrollees by trying to equalize the marginal cost of each enrollee from the insurers' perspective.

MA-PD plans participate in the bidding mechanisms and conceptually are paid the same subsidy. In practice, MA-PD plans can use the subsidy payments from the medical part of MA to reduce consumer premiums for MA-PD plans, so that consumer-facing premiums are often zero for the pharmaceutical part of the MA program (Curto et al., 2015; Starc and Town, 2015).

Data and Sample Our primary data source for analysis are Research Identifiable Files containing a 20 percent random sample of all pharmaceutical claims for Medicare beneficiaries from years 2007-2010. The dataset (Master Beneficiary Summery File) contains information about age, sex, place of residence, health conditions, and Part D enrollment for each beneficiary in the 20 percent sample. For those individuals that have purchased some type of Part D coverage, we observe detailed information about pharmaceutical claims (Part D Drug Event File). For each claim, we observe the date of purchase, the pharmacy, the prescribing physician, the total cost of transaction at the pharmacy, how much of the claim was paid by the consumer, by the plan, and by CMS. We can further observe if the consumer was eligible for low income subsidies. These data are described in detail by CMS Research Data Assistance Center (ResDAC): www.resdac.org.

We restrict the 20 percent sample as follows. We keep individuals that reside in 50 US states. We further keep individuals that either purchased a stand-alone Medicare Part D PDP plan, or a Medicare Advantage MA-PD plan, or no Part D coverage. This restriction excludes individuals who were eligible for other types of pharmaceutical coverage - primarily through employer-sponsored Part D plans. This set of individuals constitutes our analytic sample.

We enrich these data with publicly available information on the Part D plan options that were available to each consumer based on their geographic market. For each plan option we observe a vector of detailed characteristics, including deductibles, coverage in the gap, and insurer brands, as well as consumer-facing premiums. These CMS Medicare Part D Landscape files for years 2007-2010 are provided directly by CMS at www.cms.gov.

Last, we use public information on total and plan-level Part D enrollment as reported on www.cms.gov to validate the estimates based on our 20 percent sample. The same source of information provides data on average national bids, consumer premiums, and low income benchmark thresholds. We use these pieces of information to infer bids for each Part D plan, as well as which plans were eligible for LIS random assignment.

For details on data sources, file names and processing programs, please see the package with replication programs that accompanies the paper.

B Consumer Risk Types

Our supply side model allows for a plan's marginal cost to change as a function of the plan's risk pool; in other words, the supply-side model allows for adverse or advantageous selection. To facilitate differential risk sorting across plans, our demand model allows for consumer preferences to vary with their health risk. To operationalize endogenous marginal costs as an equilibrium outcome, we start by constructing a continuous one-dimensional measure of each consumer's risk. The risk measure that we construct is similar in spirit to the risk score that CMS uses for risk-adjustment. We then discretize the risk space to enable computational feasibility.

We start by generating a continuous measure of predictable risk for each Medicare beneficiary in the data. For all consumers on the market that are included in our 20 percent sample, we observe information about their health status, measured by the indicators for the presence of 66 chronic conditions. To map the indicators for chronic conditions and basic demographic information into a one-dimensional pharmaceutical spending risk score, we estimate a linear relationship between total pharmaceutical spending and information about individual's chronic conditions, age, sex, race, and eligibility for low income subsidies. The model is estimated using observations on individuals who had Part D coverage and hence their pharmaceutical claims were recorded in the data. We then use the estimated regression coefficients to predict total pharmaceutical spending among all potential Part D consumers. For each individual, we construct a risk score measure that is equal to the ratio of the individual's predicted spending to the average predicted spending in the sample. By construction, the average consumer receives a risk score of 1.

In the next step we discretize the constructed risk scores. We divide all potential Part

D enrollees into six discrete risk groups. As the distribution of risk scores has thin tails and is concentrated around the mean, we define the first and fifth risk score groups to be the bottom 5th percentile and top 5th percentile of risk scores among beneficiaries that are not eligible for low income subsidies. The second risk group are individuals with risk scores between 5th and 25th percentile; third risk group - 25th to 75th percentile, and fourth risk group - 75th to 95th percentile. We let all LIS-eligible beneficiaries be in a separate, sixth, risk group. We cannot observe individual's LIS eligiblility if the individual is not enrolled in a Part D plan. Hence, we assume that all consumers that are not observed having a Part D plan are not eligible for LIS subsidies. This appears to be a reasonable assumption, as LIS eligiblity for many individuals is determined automatically and those who are not actively choosing a Part D plan get randomly assigned to one.

For each of the six (including LIS) risk groups, we define two objects of interest. First, we compute the average risk score in each group. The resulting averages are: 0.17 risk score for the first risk group, 0.47 for the second, 0.81 for the third, 1.63 for the fourth, 2.49 for the fifth, and 1.78 for the LIS. In the process of risk-adjustment, CMS constructs similar risk scores for each individual. These "real" risk scores are then used to multiply insurers' bids before CMS determines its payments to the insurer. CMS premium subsidy is equal to the difference between the insurer bid multiplied by a risk score net of the beneficiary premium. To replicate this idea and to allow insurers to collect higher revenue when enrolling higher cost individuals, we use the average risk scores per risk type as reported above to multiply insurer bids and compute insurer revenue.

In the next step, we define how expected costs vary across five risk groups. Since we observe only one equilibrium price per plan, we have to make parametric assumptions on how costs across six risk group types relate to each other, as there is only one unknown cost parameter that is identified in the marginal cost inversion system. Critically, we assume that risk-type specific costs are related multiplicatively to each other and are the same across all plans. We normalize the marginal cost for risk group one as the baseline marginal cost c. We then assume that the marginal cost for risk groups two to six (the LIS) are equal to $\kappa_2 c$ to $\kappa_6 c$.

We estimate κ's from the claims data. For each risk type, we compute the average total insurer liability that each risk type generates in our data. For the purpose of this computation, we pool the whole universe of claims for Part D enrollees. Insurer liability is defined as the total pharmaceutical expenditures minus patient out of pocket and third-party (mostly government) payments. We estimate that the average insurer liability by risk type is

\$713 for risk type 1, \$851 for risk type 2, \$1,070 for risk type 3, \$1,557 for risk type 4,\$1,926 for risk type 5, and \$1,374 for the LIS or risk type 6.

Alternatively, one could use total expenditures across risk types rather than insurer liability as a measure of relative costs. That approach would dramatically overestimate the cost of the highest risk type enrollees, since for these enrollees the government reinsurance program significantly lowers insurer liability. In particular, while risk type 1 has an average total cost of \$1,206, risk type 5 would have an average total cost of \$4,374. This large difference, however, is misleading, since in practice a lot of the latter's expenditures would not fall under insurer liability. To avoid this misrepresentation of costs, we use average insurer liability to compute relative costs.

We normalize the resulting average insurer liability by the payment in risk type 1 (i.e. \$713). The resulting estimates of κ are: 1 for risk type 1, 1.19 for risk type 2, 1.50 for risk type 3, 2.18 for risk type 4, 2.70 for risk type 5, 1.93 for the LIS-eligible beneficiaries. Applying these estimated multipliers allows us to reduce the marginal cost inversion problem to a one equation in one unknown, while at the same time endogenizing the marginal costs of Part D plans to equilibrium allocation of risks across plans.

C Consumer Inertia

The literature has documented consumer inertia in the choice of Medicare Part D plans (Ericson, 2014; Ho et al., 2015; Polyakova, 2016; Wu, 2016; Heiss et al., 2016). Consumers tend to choose their plan when entering the Part D program for the first time, and then only infrequently make changes to their plans. To account for inertia in demand, we take a reduced form approach and include the vintage of a plan in the utility function as a proxy for inattention and switching costs. The idea is that the longer the plan has been around, the larger the proportion of its enrollees are incumbent consumers from previous years. As we illustrate in what follows, our reduced-form specification corresponds to an explicit structural model of inattention and choice.

We start by borrowing from Hortacsu et al. (2015), who posit a two-stage model of choice with inattention. In the first stage consumers make an active choice with probability α. In the second stage, attentive consumers face a standard discrete choice problem, while inattentive consumers stay in the same plan that they had in the last period. This implies that the observed share of plan j depends on its own share from the previous period as follows:

$$\hat{s}_{j,t}(p, s_{j,t-1}) = \alpha M Pr_{j,t}(p) + (1-\alpha) s_{j,t-1}, \tag{7}$$

where $\hat{s}$ is the observed share, p is the vector of plan premiums, M is market size, and $Pr_{j,t}(p)$ is the usual logit probability. In the first year of the program, this model reduces to the usual logit model, or equivalently, our discrete choice model with vintage set to zero. In year two, the observed share is a convolution of the current choice share and the set of inattentive consumers who did not make a choice. Irrespective of whether $p = 0$, where no one pays attention, or $p = 1$, where everyone is perfectly attentive, the plan accumulates consumers as time goes on and the relative share of the plan remains fixed as the rest of the world stays constant. The distinguishing feature of this model, however, is that it predicts that the firm can start raising premiums after the first year without losing as much market share as it would have in a perfectly attentive world. To see this, the derivative of Equation 7 with respect to its premium only has the current set of active choosers in it:

$$\frac{\partial \hat{s}_{j,t}(p, s_{j,t-1})}{\partial p_{j,t}} = \alpha M \frac{\partial Pr_{j,t}(p)}{\partial p_{j,t}}, \tag{8}$$

while profits are a function of the total share, of which fraction $(1-\alpha)$ are unresponsive to price changes. The key point is that as α declines, the firm can increasingly raise premiums and retain the same market share.

The mapping from this model to our model with a vintage variable is direct: as the market evolves, the share of active choosers effectively shrinks as an increasing percentage of consumers have been in the market for longer than one period. In the simplest case, assuming that no one exits the market and all pre-existing consumers are completely inattentive, $\alpha(T) = 1/(T-1)$, where T is the number of periods the market has been active. Our vintage variable proxies directly for this effect, as one can rewrite Equation 8 as:

$$\frac{\partial \hat{s}_{j,t}(p, s_{j,t-1})}{\partial p_{j,t}} = \alpha(T) M \frac{\partial Pr_{j,t}(p)}{\partial p_{j,t}} = M \frac{\partial Pr_{j,t}(p, \beta(T))}{\partial p_{j,t}}, \tag{9}$$

where $\beta(T) = F(\alpha(T))$ is a positive, monotonic transformation of $\alpha(T)$. This mapping can be generalized to allow for where $\alpha > 0$ for pre-existing consumers or where α is a function of the premium change (a la Heiss et al., 2016 and Ho et al., 2015). As such, one can view our reduced form model of demand with a vintage variable as arising from a structural two-stage model of inattention and choice.

A complete characterization of the influence of inattention and switching costs on demand

and pricing would require an equilibrium model as in Klemperer (1995) or Dubé et al. (2009). We note that this literature has conflicting predictions about the sign of pricing effects in response to switching costs: Klemperer (1995) concludes that prices are likely to be higher in equilibrium, while Dubé et al. (2009) demonstrate that prices can be lower in equilibrium. Indeed, in the analysis of these issues within the Medicare Part D setting, Ho et al. (2015) and Wu (2016) come to opposite conclusions about which pricing strategies have dominated the market in response to consumer inertia. In either environment, the fact that we are not modelling a dynamic equilibrium may lead to a bias in marginal cost estimates. In the setting of Medicare Advantage plans, Miller (2014) argues that in insurance markets that are characterized by inertial demand, the marginal cost estimates from a static Bertrand model may be around 20 percent higher or lower than the "true" dynamic values. Recognizing this concern in our setting, we have re-estimated our key counterfactual results for a 20 percent radius around our marginal cost estimates. Our qualitative conclusions are not sensitive to this specification check.

D Profit Function

This section provides a detailed description of how we arrive at the profit function in Equation 3 of the paper. We start with a description of the flow of payments in Part D and set up a general profit function that can incorporate a variety of regulatory interventions in this market. We then discuss our strategy of arriving at an empirically tractable version of the supply-side model.

Firms receive revenues across a variety of channels. For each individual that plan j enrolls, the insurer collects an enrollee premium, p_j. The premium does not vary across consumers and is determined as follows. CMS takes a (lagged) enrollment-weighted average of all bids submitted by all PDP and MA-PD plans across the country. It then declares a pre-specified share of this average for the given year (for example, 36 percent in year 2010) as the base consumer premium. The actual premium is then equal to the base premium plus the difference between the plan's bid and the average national bid.

The consumer premium is augmented with an individual-plan-specific subsidy, z_{ij}, from the government. This subsidy is equal to the bid multiplied by a measure of the enrollee's ex-ante health risk - r_i) - net of consumer premium. For an average-risk beneficiary with $r_i = 1$, the sum of the premium and government subsidy is equal to the bid that the firm submitted for that plan. For an individual with a risk score above or below the average, the insurer collects $r_i * b_j$ - out of this amount, p_j is paid by the consumer, and $z_{ij} = r_i * b_j - p_j$ is paid as the government subsidy. Since consumer premium depends on the average bid among all PDP and MA-PD plans, $\bar{b}$, we can write the subsidy as a function of the bid, the average bid, and individual-specific health risk: $z_{ij}(b_j, \bar{b}, r_i)$. The individual-level risk adjustment in the subsidy is intended to make all consumers look equally profitable to firms in order to reduce incentives for risk-based selection.

On the cost side, the *ex post* costs of a plan differ for each enrollee and depend on individual drug expenditures. Some of these costs are mitigated by the government through catastrophic reinsurance provisions, according to which the government directly pays about 80 percent of individual's drug spending for particularly high spenders. Throughout the empirical results we will refer to these reinsurance provisions as reinsurance subsidies. For an individual with a given total annual drug expenditure amount, the costs of the plan will also depend on the cost-sharing characteristics of the plan, denoted by ϕ_j. These include characteristics such as the deductible level, co-pays and co-insurance, as well as coverage in the donut hole if any. We let individual-level *ex post* costs be the function of these cost-sharing characteristics of a plan as well as the individual's measure of health risk, r_i; that is, we let the cost be $c_{ij}(r_i, \phi_j)$.

The final piece of a plan's *ex post* profit are risk corridor transfers between insurers and the federal government. These transfers happen at the end of the year, and restrict the downside (but also upside) risk of enrolling extremely costly individuals for the insurers. Medicare Part D Manual provides more details. As CMS describes in Chapter 9 of Prescription Drug Benefit Manual, risk corridors are: "Specified risk percentages above and below the target amount. For each year, CMS establishes a risk corridor for each Part D plan. Risk corridors will serve to decrease the exposure of plans where allowed costs exceed plan payments for the basic Part D benefit." (See 42 C.F.R, 423.336(a)(2).) We denote the function which adjusts a plan's *ex post* profit with Γ.

The *ex post* profit for plan j as a function of its bid b_j is then:

$$\pi_j(b_j; b_{-j}) = \Gamma\left[\sum_{i\in j}\left(p_j(\bar{b}, b_j) + z_{ij}(b_j, \bar{b}, r_i) - c_{ij}(r_i, \phi_j)\right)\right], \tag{10}$$

where the summation is taken over all individuals enrolling in the plan.

As the sum of the premium and the subsidy is by construction equal to the risk-adjusted bid submitted by insurer to Medicare, $p_j(\bar{b}, b_j) + z_{ij}(b_j, \bar{b}, r_i) = r_i * b_j$, we can re-write the ex ante expected profit of plan j for all consumers with risk level r as:

$$\pi_j^r(b_j; b_{-j}) = M_r s_{rj}(b)(rb_j - c_{rj}), \tag{11}$$

where $s_j^r(b)$ is the market share of plan j among consumers of risk r and M is the market size. We emphasize that s_j^r encapsulates all of the regulatory details involved in turning bids, b, into plan-specific market shares. To operationalize the analysis, we discretize the risk type space into five risk types among regular enrollees and LIS enrollees as a separate risk type. Let t index risk types of regular enrollees. Let θ_t denote the average risk score among type t enrollees. We can then re-write the risk-type level profit function as:

$$\pi_j^t(b_j; b_{-j}) = M_t s_{jt}(b)(\theta_t b_j - c_{jt}), \tag{12}$$

We now expand this expression to allow for multi-plan insurance organizations that offer plans to all risk types on the market, including the LIS consumers. The structure of profit from LIS enrollment is specified as entirely symmetric to the regular enrollees. We denote quantities related to regular enrollees of risk type t with superscript R, and quantities related to the LIS part of the market with superscript LIS. The profit function for insurer f offering a portfolio of $j \in J_f$ plans across all consumer types becomes:

$$\pi_f(b) = \sum_{j \in J_f} \left(\sum_{t=1}^{5} \left[M_t^R s_{jt}^R(b)(\theta_t^R b_j - c_{jt}^R) \right] + M^{LIS} s_j^{LIS}(b)(\theta^{LIS} b_j - c_j^{LIS}) \right) \tag{13}$$

Firms maximize profits by choosing bid b for each insurance plan j they offer.

Equation 13 is more complex than a standard profit function in a differentiated products market due to how the share equation $s_{jt}(b)$ is constructed. For regular enrollees, the share depends on the plan's premium, p_j^R, which is not set directly by insurers, but rather depends on the bids of other insurers in a non-linear fashion:

$$p_j^R = \max\left\{0, b_j - \bar{b} + \zeta \bar{b}\right\}, \tag{14}$$

where $\bar{b}$ is the enrollment-weighted average bid of all plans in the entire US and ζ is the share of the average bid allocated to baseline consumer premiums. The adjustment ζ is set every year by CMS and is governed by fiscal considerations and the Part D statutes; in 2010, this number was 0.36. The share equation for the low-income segment of the market is substantially more complex. It can be thought about as a piece-wise function with two components: random assignment of low-income enrollees by CMS for those plans that are eligible for random assignment, and enrollment choices by LIS consumer that make active choices of plans ("LIS choosers"). While LIS choosers are easily modeled in the standard discrete choice demand system, the eligibility requirement for random assignment introduces a discontinuity into the share function. Only plans below the average premium are eligible for random assignment, so for some choices of b_j, the share function for that portion of the market discontinuously jumps to zero.[28]

We make two assumptions to arrive at a tractable first order condition. First, we assume that the firm ignores the effect of its bidding behavior on the average bid, $\bar{b}$; this seems reasonable in light of the over 1,500 PDP plans that, along with the MA-PD plans, determine the average bid. Second, we assume that only insurers that are not competing for the randomly assigned LIS beneficiaries can be characterized as playing a Nash-Bertrand game. For these plans, the first-order condition is:

$$\frac{\partial \pi_f}{\partial b_j} = \sum_{t=1}^{5} \left[\theta_t M_t^R s_{jt}^R(b) + (\theta_t b_j - c_{jt}^R) M_t^R \frac{\partial s_{jt}^R(b)}{\partial b_j} + \sum_{k \neq j \in J_f} (b_k - c_{kt}^R) M_t^R \frac{\partial s_{kt}^R(b)}{\partial b_j} \right]$$
$$+ \theta^{LIS} M^{LIS} s_j^{LIS}(b) + (\theta^{LIS} b_j - c_j^{LIS}) M^{LIS} \frac{\partial s_j^{LIS}(b)}{\partial b_j} + \sum_{k \neq j \in J_f} (\theta^{LIS} b_k - c_k^{LIS}) M^{LIS} \frac{\partial s_k^{LIS}(b)}{\partial b_j}.$$

This expression differs from the more familiar first-order condition in the differentiated product literature in that the market size now plays an important role for the firm's decision-making. The market size affects the relative effects on profit from enrolling beneficiaries of different risk types. The incentives are further complicated by the fact that we allow for adverse or advantageous selection and endogenous marginal costs. The profitability for isnruers varies across risk types, since we do not impose that additional revenues from CMS for higher risk enrollees through the risk adjustment program offset the cost differences. To close the model, we assume that the expected marginal costs for risk types 2 to 5 and LIS are a first order polynomial with an intercept equal to zero relative to risk type 1. We denote the slope parameters with κ_t and κ_{LIS}. Collecting terms in vector notation, we can re-write the first order condition for an insurer offering plans not distorted by LIS incentives as follows:

[28] Decarolis (2015) discusses the piece-wise structure of the share function and the incentives generated by the LIS random assignment mechanism in much greater detail.

$$\sum_{t=1}^{T}\left[\theta_t M_t^R s_t^R - \Omega_t^R(\theta_t b - \kappa_t c)\right] + \theta^{LIS} M^{LIS} s^{LIS} - \Omega^{LIS}(\theta^{LIS} b - \kappa^{LIS} c) = 0. \qquad (15)$$

where

$$\Omega_{kjt}^R = \begin{cases} -M_t^R \frac{\partial s_{tj}^R(b)}{\partial b_k} & \text{if } \{j,k\} \in J_f, \\ 0 & \text{else,} \end{cases} \qquad (16)$$

and

$$\Omega_{kj}^{LIS} = \begin{cases} -M^{LIS} \frac{\partial s_j^{LIS}(b)}{\partial b_k} & \text{if } \{j,k\} \in J_f, \\ 0 & \text{else.} \end{cases} \qquad (17)$$

We use these first-order conditions to compute (by inversion for non-distorted plans and using a hedonic projection for distorted plans) for marginal costs for each plan j. These cost estimates are in turn used as inputs for computing the counterfactual equilibria.

E Welfare Function

For regular enrollees, total welfare in the Medicare Part D PDP market is comprised of three pieces: consumer surplus (CS), insurer profits (Π), and government spending (G):

$$W = CS + \Pi - \lambda G, \qquad (18)$$

where λ is the social cost of raising revenues to cover government expenditures, G. All three pieces of the welfare function are calculated relative to the outside option. For consumer surplus the normalization to the outside option (buying an MA-PD plan or not purchasing Part D insurance) follows directly from the utility model. For producer surplus, the insurer pricing decision implicitly takes into account the opportunity cost of serving the outside option. In other words, the marginal cost as recovered from the inversion of the first-order conditions incorporates the opportunity costs of potentially serving each consumer in the MA-PD market or outside of the Part D program. Consequently, the profit function is defined relative to profits that could have been made in the MA-PD program or elsewhere. Finally, since the government subsidizes both the PDP and MA-PD parts of the market, we consider government spending on PDP net of what it would have spent on the same individual elsewhere. We conservatively assume that the outside option for the government is subsidizing the same consumers in the MA-PD market. This assumption excludes the possibility that some individuals could leave subsidized insurance altogether.

Following Williams (1977) and Small and Rosen (1981), surplus for consumer i with marginal utilities ω_i from plan characteristics, including the premium, takes the following form:

$$CS(\omega_i) = \frac{1}{\alpha_i}\left[\gamma + \ln\left[1 + \sum_{j=1}^{J} \exp(v_{ij}(\omega_i))\right]\right], \tag{19}$$

where γ is Euler's constant, and v_{ij} is the deterministic component of utility for person i from plan j (utility net of the idiosyncratic shock).[29] We integrate out over the unobserved taste heterogeneity to obtain consumer surplus for each consumer risk group t:

$$CS_t = \int CS(\omega_t) dF(\omega_t). \tag{20}$$

The second piece of the welfare calculation is producer surplus that we compute using Equation 3.

The last piece of the welfare calculation is government spending. In our welfare calculations, we weigh the government spending with the shadow cost of public funds, commonly estimated at $\lambda = 1.3$.[30] We compute both the nominal government spending in the PDP program, as well as how much extra spending the PDP part of the Part D program (G^{PDP}) generates relative to the outside option of subsidizing the beneficiaries in Medicare Advantage prescription drug plans (G^{MAPD}). We allow government spending to vary by consumer risk type, which reflects differential risk adjustment and re-insurance payments across consumers of different health.

Adding the three parts of the welfare function back together, we have the following measure of total surplus for regular consumers (an analogous expression applies to LIS con-

[29] Euler's constant is the mean value of the Type I Extreme Value idiosyncratic shock under the standard normalizations in the logit model, and is approximately equal to 0.577.

[30] See, for example, Hausman and Poterba (1987).

sumers):

$$W = \overbrace{\sum_{t=1}^{t=5} M_t \int \frac{1}{\alpha_t} \left(\gamma + \ln \left[1 + \sum_{j=1}^{J} \exp(v_{jt}(\omega_t)) \right] \right) dF(\omega_t)}^{\text{Consumer Surplus (CS)}} + \overbrace{\sum_{j \in J} \left(\sum_{t=1}^{5} \left[M_t s_{jt}(b)(\theta_t b_j - \kappa_t c_j) \right] \right)}^{\text{Producer Profit } (\Pi)} - \overbrace{\lambda \left(\sum_{t=1}^{t=5} \left[\sum_{j=1}^{J} (G_{jt}^{PDP} - G_t^{MAPD}) s_{jt}(p) M_t \right] \right)}^{\text{Social Cost of Government Spending (G)}}. \quad (21)$$

This welfare function describes the surplus for the private market, where firms administer insurance contracts. To arrive at the social planner's objective function, we include "profits" in the government spending computation, as under a social planner, prices are set by fiat and there are no private profits. Social planner's problem is then to maximize the following welfare function:

$$W^{SP}(p) = \overbrace{\sum_{t=1}^{t=5} M_t \int \frac{1}{\alpha_t} \left(\gamma + \ln \left[1 + \sum_{j=1}^{J} \exp(v_{jt}(\omega_t)) \right] \right) dF(\omega_t)}^{\text{Consumer Surplus (CS)}} + \lambda \left[\overbrace{\sum_{j \in J} \left(\sum_{t=1}^{5} \left[M_t^R s_{jt}^R(p)(\theta_t^R p_j - \kappa_t^R c_j) \right] \right)}^{\text{Producer Profit } (\Pi)} - \overbrace{\left(\sum_{t=1}^{t=5} \left[\sum_{j=1}^{J} (G_{jt}^{PDP} - G_t^{MAPD}) s_{jt}(p) M_t \right] \right)}^{\text{Social Cost of Government Spending (G)}} \right] \quad (22)$$

The vector of prices that maximizes this version of the welfare function is the social planner's solution. Note that we use prices in the social planner's case, as the distinction between insurer bids and consumer premiums is not meaningful in this case. The optimal price vector is defined by the set of first-order conditions obtained by differentiating $W^{SP}(p)$ with respect to prices. The derivative of consumer surplus with respect to p_j has a conveniently simple form:

$$\frac{\partial CS(p)}{\partial p_j} = \sum_{t=1}^{t=5} \left[\int M_t \frac{1}{\alpha_t} \left[\frac{-\alpha_t \exp(v_{jt}(\omega_t))}{1+\sum_{k=1}^{J} \exp(v_{kt}(\omega_t))} \right] dF(\omega_t) \right] = -\sum_{t=1}^{t=5} M_t s_{jt}(p) \tag{23}$$

The derivative of product market profit with respect to p_j is:

$$\frac{\partial \Pi(p)}{\partial p_j} = \sum_{t=1}^{t=5} \left[\lambda M_t s_{jt}(p)\theta_t + \lambda M_t \sum_k (\theta_t p_k - \kappa_t c_k) \frac{\partial s_{kt}(p)}{\partial p_j} \right] \tag{24}$$

The derivative of government spending with respect to p_j is:

$$\frac{\partial GS(p)}{\partial p_j} = \sum_{t=1}^{t=5} \left(-\lambda \left[\sum_k (G_{kt}^{PDP} - G_t^{MAPD}) \frac{\partial s_{kt}(p)}{\partial p_j} M_t \right] \right) = \tag{25}$$

$$\sum_{t=1}^{t=5} \left(-\lambda \left[\sum_k \Delta G_{kt} \frac{\partial s_{kt}(p)}{\partial p_j} M_t \right] \right) \tag{26}$$

Summing these terms, we obtain:

$$\frac{\partial W^{SP}(p)}{\partial p_j} = \sum_{t=1}^{t=5} M_t \left((\lambda\theta_t - 1) s_{jt}(p) + \lambda \sum_k (\theta_t p_k - \kappa_t c_k - \Delta G_{kt}) \frac{\partial s_{kt}(p)}{\partial p_j} \right) \tag{27}$$

A decrease in consumer surplus in response to an increased price ($-s_{jt}(p)$) is offset, up to the cost of transferring public funds, by an increase in profit in the product market ($\lambda\theta_t s_{jt}(p)$). The degree of offset varies by consumer risk type.

The first-order conditions can be simplified using vector notation:

$$\sum_{t=1}^{t=5} M_t \left[(\lambda\theta_t - 1) s_t(p) + \lambda \Omega_t(p)(\theta_t p - \kappa_t c - \Delta G_t) \right] = 0 \tag{28}$$

where $\Omega_t(p)$ is a matrix of partial derivatives such that the element in the i-th row and j-th column is:

$$\Omega_{ijt}(p) = \frac{\partial s_j t(p)}{\partial p_i}. \tag{29}$$

It follows that optimal prices for the social planner's case are given by the following:

$$p^{SocialPlanner} = \left(\sum_{t=1}^{t=5} \lambda M_t \Omega_t(p) \theta_t \right)^{-1} \left(\sum_{t=1}^{t=5} M_t \left[(1 - \lambda\theta_t) s_t(p) + \lambda \Omega_t(p)(\kappa_t c + \Delta G_t) \right] \right) \tag{30}$$

Price is set to balance the inside and outside option enrollment for each consumer risk type. In particular, the size of each risk type market as well as the difference in risk adjustment (θ_t) versus cost factors across risk types (κ_t) play a central role in determining the social planner's allocation.

Cost of public funds A key parameter in our calculations in the social cost of government funds, which is set to 1.3 in our baseline analysis. In Table E1, we report the estimates of total surplus (accounting for the opportunity cost of government funds) for each counterfactual for $\lambda \in \{1, 1.7, 2\}$. Increasing the cost of public funds has the general effect of decreasing overall welfare in most cases. The optimal voucher shifts up by \$100 to \$900 at $\lambda = 1$ in the fixed outside option case, and remains the same at \$1,200 in the adjusted outside option case. At $\lambda = 2$, the optimal voucher shifts down by \$100 to \$700 in the fixed outside option case. In the case of adjusted outside option, making government payments more socially costly leads to a knife-edge case, where the optimal strategy is not to subsidize the market at all, since there is not enough willingness to pay for the PDP program, so that any subsidy is costlier than the utility loss from not subsidizing. The latter remains the case as long as the public cost of government funds is above 1.5 per 1 dollar of government spending.

F Outside Option Adjustment

We proceed in several steps to calculate the adjustments in the value of MA-PD coverage for consumers that allow for changes in the outside option in counterfactual equilibria. There are slight differences across different counterfactuals, so we describe them separately.

We start with counterfactuals that compute PDP subsidies via bid-averaging, similar to how the subsidies are calculated under the observed allocation. To compute MA-PD adjustments for these counterfactuals, we first compute the counterfactual premium subsidy for PDP plans. This subsidy represents the difference between PDP bids and premiums. Next, we turn to the MA-PD market. In the data, we observe MA-PD premiums, but we do not observe bids. We impute MA-PD bids using subsidization formulas that link bids and premiums. One complication is that MA-PD plans can apply additional subsidies to their bids, by pulling in resources from the medical part of the Medicare Advantage program. MA plans can use their MA subsidy to "buy down" MA-PD premiums. The data on the degree of "buy down" by plan is not publicly available. To circumvent this data gap, we turn to the MA literature, specifically Kluender and Mast (2016), who report that the average MA-PD

Table E1: Sensitivity to the Cost of Public Funds Parameter

	Counterfactuals	λ=1	λ=1.3	λ=1.7	λ=2
	(1)	(2)	(3)	(4)	(5)
Panel A	**Fixed outside option**				
(1)	Observed allocation, $M	3,306	3,441	3,620	3,755
(2)	No LIS link, $M	3,687	3,671	3,650	3,634
(3)	No LIS , no MA-PD link, $M	3,775	3,638	3,455	3,317
(4)	Independent plans, $M	3,724	3,570	3,366	3,212
(5)	Monopoly ownership, $M	3,708	3,510	3,247	3,049
(6)	$0 voucher, $M	1,121	1,165	1,222	1,266
(7)	Optimal voucher, $M	3,891	3,674	3,680	3,771
	Optimal voucher level, $	*900*	*800*	*700*	*700*
(8)	$ 1,500 voucher, $M	1,643	-1,868	-6,549	-10,060
Panel B	**Adjusted outside option**				
(9)	No LIS link, $M	4,237	4,194	4,138	4,096
(10)	No LIS , no MA-PD link, $M	4,529	4,356	4,127	3,954
(11)	Independent plans, $M	4,515	4,337	4,100	3,922
(12)	Monopoly ownership, $M	4,316	3,922	3,396	3,002
(13)	$0 voucher, $M	-889	2,332	6,628	9,850
(14)	Optimal voucher, $M	7,266	5,975	6,628	9,850
	Optimal voucher level, $	*1,200*	*1,200*	*0*	*0*
(15)	$ 1,500 voucher, $M	6,237	4,058	1,153	-1,025
Panel C	**Non-market mechanisms**				
(16)	Public option with subsidy, $M	3,178	3,229	3,298	3,350
(17)	Public option without subsidy, $M	1,165	1,223	1,301	1,360

Table reports the level of welfare (accounting for the opportunity cost of government funds) under counterfactual allocations with a fixed outside option, endogenously adjusted outside option, and non-market mechansims for different levels of the cost of public fund parameter (λ). Welfare is computed exactly as in baseline counterfactuals, only varying the cost of public funds parameter. The baseline results assume that the cost of public funds (λ) is equal to 1.3 - these results are replicated in column (2).

buy down is \$3.90 a month. We apply this adjustment to all MA-PD premiums, so that our imputed monthly MA-PD bid is equal to basic MA-PD premium observed in the data plus \$3.90 adjustment plus \$88.33 (which is the national average bid that was released by CMS in 2010) and minus \$31.94 (which was the base beneficiary premium released by CMS in 2010). With these imputed MA-PD bids in hand, we apply the counterfactual PDP subsidy to each bid. In addition, we assume that MA-PD plans would apply the same "buy-down" on top of any counterfactual Part D subsidy. Subtracting the counterfactual subsidy as well as the \$3.90 "buy-down" from each MA-PD bid gives us counterfactual MA-PD premiums for each MA-PD plan. In many cases the counterfactual PDP subsidy together with the "buy-down" are higher than the MA-PD bid. In these cases we impose a zero lower bound on MA-PD premiums — this is in line with the observed allocation in which many MA-PD plans have zero premiums.

In the next step, for each MA-PD plan we compute the difference between the observed and counterfactual premium. We take the average of these differences across all plans, which gives us one number that summarizes the average change in MA-PD premiums under the application of the counterfactual PDP subsidy. We proceed analogously in counterfactuals where subsidies are set as flat vouchers. The key difference is in how we compute the counterfactual Part D subsidy. In the case of vouchers, this is simple, since we just apply the same voucher level to imputed MA-PD bids. We then compute the resulting difference in average MA-PD premiums for consumers, which is the outside option adjustment recorded in row (26) of Table 5. We use this adjustment as our measure of the change in the value of the outside option. In practice, the increase (decrease) in the attractiveness of the outside option is computationally implemented as a symmetric decrease (increase) in the value of the inside option (the constant for the inside option choice is adjusted).

G Subsidy Design in LIS Market

Our baseline set of counterfactuals removes the LIS enrollees from the market. As discussed in the paper, this was done for both pedagogical and technical reasons. Naturally, however, determining optimal ways to provide subsidies to both markets is an important question to consider. One must confront two questions when thinking about joint subsidy setting for LIS and regular beneficiaries. First, what is the role of the LIS random assignment mechanism in determining market outcomes? Second, in the range of counterfactual mechanisms, which one would maximize total surplus across both the regular and LIS markets?

On the first dimension, LIS eligibility threshold plays an important role in disciplining prices in this market. If one does not want LIS enrollees to face premiums or cost-sharing, as is the current situation, then some kind of additional brake on premiums is needed when combining the LIS and regular markets. Without any additional brake, the LIS market by itself effectively functions as a market with 100 percent proportional subsidy, which would lead to unbounded increases in premiums. In the current system, requiring that insurers submit only one bid for both markets, combined with the LIS eligibility threshold, serves as a brake on premiums. It introduces an elasticity of demand through the discontinuity of not being assigned LIS enrollees if the plan's premium is too high, even though the LIS enrollees themselves have zero elasticity of demand for plans that are eligible for LIS random assignment. Earlier work in Decarolis (2015) explores these issues in detail.

To shed light on the counterfactual mechanisms that could maximize joint surplus of LIS and regular markets, we reintroduce LIS enrollees into the regular market in a supplementary set of counterfactuals. We focus on what we consider the most policy-relevant environment: fixed vouchers that can differ across the two market segments. We keep the bid-tying aspect, imposing that insurers set one bid for both markets, but we allow the government to set different voucher subsidies for LIS and regular consumers. We ran the analysis on a matrix of LIS and regular vouchers that range from \$0 to \$1000 in \$100 increments. We find that under the double voucher counterfactuals, the optimal voucher for regular consumers is \$800, which is the same as the optimal voucher when just considering the regular market; for the LIS consumers, lower vouchers result in general lead to higher surplus due to the very high government cost of subsidizing this market. The differences across total surplus along the LIS voucher dimension are relatively small as long as the LIS voucher is lower than \$800. It follows that setting the LIS voucher to be equal to the optimal voucher for regulars does not lead to a significant decrease in total surplus.

H Algorithm for Solving Counterfactual Equilibria

Several of our counterfactuals involve resolving equilibrium bids when the subsidy is an endogenous function of the average bid. We solve these types of equilibria in a nested fixed point algorithm. In the outer step, we first pose an average bid. We model the firms as taking this average bid as fixed. This is not an unreasonable assumption, as the marginal effect of any one firm's bids on the average bid is going to be very small, as the bid is a function of 1500 plans. Taking this average bid, $\bar{b}$, as fixed, we then solve for the vector of first-order conditions. After finding this vector of bids across all markets, we then compute the enrollment-weighted average bid. We grid search over a range of average bids until we find an average bid that correctly reflects the equilibrium average bid.

We use the sparse grids method described in Heiss and Winschel (2008) for the evaluation of all integrals. Sparse grids are efficient and accurate multidimensional quadrature methods with excellent performance. Estimation of the BLP specifications for each risk type of regular enrollees was standard with the exception of imposing the lognormality of the price coefficient. All programs and instructions on obtaining data are publicly available in the technical supplement.