NORTHWESTERN UNIVERSITY

Policy Design in Publicly Funded, Privately Provided Markets

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Paul HS Kim

EVANSTON, ILLINOIS

June 2023

ABSTRACT

Policy Design in Publicly Funded, Privately Provided Markets

Paul HS Kim

Increasingly, governments contract with private firms to provide publicly funded or subsidized goods and services, ranging from defense contracts, social insurance programs to small business loans. In such publicly funded, privately provided markets, governments set specific rules and policies to allow efficient provision or allocation of goods and services. Given the large fiscal spending in these markets, understanding the design of these policies are important. This dissertation examines policy design in the context of three such markets.

Chapter 1 studies insurers' strategic responses to regulations in government health insurance market. Publicly-funded and privately-provided health insurance programs in the U.S. are regulated to ensure a competitive marketplace. However, private firms can strategically respond to government rules and regulations that may lead to market outcomes away from the government's intended goals. I

study insurers' strategic responses to the interaction of two regulations in Medicare Part D: profit margin regulation and risk corridors (a risk sharing policy). The government utilizes insurers' self-reported cost estimates to implement both regulations. This creates a trade-off for firms; they can lower their cost report to reduce risk exposure or increase their cost report to charge higher prices. To quantify the effects of insurers' strategic responses, I estimate a structural model in which insurers are risk averse and can strategically misreport their costs. I find that insurers over-report their cost estimates by 7.5%, leading to 10% higher prices for consumers; however, by over-reporting their cost estimates, insurers are expected to pay back the government 2% of premium revenue in risk corridor payments. Thus, risk corridors limit ex-post profits more than serving as a risk sharing mechanism. I propose an alternative linear risk sharing rule to replace the existing risk corridors, which increases total surplus by 11% while maintaining insurers' risk exposure.

Chapter 2, which is joint work with Anran Li, studies the efficiency of reinsurance subsidy compared to consumer subsidy in the ACA individual health insurance market. First, we develop a model of risk averse insurers that face financial frictions in a market with adverse selection. Using the model, we show that reinsurance has two effects: i) providing cost-subsidy that reduces insurers' expected cost ii) providing insurance for insurers which reduces risk charge of risk averse insurers. As a result, the pass-through of reinsurance can be larger than one, even in an imperfectly competitive market. We further establish that in a market with adverse selection, it is unclear whether reinsurance or consumer subsidy will be more efficient. Using state-level reinsurance policies, we show empirical evidence of both financial frictions and adverse selection in the market. Many health insurers purchase private reinsurance policies despite high mark-ups. In response to public provision of reinsurance: i) premiums decrease more for insurers that buy private reinsurance ii) premiums decrease more for higher actuarial value plans. Furthermore, insurers are less likely to purchase private reinsurance, reducing insurers' indirect cost of financial frictions.

Chapter 3, which is joint work with David Stillerman, studies the design of the Paycheck Protection Program (PPP), a loan-forgiveness scheme that is implemented through private lenders and assists small businesses in keeping their employees on payroll during the COVID-19 pandemic. We develop a model of PPP lending to capture the government's tradeoff between inducing bank participation and targeting funds for use on payroll. Using the model, we establish that both increasing subsidies and relaxing forgiveness standards are effective in expanding credit access to borrowers seeking smaller loans. However, their efficacy in targeting (i.e., providing funds to businesses who will use them on payroll) depends on the correlation between loan amounts and borrowers' return to payroll. We test the implications of the model using policy variation from the PPP Flexibility Act, legislation that relaxed forgiveness standards. Consistent with the predictions of the model, the average loan amount falls by between 6 and 7% in the period following the policy change. Furthermore, marginal borrowers are more likely than inframarginal borrowers to use funds for payroll, so making forgiveness more accessible increases the average share of funds used for those purposes.

Acknowledgements

I would first like to thank the members of my committee for their invaluable support and guidance towards completing this dissertation. Thank you Robert Porter for your soft-spoken but sharp comments and steadfast advice on my work. I am especially blown away by the breadth of knowledge you've accrued over the years. Thank you David Dranove, for fostering my interest in health economics during my second year, and your ability to provide a big picture perspective on my research. Thank you Amanda Starc for all your help including but not limited to getting me access to important datasets, without which my main chapter would not have been possible and helping me navigate the complex institutional details of the Medicare market. Thank you Gaston Illanes for drawing me to the exciting field of empirical Industrial Organization in my early years of graduate school, both in your teaching and through my time as a RA. Aside from sharing your impressive technical expertise without which my work would not be complete, you've been a great mentor throughout the years. I would also like to thank Vivek Bhattacharya, who despite not being on my committee, gave up just as much time to meet with me, walk through every detail of my paper, and offer incredible advice and guidance. I'm thankful for the entire economics department, including the numerous staff & faculty and more importantly all my friends & colleagues that have helped me in so many ways and shaped my experience at Northwestern.

Second, I would like to thank the community of friends at the Graduation Christian Fellowship as well as the local church community of Evanston Bible Fellowship that has been my family away from home this past six years. I have shared life with so many of these brothers and sisters, many of whom were also going through the arduous grad school life. They provided constant encouragement that allowed me to continue my work even when things looked bleak, especially during the COVID pandemic. Through them, I've learned and gained so many things outside of research that were just as if not more important than the work I was pursuing. Without the support of this community, this dissertation wouldn't be where it is today.

Finally, I would like to thank my family for all their support throughout the years that have allowed me to be where I am today. I thank my parents and my grandparents for their sacrificial love & support, both financially and through giving up so much of their time. Last but not least, none of this would have been possible without the love and support of my wife, Lidia Kuo. But most importantly, I lift up everything to my Lord and Savior, God who ultimately put me where I am today and without his help none of this would have been possible.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Risk Corridors in Medicare Part D: Financial Risk Sharing or Profit Limiting Mechanism?[1]

1.1. Introduction

Increasingly, public health insurance programs in the US are being delivered through private insurance companies (e.g. Medicare Advantage, Medicaid, ACA Exchanges, etc.). Government spending on these programs is enormous; the US government contributes $\$0.6$ trillion in annual health insurance subsidies towards health coverage that is delivered by private insurance companies (CBO, 2020). The success of such privately-provided health insurance programs is predicated on successful competition among private firms leading to efficient provision of goods at low prices. In practice, to ensure a competitive marketplace, the government heavily regulates private firms in these settings. Examples of regulations include pricing regulations, product design regulations, and risk sharing arrangements. However, if not carefully designed these regulations also introduce strategic incentives for firms that could distort their intended goal. Understanding and evaluating how the design of these regulations affect the behavior of strategic firms

[1] I would like to thank David Dranove, Rob Porter, Gaston Illanes, and Amanda Starc for their invaluable advice and support. I also thank Vivek Bhattacharya, Bill Rogerson, and Mar Reguant as well as seminar participants at the Northwestern IO Student Seminar for their helpful comments. I am grateful for financial support from the Northwestern Graduate School Research Grant.

is crucial to a successful publicly-funded, privately-provided market, especially when the program is administered by firms with market power.

In this paper, I study insurers' strategic responses to the interaction of regulations in Medicare Part D, a US federal program administered through private insurance companies, which provides prescription drug coverage to older adults. I focus on two regulations. The first is ex-ante profit margin regulation, which puts an upper bound on price that insurers can charge relative to self-reported expected cost. The second is risk corridors (RC), a risk sharing policy that ex-post reimburses (charges) insurers for any cost overruns (underruns) relative to insurers' self-reported expected cost. While the two regulations were designed with distinct purposes, both regulations rely on insurers' self-reported cost estimates. This gives insurers a strategic incentive to misreport their costs to increase their revenue. A few recent papers study insurers' strategic responses to a single policy or regulation in isolation (Decarolis, 2015; Geruso & Layton, 2020; Sacks *et al.* , 2021a). However, the health insurance markets are often laden with numerous regulations that may affect one another. I study how the interaction of two different regulations balances insurers' strategic incentives to distort the regulatory outcomes, albeit imperfectly.

The two regulations apply to widely-used policies that the government utilizes beyond Medicare Part D. The margin regulation limits market power by constraining insurers from earning excessive profits. This is commonly used in other health insurance markets like Medicare Advantage and Medicaid. More broadly, it is comparable to the rate-of-return regulations used to regulate monopolies in the utility industry. Risk corridors protect insurers from ex-post uncertainty in cost by sharing in expenses (savings) from any cost overruns (underruns). The government utilizes risk corridors to stabilize the market (e.g. ACA exchanges and Medicaid), especially in new markets where insurers may face increased uncertainty around costs of enrollees. At the outset of the COVID-19 pandemic, the federal government contemplated extending risk corridors beyond the government programs to include the entire US health insurance market.[2] Despite the widespread use by the policymakers, there is very little research on risk corridors in the economics literature.

To illustrate how the regulations affect the market with strategic insurers, I build a stylized model with asymmetric information in which risk averse insurers can strategically report their expected cost. I show that the two sets of regulations have opposing incentives. With just the ex-ante profit margin regulation, insurers will tend to overestimate their costs so that the price they charge will appear not too high relative to their reported cost estimates. This is closely related to the theoretical literature on regulation (Baron & Myerson, 1982; Baron & Besanko, 1984), which shows that firms have incentives to report higher costs when their revenue or price is linked to cost reports. With just risk corridors, insurers want to underestimate their costs in order to increase their chance of cost "overruns",

[2]The HEROES Act, a COVID-19 relief bill passed by the House of Representatives in May 15 2020, included measures to enact risk corridors to the broader US health insurance market. The bill proposed providing a one-sided risk corridor to the Medicare Advantage, individual and small/large group health insurance market. For more details see House of Representatives (2020)

thereby increasing their likelihood of receiving reimbursements from the government. The latter result is line with Sacks *et al.* (2021a) who show that risk corridors in the ACA marketplace create similar incentives for insurers, acting as an implicit subsidy.

However, when both policies are present the two may balance, and dampen the insurers' incentives to over/underestimate their costs. If insurers overestimate their costs, they can increase the upper bound on the price that they can charge, allowing them to set higher prices; overestimating their costs, however, will also increase their chance of cost "underruns", increasing their likelihood of making ex-post risk corridor payments to the government. So when insurers overestimate their costs, the risk corridor acts as an ex-post penalty function for the insurers. On the other hand, if insurers underestimate their costs, they can increase the likelihood of receiving ex-post risk corridor reimbursements from the government; underestimating their costs, however, will also constrain the maximum price that they can charge. Which incentive dominates is an empirical question, as is quantifying the magnitude of the distortion induced by this set of policies. This paper aims to fill this gap in our understanding of the impact of the policy.

Using data from insurers' financial statements, I present descriptive evidence in line with my model implications. I compare the ex-post profit margins of insurers' Part D businesses to the margins of insurers' commercial businesses that are used as a benchmark for the Part D margin regulation. I find that insurers are much more profitable in the Part D market compared to their commercial businesses. If insurers had estimated their costs correctly, their Part D margins should be similar or lower than the commercial business margins, due to the margin regulation. Instead, most insurers overestimated their costs, allowing them to charge higher prices. Insurer level risk corridor payments show similar outcomes. The distribution is heavily skewed towards positive payments to the government, meaning most insurers have overestimated their costs. I find that these overestimates are persistent across insurers, suggesting that the overestimates are due to strategic cost reporting rather than random uncertainty in cost is playing a role. These descriptive results suggest that the current design of risk corridors acts more as a profit-limiting mechanism than it is as a risk sharing mechanism.

To quantify the degree of insurers' strategic behavior and the impact on market outcomes, I build and estimate a structural model of demand and supply. On the demand side, I build on the discrete choice model of demand estimated in Decarolis *et al.* (2020a), allowing for substantial heterogeneity across consumer types. My model of supply departs sharply from existing models in two ways: i) endogenizing the strategic self-reporting of cost estimates by the insurers and ii) allowing insurers to behave as "risk averse". That is, insurers face a disutility from taking on greater risk. Modeling insurers as "risk averse" is crucial for understanding the role of risk corridors in reducing the risk that insurers face. In the standard model of risk neutral insurers, risk sharing has no meaningful effect on outcomes. However, in reality insurers seem to exhibit risk averse behavior. Insurers face financial/regulatory frictions (Koijen & Yogo, 2015b), and often purchase reinsurance policies to lower their exposure to risk. So the amount of risk assumed by insurers impacts their marginal cost, which in turn affects their pricing decisions.

The estimates highlight a few facts. First, I estimate that most insurers have overestimated their costs by around 8% on average. In line with the descriptive evidence, insurers would like to charge a relatively high markup in their Part D business compared to their commercial market. Insurers overestimate their costs to relax the margin regulation, and charge higher prices. At the same time, by overestimating their costs, insurers face an increasing risk of having to pay money to the government in ex-post risk corridor payments. Insurers are expected to pay back the government 2% of premium revenues in risk corridor payments. Second, I find that insurers are not that risk averse. The estimated risk aversion coefficients imply that insurers face an average risk charge of $\$17.5$ for enrolling an additional enrollee, which is around 2% of insurers' marginal cost or 15% of the average margin. Third, while the magnitude of risk aversion coefficients is small, I find that the coefficients are negatively correlated with insurers' RBC ratios. That is, I find that insurers are less risk averse when they are better capitalized or more financially solvent. This suggests that insurers' risk averse behavior is driven by financial/regulatory frictions that they face (Kim & Li, 2022).

With the estimates in hand, I look at equilibrium outcomes under different market designs to quantify the effect of insurers' strategic reporting under current regulations. If insurers had correctly reported their expected costs (truthful reporting) the average price would be 10% lower, while increasing the average risk level by a factor of four. The lower price translates to a 15% higher consumer surplus. In the absence of both risk corridors and ex-ante profit margin regulation, prices would be 5.2% higher, and the risk level would be five times higher. The higher price translates to a 10% lower consumer surplus. Of the 5.2% increase in prices, I find that only 0.3% is due to increased risk level and the remaining 4.9% is from the removal of profit margin regulation. Although the risk corridor significantly reduces the amount of risk that insurers face, insurers are almost risk neutral, so the amount of risk that insurers face doesn't seem to play a large role in the market outcomes. So the intended role of the risk corridor in sharing the risks that insurers face is not significant in the current market. On the other hand, the profit margin regulation is playing a large role in keeping insurers' prices low. The risk corridor plays an important role as an ex-post transfer mechanism that penalizes insurers for over-reporting their cost estimates. As a result, risk corridor payments help enforce the profit margin regulation.

While the current market outcome yields a higher consumer surplus than would be the case without any regulations, it is much lower than the case in which insurers truthfully report their cost estimates. Given that risk corridors are in practice being used to limit ex-post profits in the market, a natural question is whether there are alternative risk corridor designs that could raise the consumer surplus. To explore this, I alter the design of risk corridors to a simple linear risk sharing rule. I vary the degree of risk sharing from 0%, indicating a fixed-price conract or no regulation to 100%, indicating a cost-plus contract that fully reimburses (charges) the insurer for any cost overruns (underruns). I find that with linear risk sharing of 58%, the government can increase its total surplus and achieve even higher surplus levels than the truthful reporting, case while maintaining the same level of risk that insurers currently face.

Related Literature

This paper is related to several distinct groups of literature on the design of social health insurance programs, supply-side frictions of insurance firms, and the regulation of private firms.

First, this work adds on to the growing body of literature on strategic responses of private firms in the health insurance markets. A closely related paper, Sacks *et al.* (2021a) studies the temporary risk corridor program in the ACA markets. The authors find that the insurers had an incentive to lower their cost benchmarks in order to increase the chance of reimbursements from the government. This is in line with the findings in this paper: when there are only risk corridors, insurers have an incentive to lower their cost benchmark. Other papers like those of Geruso & Layton (2020); Brown *et al.* (2014b) study insurers' strategic behavior in response to the design of the risk adjustment program. They find evidence of insurers/providers upcoding patient diagnoses to increase the risk adjustment payments and/or screening selectively healthy patients conditional on their risk scores. Decarolis (2015) looks at how insurers can strategically game the low income subsidy design and documents evidence of such strategic behavior leading to increased premiums. While existing papers study an insurer's strategic response to a single regulation, this paper focuses on how the interaction of two different regulations can balance insurers' strategic incentives. In heavily regulated markets like the health insurance markets, different regulations may interact with one another, and as such it is important to study the interaction of policies and not just a single policy in isolation.

Second, this paper studies risk corridors, an ex-post risk sharing policy in the health insurance markets. While there are several papers on ex-ante risk sharing or risk adjustment policies (Brown *et al.* , 2014b; Einav *et al.* , 2016; Geruso *et al.* , 2019; Carey, 2017) that make ex-ante transfers based on enrollee's predicted health risk, there is little work studying ex-post risk sharing policies. Layton *et al.* (2016a) conducts a simulated study on how risk corridors and reinsurance policies affect the distribution of insurers' costs. Sacks *et al.* (2021a) studies how the removal of the risk corridor program in the ACA exchanges led to sharp premium increases. In this paper I study how risk corridors affect the market by directly modeling and estimating "risk averse" insurers and studying how risk corridors affect insurers' pricing decisions by changing insurers' risk exposure.

Third, by modeling insurers' behavior as "risk averse", this paper adds to the literature on the supply-side frictions of insurance firms. Recent work by Koijen & Yogo (2015b, 2016a, 2022a) documents financial/regulatory frictions that life insurance companies face and how such supply-side frictions may play a significant role in the pricing of insurance contracts. Health insurance companies are also subject to similar financial regulations, and so may face similar frictions for taking on risk; in reality, health insurers will behave as if they are risk averse. To the author's knowledge, this is one of the first papers to document and incorporate such supply-side frictions in modeling the health insurance companies.

A large body of theoretical literature has studied the regulation of private firms in the context of government procurement or monopoly regulation (Baron & Myerson, 1982; Baron & Besanko, 1984; Laffont & Tirole, 1986). There is also an empirical literature on this topic.[3] This paper adds to the empirical literature by studying an empirical analogue of Baron & Besanko (1987), in which the government seeks to regulate risk averse firms in the presence of asymmetric information.[4] The paper also contributes to the literature on price regulation in healthcare markets. Cicala *et al.* (2019) studies the MLR regulation introduced by the ACA and find that it decreased incentives for insurers to control costs, thereby raising the overall costs of care. Dubois & Lasio (2018) study the price regulation of pharmaceuticals in

[3] Brocas *et al.* (2006) study the cost of asymmetric information in the rate of return regulation of water utilities, studying the empirical analog of Brocas *et al.* (2006). Abito (2020) study emissions and rate of return regulation in electric utilities, in the context of Laffont & Tirole (1986).

[4] Unlike Baron & Besanko (1987), this paper does not model the moral hazard of cost side.

France and finds that the government's price-regulation resulted in a modicum of decreases in price.

Lastly, this paper contributes to the large body of literature on the Part D program. Most of the earlier literature on Part D focuses on the demand side, looking at individuals' plan choice behaviors with respect to the rationality of plan choice, consumer myopia, and inertia.[5] Overall, this paper contributes to this literature by showing that an often overlooked regulation, risk corridors, matter when modeling the supply side. The paper finds yet another flawed market design because there is a misalignment between the objective of the government and private incentives. This leads to a worse market outcome than the government has intended.

The rest of the paper is structured as follows. In section 1.2, I provide a brief description of the Medicare Part D market, paying particular attention to the supply side policies. I then present a stylized theoretical model in section 1.3, showing the effect of both the risk corridor and the margin regulation on the market, especially in the presence of asymmetric information. In section 1.4, I detail the data used for the structural model. In section 1.5, I present the structural model of demand and

[5] Examples of this literature include: Abaluck & Gruber (2011); Kling *et al.* (2012); Ketcham *et al.* (2012, 2015); Abaluck *et al.* (2018); Einav *et al.* (2015); Dalton *et al.* (2020); Ho *et al.* (2017); Lucarelli *et al.* (2012); Polyakova (2016). A majority of this literature focuses on the demand side choice frictions, but some papers like Ho *et al.* (2017); Lucarelli *et al.* (2012); Polyakova (2016) look at insurers' strategic responses to such demand side frictions. They find that there is strong evidence for such frictions and that policies that remove such frictions will lead to welfare increases by lowering prices and decreasing drug expenditure of enrollees. On the supply-side, a few papers look at insurers' strategic benefit designs. Einav *et al.* (2018) documents within plan heterogeneity in cost sharing across different types of drugs. Lavetti & Simon (2018); Starc & Town (2020) study benefit design differences between medically-integrated (MA-PD) vs. stand-alone prescription drug (PDP) plans, finding that MA-PDs have more generous formulary designs because PDPs do not internalize spillovers between drug and medical costs.

supply and in section 1.6, the estimates. Lastly in section 1.7, I present the market equilibrium under various counterfactuals. Section 1.8 concludes.

1.2. Institutional Details

Medicare is a federal health insurance program primarily designed for Americans aged 65 and older. It provided coverage for 62 million people in 2020. Medicare Parts A&B, also known as traditional Medicare or the fee-for-service (FFS) program, directly offer hospital/medical coverage. Under Parts A&B, the government pays health care providers directly for beneficiaries' utilization or healthcare services. Alternatively, beneficiaries can get their Medicare benefits from a private heath insurance plan known as Medicare Part C or Medicare Advantage (MA). Under Part C, a private health insurer provides similar coverage benefits as those offered under Medicare Parts A&B. Enrollees may pay an additional premium to the insurer.[6] The private health plan may include additional benefits such as vision, dental and prescription drug coverage.

Medicare Part D was introduced as part of the Medicare Modernization Act (MMA) in 2003. It provides prescription drug benefits to Medicare beneficiaries. Unlike Parts A&B, in which the government provides the coverage directly, Part D benefits are provided solely by private prescription drug plans, much like part C. In 2020, 47 million Medicare beneficiaries received prescription drug benefits through Part D, costing the government $90 billion.[7]

[6] The actual premium is heavily subsidized, and insurers receive a capitated payment for each enrollee.

[7] https://www.cbo.gov/data/baseline-projections-selected-programs#10.

When it comes to the prescription drug benefits, Medicare enrollees usually choose to either: i) enroll in traditional Medicare (Part A&B) for medical coverage and enroll in a private stand-alone prescription drug coverage (PDP) or ii) enroll in a private health plan via Medicare Advantage (Part C) that provides both medical and prescription drug coverage (MA-PD).[8] This paper focuses on the stand-alone prescription drug coverage (PDP) market.[9]

The PDP market is comprised of 34 PDP regions, or groups of neighbouring states.[10] Each PDP region defines a unique market and acts as a centralized marketplace in which insurers can enter and compete by offering different prescription drug plans. Every June of the year prior to the plan benefit year, insurers will submit their "bids" to the Center for Medicare & Medicaid Services (CMS) for each plan that they're planning to offer for the following year.[11] Included in the bids are plan financial characteristics (premium, deductible, co-insurance/co-payment, actuarial value, etc.) and the formulary design (the type of drugs covered), which need to meet regulatory requirements.[12] More importantly, the bids also include

[8]Medicare enrollees could also choose to not have any prescription drug coverage via Part D whether that's enrolling in traditional Medicare and not purchasing a PDP or enrolling in a MA plan that does not provide any prescription drug coverage. This accounted for around 25% of enrollees in 2020.

[9]While most of the paper's empirical study focuses on PDPs, MA-PDs are faced with very similar if not identical policies to those addressed in this paper. In fact, some of the policies like the margin regulation exist in the MA market as well.

[10]See figure A.1 for how the regions are broken up.

[11]Note that here "bid" does not refer to a bid in an auction setting in which only one firm wins the contract. Bids refer to the premiums that insurers would like to charge. So the bids here can be thought of as prices that firms set in a standard product market setting

[12]For details on the exact requirements, refer to https://crsreports.congress.gov/product/pdf/R/R40611/19.

insurers' estimated average costs for the plan.[13] CMS reviews the insurers' bids for compliance. CMS then uses the information in insurers' bids to compute the beneficiary subsidy level, and determine the post-subsidy enrollee premiums.[14] From mid-October to December of the preceding year, enrollees choose a plan from a menu of plan options available in their region.

1.2.1. Bid Gain/Loss Margin Requirement

Aside from the requirements on plan benefit structures, insurers also face a margin regulation that limits the price that they can charge relative to the reported average cost estimate.

At the individual plan bid level, CMS scrutinizes any bids that have very high or low margins and wants to ensure that *"bids must provide benefit value in relation to the margin"*. In practice, CMS will scrutinize any plans that have negative expected margins or plans that have extraordinarily high expected margins. At the firm level, CMS requires that *"the aggregate (projected enrollment-weighted average) Part D margin as a percentage of revenue must be within 1.5 percent of the Part D sponsor's*

[13] To be more precise, insurers need to submit their expected plan-liable cost including any administrative costs as well as plan profits.

[14] The enrollee subsidy is determined by multiplying a factor by the weighted average of all plan bids, called the National Average Bid Amount. The factor was around 0.53 in 2015 i.e. the subsidy covered just over 50% of the average price, meaning enrollees only had to pay 50% of the premium to purchase an average-priced plan.

margin for all non-Medicare business, as measured by percentage of revenue".[15] Here, *non-Medicare* business refers to insurers' commercial health insurance businesses.[16,17]

CMS is intended to impose a rate-of-return type regulation by benchmarking insurers' Part D margin to their commercial business counterpart. The stated purpose of the margin requirement is

> *Gain/loss margin refers to the additional revenue requirement beyond allowed prescription drug costs and non-benefit expenses. The gain/loss requirements ensure that gain/loss margins are reasonable and that a Part D organization's Part D business is not used to subsidize its other insurance lines of business.*[18]

CMS does not want insurers to make "excessive" profit by exercising market power in the Part D business. It enforces this through the margin requirement. This regulation, however, is an ex-ante margin regulation; CMS applies this when the insurers submit their bids. The margin used is insurers' reported *expected* margin using insurers' reported cost estimate before costs have been realized. As a result, insurers' ex-post realized margin may not be in line with the margin requirement.

[15]To be more precise, insurers can choose the level at which aggregate margin can be applied. It could be at the contract level, or at the firm level. However, most firms choose requirements to be at the firm level.

[16]To be more precise, the term "non-Medicare" business refers to "all health insurance business that is not Medicare Advantage or Part D. Non-Medicare business includes, but is not limited to, the following line of business: Medicare-Medicaid, Medicare-supplemental, Medicaid and commercial."

[17]For insurers that do not have any business outside of Medicare, the margin requirement is based on a "risk-capital-surplus" approach.

[18]https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Bid-Pricing-Tools-and-Instructions-Items/BPT2015

1.2.2. Risk Corridors: Risk Sharing in Medicare Part D

When the Part D market was first introduced, policy makers were concerned about insurer participation and drug benefit affordability. To address these concerns, CMS put forth several risk sharing policies that limit insurers' financial risk. Risk corridors are one such policy.

Risk corridors are an ex-post transfer scheme between the insurer and the government. They are a function of insurers' expected cost, or *target spending* and insurers' realized cost, or *actual spending*. The government sets insurers' reported average cost estimate as *target spending* for the risk corridor program. After the contract year, insurers will report their realized cost for each plan. The government takes this as *actual spending*. The government then applies the transfers for the risk corridor program, as shown in Figure 1.1.

If the plan's actual cost is within 5% of the expected cost, there will be no transfers and the insurer will bear the full risk for that plan. If the actual cost is larger (smaller) than the expected cost by $5 - 10\%$, then the government will reimburse (charge) the insurer 50% of the difference over the 5% threshold. If the actual cost is larger (smaller) than the expected cost by more than 10%, then the government will reimburse (charge) the insurer 80% of the difference over the 10% threshold on top of the 50% cost sharing in the $5-10\%$ threshold. In short, the risk corridor is a risk sharing policy that reimburses (charges) insurers if the actual cost is higher (lower) than the expected cost.

Figure 1.1. Risk Corridor payment reconciliation as a function of actual vs. target or expected cost

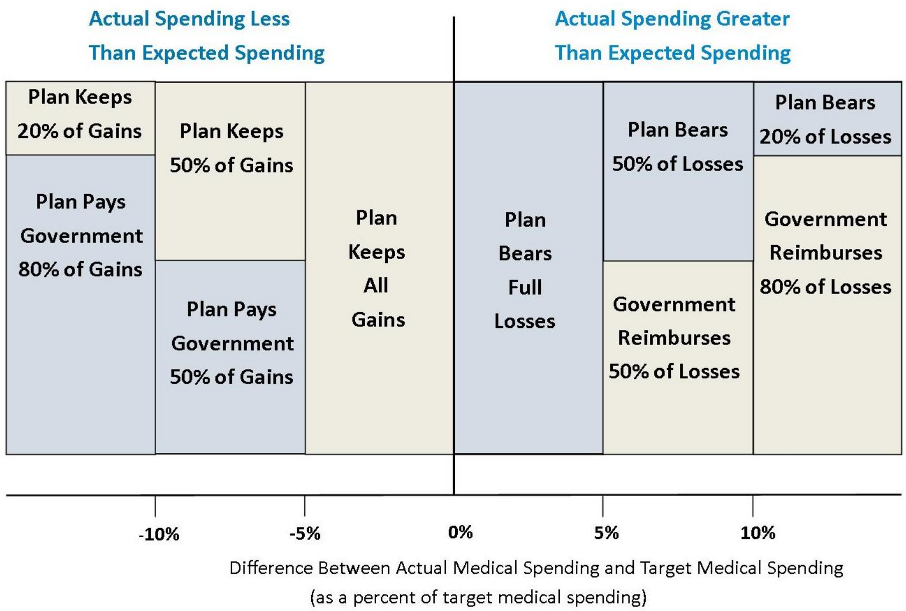


Figure shows risk corridor policy parameters in Medicare Part D.

There are two additional risk sharing policies in the premium stabilization program: risk adjustment and reinsurance. Risk adjustment is primarily there to address adverse selection. It evaluates ex-ante the riskiness of individual enrollees based on each individual's health and expected spending.[19] The plans' capitated monthly premiums (i.e. the plan "bid") are then adjusted by enrollee's risk score such that they are paid relatively more for sicker enrollees and relatively less for healthier ones. By construction, this measure is expected to be budget neutral for the government: it delivers ex-ante transfers from plans that enrolled sicker enrollees to plans that enrolled healthier enrollees.

[19] In practice, CMS takes the enrollee's historical drug expenditure as well as pre-existing medical conditions to predict drug expenditure and constructs numeric risk scores.

Reinsurance acts as an ex-post subsidy for the insurers for incurring high-cost enrollees. When an enrollee has sufficient spending to reach the out of pocket threshold, CMS will reimburse a significant portion of the cost beyond the threshold.[20] It acts as insurance for the primary insurer. In fact, this type of contract is quite common in the broader health insurance market, in which primary insurers will purchase private reinsurance from a third party, often at a high markup. Here, the government acts as a reinsurance company without collecting any premiums from the insurer, effectively providing free reinsurance.[21]

All three of the above risk sharing policies are widely used in other government-funded social insurance programs like Medicaid and the ACA Exchanges. Furthermore, ex-post risk sharing policies like risk corridors and reinsurance are commonly used in the private market.

1.2.3. "Risk Aversion" of Insurers

Here, I briefly discuss how risk affects insurance companies. While the traditional theory of the firm assumes firms to be risk neutral, in reality there are several reasons why insurance companies may be "risk averse" (Fama & Jensen, 1983).[22] First, firms are managed by individuals who may be risk averse, especially if their pay is

[20]The out of pocket threshold, also known as catastrophic cap, is a pre-defined threshold set by the government each year. In 2015, it was at about $\$7,000$.

[21]So the government's provision of reinsurance is a supply-side subsidy.

[22]The rationale for risk neutral firms is that investors can diversity their investment portfolio through diversification and minimize any firm specific risk.

tied to the firms' performance. Empirically, there is strong evidence (Hall & Liebman, 1998) of a growing correlation between manager pay and firm performance.

Second, insurance companies are subject to financial regulations. Insurance regulators (as well as rating agencies) regularly asses the financial strength of insurers, much like the capital requirements (or the solvency regulations) in the banking industry (Walter, 2019). In the U.S., while individual states have their own set of insurance regulators, most of them follow the risk-based capital (RBC) regulation set out by the National Association of Insurance Commissioners (NAIC). Risk-based capital ratios (also known as RBC ratios) determine the minimum amount of capital required for a given amount of risk assumed by the insurer and compares it to the insurer's total capital and surplus levels. For health insurers, the required capital is often some factor applied to the total claims that they are liable for.[23,24] Each year, insurance regulators review the RBC ratios of the insurers and may take action if it falls below certain standards.[25] In fact, Koijen & Yogo (2015b) document that life insurance companies face a high degree of financial/regulatory frictions due to RBC regulation and Kim & Li (2022) find that insurers' financial solvency level affects their premium setting decisions.

[23]The exact factor varies but is usually between $5-15\%$ i.e. the required capital is often set as $5-15\%$ of the total claims.

[24]Note that for health insurers, the RBC ratio is often an ex-post measure of financial solvency. This is because the minimum required capital is a function of already realized claims vs. some expected liability, as is the case for other insurance sectors.

[25]In the extreme case, the regulator will assume direct control of the insurance company (NAIC, 2011).

Third, there is an active private reinsurance market in which primary insurance companies purchase insurance products (Bovbjerg *et al.* , 2008). These private reinsurance policies are often sold by unaffiliated reinsurance companies and are purchased despite their high markups.[26] If insurers were risk neutral, there would be no market for such policies.

Finally, there is ample evidence that insurance companies take into account the amount of risk they face in setting their premiums. The actuarial literature frequently factors in "risk premium" or "risk charges", often measured by the variance or the standard deviation of the claims liability (Kahane, 1979). Furthermore, according to the American Academy of Actuaries, policies like risk corridors can reduce premiums by reducing risk charges:

> *"Risk corridors can allow insurers to reduce their risk charges, although risk charges are usually a fairly small percentage of the premium (e.g., $2\% - 4\%$). Another way risk corridors can result in lower premiums is that having a backstop can allow insurers to price using less conservative assumptions." (American Academy of Actuaries, 2020)*

As described above, there are many reasons why insurers may act as if they are risk averse. It is especially important in studying a risk sharing policy like risk corridors, as such a policy will have no real effect on risk neutral insurers. As a result, I depart from the standard model of risk neutral firms and allow insurers to behave as "risk averse".

[26] The loading factor i.e. the portion of premiums above and beyond the expected claims can be as high as 50%.

1.3. Stylized Model

Here I present a stylized model where a monopoly insurer faces some frictions for taking on risk. I introduce the two sets of regulations in the Part D market: risk corridor and margin constraint, first studying each policy by itself and later combining both together. I compare the effect of these policies in the presence of symmetric information vs. asymmetric information about costs, in which the insurer has private information about its expected cost.

1.3.1. Stylized Model

Consider a monopoly insurer facing an elastic demand for its product, $q(p)$.[27] For each individual it enrolls it faces a random marginal cost of $\tilde{c}_i = c + \epsilon_i$, where c is the expected cost, and ϵ_i is a iid zero-mean shock with $Var(\epsilon_i) = \sigma^2$.[28] With demand $q(p)$, the insurer faces a random total cost of $\tilde{C} = \sum_{i=1}^{q(p)} \tilde{c}_i$. Given the uncertainty in cost, the insurer incurs a risk charge as a function of the variance of the total cost, $V(\tilde{C})$. This can be seen as an approximation to an insurer that faces some cost of financial frictions for incurring ex-post losses.[29,30] Alternatively, the insurer

[27]Note that we are implicitly assuming that there is no uncertainty in demand.

[28]Here we assume the individual level cost shocks are independent, which seems plausible in the context of health insurance. However, we can generalize the results to a case in which individual costs are correlated.

[29]The approximation is to a model in which insurer faces some convex cost function for incurring ex-post losses. See Appendix A for more details.

[30]In reality, insurance companies do seem to face some financial/regulatory frictions regarding their solvency measures. There are state regulations on insurers' risk-based capital ratios as well as evidence of insurers purchasing private reinsurance policies to reduce risk that they face.

can be taken to be risk averse, which will be isomorphic to a model in which the insurer faces financial frictions.[31]

The insurer maximizes the following expected profit function:

$$\max_{p} \quad pq(p) - \underbrace{cq(p)}_{E[\tilde{C}]} - \underbrace{\rho V(\tilde{C})}_{\text{risk charge}} \tag{1.1}$$

where $\rho \geq 0$ is the coefficient of risk charge. The insurer's FOC yields:

$$p_0^* \left(1 + \frac{1}{\epsilon^D}\right) = c + \underbrace{\rho \frac{\overbrace{\partial V(\tilde{C})}^{\sigma^2}}{\partial q}}_{\text{marginal risk charge}} \tag{1.2}$$

where ϵ^D is the price elasticity of demand. We get a similar FOC to a standard monopoly model where marginal revenue equals marginal cost. However, here the effective marginal cost includes a marginal risk charge term that makes the marginal cost strictly higher. Note that (1.2) makes it clear that as the coefficient of risk charge and/or the uncertainty in cost increases, the marginal risk charge will increase, leading the insurer to charge higher prices.[32]

Let p_0^* denote the optimal price in (1.2) i.e. the insurer's profit-maximizing price in the absence of any regulations. In the latter sections, I compare the optimal prices under different regulations vs. p_0^*.

[31]See Appendix A for more details.

[32]This is in line with what many insurers seem to be doing when there is increased uncertainty in the cost.

1.3.2. Ex-ante Margin Constraint

Now suppose the monopoly insurer faces a margin regulation such that the insurer's margin relative to its expected cost is constrained by an upper bound of $\bar{m}$. The insurer now faces the following constrained profit function:

$$\max_{p} \quad pq(p) - cq(p) - \rho V(\tilde{C}) \qquad s.t. \qquad p \leq \bar{m}c \tag{1.3}$$

The insurer's new FOC will now be:

$$p\left(1 + \frac{1}{\epsilon^D}\right) = c + \rho\frac{\partial V(\tilde{C})}{\partial q} + \frac{\lambda}{\frac{\partial q}{\partial p}} \tag{1.4}$$

where $\lambda \geq 0$ is the Lagrange multiplier to the margin constraint. It is easy to see that the solution to the above FOC will be $p^* = p_0^*$ if the constraint does not bind (i.e. price stays the same as in (1.1)) or $p^* = \bar{m}c$ if the constraint binds.

1.3.2.1. Asymmetric Information. We explore the asymmetric information case, in which the insurer can strategically report its cost.

$$\max_{p,\delta\in[\underline{\delta},\overline{\delta}]} \quad pq(p) - cq(p) - \rho V(\tilde{C}) \qquad s.t. \qquad p \leq \bar{m}\underbrace{\delta c}_{\hat{c}} \tag{1.5}$$

The insurer can now over or underestimate its ex-ante expected cost by parameter $\delta \in [\underline{\delta}, \overline{\delta}], \overline{\delta} > 1 > \underline{\delta} \geq 0$, reporting a cost estimate of $\hat{c} = \delta c$. If $\delta < 1$ (or $\delta > 1$), then the insurer under-reports (over-reports) its expected cost, where $\delta = 1$ denotes the insurer reporting the true expected cost.[33]

[33] In other words, $\delta = 1$ can be seen as the symmetric information case in which the government knows the expected cost of the insurer.

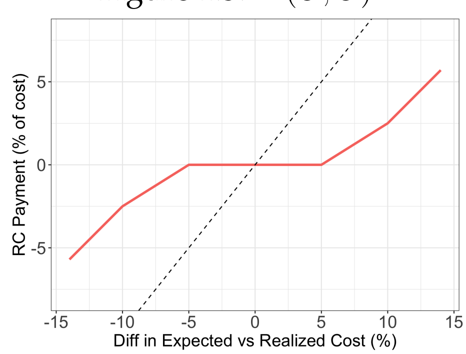
Assumption 1. *$p_0^* \leq \bar{m}\overline{\delta}c$.*

Proposition 1. *$\delta^* = \overline{\delta}$ will always be optimal for the insurer's problem in (1.5). Furthermore if assumption 1 holds, $p_m^*(\delta^*) = p_0^*$ where $p_m^*(\delta)$ denotes the insurer's profit-maximizing price in (1.5) for a given δ.*

Proposition 1 states that when we allow the insurer to strategically report its expected cost, the margin constraint plays no role in the insurer's pricing decision. This is because whatever price the insurer wants to set (i.e. p_0^*), it can set the price by reporting δ s.t. its reported cost estimate is high enough for the margin constraint to be satisfied without being penalized for misreporting its cost. In short, it will be always optimal for the insurer to report $\delta^* = \overline{\delta}$ when there is only an ex-ante margin constraint.

1.3.3. Risk Corridor

Figure 1.3. $T(\tilde{C}, C)$



Plot of $T(\tilde{C}, C)$ shown in percentage of the expected cost C. The horizontal axis shows the difference in expected cost C and the ex-post cost $\tilde{C}$ as a percentage of C. The vertical axis shows the RC payment as a percent of expected cost C.

Here, I illustrate the effect of risk corridor payments (RCP) in reducing the risk that an insurer may face. RCP act as an ex-post transfer function between the insurer and the government as a function of ex-post cost, $\tilde{C}$ and ex-ante expected cost $C = E[\tilde{C}] = cq$ (see figure 1.3). With RCP, the insurer's new ex-post cost will be:[34]

$$\tilde{C}^{rc} = \tilde{C} + \underbrace{T(\tilde{C}, C)}_{\text{rc payment}} \tag{1.6}$$

$$\text{where} \quad T(\tilde{C}, C) = \begin{cases} -0.8\tilde{C} + 0.855C & \text{if } \tilde{C} > C \text{ by more than } 10\% \\ -0.5\tilde{C} + 0.525C & \text{if } \tilde{C} > C \text{ by } 5-10\% \\ 0 & \text{if } \tilde{C} \text{ within } 5\% \text{ of } C \\ -0.5\tilde{C} + 0.475C & \text{if } \tilde{C} < C \text{ by } 5-10\% \\ -0.8\tilde{C} + 0.745C & \text{if } \tilde{C} < C \text{ by more than } 10\% \end{cases} \tag{1.7}$$

In short, the RCP is such that if the actual cost is lower (higher) than the expected cost by more than 5%, the insurer will pay (receive) a portion of the difference where the payment is determined via a kinked-linear function. If I assume that the distribution of the insurer's total cost $\tilde{C}$ is symmetrical about its mean, then the expected risk corridor payments will be zero i.e. $E[T(\tilde{C}, C)] = 0$ and hence $E[\tilde{C}^{rc}] = E[\tilde{C}] = cq$.[35] The variance of the $\tilde{C}^{rc}$ on the other hand will be directly

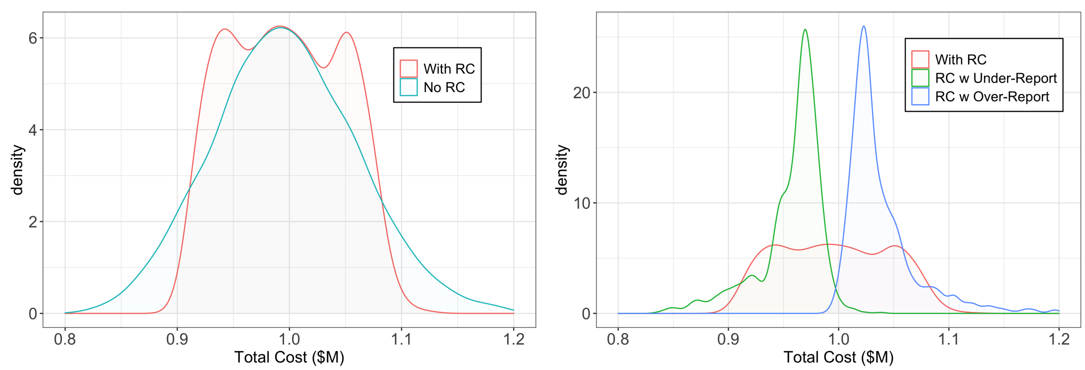
[34]Note that the function $T(x, y)$ is homogeneous of degree one. Hence, $T(\tilde{C}, C) = T(\tilde{C}/q, c)q(p)$.

[35]If we maintain the assumption that individual level costs are independent, then the central limit theorem will imply that the distribution of total cost will be approximately normal, which is symmetric.

affected by the risk corridor payments and be weakly smaller i.e. $V(\tilde{C}^{rc}) \leq V(\tilde{C})$. This is illustrated in Figure 1.6a where the distribution of the cost with RC is more condensed, and hence will have lower variance.

Figure 1.5. Distribution of Cost

(a) $\tilde{C}$ and $\tilde{C}^{rc}$ (b) $\tilde{C}^{rc}$ with cost over or underestimate



Panel (a) plots a simulated distribution of cost with and without risk corridors. Panel (b) plots a simulated distribution of cost with risk corridors when insurers over or underestimate their costs.

In the model (1.1), this implies that insurer will face lower risk charge (hence lower effective marginal cost), and charge a lower price.

1.3.3.1. Asymmetric Information. We now explore the asymmetric information case in which the insurer can strategically report its cost.[36] The insurer can now over or underestimate its ex-ante expected cost by parameter δ, hence reporting a cost estimate of $\hat{C} = \delta C$. If $\delta < 1$ (or $\delta > 1$), then the insurer under-reports (over-reports) its expected cost, where $\delta = 1$ denotes the insurer reporting the true

[36]This could be seen as the government having a uniform prior on the cost. Though in reality, the government does have some historical cost data.

expected cost.[37] The insurer maximizes a similar profit function as in (1.1) except that now the insurer's payoff is affected by the risk corridor payments through both the expected cost and the variance. Furthermore, the insurer can alter its payoff by over or under-reporting its expected cost denoted by parameter δ.

$$\max_{p,\delta\in[\underline{\delta},\overline{\delta}]} \quad pq(p) - cq(p) - \underbrace{E[T(\tilde{C},\delta C)]}_{\text{expected RC-payment}} - \underbrace{\rho V(\tilde{C};\delta)}_{\text{risk-charge}} \tag{1.8}$$

Note that the insurer's choice of δ impacts both the expected value and the variance of total cost as illustrated in Figure 1.6b. When the insurer underestimates its cost ($\delta < 1$), the distribution of cost is shifted to the left and more condensed, lowering both the expected cost and the variance. When the insurer overestimates its cost ($\delta > 1$), the distribution of cost is shifted to the right and more condensed, increasing the expected cost while lowering the variance. This is further illustrated in Figure 1.7 where I plot the expected RC payment and the variance as a function of δ for a given price. It shows that the expected RC payment is an increasing function of δ with it being 0 when $\delta = 1$. So when the insurer over (under) estimates its cost, it is expected to pay the government and vice versa. On the other hand, the variance of the cost is highest at $\delta = 1$ and hence decreases when the insurer either over or underestimates its cost. Figure 1.7 implies that the insurer will choose to underestimate its cost as much as possible. I formalize this argument below.

Assumption 2. $1 - \underline{\delta} \geq \overline{\delta} - 1$.

[37] In other words, $\delta = 1$ can be seen as the symmetric information case in which the government knows the insurer's expected cost.

Figure 1.7. Expected risk corridor payment and variance of cost as a function of δ

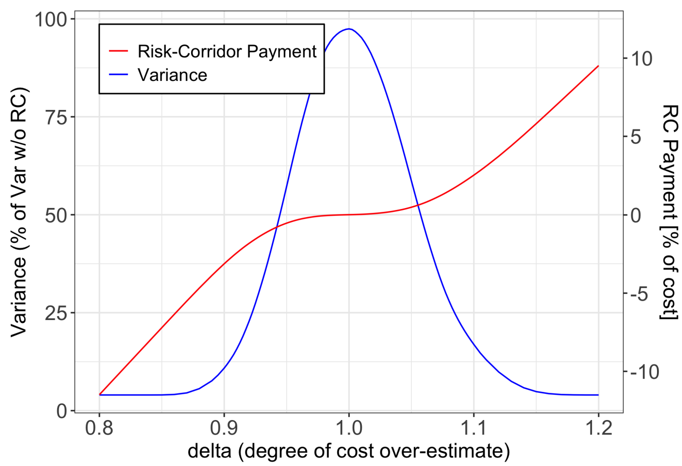


Figure plots expected risk corridor payment and variance of total cost as a function of the insurer's strategic cost reporting, δ

Assumption 2 puts restrictions on $\{\underline{\delta}, \overline{\delta}\}$. The lower bound on the degree to which the insurer can underestimate its cost is equal to or smaller in magnitude than to the upper bound of how much the insurer can overestimate its cost.

Proposition 2. *Under Assumption 2, the optimal δ^* to the insurer's problem in (1.8) will be $\underline{\delta}$ i.e. the insurer will always report the lowest possible expected cost. And furthermore $p^*_{rc}(\underline{\delta}) < p^*_{rc}(1) \leq p^*_0$ where $p^*_{rc}(\delta)$ denotes the insurer's profit-maximizing price in (1.8) for a given δ.*

The intuition for this proposition is simple. For a given price, the insurer will always want to choose δ as low as possible in order to achieve the lowest expected RCP, thereby decreasing its expected cost. The lowest δ will also minimize the variance of the cost and hence the risk that the insurer faces. Given the choice of

$\delta = \underline{\delta}$, the insurer will face strictly lower expected and variance of the cost and hence will lower its price below the price at $\delta = 1$.

1.3.4. Both Regulations

Now suppose there are both sets of regulations in place i.e. both the ex-post risk corridor payments and the ex-ante margin constraint.

The insurer's profit function will now be:

$$\max_{p,\delta} \quad pq(p) - cq(p) - \underbrace{E[T(\tilde{C}, \delta C)]}_{\text{expected RCP}} - \underbrace{\rho V(\tilde{C}; \delta)}_{\text{risk charge}} \qquad s.t. \qquad p \leq \underbrace{\bar{m}\delta}_{\bar{m}'} c \tag{1.9}$$

The endogenous cost reporting δ affects the insurer's profit in three different ways. First, δ affects the insurer's expected cost through $E[T(\tilde{C}, \delta C)]$. Second, δ affects the insurer's risk level through $V(\tilde{C}; \delta)$. Lastly, δ affects the insurer's margin constraint $p \leq \bar{m}\delta c$. From previous sections, the insurer will want to underestimate its cost ($\delta < 1$) in order to lower its expected RC payment. On the other hand, the insurer will want to overestimate its cost ($\delta > 1$) in order to increase its upper bound on the margin. When combined together, it's unclear whether the insurer will over or underestimate its cost depends on a few things. We formalize the direction of the optimal δ below.

Proposition 3. *The optimal δ^* in the insurer's problem in (1.9) will be $\underline{\delta} \leq \delta^* < 1$ or $1 < \delta^* \leq \overline{\delta}$ if the margin constraint does not bind or if the margin constraint strictly binds* *at $p^*_{rc}(1)$, respectively. Furthermore the insurer's profit-maximizing price, p^*_{both} will be s.t. $p^*_{rc} \leq p^*_{both} \leq p^*_0$.*

Proposition 3 states that the insurer will either over or underestimate its cost. The direction will depend on various primitives like the demand and marginal cost, as well as the upper bound on the margin. In general, if at $\delta = 1$ the optimal price $p^*_{rc}(1) < \bar{m}c$ then the insurer will want to underestimate its cost in order to lower its expected RC payment, decreasing its expected cost up until the margin constraint binds. Hence, the insurer will underestimate its cost but likely not all the way to $\delta = \underline{\delta}$. On the other hand, if $p^*_{rc}(1) > \bar{m}c$ (i.e. the margin constraint binds) the insurer will want to overestimate its cost in order to increase its upper bound on the margin, thereby allowing it to charge higher prices. However, the insurer will not overestimate its cost all the way to $\delta = \overline{\delta}$ as doing so would increase the expected RC payment.

In summary, the above model illustrates that when the government and the firm have symmetric information on costs, the risk corridor can reduce any frictions by reducing the risk that insurers face, and that margin constraint limits market power by constraining the price the insurer can charge. However, when there is asymmetric information (i.e. knowledge about expected cost is the insurer's private information and that the government only observes the ex-post realized cost), the risk corridor gives the insurer an incentive to underestimate its cost. On the other hand, margin regulation gives the insurer an incentive to overestimate its cost. When these incentives are both in play, the net effect is indeterminate and will depend on the context.

1.4. Data and Descriptives

1.4.1. Data

The paper uses three different types of data.

CMS Plan Data:[38] CMS's market-plan-year level data includes details on all Part D plans that were offered from 2010-2015.[39] It includes plan prices, detailed plan characteristics like deductibles, the drug formulary design and associated co-insurance/co-pay rates, as well as the plan-specific average price for each drug in the formulary.[40] It also includes total monthly enrollment, and the average risk score of the enrollees enrolled in the plan. CMS also publishes the total number of Medicare-eligible beneficiaries. I exclude plans that are employer-sponsored plans and restrict the sample to the 50 US continental states.

Although not at the plan level, I also observe year-contract level risk corridor payments to insurers from CMS's payment files. A contract is defined as a group of similar products that an insurer offers and can be thought of as firm level risk corridor payments.

[38] All CMS data are publicly available and can be found here: https://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems

[39] Specifically, I use the following data from CMS: i) Medicare Advantage/Part D Contract and Enrollment Data ii) Prescription Drug Plan Formulary, Pharmacy Network, and Pricing Information Files.

[40] However, the premium information is only available for PDP as the premium for MA-PD also include rebates from MA that insurer can apply to buy down the part D premium.

Medicare Current Beneficiary Survey Data: I use the 2012-2015 Medicare Current Beneficiary Survey (MCBS) Limited Data that includes a nationally representative sample of Medicare beneficiaries.[41] For each individual, it includes detailed demographic information including income, age, and overall health level as well as the Part D plan enrollment, which details the specific Part D plan that the individual was enrolled in, if any. The MCBS data also includes administrative claims data with information on the individual's drug purchase history for the given year; the information about the specific drug purchased and the total cost of the drug for each consumer. I restrict the sample to include individuals in the 50 U.S. states, individuals for whom I observe Part D information, and individuals for whom I observe their claims data. This results in 27,262 individual-years.[42]

Insurer Financial Statements Data: I use insurer financial statements data to get firm level Part D costs as well as the insurer's non-Medicare margins. I use two filings from the National Association of Insurance Commissioners (NAIC): 2012-2015 Medicare Part D Coverage Supplement filing and 2008-2019 5-Year Historical filing. The first filing has detailed yearly financial statements for insurers' Part D businesses (PDP-only), including the total cost incurred for each insurer in the Part D market. The second filing has firm-year level aggregate financial information like the insurer's RBC ratio, a financial solvency measure used by the insurance regulators. Lastly, I use 2010-2015 CMS's Medical Loss Ratio data that has firm-year level financial statements data across different lines of the health insurance

[41] 2014 is excluded due to MCBS missing the data for that year.
[42] I also exclude individuals enrolled in employer-sponsored Part D plans.

business. I use this to get the insurer's non-Medicare or commercial business margin used as a benchmark for the Part D margin regulation.

1.4.2. Descriptives

Table 1.1 shows summary statistics on the Part D PDP market from 2012-2015.[43] The average price for a PDP plan was $\$1,191$, of which enrollees only had to pay $\$643$ on average. The difference between the two, $\$548$, reflects the average consumer subsidy paid by the government. Plans on average had $9,600$ enrollees but there is a large variation ranging from 10 to $300,000$.

Consumers on average had 31 Part D PDP plans to choose from, offered by 13-14 different insurers. While consumers had a good number of options to choose from, the number is smaller than earlier years of the Part D market (Decarolis *et al.* , 2020a). This is in part due to CMS implementing a number of regulations that limit the total number of plans each insurer can offer.[44] But it also reflects the more concentrated market (Chorniy *et al.* , 2020). The mean Herfindahl-Hirschman index across markets from 2012-2015 was close to 2,500, which the Department of Justice regards as a highly concentrated market.[45] On average, the top three firms account for 75% of market share and the top five firms account for 90% of the market share.

[43]The summary statistics only include regular enrollees. The summary statistics for the low-income subsidy (LIS) eligible enrollees can be found in Table A.1.

[44]In 2010, CMS issued "meaningful difference" requirements in which an insurer couldn't offer two plans that were too similar to one another.

[45]The horizontal merger guidelines from the Department of Justice and the Federal Trade Commission classifies markets with HHI above 2500 as highly concentrated markets.

Table 1.1. Summary Statistics

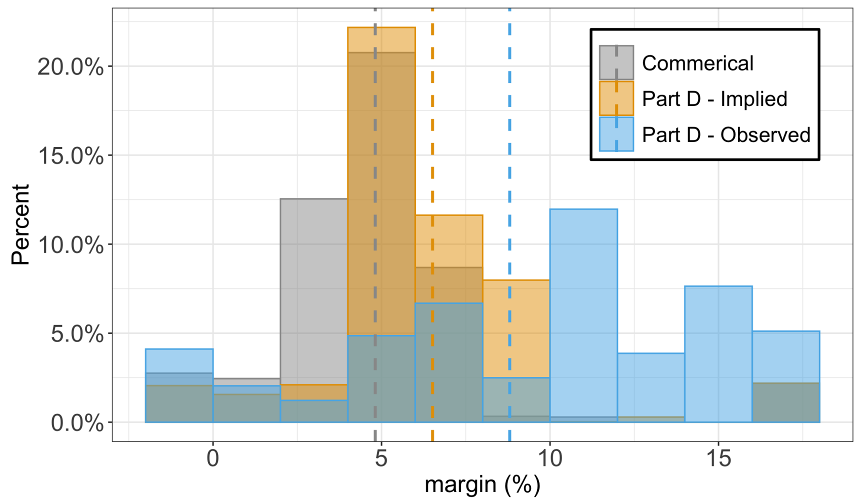
	Mean (1)	Std.Dev. (2)	Min (3)	Max (4)
A. Plan Level				
Bid ($)	1,191	367	596	2,618
Enrollee Premium ($)	643	360	150	2,096
Enrollment (000)	9.6	23.9	0.01	293.4
B. Market Level				
No of Plans	31.19	3.00	23	39
No of Insurers	13.65	1.34	10	17
Enrollment (000)	298	206	7	847
HHI Index*	2,452	489	1,801	3,822
Market Share of Top 3 Firms (%)*	74.4	5.1	64.9	84.7
Market Share of Top 5 Firms (%)*	90.0	2.5	84.4	95.5

Notes: the table shows summary statistics of the Part D stand-alone prescription drug (PDP) market from 2012-2015 in the 34 PDP regions for regular enrollees. Plan level data shows summary statistics taken across individual year-market-plan. Market level data shows summary statistics taken across year-market level. Enrollee premium refers to premium faced by regular enrollees. An insurer is defined as a unique parent organization in the CMS data. * HHI index and market share of top firms are computed using regular enrollees only.

Consistent with the relatively high level of concentration in the Part D market, figure 1.9 shows that insurers are much more profitable in their Part D business compared to their non-Medicare or commercial business. In fact, insurers' ex-post profit margins in their Part D business is higher than what the profit margin regulation would dictate. Figure 1.9 shows that insurers' observed ex-post Part D business profit margins are much higher than what's implied by the regulation i.e. within 1.5% of insurers commercial business margins. So for most insurers, the profit margin regulation likely binds and constrains the margin that they can charge in the Part D market, meaning insurers over-report their cost estimates to relax the constraint.

Figure 1.9. Distribution of Risk Corridor Payments and Insurer Margin

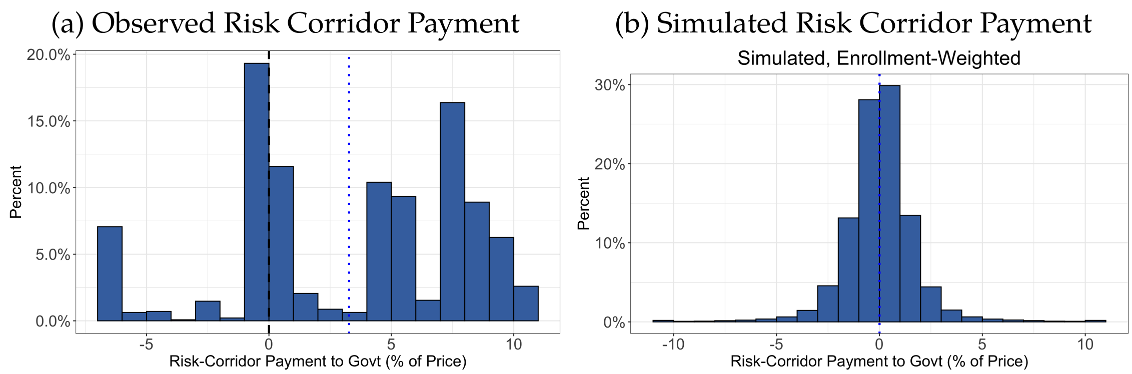
Figure 1.10. Part D vs. Non-Medicare Margin



The figure shows the distribution of 2012-2015 Part D insurers' i) non-Medicare or commercial business ex-post profit margins ii) allowable or implied profit margins under the ex-ante profit margin regulation (i.e. i) shifted by 1.5% margin) iii) observed ex-post profit margins of the PDP business. The distributions are weighted by observed enrollment.

The observed risk corridor payments in figure 1.13a further show patterns consistent with insurers over-reporting their cost estimates. The distribution is heavily skewed towards the right, implying many insurers' actual costs were much higher than their reported expected cost, and thus insurers are making large ex-post payments to the government. This is in stark contrast with figure 1.13b that shows the simulated risk corridor payments had insurers truthfully reported their cost estimates. Furthermore, the risk corridor payment patterns are persistent across years. Insurers are much more likely to have positive risk corridor payments if they had positive risk corridor payments in the prior year, and vice versa.[46]

Figure 1.12. Distribution of Observed vs. Simulated Risk Corridor Payments



Panel (a) plots plots the distribution of observed risk corridor payments from 2012-2015 for each PDP contract. Panel (b) plots the distribution of simulated risk corridor payemtns assuming insurers truthfully report their expected cost. The distributions are weighted by observed enrollment.

1.5. Empirical Model

1.5.1. Model of Demand

I model the demand for PDP coverage for Medicare beneficiaires in the 34 Part D markets over the years 2012-2013 and 2015.[47] I do so using a standard discrete choice model (Berry *et al.* , 1995) similar to Decarolis *et al.* (2020a) in which a consumer derives indirect utility from choosing a particular product and chooses the product that maximizes his or her utility. I estimate the demand separately for the

[46]See figure A.3 for more details. For more evidence of risk corridor payments being random, see A.5.

[47]I leave out demand estimation for year 2014 due to missing MCBS data in year 2014.

two populations in the market: regular enrollees and low-income subsidy (LIS) eligible enrollees.[48] Below, I detail the demand specification for the regular enrollees as the specification for the LIS enrollees follows a similar structure.[49]

Individual i's utility from choosing plan j in market m is given by:

$$u_{ijm} = \alpha_i \, p^e_{jm} + \beta_i X_{jm} + \xi_{jm} + \varepsilon_{ijm} \tag{1.10}$$

p^e_{jm} is the enrollee plan premium after government subsidy has been applied.[50] X_{jm} are other observable plan characteristics that include the plan deductible, whether the plan provides additional coverage beyond the minimum requirement (i.e. an enhanced plan), whether the plan has extra coverage in the donut hole, and the number of drugs covered in the plan's formulary. Following Decarolis *et al.* (2020a) and Starc & Town (2020), I also include plan vintage or the number of years the plan has been in the market as a reduced-form way of capturing consumer inertia.[51]

[48] LIS eligible beneficiaires receive extra assistance from the government in premiums as well as extra cost sharing in drug spending.

[49] The demand specification for the LIS enrollees is similar to that of the regular enrollees, except that I limit some preference heterogeneity within the LIS enrollees. For example, I do not partition the LIS enrollees into different income groups. I also adjust many of the plan attributes to reflect the extra cost sharing that LIS enrollees receive. For example, the deductible is zero for LIS enrollees, and many of the plan premiums are also zero if they are below the LIS benchmark.

[50] The consumer subsidy in Medicare Part D acts like a flat voucher in which government provides a pre-set $\$S$ amount regardless of which plan the enrollee chooses. The subsidy level S for a given year is usually some factor multiplied by the lagged-enrollment weighted average prices across all plans.

[51] Decarolis *et al.* (2020a) show that this approach corresponds to "structural two-stage model of inattention and choice" (Hortaçsu *et al.* , 2017).

Lastly, the observable plan characteristics include a constant, denoting the value of inside-good relative to the outside option whose utility is normalized to zero. The outside option here indicates Medicare beneficiaries enrolling in a Medicare Advantage medical plan with drug benefits (MA-PD) or opting to not purchase any prescription drug coverage through Medicare part D.[52,53]

I allow heterogeneity in preferences by allowing α_i, the price sensitivity, to vary across an individual's observable characteristics:

(1.11)

$$\alpha_i = \alpha_0 + \sum_{g=2}^{5} \alpha_g^{health}\, \mathbb{1}\{health(i) = g\} + \sum_{g=2}^{3} \alpha_g^{age}\, \mathbb{1}\{age(i) = g\} + \sum_{g=2}^{3} \alpha_g^{income} \mathbb{1}\{income(i) = g\}$$

Here, α_0 indicates the base level of price-sensitivity common for all individuals. I then allow price-sensitivity to vary by individuals' self-reported health level. The MCBS data includes survey results in which individuals are asked to select between five health levels ranging from "poor" to "excellent".[54] This is denoted by $\mathbb{1}\{health(i) = g\}$, a dummy variable equal to one if individual i's health level $health(i)$ is g and zero otherwise. Next, I allow the preferences to vary by demographic groups: I group individuals into three age bins and three income bins. This is denoted by similar dummy variables for age and income groups. Thus,

[52] In 2015, 43% of Medicare beneficiaries chose a PDP plan or the inside-good in the model, 28% chose an MA-PD plan, and the remaining 29% chose not to purchase a drug coverage through Part D. It is estimated that of the beneficiaries who do not purchase any coverage, 64% of them receive drug coverage through a third party and the remaining 36% has no drug coverage.

[53] For LIS enrollees, the outside option strictly refers to purchasing a MA-PD plan.

[54] These self-reported health levels are a good predictor of the overall medical and drug spending.

the price coefficient for the least healthy, youngest and lowest income group is the baseline coefficient α_0 whereas the price coefficient for the most healthy, oldest and highest income group is given by $\alpha_0 + \alpha_5^{health} + \alpha_3^{age} + \alpha_3^{income}$. Similarly, I allow β_i, the taste for other plan characteristics, to vary across individuals' observable characteristics in the same way. While not used in the baseline demand specification, as a robustness check I also allow for unobserved heterogeneity through random coefficients.[55]

The final component of the utility is the term: $\xi_{jm} + \varepsilon_{ijm}$. Following the literature, I assume ε_{ijm} is a $i.i.d.$ type I extreme-value distributed random taste shock. The ξ_{jm} is the unobserved plan quality specific to each market that may be correlated with the product characteristics. I first include product and market fixed effects to control for any product specific or market specific unobserved quality.[56] As is commonly the case, I assume that all non-price attributes are exogeneous but allow prices to be endogeneous.[57] I instrument for price using the number of contracts that the insurer has in nearby markets (Decarolis *et al.* , 2020a) as well as the

[55]See demand estimates section for more details on this specification.

[56]Here, the product is defined as contract plan-type pair. Within my time period, the insurer usually offers two or at most three products in each market. These products are usually vertically differentiated products in which one is a "basic" plan that offers the standard coverage and the other is an "enhanced" plan that offers additional coverage beyond the minimum level.

[57]This is motivated by the fact that plans are limited to offering two or three plans that meet certain actuarial values. Beginning in 2011, insurers are subject to meanginful difference requirements across plans that they offer i.e. the insurers are not allowed to offer two plans that are similar attributes and must pass the "meanginful difference" requirements set out by CMS. Furthermore, insurers tend to offer a stable portfolio of plans across the years I study.

insurer's RBC ratio in the prior year.[58,59] The number of contracts in nearby markets reflects potential cost-shifters in insurers' cost (e.g. negotiating prices with local pharmacies).[60] The insurer's RBC ratio in the prior year can be treated as "excluded shifter of firm markups" in Berry & Haile (2022). This is similar to Koijen & Yogo (2022a) that uses life insurers' reserve valuation as an instrument for variable annuities demand. While the existing literature uses prices in nearby markets (Hausman-style instruments) as valid instruments, it hinges on ξ_{jm} not being correlated across markets. Instead, I use supply-side instruments that do not rely on such an assumption.

I estimate the demand following Goolsbee & Petrin (2004). In the first step, I estimate the individual demographic-related coefficients and the mean utility via maximum likelihood. In the second step, I estimate the mean coefficients using two-stage least squares regression using the aforementioned instruments.

[58]RBC ratio is a commonly used measure of financial solvency level of health insurers. In the model, this could be affecting the coefficient of risk charge or the degree of how "risk averse" insurers behave. Kim & Li (2022) show evidence of such financial solveny measures affecting preimums of health insurers.

[59]One concern here could be that the RBC ratio is in part correlated to ongoing demand shocks of that insurer. While that may be true, RBC ratio concerns the insurer's financial situation across all its business lines. However, given that the Part D business usually makes up a small portion of the insurer's business it's unlikely that the demand shocks in Part D will have a large impact on the overall RBC ratio of the insurer.

[60]It might be easier for the insurers to negotiate costs with pharmacies and/or drug manufacturers by operating in larger markets.

1.5.2. Model of Supply

Accurately modeling the supply side of the Part D market is very complicated due to the numerous regulatory provisions in the market.[61] For simplicity, I present the main objective function of insurers and defer any other details to appendix A.4. The below model closely follows the stylized model presented in section 1.3 except that now insurers are multiproduct firms in an oligopoly setting, as opposed to being a single-product monopoly. The objective function of each insurer (suppressing the firm subscript) that offers a set of PDP products J_m in each market m is given by:

(1.12)

$$\Pi_{\{b\},\delta} = \sum_m \sum_{j \in J_m} \left(b_{jm} - c_{jm}\right) \underbrace{Q_{jm}(b)}_{\text{risk-adj demand}} - \underbrace{\gamma_{jm}(\delta, Q_{jm})\, c_{jm} Q_{jm}(b)}_{\text{expected rc payment}} - \underbrace{\rho V_{jm}(\delta, Q_{jm})}_{\text{risk charge}}$$

$$\text{(1.13)} \quad s.t. \quad \sum_m \sum_{j \in J_m} b_{jm} Q_{jm}(b) \leq \overline{m} \sum_m \sum_{j \in J_m} \underbrace{\delta\, c_{jm}}_{\text{cost estimate}} Q_{jm}(b)$$

The insurer maximizes the above objective by choosing the bid-vector $\{b\}$ (comprised of b_{jm}'s, one for each PDP plan), and δ, the degree of strategic cost reporting. If $\delta > 1$ (or $\delta < 1$), then the insurer chooses to overestimate (or underestimate) its

[61] Decarolis *et al.* (2020a) documents a large portion of these and tries to incorporate them as well as possible in their model. However, they do not take account of everything: in particular, the risk corridors and ex-ante margin regulations that insurers face, which is the focus of my paper. I focus on modeling these two regulations as well as possible, while incorporating the other regulatory provisions that Decarolis *et al.* (2020a) include: ex-ante risk adjustment, consumer subsidy rules, especially for the LIS enrollees, etc. However, I make some simplifying assumptions where necessary in order to make the model more tractable and focus primarily on the two regulations.

cost, and $\delta = 1$ corresponds to the insurer correctly reporting its expected cost to the government.

The insurer's objective function comprises three parts that are summed over all the plans. The first part is the standard expected profit i.e. price (or "bid" in this setting) minus the expected average cost times the demand. The second part is the expected risk corridor **transfer payments to the government**. It is the product of the plan's expected risk corridor function as a share of total expected cost, $\gamma_{jm}(\delta, Q)$, and the total expected cost, $c_{jm}Q_{jm}$. The final part of the objective function is the risk charge. It is the product of the coefficient of risk charge, ρ and the plan's variance of total cost, $V_{jm}(\delta, Q)$. Here, I am implicitly assuming that the costs across plans are independent and hence the variances can be summed across the plans.[62] Lastly, I make the assumption that $\rho \geq 0$, that is I assume insurers are not risk seeking.

$$Q_{jm}(b) = \sum_t \theta_t M_t s^t_{jm}(b) \tag{1.14}$$

Equation (1.14) shows how the risk-adjusted demand is constructed. The risk-adjusted demand is the sum of demand across individuals of different risk types, where the demand gets adjusted for different risk-types via the scaling factor θ_t.

[62] In the model, the uncertainty in cost is coming from random draws of enrollees whose costs are independent of one another. While in theory, I could allow for a more flexible correlation in costs across enrollees and/or plans, this makes the inversion of FOC much more difficult. Furthermore it makes it computationally intractable due to i) non-linearity of the risk corridor function and ii) the cost shocks need to be integrated across individual plans that could be in the order of 100+ integrals for some large insurers. I'm currently working on an approximation method that could allow a more flexible correlation structure across plan costs.

M_t and s^t_{jm} are the market size and share function of consumer of risk type t, respectively. I allow six different risk types across individuals: fove different health levels (the same health levels used in demand) across regular enrollees, and a single type for the LIS enrollees.

The risk-adjusted demand reflects two things: selection on the cost side and CMS's risk adjustment on the revenue side. On the cost side, health insurance markets typically exhibit selection in which consumers' costs may vary across different types of individuals. To model this, I allow consumers' expected marginal cost to vary by their risk types: a consumer of risk type t has a marginal cost of $\theta_t c_{jm}$ if he/she enrolls in plan j in market m. So c_{jm} is the baseline marginal cost that corresponds to the insurer's expected marginal cost of an average risk enrollee. On the revenue side CMS risk-adjusts the plan's revenue by scaling the plan's bid by the average risk score of the plan.[63] So the plan receives $\theta_t b_{jm}$ in premiums for enrolling a consumer of risk type t where b_{jm} reflects the plan's bid for an average risk enrollee. The risk adjustment inflates (deflates) the premiums of plans that enroll observably sicker (healthier) enrollees.

While I allow for selection in the model, I assume that there is perfect risk adjustment similar to Curto *et al.* (2021).[64] This is reflected by using the same risk adjustment factor for both the marginal cost and the bid. While I could allow for imperfect risk adjustment, I assume perfect risk adjustment for the following reasons. First, the focus of the paper is on ex-post risk sharing i.e. risk sharing due to

[63]For details refer to section 1.2

[64]In the appendix, I detail the full model where I allow for imperfect risk adjustment.

unpredictable uncertainty not ex-ante predictable risk, as is the case for risk adjustment. Second, assuming perfect risk adjustment simplifies the model a great deal and allows me to more reliability estimate the supply-side model without running into numerical issues.[65]

Lastly, the final component of the objective function is the ex-ante margin constraint in (1.13). The margin constraint dictates that the total revenue of the insurer compared to its reported expected total cost can't exceed the firm specific maximum margin, $\overline{m}$. It's clear that as the insurer overestimates its cost (i.e. $\delta > 1$), its reported expected total cost will increase, relaxing the margin constraint.

I make the usual conduct assumption that insurers compete via Bertrand-Nash in prices.[66] Note that here the insurer's decision to misreport its expected cost, δ, does not affect other insurers' payoffs; only their pricing decisions (through the demand) affect other insurers' payoffs.

1.5.2.1. Identification/Estimation. Estimating the supply-side model is challenging for a variety of reasons. First, as pointed out by Decarolis *et al.* (2020a) LIS-benchmark plans or the "LIS-distorted" plans have a non-linear share function that makes the standard approach of inverting first order conditions difficult.[67] Second, there are more "unknowns" than the number of first order conditions I can derive from the conduct assumption. I resolve these issues by making some reasonable

[65] As I mention in the estimation section, I sometimes run into numerical issues when I try to solve for the FOC's due to the non-linearities in the FOC's. I find that this is especially worse when I allow for imperfect risk adjustment.

[66] While this is true for the most part, in section A.4.1, I detail how for some plans this is difficult to do due to how the subsidy is set for LIS enrollees.

[67] For more details, see section A.4.1

assumptions on the marginal costs, and using a combination of first order conditions and observed data.

Table 1.2. Supply-Side Parameters and Identification

Object	Description	Inference
s^t_{jm}	market share of consumer type t	demand estimates
$\frac{\partial s^t_{jm}}{\partial b_{km}}$	derivative of market share	demand estimates
b_{jm}	plan bids	data
$\overline{m}_f$	maximum allowed margin	data
θ_t	risk adjustment multiplier for consumer type t	data*
$\gamma_{jm}(\delta, Q)$	expected risk corridor payment share function	data/simulated**
$V_{jm}(\delta, Q)$	variance of plan total cost function	data/simulated**
c_{jm}	marginal cost	estimation via FOC
δ_f	degree of strategic cost reporting	estimation via FOC
ρ_f	coefficient of risk charge	estimation via cost moment

Notes: the table shows the list of variables/functions needed to evaluate the supply-side model, and specifies how each object is constructed. *This is inferred from using claims data of individuals across different risk types. **These functions are estimated by simulating the cost distribution of plans using the claims data.

Table 1.2 shows the list of variables used in the supply-side model and categorizes them as either coming from data, demand estimation, or parameters that are to be estimated. The share functions s_{jm} are separately estimated from the demand estimation, so we can treat them as known objects. The plan bids as well as the associated enrollee premiums are observed in the CMS's plan level data. The maximum allowed margin which is at the insurer-year level is taken as the insurers' non-Medicare business margin from the insurers' financial statements.[68] The

[68] I use a combination of financial statements data from NAIC as well as the CMS's Medical-Loss Ratio data to get the prior year margins of insurers' non-Medicare business as defined by CMS.

risk adjustment factor (or the cost multiplier) across different consumer types is inferred from the claims data.[69]

The functions $\gamma_{jm}(\delta, Q)$, the expected risk corridor payment share and $V_{jm}(\delta, Q)$, the variance of total cost are obtained using the claims data and the plan level attributes data. Using the sample distribution of the enrollee's claims cost (adjusted for plan specific cost sharing), I simulate the distribution of the plan's (claims) cost for different values of strategic cost reporting, δ and the plan's demand, Q. I then approximate the functions using a 2-dimensional spline method to get a smooth function of both variables. The full details of this process can be found in Appendix A.5.

Then, I am left with the main parameters of interest: the marginal cost vector, $\mathbf{c_{jm}}$, insurers' strategic cost reporting parameter, δ_f and the coefficient of risk charge ρ_f. For the marginal cost and the strategic cost reporting parameter, I can construct associated first order conditions to the objective in (1.12) derived in section A.4.2. As mentioned above, one challenge to this is that I can't use the FOC's of the plans that are "LIS-distorted". And unlike Decarolis *et al.* (2020a), my model's FOC's are highly non-linear with respect to the marginal costs and have cross-market ties (due to the margin constraint) that makes it hard to separately estimate the marginal costs of regular plans vs. the LIS-distorted plans.[70]

[69] I take the ratio of average costs of individuals of type t to the overage cost of the overall population. For more detail, see the Appendix A.5.2

[70] Decarolis *et al.* (2020a) invert the FOC's of regular plans to back out the marginal costs of those plans first. They then project these marginal costs onto observable characteristics and predict the marginal cost of LIS-distorted using this projection on observable characteristics.

Instead, I make the following restrictions on the marginal cost of the LIS-distorted plans vs. regular plans in the same market.

$$c_{j'm}^{LIS} = \frac{1 - AV_{j'm}}{1 - AV_{jm}} c_{jm}^{regular} \tag{1.15}$$

For each LIS-distorted plan j', I find a non-distorted plan j by the same insurer in the same market.[71] I then restrict the ratio between the costs to be the same as the insurer liable average share of enrollee's costs or one minus the actuarial value of the plan. Insurer will often offer two plans in the market that are vertically differentiated by the plans' cost sharing generosity or the actuarial value. As such, the above restriction seems to be a reasonable assumption.

For the coefficient of risk charge, I construct a set of cost moments at the firm-group year level:

$$\sum_{m} \sum_{J_m} w_{jm} c_{jm}(\rho) = \underbrace{\frac{\text{Total Cost}}{\text{Total \# Enrollees}}}_{\hat{c}} \tag{1.16}$$

where w_{jm} is the enrollment weight and $c_{jm}(\rho)$ is the model-implied marginal cost given a fixed value of ρ. The right hand side of the equation is the average cost of firm or firm-group observed in the data. The above identifies ρ by trying to match model-implied cost with the observed cost data or in other words by matching the implied margins with the observed ones. Suppose $\rho = 0$ i.e. insurers are risk neutral. If I find that the observed cost (relative to premiums) is much lower

[71] For some plans that don't have other non-distorted plans in the same market, I find a plan in the nearby market.

than the model-implied cost, then insurer likely incurs risk charges and prices accordingly. So ρ will have to be some value $\rho > 0$, and by matching the above moment ρ will be estimated in such a way.

I construct the above moments at the firm-year level for the larger insurers, but group some of the smaller insurers together.[72] I estimate all the parameters jointly using a constrained GMM.[73]

$$\min_{\rho} g'Wg \tag{1.17}$$

$$\text{s.t.} \quad FOC(\rho) = 0$$

$$\text{where} \quad g = \sum_{m} \sum_{J_m} w_{jm} c_{jm}(\rho) - \hat{c}$$

1.6. Model Estimates

1.6.1. Demand Estimates

Table 1.3 shows the demand estimates for regular enrollees. Most of the coefficients follow intuitive patterns. Healthier enrollees are more price sensitive, and higher income and older individuals are less price sensitive. The implied mean premium elasticity of the demand model is -4.13, and varies from -3.7 to -4.3 depending on the health level of enrollees. These seem economically reasonable estimates and

[72] This is because some of the smaller firms don't have enough enrollees and/or plans to reliably construct the above moment. However, this means that I can't estimate ρ at the firm level but at the firm-group level.

[73] While I could solve the minimization problem by a solver, I've often ran into different numerical issues where the solver had convergence issues. In practice, I create a grid of ρ values and find associated marginal cost vectors and δ values that make the FOC hold. I then select ρ among the grid of values that minimizes the GMM objective.

are similar in magnitude compared to the elasticities estimated in other papers (-5 to -13 in Decarolis et al 2020, -2 to -6 in Lucarelli et al 2012, and -5 to -6.3 in Starc and Town 2015).[74]

Non-price coefficients also follow intuitive patterns. Healthier consumers are less likely to purchase drug coverage through the PDP market. This may be from healthier enrollees opting to enrol in MA-PD plans or that they choose not to have any drug coverage through Part D.[75] Consumers dislike higher deductibles and derive positive utility from plan generosity: they prefer enhanced plans that have higher actuarial value, likes having extra coverage in the gap and like having more drugs being covered in their plans. Lastly, the coefficient on plan age is positive and significant, meaning that existing plans are more likely to capture a larger pool of beneficiaries. The coefficient is smaller for higher income consumers, meaning they are less likely to stick with existing plans.

The demand estimates for LIS enrollees (Panel B) also show similar patterns in demand. I exclude many of the plan characteristics as LIS enrollees face little variation in those attributes due to increased cost sharing. The price coefficient for LIS enrollees is also negative and significant and follow similar patterns across age bins: younger consumers are more price sensitive. I also find positive and

[74]These papers estimate demand in the first few years of the PDP market whereas I estimate demand using more recent years.

[75]This is in line with the findings that healthier enrollees are more likely to enroll in Medicare-Advantage plan. In our model given the outside option includes MA-PDs, it may be that healthier enrollees are more likely to enroll in MA-PDs.

significant coefficient on plan age, meaning similar to the regular enrollees, LIS enrollees are more likely to stick with existing plans.

Table 1.3. Demand Estimates

	A. Regular Enrollees									B. LIS Enrollees		
	Mean Utility	Demographic Interactions								Mean Utility	Demographic Interactions	
		Health				Income		Age			Age	
	β_0	Fair	Good	VeryGood	Excellent	Medium	High	< 65	> 80		< 65	> 80
Premium ($\$000s$)	-5.95	-0.40	-0.95	-1.24	-0.97	0.39	0.85	-0.44	0.28	-4.81	-0.39	0.14
s.e.	(0.36)	(0.41)	(0.39)	(0.39)	(0.43)	(0.20)	(0.26)	(0.42)	(0.22)	(0.15)	(0.28)	(0.35)
Constant		-0.10	0.44	0.63	0.76	0.05	0.14	-0.82	-0.63		0.75	-0.10
s.e.		(0.33)	(0.30)	(0.30)	(0.31)	(0.14)	(0.18)	(0.28)	(0.16)		(0.11)	(0.16)
Deductible ($\$000s$)	-2.42	0.39	-0.40	-0.24	0.18	-0.18	-0.54	0.35	-0.70			
s.e.	(0.59)	(0.66)	(0.62)	(0.62)	(0.65)	(0.29)	(0.37)	(0.51)	(0.33)			
Enhanced	1.82	-0.10	-0.13	-0.23	-0.26	0.13	-0.09	-0.14	0.01			
s.e.	(0.17)	(0.19)	(0.18)	(0.18)	(0.19)	(0.08)	(0.10)	(0.16)	(0.09)			
Extra Coverage-Gap	0.88	0.24	0.25	0.18	0.20	-0.02	-0.04	-0.18	0.04			
s.e.	(0.25)	(0.29)	(0.27)	(0.27)	(0.30)	(0.14)	(0.17)	(0.31)	(0.14)			
No. of Drugs Covered	0.12	-0.06	-0.04	-0.09	-0.21	-0.08	-0.08	0.22	0.10	0.12	-0.03	-0.11
s.e.	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.03)	(0.04)	(0.05)	(0.03)	(0.02)	(0.03)	(0.04)
Plan Age	0.32	0.02	-0.01	-0.00	-0.05	-0.06	-0.08	0.04	0.09	0.30	0.01	0.12
s.e.	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)

Notes: the table shows demand estimates for regular and low-income subsidy (LIS) enrollees. Many of the product charactersitics for LIS enrollees are excluded as they face identical cost-sharing charactersitics like deductible across plans.

1.6.2. Supply-Side Estimates

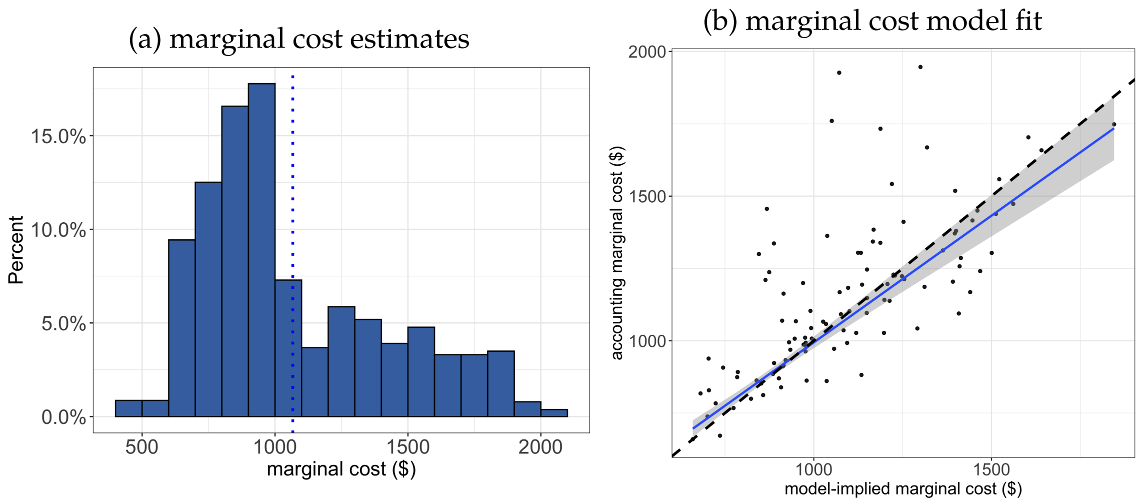
Expected Marginal Cost, c_{jm}. Figure 1.15a shows the distribution of the expected marginal cost estimates c_{jm}'s. The marginal cost is centered around $\$1,066$, but with a large variance (standard deviation of $\$354$). Much of this variation comes from the variation in plans' cost sharing characteristics. For example, plans with the standard benefit design have a mean marginal cost of $\$851$ where as plans with enhanced benefit design have a mean marginal cost of $\$1,271$. I also project the estimated plan level marginal costs onto observable plan characteristics.[76] I find intuitive patterns: higher deductible is associated with lower marginal costs, higher cost sharing (e.g. extra coverage gap, enhanced plan benefit design) is associated with higher marginal costs.

To asses whether these are reasonable marginal cost estimates, I compare the marginal cost estimates with the observed accounting cost data from the insurers' financial statements, which is shown in figure 1.15b. It shows that the estimated expected marginal cost closely follows the observed data. While there are some observations that are further from the 45-degree line, these are driven by small insurers whose costs will vary more from year to year. The enrollment-weighted average marginal cost is estimated to be $\$883$ vs. observed cost of $\$885$, suggesting that the marginal cost estimates are reasonable.[77]

[76] See Table A.2 for the results of this regression.

[77] Part of this will be "mechanical", since our estimation relies on matching the model-implied cost with the observed data. However, it doesn't guarantee that the costs will be exactly the same since for most firms, moments are aggregated across the firms.

Figure 1.14. Marginal Cost Estimates



Panel (a) plots the distribution of the marginal cost estimates, c_{jm}. Each observation is plan-year. Panel (b) plots the estimated marginal costs vs. observed accounting cost data at the firm-year level. For the model, the firm level marginal costs are computed by taking the enrollment-weighted average across all the firm's plans. The dashed line indicates the 45-degree line, and the blue line shows the best-fit line where the observations are weighted by enrollment.

The given marginal cost estimates imply that firms' implied margin is around 12.3% on average (vs. 13% observed in the data), which is much higher than the 7 percent estimated in Decarolis *et al.* (2020a). This may be because i) I use data from much later years of the program, and/or ii) because in my supply model, I endogenize the effect of strategic cost reporting on the insurers' risk corridor payments.[78] When I follow Decarolis *et al.* (2020a)'s approach to estimating the marginal cost, I get an enrollment-weighted average marginal cost estimate of $\$917$, much higher

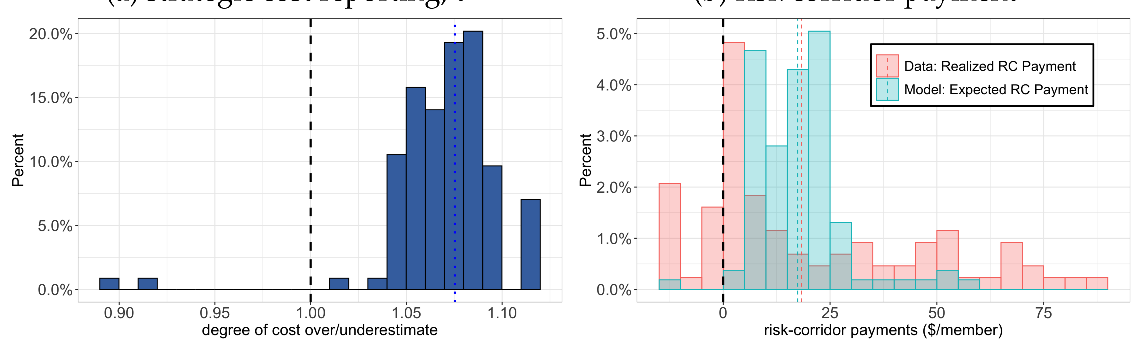
[78]Over the years, Medicare Part D market has become more concentrated. So the higher margin that I estimate may in part be reflecting higher market power that the insurers have. It could also be because I endogenize the strategic cost reporting and its effect on the two sets of regulations in the market. It's unclear in which direction the estimates will be

than what I estimate and what is observed in the data.[79] This higher marginal cost estimate implies 8.8% margin, suggesting a model that does not endogenize the strategic cost reporting and the insurers' risk charges may lead to biased marginal cost estimates.

Strategic Cost Reporting Parameter, δ. Figure 1.17a shows the distribution of firm's strategic cost reporting parameter, δ. Consistent with observed risk corridor payment patterns, I find that the insurers overwhelmingly overestimate their costs. On average, the insurers overestimated their costs by 7.5 percent. But this varies from an insurer underestimating costs by 10% to an insurer overestimating costs by 12%.

Figure 1.16. Estimated degree of strategic cost reporting, δ and implied risk corridor payments

(a) strategic cost reporting, δ (b) risk corridor payment



Panel (a) plots the distribution of the strategic cost reporting parameter, $\delta_{f,t}$. Each observation is firm-year. The dashed line indicates the mean of the distribution. Panel (b) plots the distribution of estimated expected per-enrollee risk corridor payments and observed per-enrollee risk corridor payments. Each observation is firm-year. The dashed lines indicate the mean of the distribution, respectively.

[79]In figure A.9, I assess the fit of these marginal cost estimates similar to figure 1.15b, which shows that the Decarolis *et al.* (2020a)'s approach of marginal cost estimates may lead to biased estimates.

To see whether these strategic cost reporting parameters are reasonable, I look at the expected risk corridor payments implied by the insurers' strategic cost reporting behaviors. 1.17b shows this model-implied expected risk corridor payments vs. the observed risk corridor payments by the insurers. While the two don't align perfectly, the two distributions are centered very closely to one another suggesting the model can explain the skewed distribution of the observed risk corridor payments.[80]

Coefficient of Risk Charge, ρ. Figure 1.19a shows the distribution of the coefficient of risk charge, normalized to the variance of an average enrollee. On average, the insurers face $\$17.5$ of risk charge for enrolling an additional average enrollee, however there is quite a bit of variation here as well. A large number of insurers are estimated to be risk neutral where ρ is close to zero where as some insurers have ρ implying $\$80$ of risk charge. To put the magnitude of these risk charge coefficients in perspective, on average the insurers' risk charges are around 2 percent of their expected marginal costs.[81] This magnitude is in-line with actuarial documents that suggest that insurers' risk charges are usually $2-4\%$ of their premiums (American Academy of Actuaries, 2020).

While looking at the risk charge coefficient shows the risk averseness of insurers, it doesn't show the actual risk charges that the insurers face in the current market. The insurers have risk sharing arrangements with the government through the

[80]The model implied risk corridor payment is in expectation i.e. without any shocks to the costs. As a result, we would expect the realized risk corridor payments to be noisier and have larger variance than the model-implied expected risk corridor payments.

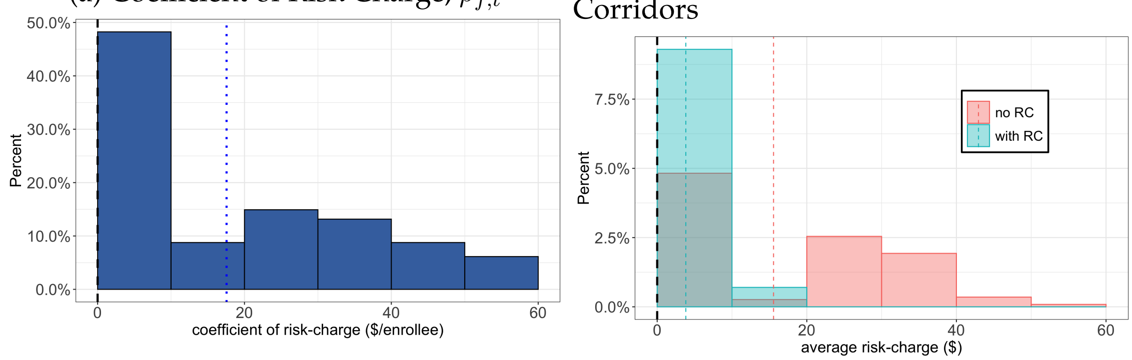
[81]I find that this ranges from 0-5 percent, depending on the insurer.

risk corridors and so will face lower risk levels. Figure 1.19b shows the realized risk charges that the insurers face with the risk corridors and compare it with the risk charges without the risk corridors. The insurers face significantly smaller risk charges with the risk corridors. With risk corridors, the insurers face average risk charge of $\$4.1$ vs. risk charge of $\$17.5$ or 76% reduction in risk charges. The large reduction in risk charges is likely amplified due to the strategic cost reporting of the insurers as shown above. Recall in section 1.3, the risk level (variance of cost) that the insurers face decreases regardless of which direction the insurers misreport their costs. And here because the insurers have overestimated their costs, their risk level is significantly reduced, lowering their risk charges.

Figure 1.18. Estimates of Coefficient of Risk Charge, $\rho_{f,t}$ and Average Risk Charge

(a) Coefficient of Risk Charge, $\rho_{f,t}$

(b) Average Risk Charge w vs. w/o Risk Corridors



Panel (a) plots the distribution of the coefficient of risk charge estimates, ρ. Each observation is firm-year. ρ is normalized to the variance of average enrollee's cost i.e. the normalized ρ represents the insurer's risk charge of enrolling an additional average enrollee. Panel (b) plots the distribution of average risk charge with risk corridors (i.e. the current risk charge level) vs. average risk charge w/o any risk corridors. Observation is at the plan-year level.

To look at heterogeneity of the coefficient of risk charge parameter across the insurers, I investigate if ρ is correlated with the insurer characteristics, especially insurer size. Columns 1 & 2 of table 1.4 show that smaller (larger) insurers tend to have lower (higher) ρ and therefore are less (more) "risk averse". However, the results are not statistically significant in part due to the lack of observations I have.[82]

Column 3 shows that ρ is negatively correlated with the RBC ratio of the insurers, meaning more financially solvent insurers have lower risk charge coefficient and therefore less "risk averse" albeit not statistically significant. However if I control for insurer fixed effects, ρ is negatively correlated with the insurer's prior year RBC ratio with statistical significance, which is shown in column 4. So if the insurer has higher RBC ratio (i.e. more financially solvent) in a given year, the lower the ρ or less "risk averse" the insurer will be in that year. To interpret the magnitude, insurers' average standard deviation of RBC ratio is 1.3, meaning that one standard deviation increase in RBC ratio is correlated with $\$5.7$ decrease in the coefficient of risk charge. This suggests that "risk averse" behavior of the insurers may be coming from financial/regulatory frictions that the insurers face (Koijen & Yogo, 2022a).

[82] This is because, for smaller firms I have to group them together and estimate a single "average" ρ for the group. As a result, while there may be more heterogeneity even among the smaller firms I am only able to estimate the average ρ for the group and so it's unclear if the above relationship will hold if I am able to observe ρ at the firm level for the smaller insurers.

Table 1.4. Cofficient of Risk Charge vs. Insurer Characteristics

	Dependent variable:			
	$\rho_{f,t}$			
	(1)	(2)	(3)	(4)
$\mathbb{1}\{\text{small firm}\}$	−2.84			
	(9.26)			
log (enrollment$_{f,t}$)		2.91		
		(2.37)		
RBC-Ratio$_{f,t-1}$			−0.81	−4.30*
			(1.30)	(2.26)
Year FE	Y	Y	Y	Y
Firm FE	N	N	N	Y
Observations	38	38	38	38
R^2	0.229	0.261	0.237	0.678

Note: *p<0.1; **p<0.05; ***p<0.01

The table shows results from regressing the coefficient of risk-charge estimates of firm (group) f in year t on firm charactersitics. ρ_{ft} is normalized to the variance of an average enrollee i.e. ρ indicates the risk-charge of insurer for enrolling an additional average enrollee. Small-firm is an indicator for firms that have $< 50,000$ enrollees. RBC-Ratio$_{f,t-1}$ is the firm f's prior year RBC-ratio. For firm-groups, the RBC-ratio represents enrollment weighted average of RBC-ratios across the firms within the group.

1.7. Alternative Market Designs

Given the structural model estimates, I run several counterfactuals to understand the effects of the two regulations: risk corridors and margin regulation. However there are several challenges to this, so I make a few abstractions.

First, as pointed out before the subsidy design of the low-income subsidy eligible consumers make it difficult to model the insurers' pricing behavior for these consumers. I instead restrict my attention to the regular enrollees and model the insurers pricing optimally targeting these consumers only.[83] Second, the outside option in my model involves a bundle of options for the consumers; opting to not purchase any drug plan or opting for an MA-PD plan. While some of the changes I make to the PDP market may also impact the MA-PD plans, I assume that these markets are separate.[84] This also implies that I assume the outside option will remain fixed throughout my counterfactual results. So the counterfactuals can be seen as a partial equilibrium setting in which I hold everything else constant and only look at changes in the PDP market.

Another challenge is in evaluating the welfare, in particular the large government spending in consumer subsidies. Throughout the counterfactuals, I keep the overall government expenditure on enrollee subsidy fixed i.e. I adjust the PDP subsidy level so that the total government subsidy expenditure is held constant through out my counterfactuals. This allows me to isolate the welfare effects on the consumer surplus and the insurer profits, as well as any changes in government spending due to the risk corridor payments but not the subsidy expenditure. Lastly, I restrict my counterfactuals to 2015.[85]

To see how the current set of regulations affect the market, I compare two alternative market designs relative to the baseline (i.e. the status quo).[86] First, I remove

[83]The regular enrollees account for a little over 60 percent of the total consumers.

[84]Most notably, the overall consumer subsidy level in Part D is computed using the bids of MA-PDs and the PDP plans. Treating these markets separately could be seen as the government severing the ties between these two markets by computing separate subsidy levels or benchmarks for each market.

[85]This is for simplicity. But in theory, I could run it for all other years in which I have the demand/supply-side model estimates.

[86]Because I restrict my attention to regular consumers I recompute the optimal bids under the baseline market design. The resulting bids remain similar to the observed bids in the data.

both sets of regulations and allow the insurers to optimally set prices in the absence of these regulations. I refer to this counterfactual as “no regulation”. Second, I include both sets of regulations but ban the insurers from strategically reporting their cost. In practice, this could reflect the government having full set of information that the insurers has in which there is no asymmetric information regarding costs. I refer to this counterfactual as “truthful reporting”. Lastly, I look at changing the design of risk corridors to a linear risk sharing rule to study the effects of different risk sharing levels. I vary linear risk sharing from no risk sharing (fixed price contract) to full risk sharing (cost plus).

1.7.1. Removal of Regulations

Table 1.5 shows market summary statistics across different counterfactuals. Column 1 shows the results for the baseline (or current) market. Column 2 shows the counterfacutal results of removing both risk corridors and margin regulation in the market. Allowing the insurers to freely choose prices without any constraints lead to higher prices of $\$1,014$ vs. $\$963$ (or 5.2%). This leads to 8 percent decrease in enrollment, and 10 percent decrease in consumer surplus. The increase in prices is in part from the increased risk level that the insurers face from no longer having risk sharing through the risk corridors. This is reflected in the average variance of cost in the baseline, which is only 17.7 percent of the level faced without any regulation, translating to 86 percent decrease in the average marginal risk charge.

Table 1.5. Counterfactual Comparisons with Baseline

	(1) Baseline	(2) No Regulation	(3) No Regulation w baseline risk	(4) Truthful Reporting $\delta = 1$
Average Bid ($)	963	$1,013.8$	$1,010.2$	864.4
Avg Marginal Risk Charge ($)	1.3	9.9	1.4	8.6
Avg Variance of Cost (%)	17.6	100	17.8	80.9
Enrollment (M)	11.1	10.2	10.4	12.5
Consumer Surplus ($M)	$2,145.3$	$1,927.9$	$1,964$	$2,470.8$
Insurer Profit ($M)	955.1	$1,508.5$	$1,444.7$	375.6
Risk Corridor Payment ($M)	136.2	0	0	0
Total Risk Charge ($M)	17.7	100.7	17.7	109.4
Total Welfare ($M)	$3,236.6$	$3,436.4$	$3,408.7$	$2,846.4$
Total Welfare w Risk Charge ($M)	$3,218.9$	$3,335.7$	$3,391$	$2,737$

Notes: the table shows various market level statistics for different counterfactuals. Column (1) shows the baseline or status-quo market. Column (2) shows counterfactual in which both risk corridors and margin regulations are removed. Column (3) shows column (2) but with the insurers' risk level reduced to the baseline level. Column (4) shows the "truthful reporting" where both regulations exist but the insurers are banned from strategically report their costs. All averages are computed using enrollment-weighted average. Avg variance of cost refers to percent of variance relative to risk level w/o any risk sharing. Risk corridor payment are payment from the insurers to the government (i.e. positive number indicates government is receiving payment from the insurers). Insurer profit equals to total revenue minus total expected cost minus the expected risk corridor payment to the government, but excludes the risk charges. Total welfare is the sum of consumer surplus, insurer profit and the risk corridor payments. Total welfare with risk charge is the total welfare numbers subtracting the total risk charge numbers.

The higher prices significantly increase the insurer profit, increasing it by 57 percent. Note that the insurer profit is smaller in the baseline, not just because of the lower prices but also from the expected risk corridor payments the insurers make to the government due to overestimating their costs. Without the risk corridor payments the insurers' profits would be 14 percent higher in the baseline, which would bring the baseline insurer profit to be within 38% of the insurer profit with no regulation. So here the risk corridor payments act as an ex-post transfer mechanism that brings down the insurers' profits.

The total welfare, which is the sum of consumer surplus, insurer profit and government earnings (via risk corridor payments) increase by 3 percent in the absence of any regulation. This reflects the large increase in the insurer profit relative to the modicum decrease in consumer surplus. The results are similar when we include the total risk charges in the welfare measure.

While comparing the baseline with no regulation counterfactual is informative, it shows combined effect of two things. First, it shows the removal of the margin constraint that allows the insurer to freely charge higher prices. Second, it also shows the removal of risk sharing arrangements via risk corridors, increasing the overall risk level that the insurers face. To decompose these two effects, I run a modified "no regulation" counterfactual in which I lower the insurers' risk levels to the same level that the insurers face in the baseline. The results are shown in column 3.[87]

The modified no regulation counterfactual shows that the prices still increase significantly, increasing by 4.9 percent vs. the 5.2 percent in the initial no regulation counterfactual. Other numbers remain at similar levels, meaning the removal of margin regulation dominates any changes brought by the removal of the risk sharing. This is because the magnitude of risk charge is relatively small even without any risk sharing arrangements.

[87] I take the estimated degree of cost reporting, δ and assume the insurers face risk level of $V(\delta, Q)$ with risk corridors vs. facing the full risk level w/o any risk corridors.

1.7.2. Truthful Reporting: No Strategic Cost Reporting

To better understand the effect of the insurers' strategic cost reporting on the market, I run a counterfactual where I ban the insurers from strategically reporting their costs. I impose the insurers' strategic cost reporting parameter, δ to be one for all the insurers while facing both sets of regulations. Column 4 of table 1.5 presents the results. The average price decreases from $\$963$ to $\$864$ (or 10.3%). The lower prices lead to 15.1 percent increase in consumer surplus, but 60 percent decrease in the insurer profit, resulting in 12 percent lower total welfare relative to the baseline.

Under no strategic cost reporting, the insurers' risk level is higher than the baseline. While in the baseline, the insurers' average variance of cost is 17% (relative to the level without any risk sharing), under the truthful reporting case, the insurers' risk level is 80%. This is due to the insurer on average having overestimated their costs in the baseline. Recall from section 1.3 that when the insurers under or overestimate their costs, not only do their expected risk corridor payments change, but their variance of cost change as well. And this is due to the non-linearity in the risk corridors function.[88]

To look at the heterogeneity in risk sharing across the insurers, I look at how the reduction in variance varies by insurer size. Figure 1.21a plots these for baseline and the counterfactual in which I disallow strategic cost reporting. It shows that

[88]When the insurers over or underestimate their costs, they also increase the probability that they trigger risk sharing payments (or reimbursements). So this decreases the variance of their overall cost.

overall level of variance is lower for most insurers in the baseline vs. the truthful reporting case, in line with the summary results in table 1.5. However, it shows that in the baseline, the variance of cost is reduced more for larger insurers compared to smaller insurers. On the other hand, when there is no strategic cost reporting the opposite is true. The variance of cost is reduced more for smaller insurers compared to larger insurers. This flipped relationship between insurer size and variance of cost is driven from larger insurers overestimating their costs more as shown in figure 1.21b. So risk corridors in the absence of strategic cost reporting is intended to reduce the risk that smaller insurers face more than the larger insurers. But due to the insurers' strategic cost reporting, in the current market larger insurers' risk is reduced more than the smaller ones. And the overall risk level that the insurers face is smaller.

If I take the truthful reporting as indicative of the government's policy goals, there are two main takeaways. One is that the government wants to combat market power by severely constraining the margin that the insurers can charge, decreasing the equilibrium prices and therefore increasing consumer surplus. Second is that the government wants to limit some but not all risk that the insurers face. Furthermore, the government wants to protect the smaller insurers against variation in cost more so than the larger insurers. And these seem to be consistent with what's stated in the initial policy goals of the regulations.

Figure 1.20. Enrollment vs. Variance of Cost, and δ across Insurers

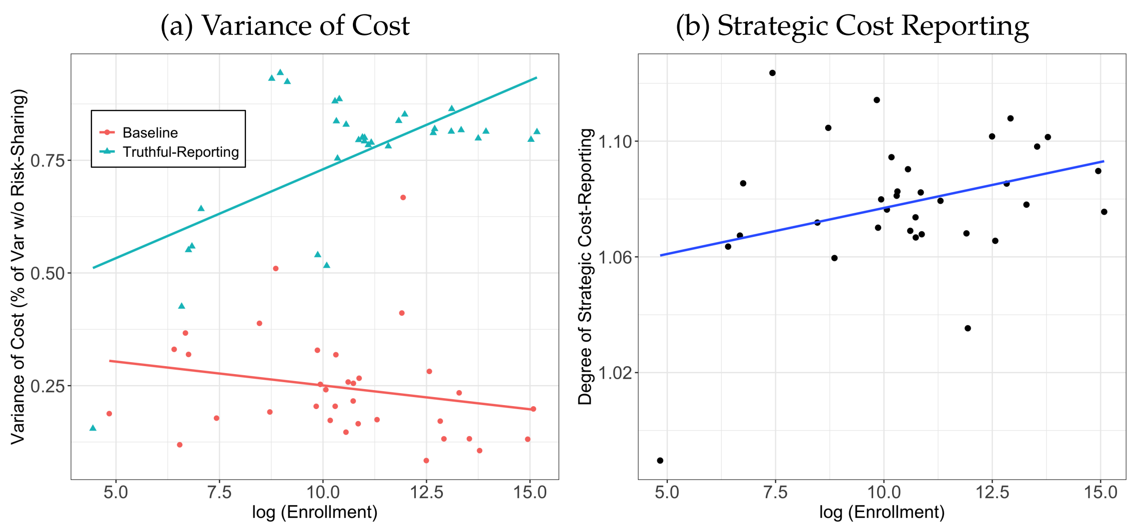


Figure plots the log enrollment numbers against variance of cost under baseline vs. truthful reporting regulations. The variance of cost is shown as a percentage of the variance of cost w/o any risk sharing regulations. Each observation is an insurer.

1.7.3. Linear Risk Sharing Rule

Here, I modify the design of risk corridors to a linear risk sharing rule. I change the ex-post risk corridor function to be:

$$T(\tilde{C}, \delta C) = \alpha(\delta C - \tilde{C}) \tag{1.18}$$

The risk sharing parameter, α governs the degree of risk sharing where $\alpha = 0$ implies no risk sharing (i.e. fixed-price contract) and $\alpha = 1$ implies full risk sharing (i.e. cost reimbursement contract). I still allow the insurers to strategically report their cost via the parameter, δ that shifts their reported expected cost. I also assume the government keeps the existing margin regulation, in which it constrains the price the insurers can charge relative to their reported expected cost.

Figure 1.22. Risk Corridor Function: Baseline vs. Linear Risk Sharing

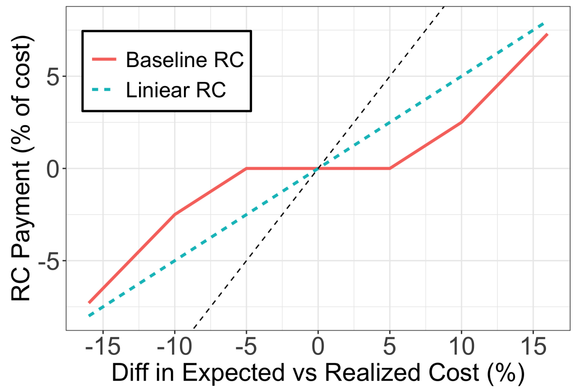
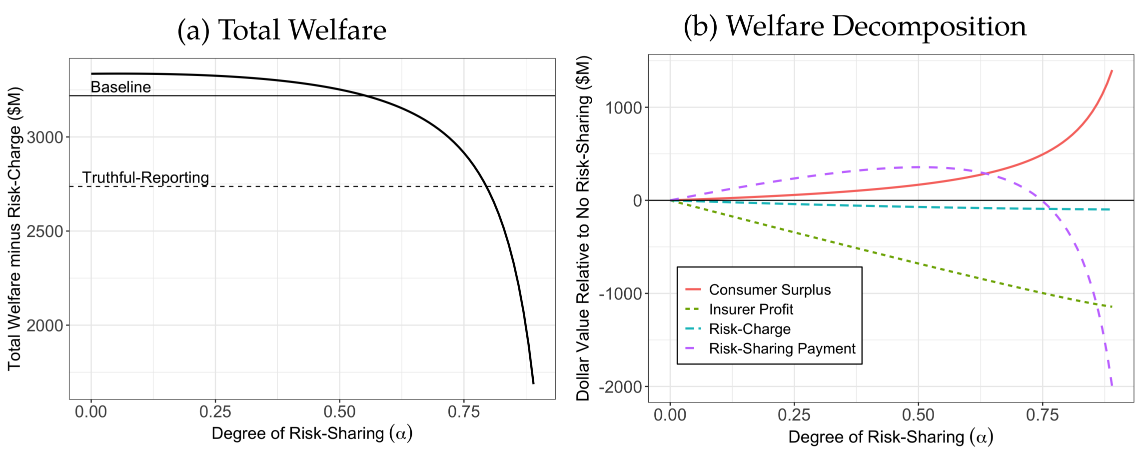


Figure plots the baseline risk corridor function vs. linear risk sharing function explored in the counterfactual.

The main difference between the linear risk sharing vs. the baseline risk corridor function is both the continuity and linearity of the transfer function with respect to the expected and realized cost. Figure 1.22 illustrates this. This means that with linear risk sharing, the insurers' strategic cost reporting will have no effect on the variance of the insurers' cost. It will be governed by the risk sharing parameter α. The specific results are derived in appendix A.6. So while the insurers' strategic cost reporting will still change the expected risk corridor payments, their variance of cost will not be affected by δ but only by the risk sharing parameter α.

Figure 1.24 shows that similar to the results found in Table 1.5, total welfare decreases as we increase risk sharing. However, the effect varies widely across different components of the welfare. Figure 1.25b shows that as risk sharing increases consumer surplus increases, but insurer profit decreases by even more. And as expected more risk sharing leads to decreased risk charges as the insurers' variance of cost is decreased. However this decrease in magnitude is very small relative to other measures, meaning the direct effect of increased risk sharing on the insurers' risk level is dominated by the indirect effect of limiting the ex-post profit of the insurers. This can be seen by the positive and initially increasing expected risk corridor payments, transferring part of the insurers' profits to the government. But as risk sharing increases further, the expected risk corridor payments decrease and turns negative, meaning the insurers are receiving payments from the government. So at higher levels of risk sharing, the risk sharing payment is acting as an indirect supply-side subsidy.

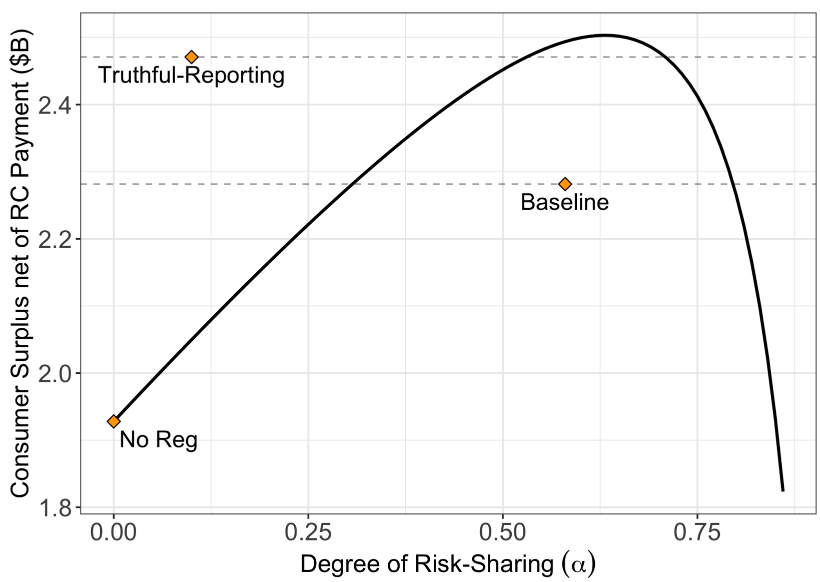
Figure 1.24. Welfare vs. Degree of Risk Sharing, α



Panel (a) plots the total welfare which is the sum of consuer surplus, expected insurer profit minus total risk charge, and risk corridor payment vs. degree of risk sharing, α. Panel (b) decomposes the total welfare into individual components. Insurer profit is the sum of total revenue minus the total expected cost net of any expected risk corridor payments, minus the total risk charges. All numbers are shown relative to when $\alpha = 0$.

Given that Medicare is a social insurance program, and the government's usage of regulations like profit margin regulation the government may be more interested in maximizing consumer surplus. Figure 1.26 shows what happens to the consumer surplus net of government expenditure on risk sharing payments and compare the values relative to other counterfactual benchmarks in table 1.5. It shows that with $\alpha = 0.64$, the total surplus measured by the sum of consumer surplus and government earnings can be maximized. In fact, this "optimal" level is just above the level of the truthful reporting case.

Figure 1.26. Consumer Surplus net of RC Payment vs. Degree of Risk Sharing, α



The figure plots the sum of consumer surplus and risk corridor payments to the government for varying degrees of linear risk sharing, α. The orange points denotes where different counterfactual scenarios. For example, the status quo shows

While choosing a risk sharing level that yields high total surplus may be ideal, it may also over protect the insurers. Figure 1.26 shows that the truthful reporting case is comparable to relatively low level of risk sharing. Although not modeled in this paper, over protecting the insurers may decrease the insurers' incentive to contain their costs (Cicala *et al.* , 2019). So here, there's a trade-off between over-insuring the insurers vs. achieving high levels of surplus. While the linear risk sharing rule won't be able to achieve comparable surplus levels at the truthful reporting risk level, it can still improve on the baseline policy. The insurers' risk level in the baseline policy is comparable to $\alpha = 0.58$. At this risk sharing level, similar total surplus levels as the truthful reporting case can be achieved. So changing the baseline risk corridor to a linear risk sharing rule with $\alpha \in [0.3, 0.58]$ can yield higher total surplus while not lowering the insurers' risk levels any further.

1.8. Conclusion

I study how insurers' strategic responses to regulations can distort the intended purpose of both the risk corridors and margin regulation in Medicare Part D. Both regulations use the insurers' self-reported cost estimate where insurers have a strategic incentive to over or underestimate their costs to increase their revenue. However insurers have conflicting incentives to misreport under each regulation. Under risk corridors, insurers want to underestimate to receive payments. Under margin regulation, insurers want to overestimate to charge higher prices. Having both will have a balancing effect. Using a structural model, I estimate that insurers have overestimated their costs by 8% on average. I find that insurers are not that risk averse and so the impact that risk corridors can have as a risk sharing policy may be limited in the current market. Instead, risk corridors act more as an ex-post penalty function for insurers that overestimate their costs. Risk corridors therefore help enforce the margin regulation, keeping the prices lower than without the regulation. Given the findings, I propose a linear risk sharing function to replace the current risk corridors, which increases total surplus while maintaining the same level of risk for insurers.

Neither regulation is unique to Medicare Part D and they are widely used in other publicly-funded health insurance markets, such as Medicaid and ACA exchanges. This is especially true for risk corridors. During the heightened uncertainty brought on by COVID, Congress discussed implementing risk corridors at a national level, affecting all health insurance markets.[89] As such, ensuring careful design of these policies without causing other distortions is crucial, especially when they are being implemented in a much broader scope.

More generally, these findings highlight two challenges that the government should consider in designing regulations for private firms. One is carefully examining private firms' incentives and determining whether those are aligned with the government's objectives. But more importantly, government and researchers alike should also examine the interaction of different policies in the market. While studying a single policy in isolation may be valid in certain settings, markets are

[89]The Heroes Act, passed by the House of Representatives on May 15, 2020 included provision that would establish risk corridor program to stabilize premiums for the individual and commercial markets as well as the Medicare Advantage markets, essentially covering most if not all health insurance markets.

often laden with several different regulations. Failing to account for the interaction between regulations may have unintended consequences in the market, and in some cases bring more harm than good.

CHAPTER 2

Pass-Through of Reinsurance vs Consumer Subsidy: Evidence from the Individual Health Insurance Market[1]

2.1. Introduction

Government subsidies play an important role in many markets today. These subsidies span many industries ranging from agriculture, automobile to the housing market. In these markets, governments often have a choice in how they implement such subsidies. They could either employ demand-side subsidy that directly subsidizes consumers for purchasing goods or supply-side subsidy that subsidizes the cost of production. Both types of subsidies ultimately seeks to increase access to certain goods & services by lowering the consumer prices. This is particularly pertinent in the US health insurance market, where government plays a huge role in providing subsidies; in 2021, the US government spent close to $920 billion in net federal health insurance subsidies.[2] A natural question rises: which form of subsidy is more effective and efficient for the government?

[1]This chapter is joint work with Anran Li. We would like to thank Vivek Bhattacharya, David Dranove, Gaston Illanes, Rob Porter for helpful comments. We thank seminar participants at Northwestern University for valuable suggestions. We thank Frank Limbrock for assistance with data application and access. We thank the Kellogg Research Support team for their assistance with data storage and management. We gratefully acknowledge financial support from the Graduate School Research Grant and the Center of Industrial Organization at Northwestern University.

[2]https://www.cbo.gov/publication/56571

In this paper, we study the two different forms of subsidy mechanisms in the ACA individual health insurance market. Since its inception the individual health insurance market has faced many issues including but not limited to: insurer exits, huge premium hikes and various uncertainty in federal policies. As a result, individual states have increasingly employed their own policies to stabilize the market. One such policy that is being increasingly adopted is reinsurance subsidy that subsidizes the health insurers ex-post for incurring high-cost enrollees. However, there is some debate among the policymakers on the effectiveness of such supply-side subsidy vs. existing direct-to-consumer demand-side subsidy. We study both sets of policies and determine which form of subsidy may be more effective in increasing access to health insurance. In particular, we show that with adverse selection and the existence of financial frictions for the insurers, the answer depends on the primitives of both the demand-side and the supply side.

To illustrate the role of both demand-side and supply-side forces in the efficiency of the two subsidy mechanisms, we begin by building a theoretical model of the demand with selection and supply in which insurers face frictions for taking on risk. In a stylized risk averse monopoly insurer model, we incorporate the cost of financial frictions, resulting in insurers charging higher prices for taking on larger risk. We then examine both the demand-side subsidy that directly subsidizes the enrollee at the time of purchase, and the reinsurance subsidy that ex-post subsidizes a portion of the costs that the insurance company incurs. Using the model, we study their effects on the market equilibrium as well as comparative statics on which subsidy mechanism may be more effective for the government.

Consumer subsidy directly subsidizes the price of the insurance, thereby boosting the demand and lowering the equilibrium price. Reinsurance subsidy has two different effects: i) direct effect of lowering the expected cost of insurer and ii) indirect effect of providing insurance for the insurers thus lowering the cost of financial frictions. These two effects works together to lower the effective marginal cost of the insurer and thus equilibrium price. Because of reinsurance's indirect effect of lowering insurer risk, the pass-through of reinsurance subsidy can be larger than one even in a monopoly market. In comparing the two subsidy mechanisms our theretical model predicts that in a market with just financial frictions (i.e. without any selection), reinsurance subsidy is more efficient for the government. That is for a dollar spent on subsidy, the reinsurance subsidy will haver greater impact on price decreas. However, when there is adverse selection in which marginal enrollee's cost is lower than the average enrollee's cost, it is ambiguous as to which form of subsidy mechanism would be more efficient for the government. In such a setting, the answer will depend on two key model primitives: the degree of adverse selection and the degree of risk aversion (or indirect cost of financial frictions).

We then present empirical evidence of both frictions in the market. First, we document that many primary health insurers purchase private reinsurance policies despite having to pay high markups. We find that smaller, more financially constrained and non-profit insurers are more likely to purchase private reinsurance policies, suggesting such behavior is driven by financial/regulatory frictions, similar to those of banks. Second, using event study of state-level government reinsurance programs we show that reinsurance decreases premiums significantly, with some evidence of pass-through of greater than one. Furthermore, we find that premium decreases are larger for insurers that may face greater financial frictions, suggesting financial frictions do affect insurer pricing decisions. Additionally, premium decreases are larger for plans with higher actuarial value, suggesting there is adverse selection in the market in which sicker enrollees select into plans with more generous plan benefit designs. Lastly, we find evidence of government reinsurance decreasing the insurers' purchasing of private reinsurance on both the intensive and extensive margins.

Related Literature

Our paper contributes to several strands of literature. First, our paper adds on to studies on financial/regulatory frictions of insurers. Majority of this work (Koijen & Yogo, 2015a, 2016b, 2022b) focuses on studying financial/regulatory frictions that life insurance companies face and how such frictions may play a significant role in pricing of insurance contracts. On the health insurance side, Kim (2022) estimates a model of risk averse insurers to study the role of government's risk sharing policy. Our paper takes a similar approach, but offers potential improvement by leveraging novel data on insurers' private reinsurance purchasing behavior.

Second as our paper studies reinsurance policy it is closely related to papers that study risk sharing policies in the health insurance markets. A large number of these papers (Brown *et al.* , 2014a; Layton *et al.* , 2018; Layton, 2017) study ex-ante risk adjustment that adjusts insurer payments according to health risk profile of its enrollees. On the ex-post risk sharing side, a small number studies risk corridors (Layton *et al.* , 2016b; Sacks *et al.* , 2021b; Holmes, 2021), as well as reinsurance (Polyakova *et al.* , 2021; Drake *et al.* , 2019; McGuire *et al.* , 2020). To our knowledge our paper is the first to directly model and study how reinsurance, an ex-post risk adjusting insurer subsidy affects a "risk averse" insurer and how such policy affects insurers' pricing decisions.

Third, our paper adds onto the literature on subsidy design in health insurance markets. Tebaldi (2017); Polyakova & Ryan (2019), and Decarolis *et al.* (2020b) study subsidy design in the individual health insurance market, and Medicare Part D, respectively. These papers focus on optimal consumer subsidy design, focusing on adverse selection across different demographics as well as imperfect competition in the market. Closer to our paper is Einav *et al.* (2019) that compares consumer subsidy and ex-ante risk adjustment payment subsidy to insurers. Our paper, however, studies reinsurance subsidy which is an ex-post supply side subsidy to the insurer that may have additional impact when insurers are risk averse or face financial frictions.

More broadly, our paper relates to the market design literature on the choice of the optimal regulation instrument, including allocating consumer and production subsidies in electrical vehicle (Springel, 2021) and solar panel industries (De Groote & Verboven, 2019), granting production, investment and entry subsidies in ship building industries (Barwick *et al.* , 2021).

The remaining paper is organized the following way. Section 2.2 describes the institutional setting and background. In section 2.3, we provide a stylized model to illustrate relative effectiveness of consumer subsidies and reinsurance subsidies. Section 2.4 presents the data and some empirical evidence of the existence of risk frictions that insurers face as well as adverse selection in the market. Lastly, section 2.5 concludes.

2.2. Institutional Setting and Background

2.2.1. ACA Exchanges and State Reinsurance Programs

As part of Affordable Care Act (ACA), in 2014 the individual health insurance or ACA exchanges began offering private health insurance plans for consumers in a centralized platform. In these state-level marketplaces, private health insurers entered and offered various health insurance plans that met the benefit design requirements laid out by the ACA. Individual consumers without employer-sponsored health insurance coverage could purchase their own health insurance in the ACA exchanges. Many of these consumers received federal subsidy that significantly brought down their effective enrollee premium.

Despite the well-intentioned ACA, the individual market experienced numerous challenges in the first few years. Many of the markets experienced huge premium increases, accompanied by insurer exits where some markets were left with one or in the worst case zero insurers. Some of the factors that contributed to such instability in the market include: i) large uncertainty regarding medical usage of consumers, many of whom were previously uninsured ii) discontinuation of the federal reinsurance program that partly shielded insurers from large costs iii) heightened political uncertainty regarding the ACA (e.g. threat of repealment of ACA).

In response to the destabilizing individual health insurance market, many state governments started taking their own actions to stabilize the market and prevent it from further unraveling. One such measure that is increasingly being adopted by many states is reinsurance program. As stated in Table B.1, 14 states have implemented the reinsurance program as of 2021. At its core, reinsurance works as a secondary insurance for the primary insurers. That is when the insurer enrolls a very costly enrollee, the government shares in the cost with the insurer beyond some threshold.[3]

State reinsurance programs reimburse insurers for high-cost claims in different ways: some programs pay a portion of claims for consumers with certain medical conditions, while other programs reimburse a percentage of claims between specified dollar amounts. States fund the reinsurance program with both federal

[3] https://www.cbpp.org/sites/default/files/atoms/files/4-3-19health.pdf

pass-through funds from ACA 1332 waiver and states' own general fund revenues. We obtain the initiation year and program structure of each state from the Center of Medicaid and Medicare Services (CMS) and report this information in Table B.1.

While state reinsurance programs are a popular tool being adopted to help bring down enrollee premiums, there is some debate amongst the policymakers as to the effectiveness of such policy. In particular the efficiency of public reinsurance depends on whether insurers pass-through the cost savings onto the consumers by lowering the premiums. Instead some policymakers argue that a direct-to-consumer subsidy should be used to better target individual consumers, brining down their effective premiums. In section 2.3, we highlight the importance of frictions in the market in determining the relative efficiency of the two subsidy mechanisms: adverse selection and risk aversion of insurers stemming from financial/regulatory frictions.

2.2.2. Insurer Regulation and Private Reinsurance

In this section, we provide details on insurance company regulation in the U.S. and how such regulations may lead insurers to behave as if risk averse such as purchasing private reinsurance policies.

Risk-based capital regulation: Insurance regulators like the National Association Of Insurance Commissioners (NAIC) in the U.S. use risk-based capital as a method to evaluate financial strength of insurance companies. risk-based capital ratio essentially captures ratio of capital surplus to required capital for its liabilities.

$$\text{RBC-Ratio} = \frac{\text{Asset - Liabilities}}{\text{Required Capital}} \tag{2.1}$$

The required capital is also known as risk-based capital and is usually some exogenous multiple of the liabilities. Let A denote the asset, L denote the liabilities of an insurance company. Then risk-based capital is defined as ϕL, where ϕ is the risk-based capital multiplier. Then we can define the statutory capital as:

$$K = A - L \quad - \quad \underbrace{\phi L}_{\text{risk-based capital}} \tag{2.2}$$

i.e. statutory capital, K is the amount of capital surplus that the insurance company has above and beyond the very minimum that is required.

Insurance regulators deem certain RBC-ratios as meeting financial solvency requirements of the insurance companies. For example, NAIC scrutinizes any companies that have RBC-ratios below 200% and starts taking various actions once it falls below 200% ranging from company-level warning to full control of the company.[4]

While insurance companies regardless of their line of business are evaluated using the RBC-ratio measures, the exact way the RBC-ratio is computed differs across insurers' line of business. For health insurers, most of its risk or liability

[4] https://shorturl.at/gjCS7

stems from their underwriting risk or the enrollees' claims cost that the insurer is liable for. So for health insurance companies, RBC-ratio is an ex-post solvency measure that evaluates how much extra capital insurers have in relation to their claims liability. This implies insurers most hold or raise certain level of capital for the amount of risk that they are assuming. One way that primary health insurers in particular small insurers increase their underwriting ability without raising additional capital is through purchasing private reinsurance, decreasing their risk level or balance sheet liability.

Private reinsurance: Given the risk-based capital regulation by the insurance regulators, and associated financial frictions that insurers may face, primary insurers often purchase private reinsurance from a third party. At the basic level, reinsurance is "insurance for insurance companies" and acts as a back-stop against large losses. According to NAIC, primary insurers will purchase private reinsurance for many reasons: "1) expanding the insurance company's capacity; 2) stabilizing underwriting results; 3) financing; 4) providing catastrophe protection; 5) withdrawing from a line or class of business; 6) spreading risk; and 7) acquiring expertise."[5]

Historically, reinsurance policies have been more widely used in the property & casualty insurance market, in which primary insurers face risk of small probability of large catastrophic event. However, more and more reinsurance companies that traditionally offer reinsurance policies in the property & casualty insurance market

[5] https://content.naic.org/cipr-topics/reinsurance

are also offering reinsurance in the health insurance markets, especially for smaller insurers or in some cases health care providers. In section 2.4, we show that many health insurers do in fact purchase private reinsurance.

2.3. Theoretical Model

Here we present a theoretical model to highlight the role of selection and risk frictions in pass-through of reinsurance and consumer subsidy.

Suppose there is a risk averse monopoly insurer that sells a single insurance plan. There are two types of individuals in the market with $t \in \{\ell, h\}$. The insurer faces an elastic demand of $q_t(p)$ for individuals of type t. We assume that type $t = \ell$ individuals have more elastic demand i.e. $\varepsilon_\ell(p) \geq \varepsilon_h(p) \, \forall \, p$ where $\varepsilon(p)$ is the price elasticity of demand.

For each individual i of type t the insurer enrolls, it faces a random marginal cost of $\tilde{c}_i^t$ where $\tilde{c}_i^t \sim F_t$. We assume $\tilde{c}_i$ is independently distributed regardless of individual's type. Let $c_t = E[\tilde{c}_i^t]$, and $\sigma_t^2 = \text{Var}(\tilde{c}_i^t)$. We allow for the possibility of (adverse) selection in the market by allowing F_t to be different across the two individual types.

The monopoly insurer faces the following objective function in which it seeks to maximize its expected profit subject to risk charge.

$$\max_p \quad \underbrace{p\,\overbrace{\left(q_\ell(p) + q_h(p)\right)}^{Q(p)}}_{\text{premium revenue}} - \underbrace{\left(c_\ell q_\ell(p) + c_h q_h(p)\right)}_{\text{expected cost}} - \underbrace{\rho\,\overbrace{\left(\sigma_\ell^2 q_\ell(p) + \sigma_h^2 q_h(p)\right)}^{\text{total variance}}}_{\text{risk charge}} \tag{2.3}$$

Here we model the insurer's risk aversion behavior by insurer incurring a risk charge from the uncertainty in its total cost. As shown in (2.3), the risk charge is the product of ρ, the coefficient of risk charge, and the variance of total cost. ρ here could be thought of as the risk aversion parameter where the insurer faces a CARA utility function. Given the above objective, insurer's first order condition is

$$\underbrace{p + \frac{Q(p)}{\frac{\partial Q(p)}{\partial p}}}_{MR} = \underbrace{(\lambda(p)c_\ell + (1-\lambda(p))c_h)}_{MC} + \underbrace{\rho\left(\lambda(p)\sigma_\ell^2 + (1-\lambda(p))\sigma_h^2\right)}_{\text{marginal risk charge}} \tag{2.4}$$

where $\quad \lambda(p) = \dfrac{\frac{\partial q_\ell(p)}{\partial p}}{\frac{\partial Q(p)}{\partial p}}$

Hence the insurer faces an effective marginal cost that is the sum of its marginal cost and marginal risk charge. All else equal insurer facing a heightened risk frictions (i.e. higher ρ or variance of cost) will charge higher price. Let p_0^* denote the optimal price that insurer sets from (2.3).

2.3.1. Reinsurance Subsidy

Given the model, we examine how reinsurance subsidy affects the insurer's pricing behavior and its associated pass-through to the consumers. Suppose the government offers a stop-loss reinsurance that fully reimburses the insurer for any costs beyond the deductible $\theta > 0$. That is if an individual's ex-post cost $\tilde{c}_i > \theta$ then the government fully reimburses the insurer for any cost that exceeds θ. Given such a reinsurance scheme, insurer's ex-post cost for an individual i will be

$$\tilde{c}_i(\theta) = \begin{cases} \tilde{c}_i & \text{if } \tilde{c}_i \leq \theta \\ \\ \theta & \text{if } \tilde{c}_i > \theta \end{cases}$$

The reinsurance policy will decrease both the expected cost and the variance of each individual's cost. Let $c_t(\theta)$, and $\sigma_t^2(\theta)$ denote the insurer's expected cost, and the variance of type t individual for reinsurance policy of θ, respectively. With the reinsurance, insurer's FOC will now be

$$p + \frac{Q(p)}{\frac{\partial Q(p)}{\partial p}} = \underbrace{(\lambda(p)c_\ell(\theta) + (1-\lambda(p))c_h(\theta))}_{MC} + \underbrace{\rho\left(\lambda(p)\sigma_\ell^2(\theta) + (1-\lambda(p))\sigma_h^2(\theta)\right)}_{\text{marginal risk charge}} \tag{2.5}$$

The FOC in (2.5) shows that reinsurance will decrease the effective marginal cost in two ways. First, reinsurance decreases the expected cost of each individual and as a result decreases the marginal cost of the insurer. Furthermore because reinsurance acts as an insurance for the insurer, it decreases the variance of the insurer's total cost, decreasing the marginal risk charge of the insurer. Given that reinsurance will unilaterally decrease the right hand side or the insurer's effective marginal cost, it will decrease insurer's optimal price. That is $p^*(\theta) < p_0^*$ where $p^*(\theta)$ denotes the insurer's optimal price for reinsurance level θ.

While the decrease in price is not surprising as reinsurance is essentially providing a cost subsidy for the insurer, we look at how much reinsurance can decrease the price. That is for a given amount of reinsurance subsidy, we are interested in quantifying the pass-through rate. For a risk neutral government, the cost of providing reinsurance is the expected amount the government is expected to reimburse the insurer. For individual of type t, this is given by $r_t(\theta) = c_t(0) - c_t(\theta)$. Then the total expected reinsurance expenditure as well as the average expected reinsurance cost per consumer are

$$
\begin{aligned}
R(\theta) &= r_\ell(\theta) q_\ell(p) + r_h(\theta) q_h(p) \\
r(\theta) &= \alpha(p) r_\ell(\theta) + (1 - \alpha(p))\, r_h(\theta), \quad \text{where} \quad \alpha(p) = \frac{q_\ell(p)}{Q(p)}
\end{aligned}
\tag{2.6}
$$

So for the government the average reinsurance cost is given by $r(\theta)$ in (2.6). The reinsurance pass-through rate is $(p_0 - p^*(\theta))/r(\theta)$.

Proposition 4. *If insurer is risk averse i.e. $\rho > 0$, then the reinsurance pass-through rate, $(p_0 - p^*(\theta))/r(\theta)$ can be greater than 1.*

Proposition 4 states that if insurer is risk averse, then the pass-through rate of reinsurance subsidy could be larger than 1. In a standard monopoly setting, the pass-through of cost subsidy is often smaller than one due to market power. However when the monopoly insurer is risk averse, reinsurance not only affects its expected cost, but reduces the risk that the insurer faces. This indirect effect of reinsurance for a risk averse insurer is why the pass-through rate could be greater than one.

2.3.2. Reinsurance vs Consumer Subsidy

While reinsurance as an ex-post cost subsidy can lead to pass-through of greater than one, we are interested in comparing such subsidy to a more straightforward direct-to-consumer subsidy. To do so, in this section we compare the pass-through rate of reinsurance subsidy vs. consumer subsidy through the same theoretical model. In particular, we look at the role of adverse selection and the insurer's risk frictions in determining the efficiency of each subsidy mechanism.

Let s be a per-quantity (or per-enrollee) demand-side consumer subsidy that the government hands out to each consumer. Then the price that consumers face will be given by

$$p^e = p - s$$

and the demand for insurance will be given by $Q(p^e) = q_\ell(p^e) + q_h(p^e)$.

To compare the pass-through rate of the two subsidy mechanisms, we look at the government expenditure under consumer subsidy vs. reinsurance subsidy that yield the same price for consumers. So holding the price change constant, we compare how costly each subsidy mechanism is for the government.

Let $p_r^*(\theta)$ be the equilibrium price under reinsurance of level θ. And let p_s^* be the equilibrium price under demand subsidy of level s. For a given θ we can solve for the s such that

$$p_r^*(\theta) = p^e = p_s^* - s$$

That is consumers face the same price under both reinsurance and demand-side subsidy. The corresponding subsidy level $s(\theta)$ that yields the same price for consumer as reinsurance of level θ is

$$\begin{aligned} s(\theta) &= p_s^* - p_r^*(\theta) \\ &= \underbrace{\lambda(p) r_\ell(\theta) + (1-\lambda(p))\, r_h(\theta)}_{\text{marginal reinsurance cost}} + \underbrace{\rho\left(\lambda(p)\Delta\sigma_\ell^2(\theta) + (1-\lambda(p))\,\Delta\sigma_h^2(\theta)\right)}_{\text{marginal change in risk charge}} \end{aligned} \tag{2.7}$$

where $\Delta\sigma_t^2(\theta)$ denotes the change in variance of cost for reinsurance level θ. Given the above expression for $s(\theta)$, we want to determine the relative magnitude of $r(\theta)$ vs. $s(\theta)$.

Proposition 5. *Let adverse selection in the market be defined as $F_\ell(t) \leq F_h(t)\ \forall\, t, c_\ell < c_h,\ \sigma_\ell^2 < \sigma_h^2$. Then the relative magnitude of $r(\theta)$ and $s(\theta)$ will depend on the following:*

(1) ***No risk frictions and no selection:*** *If insurer is risk neutral i.e. $\rho = 0$, and there is no selection i.e. $F_\ell = F_h$ then $s(\theta) = r(\theta)\,\forall\,\theta$.*
(2) ***Risk aversion, no selection:*** *If insurer is risk averse i.e. $\rho > 0$, and there is no selection, then $s(\theta) > r(\theta)\,\forall\,\theta$.*
(3) ***Adverse selection, no risk frictions:*** *If insurer is risk neutral i.e. $\rho > 0$, and there is adverse selection in the market, then $s(\theta) < r(\theta)\,\forall\,\theta$.*
(4) ***Risk aversion, and adverse selection:*** *If insurer is risk averse i.e. $\rho > 0$, and there is adverse selection in the market, then the relative magnitude of $s(\theta)$ and $r(\theta)$ is ambiguous.*

Proposition 5 states that without any risk frictions nor selection, the pass-through rate of reinsurance and consumer subsidy is the same. However, when insurer is risk averse and/or there is adverse selection in the market, the relative efficiency of each subsidy mechanism will vary. Under risk aversion, reinsurance subsidy will generate large pass-through due to its ability to reduce the risk that insurer may face, further lowering the effective marginal cost. Under adverse selection the marginal reinsurance cost will be smaller than the average reinsurance cost, making the consumer subsidy to have a larger pass-through rate. However when both frictions exist in the market, the relative magnitude of the pass-through rates will be ambiguous as it will depend on the relative magnitude of each friction.

The comparative results above highlight the importance of both insurer's risk frictions as well as the degree of adverse selection in the market. In section 2.4, we provide empirical evidence of insurers facing risk frictions as well as adverse selection in the market. Using several states that have alreay implemented state-level reinsurance policies, we examine the pass-through rate of the policy and whether risk frictions and/or adverse selection seem to play a role.

In ongoing work, we plan to take the model to data and estimate an empirical model allowing for both risk frictions and adverse selection in the market. Having empirically estimated the degree of both types of frictions, we plan to compare the relative efficiency of reinsurance and consumer subsidy in lowering the enrollee premiums.

2.4. Empirical Results

Given the importance of both the risk frictions and adverse selection in the market in determining the efficiency of the two subsidy mechanisms, in this section we empirically document and show evidence of both types of frictions in the market.

2.4.1. Data

We collect information on insurer-level health insurance contracts, reinsurance contracts, and capital adequacy levels from several proprietary and public datasets.

Health insurance contracts. We collect information on health insurance product from CMS Health Insurance Exchange Public Use Files in 2014-2019. This dataset is a publicly available dataset of the universe of plans launched through the federally facilitated exchanges marketplaces. We supplement this dataset with the publicly available plan attributes files for state-based exchanges marketplaces from Center for Consumer Information and Insurance Oversight (CCIIO) [6]. These two datasets are both at plan-year level and have the same layout: we observe premium, financial attributes, including deductibles, coinsurance rates, co-pays, out-of-pocket limits for every plan.

[6] 38 states use federally facilitated exchanges marketplaces: AK, AL, AR, AZ, DE, FL, GA, HI, IA, IL, IN, KS, KY, LA, ME, MI, MO, MS, MT, NC, ND, NE, NH, NJ, NM, NV, OH, OK, PA, SC, SD, TN, TX, UT, VA, WI, WV, WY. The remaining states use state-based exchanges marketplaces. We collect information on plan characteristics for these 13 states, including CA, CO, CT, DC, ID, MA, MD, MN, NY, OR, RI, VT, WA

Reinsurance contracts. Our primary private reinsurance data comes from National Association of Insurance Commissioners (NAIC), which is a standard setting and regulatory support organization governed by insurance regulators from each state. As part of its mandate to coordinate regulatory oversight, NAIC collects and publishes operational data from insurance companies. The data used in this study come from the Schedule S reports for all insurers in the life and health line of business from 2016-2019. The data is at unique reinsurance contract level, and have detailed information on seller identity, buyer identity, contract effective data, reinsurance premiums, and realized reinsurance claims.

We supplement the NAIC reinsurance records with Medical Loss Ratio (MLR) database, a regulatory database maintained by the Center for Medicare and Medicaid Services (CMS). The MLR database contains detailed financial information about health insurers across different market segments (individual, small group, and large group) at the state level. Specially, the MLR database has information on medical claims costs, health insurance premium revenue, enrollment, and the sum of public and private reinsurance premiums and realized claims at the insurer-year level for all insurers from 2010-2019.

Table 2.1 provides an overview of reinsurance-purchasing pattern in our data. 82% of insurers on ACA exchanges purchase private reinsurance, and expenses on reinsurance premiums account for 5.47% of health insurance premiums income over our sample period.

Table 2.1. Insurer summary statistics

	All	Exchange
Mean health insurance premium income, all market segments	315.34	503.86
Mean health insurance premium income, all market segments	315.34	503.86
Mean health insurance claims costs, all market segments	267.83	427.92
Mean health insurance member-months, market segments	1.38	1.95
Mean health insurance premium income, Exchanges	47.72	79.69
Mean health insurance claims costs, Exchanges	39.19	65.44
Mean health insurance member-months, Exchanges	0.1	0.17
Mean reinsurance insurance premium expenses, all market segments	17.81	27.6
Mean reinsurance insurance realized claims costs, all market segments	0.21	0.36
Percent of insurer-year pairs that purchase reinsurance	0.83	0.82
Number of insurer-year pairs	7115	4261

Notes: Premiums, claims costs and member-months statistics are in millions. The sample includes all insurers in MLR dataset in 2016-2019. The first column reports statistics for all insurers that sell health insurance, while the second column focuses on insurers that have positive health insurance premium income from the Exchanges market for both datasets.

Financial Solvency. Our primary data on insurers' financial solvency/capital adequacy measures come from National Association of Insurance Commissioners (NAIC) that collects various financial statements data from insurance companies in the US.[7] More specifically, we use data from the 5-yr Historical Data for all insurers in the life and health line of businesses from 2006-2020. The data is at the firm-year level and for each unique insurer, shows summary financial statements data for the past 5 years starting from the year of filing. Most importantly, it includes data on insurers' statutory capital level as well as the authorized control level of capital that can be used to construct the RBC-ratio of insurers as described earlier.

[7] There is one state missing in this data. California does not require insurers to submit such filings to NAIC and such only includes insurers that submit them on a voluntary basis.

2.4.2. Private reinsurance purchasing by health insurers

Table 2.2. Characteristics of Insurers with vs. without Private Reinsurance

Variable	All	Has Reins	No Reins
has private reins	0.681 (0.466)	1 (0)	0 (0)
private reinsurance share of premium	0.03 (0.119)	0.056 (0.16)	0 (0)
# of members (million)	0.554 (2.134)	0.366 (1.271)	0.943 (3.285)
insurer operates > 5 states	0.103 (0.304)	0.1 (0.3)	0.139 (0.347)
RBC ratio	7.176 (5.47)	6.527 (4.448)	8.201 (6.374)
Non-profit	0.325 (0.469)	0.344 (0.475)	0.266 (0.442)
share of individual market revenue	0.164 (0.275)	0.191 (0.308)	0.128 (0.222)

Notes: This table reports various average and standard deviation (shown in parenthesis) of insurer characteristics for all insurers, insurers that do purchase private reinsurance and insurers that do not purchase any private reinsurance policies. The data includes all health insurers from 2017-2019, where each observation is a unique insurer-year. Private reinsurance share of premium is the ratio of insurer's private reinsurance premium paid to the reinsurer to insurer's total earned premium revenue. Share of individual market revenue is the insurer's total revenue earned in the individual market divided by the insurer's total revenue across all markets.

In this section, we document primary health insurers' private reinsurance purchasing behavior. Table 2.2 shows that 68% of health insurers purchase some amount of private reinsurance, showing widespread use of private reinsurance amongst health insurers. On average, insurers spend 3% of their premium revenue to buy private reinsurance. Conditional on insurer purchasing private reinsurance, this number goes up to 5.6%.

Table 2.2 also looks at the characteristics of insurers that buy private reinsurance vs. those that do not. Insurers that do buy private reinsurance tend to be smaller, regional insurers. They are more likely to be financially constrained as evidenced by the lower average RBC ratio, suggesting that insurers may purchase private reinsurance due to financial/regulatory frictions that they face. We also see that non-profit insurers that may be have more limited capital market access are more likely to purchase private reinsurance. Lastly, insurers whose individual market business makes up a greater share of their overall revenue are more likely to purchase private reinsurance. As a result, government reinsurance may be especially relevant in the individual health insurance market.

2.4.3. Effect of state-level reinsurance programs

Given that several states have implemented their own reinsurance policies in the recent years, we examine the impact of such policies in the market. These state-level reinsurance programs are essentially free reinsurance contracts with zero premiums, which lower both the expected cost and any risk charges that insurers may be internalizing due to financial frictions. Using an event study framework, we first look at the overall impact of reinsurance on insurer premiums.

We run the following analysis to quantify the impact of the reinsurance policies.

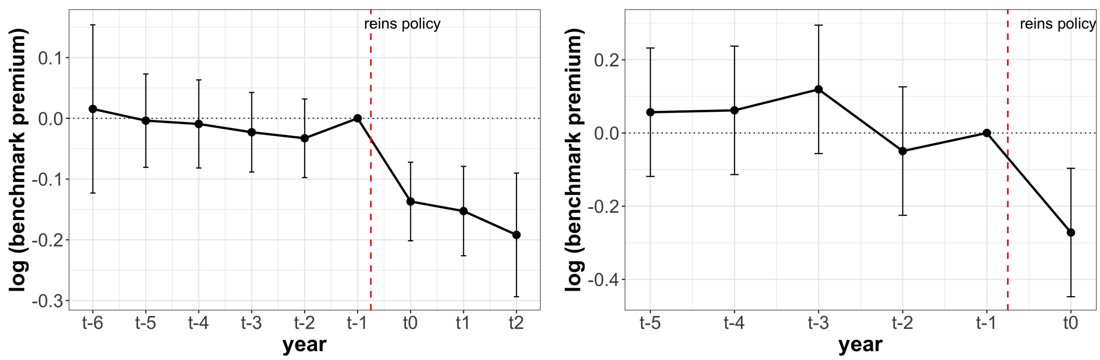
$$y_{rt} = \beta_0 + \sum_t \beta_t D_{rt} + \gamma_t + \gamma_r + \varepsilon_{rt} \tag{2.8}$$

where r denotes the rating region, and t denotes the year.[8] D_{rt} is an indicator for whether rating region r (or the state that it is part of) has reinsurance. The coefficients of interest β_t show effect of having reinsurance policy for the given year t.

[8]Rating region is a pre-defined geographic area, which is often a groups of counties, that insurers set their prices in. That is for a given insurance plan, the insurer has to set the same price within the rating region.

Figure 2.1. Effect of State Reinsurance Policy on Premiums

(a) All States (b) CO



The above figures show the yearly effect of state reinsurance on average benchmark premiums. Panel (a) plots the results from pooling all states i.e. it utilizes staggered event study framework. Panel (b) shows the results comparing CO and all other control states without the reinsurance program.

Figure 2.1 shows that large and statistically significant impact of government reinsurance on the premiums. Across all states, we find that reinsurance decreased premiums around 13.5% and find no evidence of any pre-trends suggesting that these decreases are unlikely due to any heterogeneous changes occurring in states with government reinsurance.

Figure 2.2b shows the results just for the state of Colorado, which implemented its reinsurance policy in 2020. Using the state's estimate of the cost of reinsurance, we can calculate a back of the envelope pass-through rate of CO reinsurance policy. The state estimates that the reinsurance cost the government $\$220$ million, or $\$1,373$ per enrollee. Our estimated impact of reinsurance on Colorado premiums is 23% or $\$1,812$ per enrollee. This translates to a pass-through rate of 1.31. That is for the CO government, for every $\$1$ spent on reinsurance, the premium decreased by $\$1.31$. This is contrary to a typical pass-through rate of smaller than one in an imperfectly competitive market. Given that firms likely have some degree of market power in this market, this finding suggests that consistent with our theoretical model insurers do face financial frictions. Next, we show further evidence of insurers internalizing financial frictions in their pricing decisions.

2.4.3.1. Evidence of financial frictions. We showed that government reinsurance programs decrease the average premium levels. Next, we examine whether there are heterogeneous impact of the policy. In particular, we look at how the impact varies across different insurer characteristics. Equation (2.9) shows the specification where instead of just looking at the impact of reinsurance via D_{rt}, we interact the dummy with region specific variable, s_{rt}. We use several measures that are related to the degree of financial frictions of insurers in that region.

$$y_{rt} = \beta_0 + \beta_1 D_{rt} + \beta_2 s_{rt} D_{rt} + \gamma_t + \gamma_r + \varepsilon_{rt} \tag{2.9}$$

Table 2.3 shows that across several measures of insurers' financial frictions, the effect of reinsurance is greater for regions with more insurers that are likely to face higher degree of financial frictions. We see that if the region has more insurers with private reinsurance, the premium decrease is larger. If there are more insurers that have really low RBC-ratio and are likely to be financially constrained, the premium decreases is much greater. Lastly, we find that places with more

Table 2.3. Heterogeneous Effects of State Reinsurance Policy on Premiums

	Dependent variable:			
	log(premium_avg)			
	(1)	(2)	(3)	(4)
Reins	-0.135^{***}	-0.087^{***}	-0.127^{***}	-0.037
	(0.013)	(0.031)	(0.013)	(0.031)
Reins $\times$ share with private reins		-0.071^{*}		
		(0.042)		
Reins $\times$ share with RBC-ratio < 2			-0.226^{**}	
			(0.097)	
Reins $\times$ share of non-profit				-0.150^{***}
				(0.043)
Observations	2,881	2,881	2,881	2,881
Adjusted R^2	0.837	0.837	0.837	0.838

Notes: This table reports the effect of state-level reinsurance programs on premiums using difference-in-difference framework. Columns 2 through 4 also estimate hetrogeneous impact of the reinsurance across different types of insurers. Column 2 uses the share of insurers that had private reinsurance in 2017. Column 3 uses share of insurers with RBC-ratio of below 2, which is a threshold that insurance regulators often use. Column 4 uses share of non-profit insurers in a given rating region. All specifications include year, and rating region FEs. Standard errors are clustered at the rating region level. *p<0.1; **p<0.05; ***p<0.01

non-profit firms that have limited capital market access also experiences greater premium decreases.

Next, we look at whether government reinsurance has any impact on insurers' private reinsurance purchasing behavior. Given that government reinsurance lowers insurers' risk much like private reinsurance policies, we expect insurers to substitute away from purchasing private reinsurance as a response to free public provision of reinsurance. We consider two primary outcomes at the insurer-year level: (1) on the extensive margin, whether the insurer purchase private insurance;

(2) on the intensive margin, the share of the health premium income spent to purchase private reinsurance contracts.

Let f denote insurer, t denote year, s denote the state where insurer f operates. We estimate the following specification:

$$y_{ft} = \beta D_{s,t} + \beta X_{ft} + \gamma_f + \gamma_t + \epsilon_{ft}, \tag{2.10}$$

where $D_{s,t}$ is the indicator state s has reinsurance policy in place; X_{ft} is the logarithm of per-member claims costs in the previous year; γ_f, γ_t are insurer and year fixed effects, respectively. The coefficient of interest β, which measures the changes in private reinsurance purchase in response to the initiation of state reinsurance programs, holding claims costs level fixed. We hypothesize that initiation of the state reinsurance programs will reduce both the probability of purchasing private reinsurance and the amount of private reinsurance purchased.

To verify that the identified changes in private reinsurance purchases are driven by changes in financial constraints, we conduct an additional heterogeneity analysis. We examine whether the effects are more pounced for insurers whose financial constraints are binding. We divide insurers into two groups, by whether previous years' RBC ratio is in the bottom 5th. percentile[9] We then estimate analogues of equation (2.10) in which we interact the treatment indicator with an indicator denoting whether the insurer has previous years' RBC ratio above or below the 5th percentile.

[9] 3.1 is 5th percentile of RBC distribution. An RBC ratio of 2 is the minimum surplus level needed for a health insurer to avoid regulatory action.

Table 2.4 reports the regression results. Column (1) and (3) report the average treatment effects of reinsurance programs. The initiation of state-level reinsurance programs reduce the probability of purchasing private reinsurance by 3% and the share of premium income used to purchase reinsurance by 0.2%. Column (2) and (4) examine treatment effect heterogeneity. The estimated effects are concentrated in insurers with low RBC ratio, as expected. The statistically significant difference in the impact of reinsurance programs between insurers with the same previous claims costs but high and low RBC ratios, provides suggestive evidence that financial frictions are at work.

Table 2.4. Effect of state-reinsurance program on private reinsurance purchase

	Probability of purchasing private reinsurance		Share of premium income used to purchase reinsurance	
	(1)	(2)	(3)	(4)
Reinsurance policy	-0.0380**	-0.261***	-0.002	0.000
	(0.015)	(0.093)	(0.004)	(0.011)
Reinsurance policy		0.229**		-0.002
$\times$ high RBC		(0.090)		(0.013)
N	2787	2787	2787	2787
Baseline mean	0.872	0.872	0.205	0.205
High RBC grp mean		0.869		0.218

Notes: This table reports the effects of reinsurance programs on private reinsurance purchase. The regression sample is at the insurer-year level in 2017-2019 for all insurers nationwide that have positive health premium income on the Exchanges. The regression specification controls for the logarithm of realized claims costs in the previous year, state fixed effects, and year fixed effects. Insurers are divided into high and low RBC ratio group, by whether previous years' RBC ratio is in the bottom 5th percentile. Standard errors are clustered at state level. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

2.4.3.2. Evidence of adverse selection. In the previous section, we've showed ample empirical evidence that insurers do face financial frictions in this market. However, as illustrated in section 2.3, the role of adverse selection is just as important in determining the efficiency of reinsurance vs. consumer subsidy. While adverse selection is a well-documented phenomenon in the health insurance market (), we offer additional evidence that suggests adverse selection does matter in this market. To do so, we look at whether the impact of reinsurance policies vary across different types of insurance products. More specifically, we look at the heterogeneous impacts across different metal tiers, that have different actuarial value of insurance.

In the absence of selection, we'd expect consumers to be equally represented across the different metal tiers. With adverse (advantageous) selection, we typically expect sicker (healthier) consumers to select into plans with higher actuarial value. Given that reinsurance only reimburses the insurer when the enrollee's claims cost is beyond some high threshold, we'd expect reinsurance to have heterogeneous impact on different consumers. That is reinsurance would decrease insurers' expected cost much more for sicker enrollees. As a result, with adverse selection we'd expect greater change in premiums for higher metal tier products.

Table 2.5 shows suggestive evidence of strong adverse selection in the market. The effect of reinsurance on premiums is much larger for higher actuarial value plans. For example, the highest actuarial valued Platinum plans' premiums

Table 2.5. Effects of State Reinsurance Policy across Different Metal Tiers

	Dependent variable:				
	log(premium)				
	Catastrophic	Bronze	Silver	Gold	Platinum
	(1)	(2)	(3)	(4)	(5)
Reins	-0.068^{***}	-0.095^{***}	-0.143^{***}	-0.161^{***}	-0.213^{***}
	(0.018)	(0.014)	(0.013)	(0.014)	(0.027)
Observations	2,517	2,881	2,881	2,877	1,366
Adjusted R^2	0.645	0.795	0.858	0.814	0.907

Notes: this table reports the effect of state-level reinsurance programs on premiums of different actuarial valued plans using difference-in-difference framework. Each observation is at the year-rating region level, and denotes the average premiums of specified metal level, where catastrophic indicates the least generous plan benefit design and platinum is the most generous one. All specifications include year, and rating region FEs. Standard errors are clustered at the rating region level. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

decrease 21% whereas the lowest actuarial valued Catastrophic plans' premiums decrease only by 6.8%.

The above empirical results illustrate that both financial frictions and adverse selection are clearly present in the market, and will impact insurers' pricing decisions thereby affecting enrollee's premiums in the market. Taken together, these results highlight the need for an empirical model that incorporates both forces. Estimation of the model is currently in progress, and we hope to shed light on the relative importance of both forces in the market. Given the estimates, we aim to determine the efficiency of the two subsidy mechanisms: reinsurance and consumer subsidy.

2.5. Discussion and Conclusion

Government subsidy plays a major role in the US health insurance markets. Given the large magnitude of fiscal spending in the area, it is important that the government utilizes an efficient subsidy mechanism to deliver affordable insurance coverage. In this paper, we examine the efficiency of two widely used subsidy mechanisms in the individual health insurance market: government reinsurance and direct-to-consumer subsidy. Reinsurance subsidizes the insurers by ex-post covering some of high-cost enrollee's costs, which in turn will lead insurers to decrease its premium. Consumer subsidy on the other hand, directly subsidizes the purchase of insurance for consumers, lowering the effective enrollee premium.

By building a theoretical model in which insurers face financial frictions in adverse selection market, we show that both forces play an important role in determining the efficiency of the two subsidy mechanisms. With financial frictions, reinsurance not only decreases the expected cost of insurers but lowers insurer's risk charge stemming from financial frictions. With adverse selection, the cost of a marginal enrollee tends to be smaller than the cost of an average enrollee. Because consumer subsidy targets the marginal enrollee whereas reinsurance targets the average enrollee, it may be a more efficient mechanism in the absence of any financial frictions. However with both forces in the market, it is unclear which subsidy mechanism will be more efficient for the government.

Using state-level reinsurance policies, we show empirical evidence that both frictions exist in the market. We first document the widespread use of private reinsurance by health insurers despite the high markup. With government reinsurance, insurers are less likely to purchase private reinsurance. Furthermore, reinsurance decreases premiums more for insurers that tends to buy private reinsurance. Lastly, we show that government reinsurance decreases the premiums of higher actuarial valued plans more than lower actuarial valued plans, suggesting evidence of adverse selection in the market.

While we can't quantify which force plays a more important role, our results highlight the importance of both in studying different subsidy mechanisms in the market. As such, we are in the process of building and estimating an empirical model of demand & supply allowing for both financial frictions and adverse selection in the market. With the estimates, we hope to quantify and evaluate the efficiency of the two subsidy mechanisms.

CHAPTER 3

Incentive Structures and Borrower Composition in the Paycheck Protection Program[1]

3.1. Introduction

Public policy is frequently implemented through private actors (e.g., Affordable Care Act Exchanges, Small Business Administration Loan Guarantee Programs). In such situations, the government faces a tradeoff between inducing the private actors to participate and achieving policy goals. To that end, policymakers typically leverage direct subsidies and other forms of incentives to influence the behavior of participants on both sides of the market. Understanding the responsiveness of actors to these incentives is key in policy design.

In this paper, we study the incentive design of the Paycheck Protection Program (PPP). The PPP was a $953 billion loan forgiveness program that aimed to assist small businesses in keeping their employees on payroll during the Coronavirus pandemic.[2] Loans were issued by banks, but borrowers are eligible for forgiveness if they use a sufficient share of the loan to support payroll expenses.

[1]This chapter is joint work with David Stillerman. We would like to thank Vivek Bhattacharya, Gaston Illanes, Robert Porter, and Mar Reguant, as well as participants at the Northwestern IO Student Seminar, for helpful comments and suggestions.

[2]For more details on the program, see https://www.sba.gov/funding-programs/loans/coronavirus-relief-options/paycheck-protection-program

Should the borrower not meet this threshold, a portion of the loan is forgiven and the remaining balance must be repaid at an interest rate and loan term set by the government.[3] A large amount of recent work has analyzed the targeting of the PPP (see, e.g., Granja *et al.* (2021) and Bartik *et al.* (2020)) and its impact on employment and business survival (Hubbard & Strain (2020)). Joaquim & Netto (2021) study lender incentives in the program, but, to the best of our knowledge, our work is the first to analyze the role of the PPP's policy design in determining the allocation of loans and the subsequent decision of whether to use the funds on payroll expenses.

In the PPP, when lenders issued a loan, they received a subsidy equal to a pre-specified percentage of the loan amount. Outside of the subsidy, there were two other categories of incentives that influenced lenders' decisions. First, the SBA set the standard for loan forgiveness by fixing a minimum share of the loan amount that must be used on payroll expenses. Second, it set the interest rate and maturity of non-forgiven loans. These interventions indirectly induced participation of lenders by influencing the behavior of borrowers. The presence of both sets of incentives in the PPP makes it a fruitful environment in which to study the design of policies implemented through the private sector. In particular, we seek to address the following question: are subsidies or borrower-side incentives more effective in

[3]For first-draw loans, the interest rate is 1% and the maturity is either two or five years depending on the date of issuance. See https://www.sba.gov/funding-programs/loans/coronavirus-relief-options/paycheck-protection-program/first-draw-ppp-loans.

(1) increasing credit access and (2) targeting (i.e., ensuring loans are used for forgivable purposes)? Importantly, we show that the answer to this question depends on a number of primitives underlying both lender and borrower decisions.

The design of the program has received considerable attention from legislators and the media. In particular, politicians have questioned the structure of the ex-ante subsidy. Speaker of the House Nancy Pelosi stated: âĂIJWe have to take a look at how banks are compensated. They get a higher percentage for a small loan, but if you get 5% on a $50,000 loan, thatâĂŹs a lot less than getting 1% on a $5 million loan.âĂİ[4] Despite the attention paid in the policy arena, there is limited work in the economics literature on the design of the program. We bridge this gap.

To illustrate the role of the underlying borrower and lender primitives in determining the efficacy of policy design, we develop a model of PPP lending. Borrowers are differentiated by their observable loan amount, their unobserved propensity to use funds for payroll purposes (i.e., the relative return of payroll versus other uses), and their unobserved level of cash on hand. With knowledge of these primitives, the borrowers choose the share of funds to use on payroll, given the forgiveness standards and other incentives specified by the program. Lenders are differentiated by their fixed cost of lending to a borrower (e.g., up-front administrative costs, verifying borrower materials) and their marginal cost of an additional dollar. They decide whether to approve a borrower's application for a loan. The

[4] https://www.marketwatch.com/story/pelosi-suggests-banks-making-loans-in-small-business-program-shouldnt-get-paid-more-for-serving-bigger-companies-2020-04-27

relationship between the loan amount and other primitives underlying the borrowers' loan-use decisions, as well as the lenders' fixed and marginal costs of issuing the loan, determine how the policy design affects access to funds and the targeting of the program.

The model has two key implications. First, we show that increasing subsidies and relaxing forgiveness standards induce lenders to offer loans to borrowers seeking smaller amounts. This result implies that policymakers have multiple levers at their disposal to induce an expansion of credit. Second, the correlation between loan amounts and borrowers' tendency to use funds for forgivable purposes determines the relative efficacy of ex-ante and ex-post incentives in improving the program's targeting.

Given that one set of incentives need not always be better than the other, we evaluate the observed design of the PPP. We exploit temporal variation in the stringency of forgiveness standards through the passage of the PPP Flexibility Act to provide descriptive evidence to (1) validate the model's key implications and (2) assess the program's efficacy in targeting funds to borrowers who use them on payroll. This piece of legislation made forgiveness standards less stringent, decreasing the minimum share allocated to payroll from 75% to 60%. While our current research design does not allow us to recover causal estimates of the policy response, we find evidence consistent with the predictions of the model. The average loan amount falls by between 6 and 7% in the period following the implementation of the Flexibility Act, which is consistent with lenders being willing to issue smaller loans when forgiveness is easier to obtain. This decline in loan amounts disproportionately benefits businesses in wholesale and retail trade, sole proprietors, and those in urban areas. This result implies that the credit expansion induced by changes to borrower-side incentives could alter the composition of borrowers who benefit from the public program.

To assess the program's targeting, we analyze the relationship between loan size and borrowers' propensity to use funds for payroll. We find that the marginal borrowers (i.e., those who receive loans in the post period but would not have prior to the legislation change) are more likely to use funds on payroll than the inframarginal borrowers. Again, while we do not attribute a causal interpretation to this result, it suggests that the observed relaxation of forgiveness standards improved the targeting of the program.

In aggregate, we find empirical evidence consistent with the claim that the policy change improved both access and targeting. The improvement in access was particularly beneficial for a set of borrowers who typically seek smaller loans and for whom banks may face larger fixed costs of loan issuance. In ongoing work, we estimate the primitives underlying the borrowers' and lenders' decision problems. The goal of such an analysis is to consider counterfactual policy designs – for example, a policy in which banks are subsidized per loan instead of per dollar lent – which may have different implications for access and targeting.

The remainder of the paper proceeds as follows. Section 3.2 reviews the related literature, Section 3.3 describes the Paycheck Protection Program, Section 3.4

presents the model and discusses the main comparative statics results, Section 3.5 presents the data and empirical analyses, and Section 3.6 concludes.

3.2. Related Literature

There are a growing number of papers that study the Paycheck Protection Program. These can be divided into four main strands.

A large number of papers focus on the casual impact of PPP on labor market outcomes. Autor *et al.* (2020) and Chetty *et al.* (2020) look at the unemployment rate of PPP-eligible vs. non-eligible firms around the 500 employee cut-off and find that the PPP did boost employment although their magnitudes differ. Barraza *et al.* (2020), Faulkender *et al.* (2021) and Granja *et al.* (2021) use regional variation in lender composition to estimate impact of PPP on labor market outcomes in different geographic regions and find mixed results. Barraza *et al.* (2020) finds that during the first month, the PPP reduced unemployment by 1.4%, and Faulkender *et al.* (2021) finds that 10% increase in eligible payroll covered by the PPP resulted in 1-2% decrease in weekly initial unemployment insurance claims. However, Granja *et al.* (2021) finds no significant evidence that the PPP had a substantial effect on local employment outcomes. Doniger & Kay (2021) uses the 10-day delay between first and second round of the PPP and finds that regions with one percentage point fewer delayed loans within the event window have lower unemployment rates by over 10 basis points.

Related to the above, several papers study and evaluate PPP more comprehensively, especially focusing on targeting. Humphries *et al.* (2020) finds that, despite smaller firms showing largest improvements upon receiving PPP loans, they were less aware/less likely to apply and less likely to get approved. Bartik *et al.* (2020) and Hubbard & Strain (2020) both find that PPP approval on net is beneficial for the small businesses, increasing their chance of survival. Bartik *et al.* (2020), however, notes that banks are less likely to approve higher-distressed firms, and more likely to approve firms with existing connections, even though PPP loans do not seem to be more effective for that group. Similarly, Granja *et al.* (2021) finds no evidence that PPP funds flowed to areas more adversely affected by COVID-19 and that banks played a major role in targeting - bank participation in the initial phase depends largely on bank characteristics, which explains spatial differences in loan disbursements. Joaquim & Netto (2021) also finds that, during the first phase of PPP, firms that were less affected by COVID-19 received loans earlier but the opposite is true for the second phase. The paper also builds a model of PPP allocation with firms and banks, and finds that the PPP saved 7.5 million jobs.

A few papers specifically study how role of different banks in distributing PPP loans. Li & Strahan (2020) shows that relationship banks (i.e., banks that typically associated as having close relationships with their borrowers) played a big role in disbursing PPP loans. James *et al.* (2021) similarly finds that community banks made loans faster and lent more relative to their assets compared to larger banks. Lastly, Erel & Liebersohn (2020) studies the role of FinTech in PPP lending and finds that FinTech is more likely to be used in areas with fewer bank branches, lower incomes, larger minority shares, and places more severely impacted by COVID-19.

The paper also estimates that FinTech expanded the supply of PPP credit rather than substituting borrowers away from banks. Finally, Lopez & Spiegel (2021) examine the role of the PPP Liquidity Facility, and find that it, along with the PPP itself, increased the growth rate of small-business lending.

Next, a number of papers study the impact of PPP take-up on firms, focusing on the impact on firm valuation and other frictions that some firms may be facing. Balyuk *et al.* (2020) focuses on small, publicly listed firms and finds that although firms with PPP funds experience positive valuation effects, many firms end up returning the funds, which is also associated with increase in the firms' valuation. Cororaton & Rosen (2021) looks at set of all PPP-eligible public firms and documents that firm value declines after the PPP loan announcement and increases after some of the firms return their PPP loans. The paper suggests that reputational harm and negative signaling limit public firms' participation in the PPP.

Lastly, there is a significant body of work on racial disparities in PPP recipients. Fairlie & Fossen (2021) and Wang & Zhang (2021) use regional variation in minority population to show that places with higher minority share received disproportionately fewer PPP loans. Howell *et al.* (2021) uses PPP loan-level data to show that Black-owned businesses were more likely to obtain their loan from a FinTech lender versus a traditional bank. Among banks, smaller banks were much less likely to lend to Black-owned firms. Chernenko & Scharfstein (2021) studies a large sample of Florida restaurants and finds significant racial disparities between Black/Hispanic-owned firms vs. White-owned firms. It also finds evidence that this is driven by bank-lending compared to non-bank lending and that Black-owned businesses are more likely to substitute away from bank-administered PPP loans to SBA-administered EIDL loans.

Much of the above literature focuses on finding the overall (or lack there of) impact of PPP. Our paper differs in that we focus on the PPP program design and how different policy parameters could affect both the allocation and targeting of loans. While some of the papers do note that banks play a major role in targeting and allocating loans, most of the papers do not model or study how changing the program parameters could impact the effectiveness of the program. To our knowledge, we are the first paper to model and study how the program design, in particular the lender subsidy and borrower forgiveness standards, could alter both the borrowers and lenders' behavior, shifting the equilibrium allocation of loans.

3.3. Institutional Background

3.3.1. Program Description

The Paycheck Protection Program was originally established by the CARES Act, which allocated $349 billion of funding for the loan forgiveness program between April 3 and April 16, 2020.[5] Later, the Paycheck Protection Program and Health

[5]https://www.npr.org/sections/coronavirus-live-updates/2020/04/16/835958069/small-business-emergency-relief-program-hits-349-billion-cap-in-less-than-2-week

Care Enhancement Act allocated an additional $320 billion[6] with the first applications accepted on April 27, 2020.[7] These two acts served as the primary funding sources for the first-draw loans.

The program aimed to provide forgivable loans to businesses with 500 or fewer employees worldwide, though this restriction was relaxed for businesses in NAICS 72, Accommodation and Food Services (Bartik *et al.* (2020)). Businesses were also required to meet specified SBA size standards and have tangible net worth less than or equal to $15 million as of March 27, 2020. Importantly, these businesses were not required to satisfy a "credit elsewhere" test, meaning they may have been able to receive financing from other sources.[8]

As mentioned in Section 3.1, the program consisted of two sets of incentives. First, lenders received subsidies, as a share of the loan amount, for participating in the program. For loans less than or equal to $350 thousand, lenders received a subsidy equal to 5% of the loan amount. The subsidy rate declined for larger loans – 3% for loans between $350 thousand and $2 million and 1% for loans greater than $2 million.[9] Second, the policy stipulated forgiveness standards. Prior to

[6] For further details, see https://www.ama-assn.org/delivering-care/public-health/summary-paycheck-protection-program-and-health-care-enhancement-act.

[7] https://fortune.com/2020/04/23/ppp-sba-paycheck-protection-program-loans-applying-round-2-what-to-know-small-business-application-congress-funding/

[8] See https://home.treasury.gov/system/files/136/Paycheck-Protection-Program-Frequently-Asked-Questions.pdf. The credit-elsewhere test is an important feature of other SBA lending programs, including the SBA 7(a) Program, the agency's largest loan guarantee scheme. Therefore, the subset of borrowers who received funding through the PPP differed from the subset who participated in other lending programs.

[9] https://www.sba.gov/sites/default/files/2021-02/Procedural%20Notice%205000-20091%20-%202nd%20Updated%20PPP%20Processing%20Fee%20and%201502%20Reporting-508.pdf.

the passage of the PPP Flexibility Act, forgiveness required the borrower use 75% of loan proceeds on payroll and associated expenses, including gross salary and wages, tips, vacation and sick leave, holiday pay, health insurance, and retirement benefits.[10] Following the policy change, the borrower was required to use only 60% on payroll.

Businesses were given a pre-specified amount of time, called the *covered period*, during which they were required to allocate the funds. This stipulation limited businesses from reallocating funds between profits and loan proceeds to keep the loan open indefinitely. Again, as we detail in the next subsection, the PPP Flexibility Act changed the program's covered period.

To obtain loan forgiveness, the borrower must complete an application, either directly with the SBA or through their lender, and compile documentation to prove a sufficient share of funds was allocated to payroll. This documentation includes bank account statements or reports from third-party payroll services to confirm compensation amounts, as well as payroll tax forms. With this information, the SBA conducts a review and determines how much of the loan to forgive.[11]

[10]There were a number of contingencies associated with the forgiveness standards. For loans above $50 thousand, the forgiveness rate decreased if the borrower reduced wages by more than 25%. Also, if borrowers were unable to spend the stipulated amount on payroll but demonstrated good-faith effort in rehiring or reduced employees' hours in response to public-health guidance, they were still eligible for forgiveness.

[11]https://www.sba.gov/funding-programs/loans/covid-19-relief-options/paycheck-protection-program/ppp-loan-forgiveness#section-header-4

3.3.2. PPP Flexibility Act

After the initial rollout of the program, Congress amended the SBA's guidance through the issuance of the PPP Flexibility Act. This policy change occurred in response to a number of perceived shortcomings of the original program's structure. For example, due to public-health guidance, businesses such as restaurants and bars did not anticipate being able to reopen in time to spend a sufficient share of funds during an eight-week covered period.[12] The legislation relaxed the stringency of forgiveness standards.

The PPP Flexibility Act was passed on June 5, 2020, and this piece of legislation altered a number of borrower incentives. It decreased the minimum payroll share from 75% to 60%, increased the repayment period for the non-forgiven portion of the loan from 2 to 5 years, and increased the covered period from 8 to 24 weeks. This policy change provides us with variation in the generosity of forgiveness, primarily through the changes to the minimum payroll share and covered period, with which we analyze equilibrium responses by both borrowers and lenders.

3.4. Model

We model the PPP lending process in two steps, which capture the main decision problems faced by borrowers and lenders. First, a borrower i and lender j are paired, and, after observing the loan amount, borrower characteristics, and the borrower's report of the share it expects to use for payroll purposes, the lender

[12] https://www.jdsupra.com/legalnews/flexibility-act-significantly-improves-40387/

decides whether to issue a loan to the borrower. If the loan is issued, the borrower then receives a shock to its reported share, which determines its final allocation to payroll.

This model captures the responses of both lenders and borrowers to changes in the subsidy rate, forgiveness standards, and loan characteristics (i.e., interest rate and maturity). From the model, we derive two comparative statics results to illustrate the relative efficacy of the two sets of policy levers in (1) expanding access to PPP loans and (2) targeting the loans to businesses that are likely to use the funds on payroll.

3.4.1. Preliminaries

Suppose borrowers are differentiated in three dimensions: (1) b_i, the loan amount, which is a constant multiple of the previous year's profits and is observable and verifiable by the bank, (2) θ_i, a parameter that determines the relative return on funds used for non-forgivable purposes[13] compared to it being used for forgivable (payroll) purposes, and (3) w_i, the borrower's per-period cash on hand or, equivalently, an unobserved profit shifter. Suppose $\theta_i, b_i \sim H_{\theta,b}$, and assume b_i is common knowledge, while θ_i is known only by the borrower. The level of cash on hand is independent of θ_i and b_i, and $w_i \sim H_w$. Lenders j are differentiated by their fixed cost of issuing a loan to borrower i, c_{ij}.

[13] To be more precise, non-forgivable purposes means not just non-payroll expenses but funds used for other allowed business expenses that fall under broad categories as defined by PPP.

The policymaker sets a subsidy rate, S, which is a share of the loan amount, and three other incentives: $\underline{f}$, the minimum share of funds used for forgivable purposes, r, the net present value of interest paid on the loan should it not be forgiven, and T, the length of the covered period. If the borrower satisfies the forgiveness criteria, then it receives full loan forgiveness. If it does not satisfy the criteria, the loan is partially forgiven, and the borrower must repay the balance at the stipulated interest rate.[14]

Upon receiving the loan, the borrower has a sum of funds equal to the loan amount plus the per-period cash on hand times T, the length of the covered period. At this point, borrowers choose the share of total funds ($T \cdot w_i + b_i$) to use for payroll purposes, f_i. Upon obtaining the loan, the borrower then receives a shock to its payroll share, ϵ_i. This shock follows a mean-zero, symmetric distribution: $\epsilon_i \sim F_\epsilon$. Note that the government and the econometrician do not observe this final share. Instead, they observe the reported share of the loan amount (b_i) used for payroll. This is observed both at loan origination, $\tilde{f}_i$, and after repayment, $\tilde{f}_i^{post}$:

$$\tilde{f}_i = \min\left\{1, f_i\left(1 + \frac{T \cdot w_i}{b_i}\right)\right\}$$

$$\tilde{f}_i^{post} = \max\left\{\min\left\{1, \tilde{f}_i + \epsilon_i\right\}, 0\right\}$$

[14]For more information on the terms of forgiveness, see https://www.sba.gov/funding-programs/loans/coronavirus-relief-options/paycheck-protection-program/ppp-loan-forgiveness#section-header-0.

If $\tilde{f}_i^{post} \geq \underline{f}$ then all of the initial funds are forgiven, otherwise only a portion of the funds are forgiven. Specifically, for $\tilde{f}_i^{post} < \underline{f}$, the share forgiven is $1 - \frac{\tilde{f}_i^{post}}{\underline{f}}$.

3.4.2. Borrower's Problem

Borrower i, seeking a loan of size b_i, is informed of its propensity to use funds for payroll purposes, θ_i, its cash on hand, w_i, the forgiveness standards of the government, $\underline{f}$, the net present value of all interest paid on the loan should it not be forgiven, r, and the length of the covered period, T. The borrower sets its share of funds used for payroll by solving:

$$\begin{aligned} \max_{f_i \in [0,1]} \; & \gamma \left(f_i + (1 - f_i) R(f_i, \theta_i) \right) (b_i + T \cdot w_i) \\ & - \mathbb{1} \left\{ f_i \left(1 + \frac{T \cdot w_i}{b_i} \right) < \underline{f} \right\} (1 + r) b_i \left(1 - \frac{f_i}{\underline{f}} \left(1 + \frac{T \cdot w_i}{b_i} \right) \right), \end{aligned} \tag{3.1}$$

where $R(\cdot, \cdot)$ is the return on funds used for non-payroll purposes relative to that on funds used for payroll purposes (i.e., $R(\cdot, \cdot) = 1$ means that borrowers are indifferent between using the funds on forgivable and non-forgivable purposes) and γ is the per-dollar value or return on funds used for payroll.

To ensure the borrower's problem has a unique solution, conditional on T and w_i, and that the optimal share assigned to payroll is monotonic in the borrower's type, θ, we make a number of assumptions on the form of $R(\cdot, \cdot)$. In particular, we assume the function is twice differentiable in each of its arguments and impose three further conditions:

Assumption 3. *$R(f, \theta)$ is twice differentiable in both of its arguments with (i) $\frac{\partial R}{\partial \theta} > 0$, (ii) $\frac{\partial R}{\partial f} > 0$, and (iii) $\frac{\partial R^2}{\partial \theta \partial f} < 0$.*

Assumption 4. *$R(f, \theta)$ is concave in its first argument.*

Assumption 1(i) is without loss of generality and imposes monotonicity of R in θ, Assumption 1(ii) assumes decreasing returns to scale, while Assumption 1(iii) imposes a single-crossing condition. Assumption 2 is more restrictive and is sufficient, along with the other assumptions, to guarantee uniqueness and monotonicity. However, it may not be a necessary condition, and work is in progress to weaken this assumption.

Under the above assumptions, the borrower's optimal share f_i^* is weakly decreasing in its type, θ_i. We summarize this result in the following proposition, the proof of which is in Appendix C.1.

Proposition 6. *Under Assumptions 1–2, for each θ_i, (i) there exists a unique $f^*(\theta_i) \in [0, 1]$ that solves the borrower's optimization problem, and (ii) $f^*(\theta_i)$ is weakly decreasing in θ_i.*

Figure 3.1 displays the mapping from borrower type to the optimal payroll share given by the solution to the borrower's maximization problem. The plot illustrates the quantity both as a share of total funds (f_i^*) and as a share of the loan amount reported at origination ($\tilde{f}_i^*$). There are four distinct cutoff types at which the mapping discretely changes. These points of non-differentiability signify changes between corner solutions (i.e., $f^* = 1$, $f^* = \underline{f}$, and $f^* = 0$) and interior solutions.

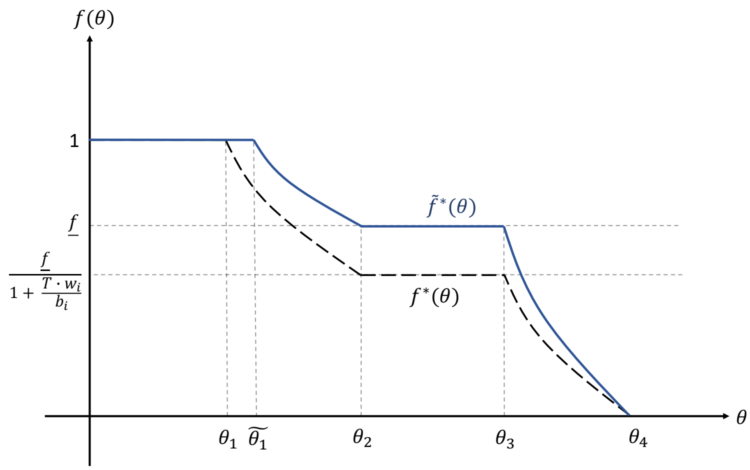


Figure 3.1. Borrower's Optimal Payroll Share

There are two main takeaways from Figure 3.1. First, the result of Proposition 1 is apparent. Conditional on the level of cash on hand, w_i, the optimal payroll share is weakly decreasing in the borrower's type. Second, the level of cash on hand – in particular, its magnitude relative to the loan amount – determines the share of borrower types who choose a payroll share equal to one or a share equal to the policy cutoff. It does not, however, influence the share of borrowers who choose a share of zero. When estimating the model, which is in progress, this type of variation informs the distribution of cash on hand.

The mapping from Figure 3.1 fully characterizes the optimal decision of a given borrower. The lender has knowledge of this rule, as well as a number of borrower observables. We now move to describe the lender's side in more detail and, importantly, its decision of whether to issue a loan to a borrower.

3.4.3. Lender's Problem

Suppose the borrower truthfully reports the payroll share at origination, $\tilde{f}_i^*$, to the lender[15], and the lender observes the loan amount, b_i, and other borrower covariates. The lender does not observe the realization of the borrower's cash on hand, w_i, but knows the distribution from which it is drawn. With this information in hand, lender j decides whether to approve the loan application of borrower i.

The lender's profit from issuing a loan is given by:

$$\pi_{ij} = S(b_i) - c_{ij} + s(\tilde{f}_i^*)(r - \delta_j)b_i,$$

where $S(b_i)$ is the subsidy offered to the lender for issuing a loan of size b_i (in this case, $S(b_i) = Sb_i$), c_{ij}[16] is the borrower-lender specific fixed cost of processing the loan application, r is the net present value of interest payments, δ_j is the net present value of the marginal cost of lending, and $s(\tilde{f}_i^*)$ is the expected share of the initial

[15]Borrowers are legally constrained from lying on their loan application. The borrower must "certify that the information provided in [the] application and the information provided in all supporting documents and forms is true and accurate in all material respects." False statements are punishable by imprisonment of up to five years or a fine of up to $250,000. See https://www.sba.gov/sites/default/files/2021-03/BorrowerApplication2483ARPrevisions%20%28final%203-18-21%29-508.pdf.

[16]This cost captures, for example, whether the borrower and lender have a pre-existing relationship. Lenders may face lower costs of loan issuance if the business has a checking account with that bank.

loan that is not forgiven and remains a loan. The payoff in the event of not issuing the loan is normalized to zero, so the lender approves the loan if $\pi_{ij} \geq 0$.

Given the structure of the ex-post shock to the payroll share, ϵ_i, the lender's expected share forgiven takes the following form:

$$s(\tilde{f}_i^*) = E_{\epsilon_i}\left[\mathbb{1}\{\tilde{f}_i^{post}(\epsilon_i) < \underline{f}\}\left(1 - \frac{\tilde{f}_i^{post}(\epsilon_i)}{\underline{f}}\right)\right]$$

To ensure the solution to the lender's problem is a unique cutoff rule, we require one further assumption.

Assumption 5. *For all lenders j, (i) $r - \delta_j < 0$ and (ii) $S + s(\tilde{f}_i^*)(r - \delta_j) > 0$.*

The first part of the assumption implies that lenders earn a loss if a loan were not forgiven at the observed interest rate, r. The program stipulates a rate of 1%, and anecdotal evidence suggests lenders prefer loans to be forgiven. If this assumption did not hold, lenders would earn a higher return on non-forgiven loans, given they are fully guaranteed. The second part of the assumption implies the subsidy rate is high enough such that a lender would issue a loan of some amount. Again, this assumption is consistent with the lenders' decisions to participate in the program.

Proposition 2 summarizes the solution to the lender's decision problem and describes a number of comparative statics results. A proof is available in Appendix C.1.

Proposition 7. *Under Assumption 3, there exists a minimum loan amount, $\underline{b}$, such that lender j approves all loans with $b_i > \underline{b}$. The minimum loan amount is: (i) decreasing in S, (ii) decreasing in r, and (iii) increasing in $\underline{f}$. Specifically, $\underline{b}$ is given by:*

$$\underline{b}(\tilde{f}_i^*) = \frac{c_{ij}}{S + (r - \delta_j)s(\tilde{f}_i^*)}$$

This proposition illustrates the mechanisms available to the policymaker to induce lenders to extend funds to borrowers seeking smaller amounts. In particular, the policymaker may increase the subsidy rate, increase the interest rate on the non-forgiven portion of the loan, or make forgiveness standards less stringent. However, despite all levers leading to an expansion of lending activity, they have different implications for the targeting of the program. In the next subsection, we illustrate two comparative statics results that show that the relative efficacy of changes to ex-ante subsidies and ex-post interventions (i.e., interest-rate or forgiveness-standard changes) in program targeting depends critically on the relationship between loan amount and the borrower's propensity to use funds on payroll.

3.4.4. Policy Design

To examine the relative efficacy of the policies available to the regulator, we first define its objective. We do not take a stance on the social welfare function – instead, we examine comparative statics on one metric of use for the regulator, the average share forgiven. Because the stated aim of the PPP is to provide businesses with the funding required to keep employees on payroll, we consider this metric a measure of how well the program is targeted. Specifically, define the average payroll share as:

$$PS = \frac{\int_{\underline{b}}^{\infty} b_i \tilde{f}^*(\theta_i)\, dH_{\theta_i,b_i}}{\int_{\underline{b}}^{\infty} b_i\, dH_{b_i}}$$

With a slight abuse of notation, we do not explicitly consider the relationship between $\tilde{f}^*$ and w_i. Because $\tilde{f}^*$ is monotonic in θ_i for all w_i and the comparative statics results hold for each w_i, then they hold when integrating over w_i.

With the objective function in hand, we first consider the impact of the ex-ante subsidy on the average payroll share. From the previous subsection, we know this policy lever is effective in expanding access to PPP funds. Now, we show that making subsidies more generous has an ambiguous effect on the program's targeting. Depending on the primitives underlying the borrowers' decisions, this change could either increase or decrease the average share of funds allocated to payroll.

Consider a change in the subsidy rate from S' to S'' where $S'' > S'$. The change in the average payroll share takes the form:

$$PS(S'') - PS(S') = \frac{\int_{\underline{b}(S'')}^{\underline{b}(S')} b_i\, dH_{b_i}}{\int_{\underline{b}(S'')}^{\infty} b_i\, dH_{b_i}} \left(\underbrace{\frac{\int_{\underline{b}(S'')}^{\underline{b}(S')} b_i \tilde{f}^*(\theta_i)\, dH_{\theta_i,b_i}}{\int_{\underline{b}(S'')}^{\underline{b}(S')} b_i\, dH_{b_i}}}_{\text{Forgiveness rate of marginal borrower}} - \underbrace{\frac{\int_{\underline{b}(S')}^{\infty} b_i \tilde{f}^*(\theta_i)\, dH_{\theta_i,b_i}}{\int_{\underline{b}(S')}^{\infty} b_i\, dH_{b_i}}}_{\text{Forgiveness rate of average borrower}} \right)$$

The two forces determining the aggregate impact of a change in the subsidy rate are apparent in the above expression. Whether an increase in the generosity of the subsidy leads to more funds allocated to payroll depends on the relative strength of the contribution of the marginal borrower and that of the average borrower. If the marginal borrower is more likely than the average borrower to use funds for payroll, then a larger subsidy leads to a greater average share allocated to payroll. The opposite is true if the marginal borrower has a lower propensity to use funds on payroll than the average borrower. The primitive relationship between the borrowers' propensity to use funds for payroll and the loan amount determines which of the two forces dominates.

The effect of a change to the stringency of the forgiveness standards instead depends not only on the primitive relationship between loan amounts and the borrowers' propensity to use funds on payroll but also on how responsive inframarginal borrowers are to changes in the forgiveness rules. Consider a move from a threshold of $\underline{f}'$ to $\underline{f}''$ where $\underline{f}'' < \underline{f}'$. In this case, the change to the average payroll share is:

$$PS(\underline{f}'') - PS(\underline{f}') = \frac{\int_{\underline{b}(\underline{f}'')}^{\underline{b}(\underline{f}')} b_i \, dH_{b_i}}{\int_{\underline{b}(\underline{f}'')}^{\infty} b_i \, dH_{b_i}} \left(\underbrace{\frac{\int_{\underline{b}(\underline{f}'')}^{\underline{b}(\underline{f}')} b_i \tilde{f}^*(\theta_i, \underline{f}'') \, dH_{\theta_i, b_i}}{\int_{\underline{b}(\underline{f}'')}^{\underline{b}(\underline{f}')} b_i \, dH_{b_i}}}_{\text{Forgiveness rate of marginal borrower}} - \underbrace{\frac{\int_{\underline{b}(\underline{f}')}^{\infty} b_i \tilde{f}^*(\theta_i, \underline{f}') \, dH_{\theta_i, b_i}}{\int_{\underline{b}(\underline{f}')}^{\infty} b_i \, dH_{b_i}}}_{\text{Forgiveness rate of average borrower}} \right)$$

$$+ \underbrace{\frac{\int_{\underline{b}(\underline{f}')}^{\infty} b_i \left(\tilde{f}^*(\theta_i, \underline{f}'') - \tilde{f}^*(\theta_i, \underline{f}') \right) \, dH_{\theta_i, b_i}}{\int_{\underline{b}(\underline{f}'')}^{\infty} b_i \, dH_{b_i}}}_{\text{Change in forgiveness rate for inframarginal borrowers}}$$

Given $\underline{f}'' < \underline{f}'$, the final term is necessarily negative. Inframarginal borrowers have an incentive to weakly decrease their share allocated to payroll, as lower allocations still receive full forgiveness. Thus, the aggregate effect of a decline in the stringency of forgiveness standards on payroll share is negative if the average borrower is more likely than the marginal borrower to use funds on payroll. If, instead, the average borrower has a lower propensity to allocate funds to payroll, then the aggregate impact of the policy change is ambiguous and depends on the relative strength of the three components described in the expression above. Importantly, in contrast to the case of a subsidy, the adjustment in behavior of inframarginal borrowers can play a pivotal role.

The comparative statics results described in the subsections above provide one key testable implication of the model – the monotonicity of $\underline{b}$ in ex-ante subsidies and forgiveness standards. Furthermore, the model suggests that the correlation of the loan amounts and the share allocated to payroll is of first-order importance when judging the impact of policy changes on program targeting. In Section 3.5, we test the main implication of the model using policy variation from the PPP Flexibility Act. We then provide evidence of the primitive correlation between the loan amounts and payroll shares, showing the observed decrease in forgiveness stringency led to more funds allocated to payroll.

In ongoing work, we plan to take this model to data, estimating the primitives underlying the borrowers' decisions. Using these estimates, we plan to recover the fixed costs of lending. With the primitives, we can simulate counterfactual policy designs to characterize optimal policies for different government objectives, including the objective we have stressed in this section.

3.5. Empirical Results

3.5.1. Data

For our empirical analysis, we rely on loan-level data from the U.S. Small Business Administration (SBA), which maintains a public dataset containing each approved PPP loan. The dataset contains loan characteristics, including the loan amount, the date the loan was funded, and the loan term. It also contains the borrower's name and address, as well as characteristics such as the NAICS code of the business, a coarse indicator of the business age, and the business type (i.e., individual, corporation, etc.). Finally, it provides the name of the originating and servicing lender.

Outside of characteristics, we observe measures of expected loan use and ex-post loan performance. Specifically, the dataset lists, at origination, the expected share of loan proceeds allocated to utilities, payroll, mortgage interest, and rent, among other uses. The ex-post outcome of interest is the total amount forgiven, from which we calculate the ex-post forgiveness share.

We augment the loan-level data with bank balance sheet information from the Federal Financial Institutions Examination Council's (FFIEC) Uniform Bank Performance Reports. We use data from March 31, 2020 and match using the bank name. From these reports, we obtain the banks' Tier I Leverage Ratio as a proxy for the shadow cost of lending.

Because our empirical analysis focuses on lenders' responses to the PPP Flexibility Act, we restrict attention to a small window around the policy change. We consider loans issued up to four weeks before the event and up to eight weeks after the event. We analyze lending decisions within this window for two main reasons. First, the composition of borrowers who received loans in the first four weeks after the implementation of the PPP (which occurred eight weeks before the passage of the PPP Flexibility Act) differed from the composition who received them later on. To isolate the impact of the policy change, we seek to standardize the set of borrowers. Second, the take-up of the policy change occurred over time. A number of loans issued in the first four weeks after the policy change indicated an expected payroll share of 75%, the pre-period threshold. The share of loans issued with this expected share declines to close to zero by the fifth week after the policy range. Therefore, our window is not symmetric, and we include loans issued up to eight weeks after the change. We provide empirical support for our window definition in Appendix C.3.

We make a few further restrictions to isolate the equilibrium response to the change in forgiveness standards. We restrict to loans of up to $350,000. All loans in this category received the same ex-ante subsidy of 5%. This restriction does not eliminate a large number of loans, as only 4.3% of issuances in our twelve-week window are larger than this amount.

Table 3.1 displays summary statistics for our window of interest – twelve weeks around the policy change. There is considerable variation in the size of loans issued. The average loan amount is $22.7 thousand; however, almost 75% of loans are issued for less than $20 thousand. This heterogeneity is mirrored in the number of jobs supported, ranging from a minimum of zero to a maximum of 500. The

Table 3.1. Descriptive Stats on Approved Loans - 12 Weeks Around Policy

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Loan Amount ($000)	1,013,491	22.7	35.1	0.1	5.8	20.8	350.0
# Jobs	1,013,491	3.5	9.0	0	1	3	500
Urban	1,013,491	0.84	-	-	-	-	-
Non-Profit	1,013,491	0.02	-	-	-	-	-
Existing Business ($>$ 2 Years)	833,451	0.83	-	-	-	-	-
Payroll Share (At Orig.)	1,013,491	0.97	0.10	0	1	1	1
Forgiveness Share (Ex-Post)	556,822	0.99	0.07	0	1	1	1

Note: This table presents descriptive statistics for the main analysis sample. This sample includes all loans up to $350,000 issued up to four weeks before and eight weeks after the passage of the PPP Flexibility Act. The sample size differs for two variables due to missing data. In the regressions, when we include fixed effects for business age, we include a category for unknown rather than dropping these loans.

PPP funds support businesses of different ages and geographies – 84% of loans are issued to urban businesses and 83% are issued to businesses that have existed for at least 2 years. Lastly, we observe the share of the loan committed to payroll at origination and the share of the loan forgiven ex-post. These final two distributions suggest that forgiveness is common, as the majority of loans are committed to payroll at origination and end up being completely forgiven.

3.5.2. Empirical Tests - PPP Flexibility Act

To recap, our model yields three empirical implications, two that follow directly from the structure of the model, and a third that evaluates whether the observed policy change (i.e., the PPP Flexibility Act) improved program targeting.

Empirical Implication 1. *The minimum loan amount, $\underline{b}$, decreases when forgiveness standards become more generous. Thus, average loan amounts decline in response to the policy change.*

Empirical Implication 2. *The response in $\underline{b}$ is stronger for lenders who face a higher fixed cost (or higher marginal cost) of loan issuance.*

Empirical Implication 3. *Whether the response to the policy change improves program targeting (i.e., the share allocated to payroll) depends on (1) the relationship between borrowers' propensity to use funds for payroll and the loan amount, and (2) the response of inframarginal borrowers to the policy change.*

In the subsections below, we describe the ways in which we empirically test the above predictions of our model. While our research design does not lend itself to recovering causal estimates of the policy impact, the results we present provide support for the structure of the model and suggestive evidence of the efficacy of the policy change. Furthermore, we highlight how the lenders' responses to the policy change alter the composition of borrowers who receive funding under the PPP. In total, these results have important implications for equality in credit access across (1) demographic groups and (2) business types.

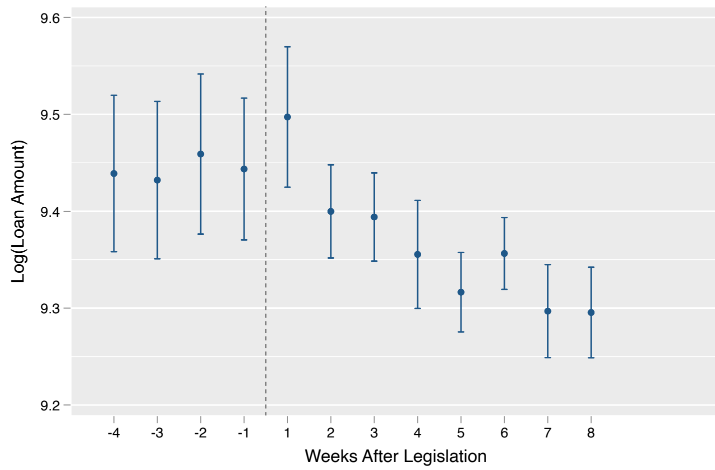
3.5.2.1. Aggregate Response in Loan Size. We begin by considering Testable Implication 1, which implies that lenders respond to the decline in the stringency of forgiveness rules by issuing smaller loans. We first analyze the equilibrium impact of the policy change, testing whether the average loan amount falls in the period following the passage of the PPP Flexibility Act. We estimate event-study specifications of the following form:

$$\log(b_{ijt}) = \alpha \mathbb{I}(t = Post) + \beta X_{ijt} + \epsilon_{ijt}, \tag{3.2}$$

where b_{ijt} is the size of the loan issued to borrower i by lender j in period t and X_{ijt} is a vector of controls, including fixed effects for lender, two-digit NAICS, business type (sole proprietorship, corporation, etc.), borrower state, urban/rural, and business age. Note that the results of this specification could be confounded by underlying time trends in loan amounts. Figure 3.2 displays the path of the average loan amount over time, and this plot suggests the absence of a time trend in the four weeks preceding the policy change. That being said, the pre-period consists of only four weeks, so further support is necessary. The analysis in the next subsection provides further suggestive evidence that a time trend is not the only cause of the observed loan-amount response, and work is in progress to improve the research design to address this shortcoming.

Table 3.2 displays results for the event-study specifications. The first column displays results with no controls, the second column includes lender fixed effects, while the final column adds the remaining borrower controls. Following the passage of the PPP flexibility act, the average loan amount is approximately 7.2% lower than in the baseline. A change in lender composition explains approximately 15% of this decline. The remainder is explained, almost completely, by a change in the

Figure 3.2. Log(Loan Amount) Across Time – 12 Week Window



Note: This figure displays the time trend Log(Loan Amount). In particular, it plots the coefficient of a regression of Log(Loan Amount) on an indicator for the week of loan issuance. The bars denote the 95% confidence interval of the estimate, calculated using standard errors clustered by borrower state.

composition of borrowers who receive loans. When we include the full set of controls, the loan amount is only 2.0% lower following the policy change, and this decline is not statistically significant. This result suggests that increasing the generosity of forgiveness may disproportionately benefit certain types of borrowers, namely those who typically seek smaller loans. Sole proprietors and self-employed individuals are two groups whose share of loans is higher following the policy change than in the baseline.

Figure 3.3 further unpacks the impact of the program across borrower covariates and displays the change in the share of loans issued to given types of borrowers in the post-legislation period. Panel (a) displays changes in share by one-digit NAICS. Businesses associated with one-digit NAICS codes 4 and 8 are more likely

Table 3.2. Aggregate Loan Amount Changes - 12 Weeks Around Policy

	(1) Log(Loan Amount)	(2) Log(Loan Amount)	(3) Log(Loan Amount)
Post-Legislation	-0.0723***	-0.0613***	-0.0197
	(0.0205)	(0.0184)	(0.0179)
Observations	1,013,491	1,013,319	1,013,316
Lender FE	No	Yes	Yes
Borrower Controls	No	No	Yes

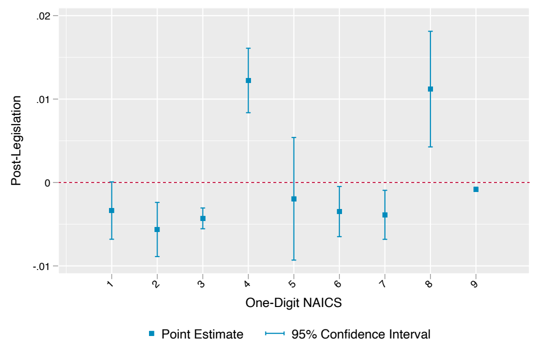
* p<0.1, ** p<0.05, *** p<0.01

Note: This table presents results for the aggregate event-study specifications defined by Equation (2). Standard errors clustered by borrower state are shown in parentheses. Lender FEs are defined as a combination of a lender name, lender city, and lender state. Borrower controls include fixed effects for the two-digit NAICS code, business type (corporation, LLC, sole proprietorship, other), urban/rural, business age ($>$ 2 years, $\leq$ 2 years, unanswered), and borrower state.

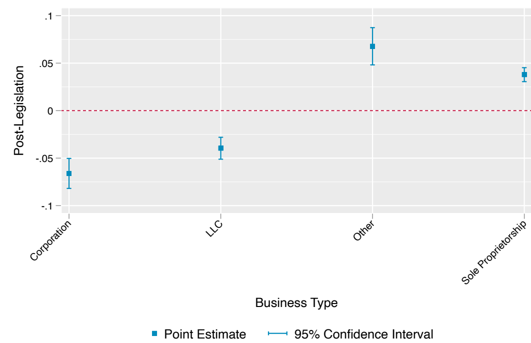
to receive funding following the policy change than they were prior. The former includes wholesale and retail trade, while the latter includes services such as salons and barbershops. Part of the media response to the PPP Flexibility Act, as described in Section 3.3 centered on expanding credit to businesses in industries most affected by public-health measures. This result suggests the policy change may have succeeded in that aim.

Panel (b) examines the pre- and post-legislation shares by business type. In the period following the legislation, sole proprietors received a larger share of loans than corporations and LLCs. This disproportionate impact likely operates through the loan-size channel. Sole proprietors operate smaller businesses than their counterparts. Because loan sizes were a function of prior-year profits, these businesses

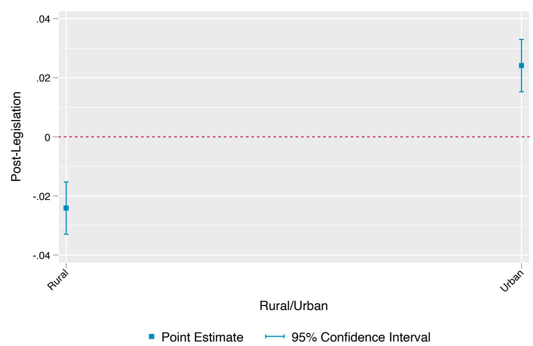
Figure 3.3. Heterogeneity in Aggregate Loan Amount Changes



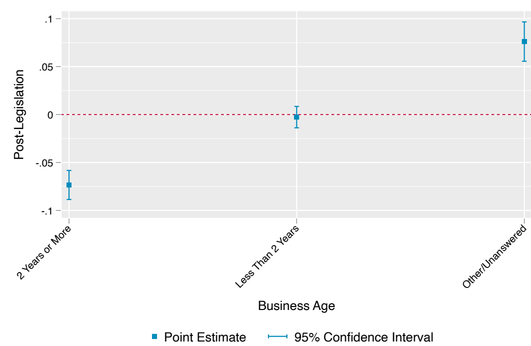
(a) One-Digit NAICS



(b) Business Type



(c) Urban/Rural



(d) Business Age

Note: These plots display the coefficient on a regression of an indicator of the specified business characteristic on a dummy for the post-legislation period. The bars indicate the 95% confidence interval calculated using standard errors clustered by borrower state.

were constrained, by the program rules, from obtaining large loans. Thus, increasing lenders' willingness to issue these smaller loans confers disproportionate benefits to sole proprietors.

Similar trends emerge in panels (c) and (d). In panel (c), we examine shares by urban/rural status, showing that urban borrowers comprise a larger share in the post-legislation period. Finally, panel (d) shows that well-established businesses, those established more than two years prior, are more likely to receive a loan prior to the legislation change. In total, these results suggest that the PPP Forgiveness Act altered the composition of borrowers receiving loans.

It is important to note that these event-study specifications illustrate equilibrium effects of the policy change and, as mentioned previously, could be confounded by underlying time trends in loan amounts. However, taken at face value, this analysis implies a differential impact in the policy change across borrower covariates. In the next subsection, we look instead at heterogeneity across types of lenders to determine whether our model's predictions regarding lending costs hold in the data.

3.5.2.2. Heterogeneity by Cost. In this subsection, we examine heterogeneity in the policy impact by the cost of loan issuance. This analysis provides evidence in support of Testable Implication 2, which implies that the average loan size decreases by more for lenders who face a high cost of lending. It is important to note that we do not ascribe a causal interpretation to these results. Callaway *et al.* (2021) show that the standard two-way fixed effects estimator of a generalized difference-in-differences model with continuous treatment does not recover the average causal response under the standard common trends assumption. Instead, the estimate includes a bias term whose magnitude depends on the extent of heterogeneity in treatment effects. Put another way, if the average treatment effect on the treated is more negative for lenders with high issuance costs than for lenders with low costs, then the estimate we recover is downward biased.

In our context, we are most interested in testing whether lenders who find it costlier to issue loans are more responsive to the PPP Flexibility Act. In effect, we seek to pin down the direction, but not necessarily the magnitude, of the average causal response. In this setting, it is reasonable to assume that the average treatment effect on the treated has the same sign for all "doses" of treatment. Under this assumption, the estimate from the standard two-way fixed effects specification has the same sign as the average causal response.

With these caveats in mind, we estimate the following two-way fixed effects specification:

$$\log(b_{ijt}) = \delta_j + \delta_t + \alpha\mathbb{I}(t = Post) \times \tilde{T}1_j + \beta X_{ijt} + \epsilon_{ijt}, \tag{3.3}$$

where b_{ijt} is again the size of the loan issued to borrower i by lender j in period t and X_{ijt} is a vector of controls, including fixed effects for two-digit NAICS, business type (sole proprietorship, corporation, etc.), borrower state, urban/rural, and business age. $\tilde{T}1_j$ is the standardized tier-one leverage ratio for lender j.[17] To remove outliers, we estimate this specification for ratios that fall, inclusively, between the 1st and 99th percentile of the distribution across all lenders.

The tier-one leverage ratio acts as a proxy for a bank's lending cost. This ratio is the sum of tier-one regulatory capital divided by total consolidated assets, and

[17] To compute the standardized variable, we subtract the mean and divide by the standard deviation across all lenders in the sample. This calculation does not weight by the number of loans issued by a given lender.

banks must maintain a tier-one leverage ratio of at least 4% to be considered "adequately capitalized" by the banking regulators.[18] PPP loans receive a zero risk weight but are included in the calculation of total consolidated assets unless they are pledged as collateral for a loan from the Federal Reserve's Paycheck Protection Program Liquidity Facility.[19] Thus, banks face a shadow cost of lending due to these regulatory frictions, and that shadow cost is higher for banks closer to the tier-one leverage threshold.

The specification detailed in equation (3) tests whether high-leverage ratio lenders (i.e., those who face a lower shadow cost of lending) are less responsive to the implementation of the PPP Flexibility Act. Table 3.3 displays results for this specification. In the first column, we show results for the specification with no borrower controls, while, in the second column, we add controls, including fixed effects for two-digit NAICS, business type (sole proprietorship, corporation, etc.), borrower state, urban/rural, and business age.

These results illustrate the differential impact of the policy change across bank types. In particular, consistent with the prediction of our model, banks who face a lower shadow cost of lending (and, therefore, a lower marginal cost of lending) are less responsive to the policy change. A one standard deviation increase in a bank's tier-one leverage ratio is associated with a 4 to 6 percentage point decrease in the change in loan amount between the pre- and post-legislation periods. Given the

[18] https://www.moodysanalytics.com/-/media/article/2011/11-01-03-dodd-frank-act-regulations-minimum-capital-requirements.pdf

[19] For more details, see https://www.elliottdavis.com/ppp-loans-pplf-capital-ratios/. In future work, we plan to exploit the variation across banks in their participation in the PPPLF.

Table 3.3. Heterogeneity by Tier 1 Leverage Ratio - 12 Weeks Around Policy

	(1) Log(Loan Amount)	(2) Log(Loan Amount)
Post-Legislation × Standardized T1 Leverage	0.0479***	0.0547***
	(0.0088)	(0.0086)
Observations	743,215	743,214
Borrower Controls	No	Yes

* p<0.1, ** p<0.05, *** p<0.01

Note: This table presents results for the event-study specifications, allowing for heterogeneity by the lender's tier-one leverage ratio, defined by Equation (3). Standard errors clustered by borrower state are shown in parentheses. Results in both columns include lender FEs are defined as a combination of a lender name, lender city, and lender state. Borrower controls include fixed effects for the two-digit NAICS code, business type (corporation, LLC, sole proprietorship, other), urban/rural, business age ($>$ 2 years, $\leq$ 2 years, unanswered), and borrower state.

mean change of between 6 and 7 percent, this is a substantial amount of heterogeneity across lender types. We are in the process of unpacking this result, but, at the very least, it indicates that lender-side heterogeneity should be a consideration when evaluating the efficacy of the policy.

3.5.3. Program Targeting

The preceding empirical results serve two purposes. First, they validate two main empirical implications of the model. Second, they provide evidence of the types of borrowers who benefit from a more generous forgiveness policy. But, we have yet to address whether the policy change improves the program's targeting. In this section, we examine the targeting question through the final empirical implication. In particular, we examine whether the borrowers brought into the program after the policy change are more or less likely to use funds for payroll purposes.

We first estimate regressions to assess changes to the aggregate payroll share before and after the policy change. Specifically, we estimate specifications of the form:

$$f_{ijt} = \alpha\mathbb{I}(t = Post) + \beta X_{ijt} + \epsilon_{ijt}, \tag{3.4}$$

where f_{ijt} is the share of the loan allocated to payroll at origination. The other variables are defined as before.

Table 3.4 displays results for this specification. Following the policy change, the average payroll share is 0.6 percentage points higher than in the pre period. These results suggest that, in aggregate, a decline in the stringency of forgiveness standards is associated with better targeting. It is important to note that the aggregate estimates capture two sets of responses. They combine (1) a change to borrower composition (through lenders' responses in the threshold loan amount, $\underline{b}$) and (2) changes to behavior of inframarginal borrowers. Because we observe a larger share allocated to payroll when forgiveness standards are more lenient, assuming a static distribution of borrowers, the first channel must necessarily dominate. However, we conduct one further analysis to validate the role of this channel, analyzing the relationship between the loan amount and borrowers' tendencies to use funds for payroll purposes.

Figure 3.5 presents a binned scatter plot of payroll share versus the loan amount. In this plot, the negative relationship between these two quantities is apparent. It

Table 3.4. Aggregate Payroll Share Response - 12 Weeks Around Policy

	(1) Payroll Share (At Orig.)	(2) Payroll Share (At Orig.)
Post-Legislation	0.0063***	0.0055***
	(0.0011)	(0.0013)
Observations	1,013,491	1,013,483
Borrower Controls	No	Yes

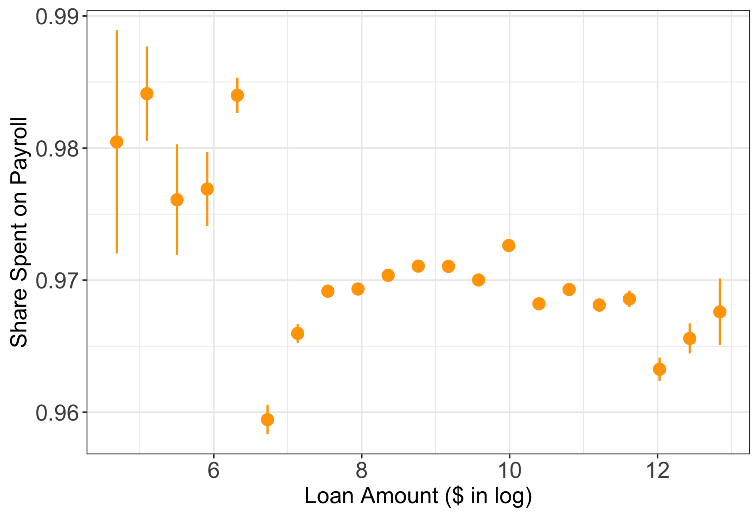
* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Note: This table presents results for the event-study specifications with payroll share on the lefthand side, defined by Equation (3). Standard errors clustered by borrower state are shown in parentheses. Borrower controls include fixed effects for the two-digit NAICS code, business type (corporation, LLC, sole proprietorship, other), urban/rural, business age ($>$ 2 years, $\leq$ 2 years, unanswered), and borrower state.

follows that the marginal borrowers, who receive loans in the post-legislation period, are more likely to use funds on payroll than the inframarginal borrowers. The compositional change leads to an increase in the average propensity, across all borrowers, to use funds for forgivable purposes. Thus, we observe a larger aggregate share dedicated to payroll, as this channel (i.e., the compositional shift) outweighs any adjustment in incentives for the inframarginal borrowers.

There are two important caveats to this analysis. First, as before, the event-study results could be confounded by underlying time trends in the payroll share. Second, we observe only a local change to borrower incentives. For example, the threshold payroll share decreases from 75% to 60%. Our results suggest that inframarginal borrowers do not respond strongly enough to this local adjustment to outweigh the first channel. However, larger changes to the incentive structure of

Figure 3.5. Binned Scatter of Payroll Share vs. Log(Loan Amount) – 12 Week Window



Note: This figure displays the binned scatter plot of payroll share vs. loan amount (in log $). Each point represents the average payroll share of observations within each 20 equally spaced bins of log loan amounts. The error bars represent the standard error of the mean of observations within each bin.

the program could have different effects. This is one motivation for the estimation of a structural model, which is currently in progress, and the consideration of counterfactual policies.

3.5.4. Discussion

The empirical results, despite their limitations, have a number of implications for the design of the PPP. First, they provide evidence to validate the efficacy of the use of borrower-side incentives (e.g., forgiveness standards and the length of the covered period) in expanding credit to borrowers seeking smaller loans. If lenders prefer that loans be fully forgiven, then easing the burden of achieving forgiveness not only affects borrower decisions but also induces lenders to be more generous. Furthermore, because lending-cost heterogeneity drives a substantial portion of the variation in policy responses, the existence of an institution like the PPP Liquidity Facility can play an important role in the policy design. This setup allows the policymaker to effectively subsidize lenders' marginal costs, whereas the subsidy instead addresses up-front fixed costs of issuance.

The second takeaway from the empirical analysis involves the relationship between loan amount and the borrowers' propensity to use funds on payroll. The observed decline in the stringency of forgiveness standards is associated with an increase in the share of funds allocated to payroll. This change suggests a negative relationship loan size and the propensity to use funds for payroll purposes. The marginal borrowers use funds on payroll, and this change in composition offsets any declines in targeting for the inframarginal borrowers. But, as mentioned above, the relative magnitudes of these two effects may differ under alternative policy designs (e.g., different subsidy rates or larger changes to borrower incentives).

Taken together, these results motivate the use of a structural model to analyze these counterfactual designs. While the evidence in this paper suggests that borrower incentives and marginal lending costs are important drivers of lenders' decisions of whether to issue loans, the descriptive evidence is insufficient to quantify the borrower- and lender-side responses to alternative policies. Estimation of our model is currently in progress, and we aim to quantify the separate impact of fixed-cost subsidies, changes to marginal cost, and forgiveness standards on the composition of borrowers receiving loans through the PPP.

3.6. Conclusion

Quantifying the responsiveness of private actors to the incentives provided by public policies is crucial in determining (1) who participates in the programs, (2) who benefits from them, and (3) how well the programs achieve their targeted aims. In this paper, we analyze the design of the PPP, which aims to provide funds to businesses to keep employees on payroll. We develop a model of PPP lending to illustrate the levers available to policymakers to induce program participation and encourage borrowers to use the loans for their intended purpose. We show that a number of underlying primitives determine which of the levers is most effective in targeting loans to those who seek to use funds for payroll purposes. Importantly, the relationship between the loan amount the borrower is seeking and its propensity to use that fund for forgivable purposes (i.e., payroll) is critical in determining the efficacy of the program.

We then exploit variation in the stringency of the program's forgiveness standards to validate a number of salient features of the model and evaluate whether the policy change improved the targeting of the program. Consistent with our model, we find that average loan amounts are lower in the period following the policy change, which is suggestive of an expansion of lending. This expansion disproportionately accrues to newer businesses, as well as sole proprietors, indicating the importance of policy design in determining the program's impact across demographic groups (e.g., business types). Finally, we find that the program's targeting is better in the period following the decline in forgiveness standards, resulting in an increase in the average propensity to use funds on payroll.

We are in the process of taking the model to data with the goal of considering counterfactual policy designs. While our current framework allows us to analyze a single change to borrower incentives, it does not allow us to consider interactions between fixed-cost subsidies, marginal-cost subsidies, and borrower incentives. The counterfactual exercises will illuminate the tradeoffs faced by policymakers and could guide future policy decisions for programs implemented through private actors.

References

Abaluck, Jason, & Gruber, Jonathan. 2011. Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *American Economic Review*, **101**, 1180–1210.

Abaluck, Jason, Gruber, Jonathan, & Swanson, Ashley. 2018. Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets. *Journal of Public Economics*, **164**(aug), 106–138.

Abito, Jose Miguel. 2020. Measuring the Welfare Gains from Optimal Incentive Regulation. *Review of Economic Studies*, **87**, 2019–2048.

American Academy of Actuaries. 2020. Health Insurance Risk Mitigation Mechanisms and COVID-19. *Issue Brief*.

Autor, David, Cho, David, Crane, Leland D, Lutz, Byron, Montes, Joshua, Peterman, William B, Ratner, David, Villar, Daniel, & Yildirmaz, Ahu. 2020. An Evaluation of the Paycheck Protection Program Using Administrative Payroll Microdata. *NBER Working Paper*.

Balyuk, Tetyana, Prabhala, Nagpurnanand, & Puri, Manju. 2020. Indirect Costs of Government Aid and Intermediary Supply Effects: Lessons From the Paycheck Protection Program. *NBER Working Paper*.

Baron, David P, & Besanko, David. 1984. Regulation, Asymmetric Information, and Auditing. *The RAND Journal of Economics*, **15**(4), 447–470.

Baron, David P, & Besanko, David. 1987. Monitoring, Moral Hazard, Asymmetric Information, and Risk Sharing in Procurement Contracting. *The RAND Journal of Economics*, **18**(4), 509–532.

Baron, David P, & Myerson, Roger B. 1982. Regulating a Monopolist with Unknown Costs. *Econometrica*, **50**(4), 911–930.

Barraza, Santiago, Rossi, Martin, & Yeager, Timothy J. 2020. The Short-Term Effect of the Paycheck Protection Program on Unemployment. *SSRN Electronic Journal*.

Bartik, Alexander W, Cullen, Zoe B, Glaeser, Edward L, Luca, Michael, Stanton, Christopher T, Sunderam, Adi, & Bartik, W. 2020. The Targeting and Impact of Paycheck Protection Program Loans to Small Businesses Alexander. *NBER Working Paper*.

Barwick, Panle Jia, Kalouptsidi, Myrto, & Zahur, Nahim Bin. 2021. *Industrial Policy Implementation: Empirical Evidence from China's Shipbuilding Industry*. Cato Institute.

Berry, Steven, Levinsohn, James, & Pakes, Ariel. 1995. Automobile Prices in Market Equilibrium. *Econometrica*, **63**(4), 841–890.

Berry, Steven T, & Haile, Philip A. 2022. Nonparametric Identification of Differentiated Products Demand Using Micro Data *. *Working Paper*.

Bovbjerg, Randall R, Garrett, A Bowen, Clemans-Cope, Lisa, & Masi, Paul. 2008. Reinsurance in State Health Reform State Coverage Initiatives. *Robert Wood Johnson Foundation*.

Brocas, Isabelle, Chan, Kitty, & Perrigne, Isabelle. 2006. Regulation under Asymmetric Information in Water Utilities. *American Economic Review*, **96**(2), 62–66.

Brown, Jason, Duggan, Mark, Kuziemko, Ilyana, & Woolston, William. 2014a. How does risk selection respond to risk adjustment? New evidence from the Medicare Advantage Program. *American Economic Review*, **104**(10), 3335–64.

Brown, Jason, Duggan, Mark, Kuziemko, Ilyana, & Woolston, William. 2014b. How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program âĂă. *American Economic Review*, **104**(10), 3335–3364.

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P.H.C. 2021. Difference-in-Differences with a Continuous Treatment. *Working Paper*.

Carey, Colleen. 2017. Technological change and risk adjustment: Benefit design incentives in medicare part D. *American Economic Journal: Economic Policy*, **9**(1), 38–73.

CBO. 2020. *Federal Subsidies for Health Insurance Coverage for People Under 65 CONGRESS OF THE UNITED STATES CONGRESSIONAL BUDGET OFFICE*. Tech. rept. Congressional Budget Office.

Chernenko, Sergey, & Scharfstein, David S. 2021. Racial Disparities in the Paycheck Protection Program. *SSRN Electronic Journal*.

Chetty, Raj, Friedman, John N., Hendren, Nathaniel, Stepner, Micahel, & The Oppourtnity Insigits Team. 2020. How Did Covid-19 and Stabilization Policies Affect Spending and Employment? a New Real-Time Economic Tracker Based on Private Sector Data. *NBER Working Paper*.

Chorniy, Anna, Miller, Daniel, & Tang, Tilan. 2020. Mergers in Medicare Part D: Assessing market power, cost efficiencies, and bargaining power. *International Journal of Industrial Organization*, **68**(jan).

Cicala, Steve, Lieber, Ethan M J, & Marone, Victoria. 2019. Regulating Markups in US Health Insurance. *American Economic Journal: Applied Economics*, **11**(4), 71–104.

Cororaton, Anna, & Rosen, Samuel. 2021. Public Firm Borrowers of the US Paycheck Protection Program. *Working Paper*.

Curto, Vilsa, Einav, Liran, Levin, Jonathan, & Bhattacharya, Jay. 2021. Can health insurance competition work? Evidence from medicare advantage. *Journal of Political Economy*, **129**(2), 570–606.

Dalton, Christina M, Gowrisankaran, Gautam, & Town, Robert J. 2020. Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D. *Review of Economic Studies*.

De Groote, Olivier, & Verboven, Frank. 2019. Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems. *American Economic Review*, **109**(6), 2137–72.

Decarolis, Francesco. 2015. Medicare Part D: Are Insurers Gaming the Low Income Subsidy Design? *American Economic Review*, **105**(4), 1547–1580.

Decarolis, Francesco, Polyakova, Maria, & Ryan, Stephen P. 2020a. Subsidy Design in Privately Provided Social Insurance: Lessons from Medicare Part D. *Journal of Political Economy*.

Decarolis, Francesco, Polyakova, Maria, & Ryan, Stephen P. 2020b. Subsidy design in privately provided social insurance: Lessons from Medicare Part D. *Journal of Political Economy*, **128**(5), 1712–1752.

Doniger, Cynthia, & Kay, Benjamin. 2021. Ten Days Late and Billions of Dollars Short: The Employment Effects of Delays in Paycheck Protection Program Financing. *Fed Working Paper*.

Drake, Coleman, Fried, Brett, & Blewett, Lynn A. 2019. Estimated Costs of a Reinsurance Program to Stabilize the Individual Health Insurance Market: National- and State-Level Estimates. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, **56**, 0046958019836060.

Dubois, Pierre, & Lasio, Laura. 2018. Identifying Industry Margins with Price Constraints: Structural Estimation on Pharmaceuticals. *American Economic Review*, **108**(12), 3685–3724.

Einav, Liran, Finkelstein, Amy, & Schrimpf, Paul. 2015. THE RESPONSE OF DRUG EXPENDITURE TO NONLINEAR CONTRACT DESIGN: EVIDENCE FROM MEDICARE PART D. *Quarterly Journal of Economics*.

Einav, Liran, Finkelstein, Amy, Kluender, Raymond, & Schrimpf, Paul. 2016. Beyond Statistics: The Economic Content of Risk Scores. *American Economic Journal: Applied Economics*, **8**(2), 195–224.

Einav, Liran, Finkelstein, Amy, & Polyakova, Maria. 2018. Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D. *American Economic Journal: Economic Policy*.

Einav, Liran, Finkelstein, Amy, & Tebaldi, Pietro. 2019. Market design in regulated health insurance markets: Risk adjustment vs. subsidies. *Unpublished mimeo, Stanford University, MIT, and University of Chicago*, **7**, 32.

Erel, Isil, & Liebersohn, Jack. 2020. Does Fintech Substitute for Banks? Evidence from the Paycheck Protection Program. *SSRN Electronic Journal*.

Fairlie, Robert W., & Fossen, Frank M. 2021. Did the $660 Billion Paycheck Protection Program and $220 Billion Economic Injury Disaster Loan Program Get Disbursed to Minority Communities in the Early Stages of COVID-19? *NBER Working Paper*.

Fama, Eugene F, & Jensen, Michael C. 1983. Separation of Ownership and Control SEPARATION OF OWNERSHIP AND CONTROL*. *The Journal of Law & Economics*, **26**(2), 301–325.

Faulkender, Michael, Jackman, Robert, Miran, Stephen, Jackman, Robert, & Miran, Stephen. 2021. The Job-Preservation Effects of Paycheck Protection Program Loans. *Office of Economic Policy Working Paper*.

Geruso, Michael, & Layton, Timothy. 2020. Upcoding: Evidence from Medicare on Squishy Risk Adjustment. *Journal of Political Economy*.

Geruso, Michael, Layton, Timothy, & Prinz, Daniel. 2019. Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges. *American Economic Journal: Economic Policy*, **11**(2), 64–107.

Goolsbee, Austan, & Petrin, Amil. 2004. The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV. *Econometrica*, **72**(2), 351–381.

Granja, João, Makridis, Christos, Yannelis, Constantine, & Zwick, Eric. 2021. Did the Paycheck Protection Program Hit The Target? *NBER Working Paper*.

Hall, Brian J, & Liebman, Jeffrey B. 1998. Are CEOS Really Paid Like Bureaucrats? *The Quarterly Journal of Economics*, **113**(3), 653–691.

Ho, Kate, Hogan, Joseph, & Scott Morton, Fiona. 2017. The impact of consumer inattention on insurer pricing in the Medicare Part D program. *The RAND Journal of Economics*, **48**(4), 877–905.

Holmes, Jonathan. 2021. Selection Across US Health Insurance Markets.

Hortaçsu, Ali, Madanizadeh, Ali, & Puller, Steven L. 2017. Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market. *American Economic Journal: Economic Policy*.

House of Representatives. 2020. *HEALTH AND ECONOMIC RECOVERY OMNIBUS EMERGENCY SOLUTIONS (HEROES) ACT*.

Howell, Sabrina T, Kuchler, Theresa, Snitkof, David, Stroebel, Johannes, & Wong, Jun. 2021. Racial Disparities in Access to Small Business Credit: Evidence from the Paycheck Protection Program. *NBER Working Paper*.

Hubbard, Glenn, & Strain, Michael R. 2020. Has the Paycheck Protection Program Succeeded? *NBER Working Paper*.

Humphries, John Eric, Neilson, Christopher A., & Ulyssea, Gabriel. 2020. Information frictions and access to the Paycheck Protection Program. *Journal of Public Economics*, **190**, 104244.

James, Christopher, Lu, Jing, & Sun, Yangfan. 2021. Time is Money: Real Effects of Relationship Lending: Evidene from Community Bank Lending in the Paycheck Protection Program. *Working Paper*.

Joaquim, Gustavo, & Netto, Felipe. 2021. Bank Incentives and the Effect of the Paycheck Protection Program. *Working Paper*.

Kahane, Yehuda. 1979. THE THEORY OF INSURANCE RISK PREMIUMS-A RE-EXAMINATION IN THE LIGHT OF RECENT DEVELOPMENTS IN CAPITAL MARKET THEORY. *ASTIN Bulletin*.

Ketcham, Jonathan D, Lucarelli, Claudio, Miravete, Eugenio J, & Roebuck, M Christopher. 2012. Sinking, Swimming, or Learning to Swim in Medicare Part D. *American Economic Review*, **2012**(6), 2639–2673.

Ketcham, Jonathan D, Lucarelli, Claudio, & Powers, Christopher A. 2015. Paying Attention or Paying Too Much in Medicare Part D. *American Economic Review*, **105**(1), 204–233.

Kim, Paul, & Li, Anran. 2022. Consumer Subsidy vs. Reinsurance Subsidy in ACA Individual Health Insurance Market. *Working paper*.

Kim, Paul HS. 2022. Risk-Corridors in Medicare Part D: Financial Risk-Sharing or Profit-Limiting Mechanism? *Working Paper*.

Kling, Jeffrey R, Mullainathan, Sendhil, Shafir, Eldar, Vermeulen, Lee C, & Wrobel, Marian V. 2012. COMPARISON FRICTION: EXPERIMENTAL EVIDENCE FROM MEDICARE DRUG PLANS. *Quarterly Journal of Economics*.

Koijen, Ralph SJ, & Yogo, Motohiro. 2015a. The cost of financial frictions for life insurers. *American Economic Review*, **105**(1), 445–75.

Koijen, Ralph S.J., & Yogo, Motohiro. 2015b. The Cost of Financial Frictions for Life Insurers. *American Economic Review*, **105**(1), 445–475.

Koijen, Ralph S.J., & Yogo, Motohiro. 2016a. Shadow Insurance. *Econometrica*, **84**(3), 1265–1287.

Koijen, Ralph S.J., & Yogo, Motohiro. 2016b. Shadow Insurance. *Econometrica*, **84**(3), 1265–1287.

Koijen, Ralph S.J., & Yogo, Motohiro. 2022a. The Fragility of Market Risk Insurance. *Journal of Finance*, **77**(2), 815–862.

Koijen, Ralph S.J., & Yogo, Motohiro. 2022b. The Fragility of Market Risk Insurance. *Journal of Finance*, **77**(2), 815–862.

Laffont, Jean-Jacques, & Tirole, Jean. 1986. Using Cost Observation to Regulate Firms. *Journal of Political Economy*, **94**(3), 614–641.

Lavetti, Kurt, & Simon, Kosali. 2018. Strategic Formulary Design in Medicare Part D Plans. *American Economic Journal: Economic Policy*.

Layton, Timothy J. 2017. Imperfect risk adjustment, risk preferences, and sorting in competitive health insurance markets. *Journal of health economics*, **56**, 259–280.

Layton, Timothy J., Mcguire, Thomas G., & Sinaiko, Anna D. 2016a. RISK CORRIDORS AND REINSURANCE IN HEALTH INSURANCE MARKETPLACES. *American Journal of Health Economics*.

Layton, Timothy J, McGuire, Thomas G, & Sinaiko, Anna D. 2016b. Risk corridors and reinsurance in health insurance marketplaces: insurance for insurers. *American journal of health economics*, **2**(1), 66–95.

Layton, Timothy J, McGuire, Thomas G, & Van Kleef, Richard C. 2018. Deriving risk adjustment payment weights to maximize efficiency of health insurance markets. *Journal of health economics*, **61**, 93–110.

Li, Lei, & Strahan, Philip E. 2020. Who Supplies PPP Loans (And Does It Matter)? Banks, Relationships and the COVID Crisis. *NBER Working Paper*.

Lopez, J.A., & Spiegel, M.M. 2021. Small Business Lending Under the PPP and PPPLF Programs. *Federal Reserve Bank of San Francisco Working Paper*.

Lucarelli, Claudio, Prince, Jeffrey, & Simon, Kosali. 2012. The welfare impact of reducing choice in medicare part D: A comparison of two regulation strategies. *International Economic Review*, **53**(4), 1155–1177.

McGuire, Thomas G, Schillo, Sonja, & Van Kleef, Richard C. 2020. Reinsurance, repayments, and risk adjustment in individual health insurance: Germany, the Netherlands, and the US marketplaces. *American Journal of Health Economics*, **6**(1), 139–168.

NAIC. 2011. *Risk-Based Capital (RBC) For Health Organizations Model Act*.

Polyakova, Maria. 2016. Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D. *American Economic Journal: Applied Economics*, **8**(3), 165–195.

Polyakova, Maria, & Ryan, Stephen P. 2019. *Subsidy targeting with market power*. Tech. rept. National Bureau of Economic Research.

Polyakova, Maria, Bhatia, Vinayak, & Bundorf, M Kate. 2021. Analysis of publicly funded reinsuranceâĂŤgovernment spending and insurer risk exposure. *Pages e211992–e211992 of: JAMA Health Forum*, vol. 2. American Medical Association.

Sacks, Daniel W., Vu, Khoa, Huang, Tsan Yao, & Karaca-Mandic, Pinar. 2021a. How do insurance firms respond to financial risk sharing regulations? Evidence from the Affordable Care Act. *Health Economics*, **30**(6), 1443–1460.

Sacks, Daniel W, Vu, Khoa, Huang, Tsan-Yao, & Karaca-Mandic, Pinar. 2021b. How do insurance firms respond to financial risk sharing regulations? Evidence from the Affordable Care Act. *Health Economics*, **30**(6), 1443–1460.

Springel, Katalin. 2021. Network externality and subsidy structure in two-sided markets: Evidence from electric vehicle incentives. *American Economic Journal: Economic Policy*, **13**(4), 393–432.

Starc, Amanda, & Town, Robert J. 2020. Externalities and Benefit Design in Health Insurance. *Review of Economic Studies*, **87**, 2827–2858.

Tebaldi, Pietro. 2017. Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA. *Becker Friedman Institute for Research in Economics Working paper*.

Center for Medicare & Medicaid Services. 2021. State Innovation Waivers: State-Based Reinsurance Programs.

Walter, John. 2019. US Bank Capital Regulation: History and Changes Since the Financial Crisis. *Economic Quarterly*, **105**(1), 1–40.

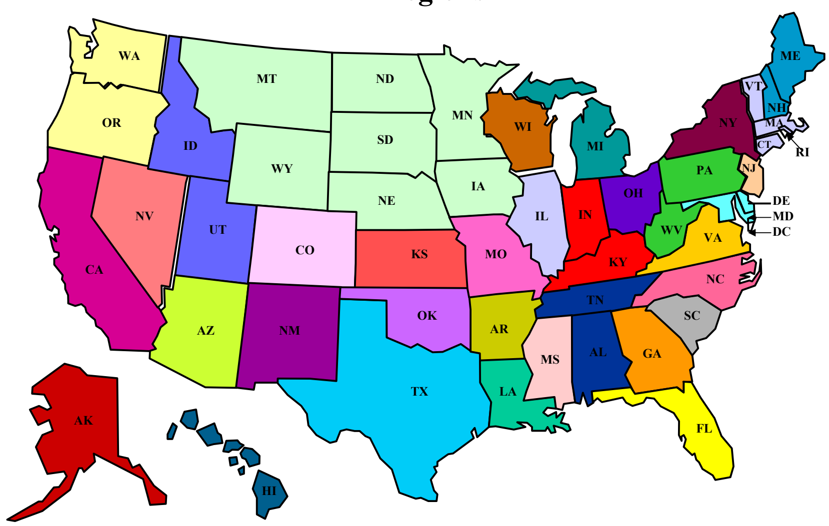
Wang, Jeffrey, & Zhang, David Hao. 2021. The Cost of Banking Deserts : Racial Disparities in Access to PPP Lenders and their Implications. *SSRN Electronic Journal*.

APPENDIX A

Additional Details for Chapter 1

A.1. Additional Figures

Figure A.1
PDP Regions



Note: Each territory is its own PDP region.

Figure shows the 34 PDP regions in the U.S.

Figure A.3. Persistence of Risk-Corridor Payment

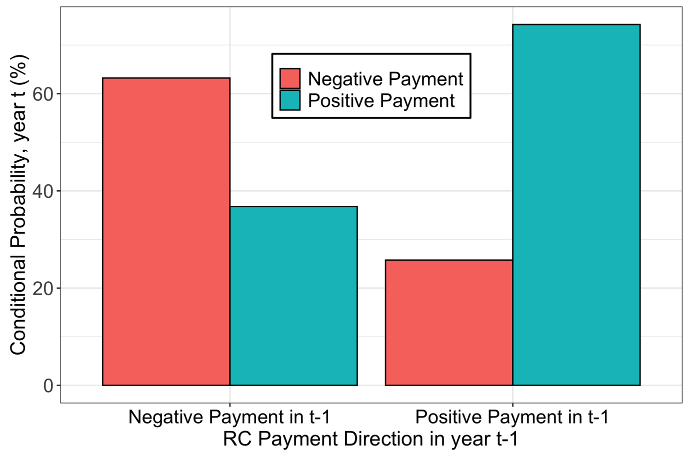
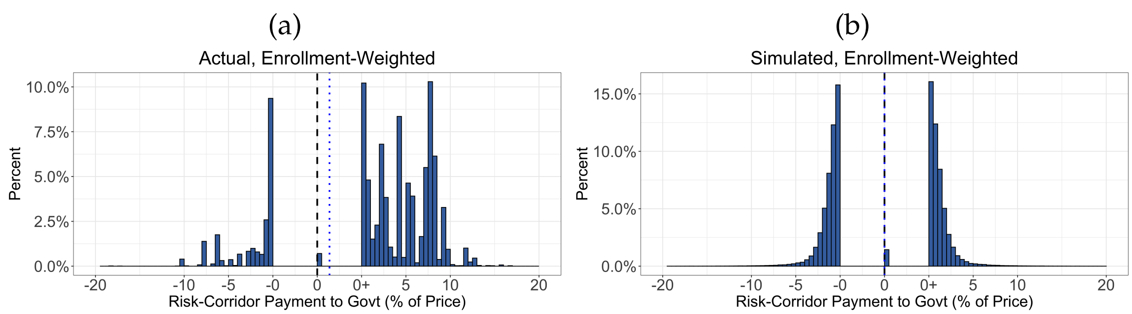


Figure plots the conditional probably of risk-corridor payment being positive vs. negative as a function of insurer's risk-corridor payment direction in the prior year.

Figure A.5. Distribution of Risk-Corridor Payments



Panel (a) plots the distribution of observed contract-level risk-corridor payments from 2009-2015. Panel (b) plots the distribution of simulated risk-corridor payments from 2009-2015 using the claims data. The distributions are weighted by observed enrollment.

Figure A.7. Marginal Cost Estimates: Standard vs. Enhanced Plans

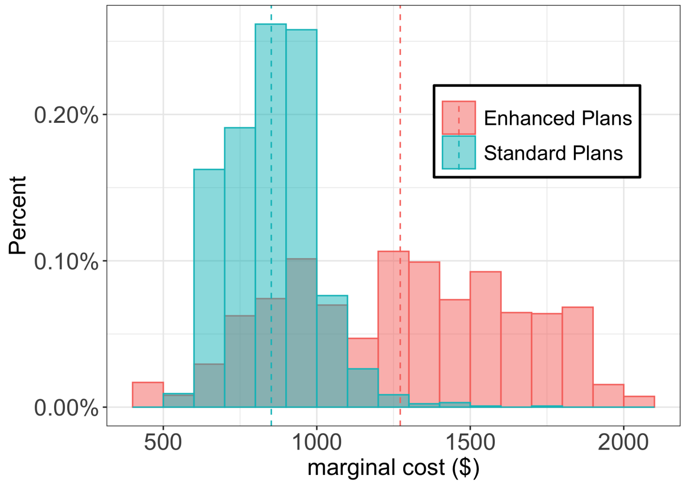


Figure plots the distribution of marginal cost estimates separately for standard plans vs. actuarially enhanced plans. Standard plans refer to plans that meet the basic/minimum benefit design and enhanced plans refer to plans that have increased cost-sharing benefits above the standard benefit design. Each observation is plan-year.

Figure A.9. Model fit of MC using Decarolis et al

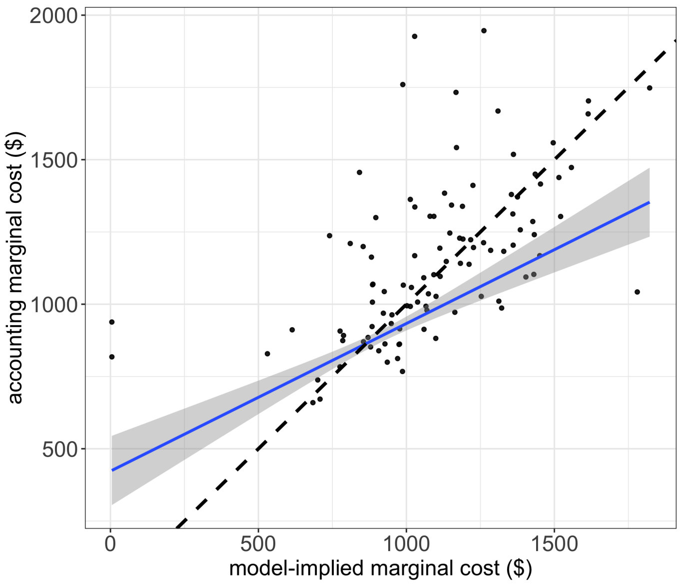


Figure plots the marginal cost estimates obtaining using Decarolis *et al.* (2020a) approach vs. the observed per-enrollee risk-corridor payments at the firm-year level. The dashed line indicates the 45-degree line, and the blue line shows the best-fit line.

Figure A.11. Model fit of risk-corridor payments

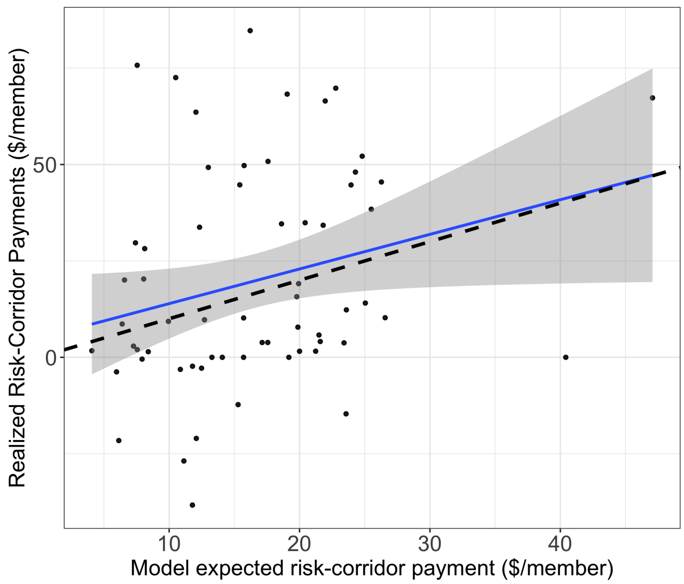
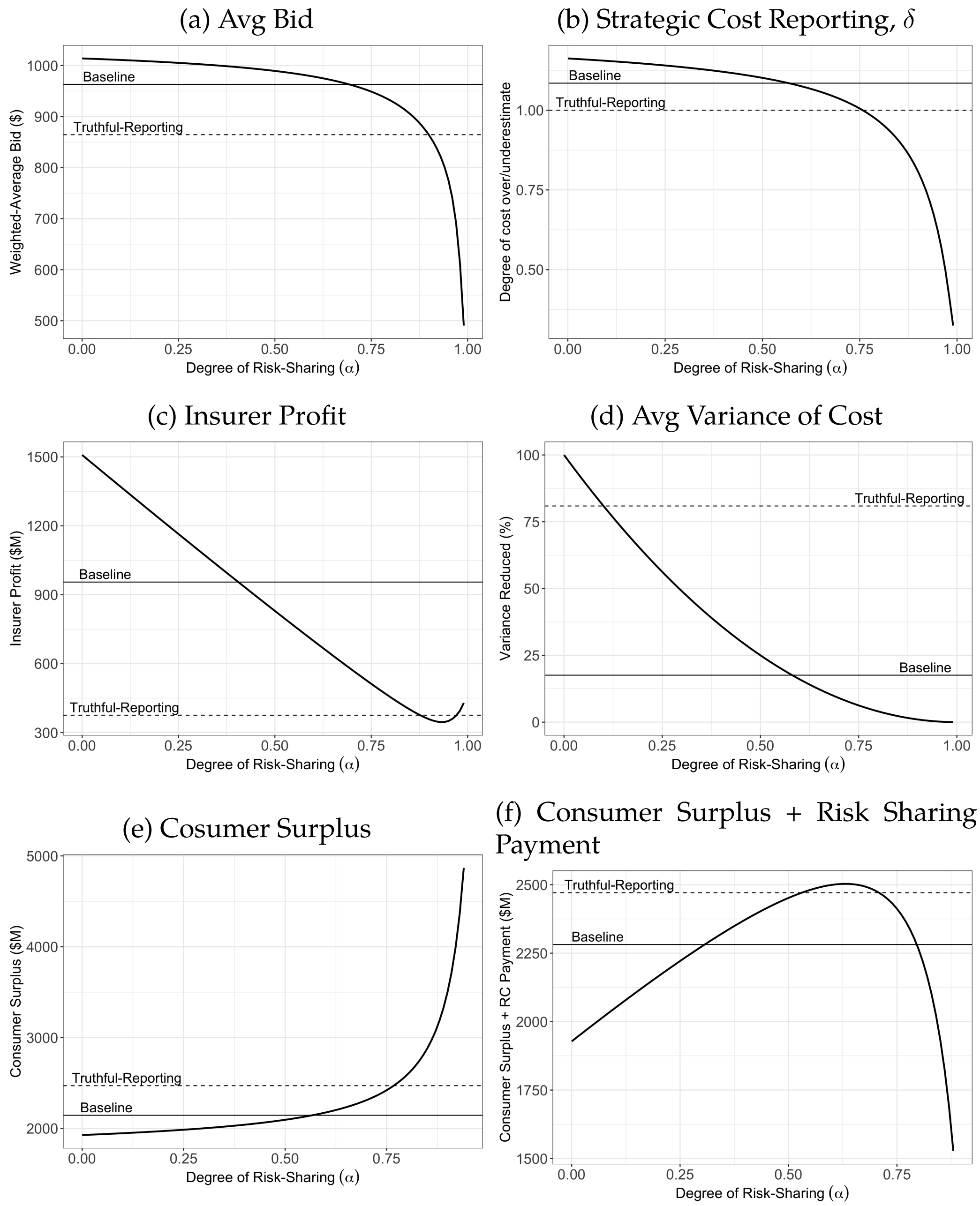


Figure plots the model implied expected per-enrollee risk-corridor payments vs. the observed per-enrollee risk-corridor payments at the firm-year level. The dashed line indicates the 45-degree line, and the blue line shows the best-fit line.

Figure A.13. Market level Variables vs. Degree of Risk Sharing, α



Panel (a) plots the average bid. Panel (b) plots average δ, degree of strategic cost reporting parameter across firms. Panel (c) plots the total insurer profit. Panel (d) plots the average variance of cost relative to the case w/o any risk sharing (i.e. $\alpha = 0$). All averages are computed by taking the enrollment-weighted average across all plans in the market. Panel (e) plots the total consumer surplus, and panel (f) plots the sum of total consumer surplus and risk corridor payments from insurers to the government. The solid horizontal line indicates the baseline numbers and the dashed horizontal line indicates the truthful reporting counterfactual.

A.2. Additional Tables

Table A.1. Summary Statistics for Low-Income Subsidy Elligible Enrollees

	Mean (1)	Std.Dev. (2)	Min (3)	Max (4)
A. Plan-Level				
Bid ($)	1,191	367	596	2,618
Enrollee Premium ($)	298	342	0.0	1,831
Enrollment (000)	7.7	21.4	0.01	409.0
B. Market-Level				
No of Plans	31.19	3.00	23	39
No of Insurers	13.65	1.34	10	17
Enrollment (000)	240	195	14	1,021
HHI Index*	1,965	593	1,106	4,252
Market Share of Top 3 Firms (%)*	64	11	44	91
Market Share of Top 5 Firms (%)*	82	8	60	97

Notes: the table shows summary statistics of the Part D stand-alone prescription drug (PDP) market from 2012-2015 in the 34 PDP regions for low-income subsidy (LIS) eligible enrollees. Plan-level data shows summary statistics taken across individual year-market-plan. Market-level data shows summary statistics taken across year-market level. Enrollee premium refers to premium faced by LIS enrollees. An insurer is defined as a unique parent organization in the CMS data. * HHI index and market share of top firms are computed using LIS enrollees only.

Table A.2. Marginal Cost vs. Plan Characteristics

	Dependent variable:			
	mc_{jm}			
	(1)	(2)	(3)	(4)
deductible	-0.55^{***}	-0.47^{***}	-0.46^{***}	-0.48^{***}
	(0.05)	(0.05)	(0.05)	(0.05)
isExtraCovgGap	493.41^{***}	481.14^{***}	481.00^{***}	468.55^{***}
	(11.85)	(11.35)	(11.24)	(11.12)
isEnhanced	3.83	35.27^{**}	36.22^{**}	45.83^{***}
	(16.30)	(15.60)	(15.44)	(15.17)
n_drugs_tier1	0.03^{***}	0.01	0.01	0.001
	(0.004)	(0.004)	(0.004)	(0.01)
Observations	2,661	2,661	2,661	2,661
R^2	0.64	0.67	0.68	0.75
Year FE	N	Y	Y	Y
Market FE	N	N	Y	Y
Firm FE	N	N	N	Y
Observations	2,661	2,661	2,661	2,661
R^2	0.64	0.67	0.68	0.75

Note: *p<0.1; **p<0.05; ***p<0.01

The table shows a hedonic regression of marginal cost estimates on observable plan characteristics.

A.3. Details on the Stylized Model

A.3.1. Generalization of Stylized Model

Here we show that the stylized model in 1.3.1 is an approximation to a model in which insurer faces a financial frictional loss function. Consider the same setting, but now the insurer faces some convex financial frictional loss function:

$$\max_{p} \quad pq(p) - \underbrace{cq(p)}_{E[\tilde{C}]} - \underbrace{E_{\tilde{C}}\left[L(\overbrace{pq(p) - \tilde{C}}^{\pi})\right]}_{\text{expected financial frictions cost}} \tag{A.1}$$

where $L()$ is a continuous, non-decreasing, and convex function. Taking the FOC yields:

$$p^{*}\left(1 + \frac{1}{\epsilon^{D}}\right) = c + \underbrace{\frac{E\left[L(\pi)\frac{\frac{\partial f(\tilde{C})}{\partial \tilde{C}}}{f(\tilde{C})}\right]}{1 - E[L'(\pi)]}}_{\text{marginal financial frictional cost}} \tag{A.2}$$

The FOC in (A.4) yields similar form as (1.2), except that the marginal financial frictional cost takes the place of the marginal risk charge term in the original model. The marginal financial frictional cost term above is a function of the loss function $L()$ and the distribution of the total cost $\tilde{C}$. Hence, if we parametrize $L()$ function upto some parameter ρ and take the second moment of the cost distribution $\tilde{C}$ i.e. $V(\tilde{C})$ to describe the cost distribution then we can take the original stylized model to be an first order approximation of the above model in (A.1).

Similarly, if we assume a model in which the insurer is risk-averse where the insurer's objective function is now:

$$\max_{p} \; E_{\tilde{C}}\big[\overbrace{pq(p) - \tilde{C}}^{\pi}\big] \tag{A.3}$$

where $U()$ is some continuous, non-decreasing and concave utility function. The FOC yields:

$$p^* \left(1 + \frac{1}{\epsilon^D}\right) \; = \; c \; + \; \underbrace{\frac{E\left[U(\pi)\frac{\frac{\partial f(\tilde{C})}{\partial \tilde{C}}}{f(\tilde{C})}\right]}{-E[U'(\pi)]}}_{\text{marginal risk disutility}} \tag{A.4}$$

A.3.2. Stylized Model Proof

Proposition 1: $\delta^* = \overline{\delta}$ will always be optimal for the insurer's problem in (1.5). And furthermore if assumption 1 holds, $p_m^*(\delta^*) = p_0^*$ where $p_m^*(\delta)$ denotes the insurer's profit-maximizing price in (1.5) for a given δ.

Proof: I begin by proving that it is always optimal for the insurer to choose $\delta = \overline{\delta}$. First note that insurer's choice of δ does not directly affect its objective function. δ only affects insurer's margin constraint which is relaxed the most when $\delta = \overline{\delta}$, allowing the insurer to choose any price $p \leq \bar{m}\overline{\delta}c$. Hence, it is always optimal for the insurer to choose $\delta^* = \overline{\delta}$.

Next, if assumption 1 holds then insurer can choose the maximum $\delta = \overline{\delta}$ in which case $p_0^* \leq \bar{m}\overline{\delta}c$ by the assumption and as a result continue to charge its optimal price without the margin constraint of p_0^*. ■

Proposition 2: Under Assumption 2, optimal δ^* to the insurer's problem in (1.8) will be $\underline{\delta}$ i.e. insurer will always report lowest possible expected cost. And furthermore $p_{rc}^*(\underline{\delta}) < p_{rc}^*(1) \leq p_0^*$ where $p_{rc}^*(\delta)$ denotes insurer's profit-maximizing price in (1.8) for a given δ.

Proof: Without loss of generality, I restrict my attention to a simplified risk-corridor function. A simple risk-corridor function can be written as:

(A.5)
$$T(\tilde{C}, C) = \begin{cases} \alpha(0.95C - \tilde{C}) & \text{if } \tilde{C} < 0.95C \\ 0 & \text{if } 0.95C \leq \tilde{C} \leq 1.05C \\ \alpha(1.05C - \tilde{C}) & \text{if } \tilde{C} > 1.05C \end{cases}$$

where $\alpha \in [0, 1]$ is the degree of risk-sharing parameter. I can re write the ex-post total cost with strategic cost-reporting parameter δ as

$$\tilde{C}^{rc}(\delta) = \tilde{C} + T(\tilde{C}, \delta C) = C(x + T(x, \delta))$$

where $x = \tilde{C}/C$. I begin by showing that $\frac{\partial E[T(x,\delta)]}{\partial \delta} > 0 \, \forall \, \delta$.

$$
\begin{aligned}
E[T(x,\delta)] &= \int_0^{0.95\delta} \alpha(0.95\delta - x)dF(x) + \int_{1.05\delta}^{\infty} \alpha(1.05\delta - x)dF(x) \\
&= E[\alpha(0.95\delta - x)|x < 0.95\delta]Pr[x < 0.95\delta] \\
&\quad + E[\alpha(1.05\delta - x)|x < 1.05\delta]Pr[x > 1.05\delta] \\
\frac{\partial E[T(x,\delta)]}{\partial \delta} &= \alpha \int_0^{0.95\delta} 0.95dF(x) + \alpha \int_{1.05\delta}^{\infty} 1.05dF(x) \\
&= \alpha \left(0.95Pr[x < 0.95\delta] + 1.05Pr[x > 1.05\delta]\right) > 0
\end{aligned}
$$

Next, I show how the variance of cost changes as a function of δ. The variance of total cost with RC can be decomposed into three components:

$$
Var(x + T(x,\delta)) = Var(x) + \underbrace{Var(T(x,\delta)) + 2Cov(x, T(x,\delta))}_{V(\delta)}
$$

The first term is variance of x, where as the last two terms are the variance of the RC payment and the covariance between x and the RC payment.

$$
\begin{aligned}
V(\delta) &= E[\alpha(0.95\delta - x)\big(\alpha(0.95\delta - x) + 2(x-1)\big)|x < 0.95\delta]Pr[x < 0.95\delta] \\
&\quad + E[\alpha(1.05\delta - x)\big(\alpha(1.05\delta - x) + 2(x-1)\big)|x > 1.05\delta]Pr[x > 1.05\delta] - E[T(x,\delta)]^2 \\
\frac{\partial V(\delta)}{\delta} &= 1.9\alpha \int_0^{0.95\delta} \alpha(0.95\delta - x) + (x-1)dF(x) + 2.1\alpha \int_{1.05\delta}^{\infty} \alpha(1.05\delta - x) + (x-1)dF(x) \\
&\quad - 2E[T(x,\delta)]\frac{\partial E[T(x,\delta)]}{\partial \delta} \\
&= 1.9\alpha E[\alpha(0.95\delta - x) + (x-1) - E[T]|x < 0.95\delta]Pr[x < 0.95\delta] \\
&\quad + 2.1\alpha E[\alpha(1.05\delta - x) + (x-1) - E[T]|x > 1.95\delta]Pr[x > 1.95\delta]
\end{aligned}
$$

Given the above derivations, I present the following lemma.

Lemma 1. *There exists a unique* $\delta_0 \in (\underline{\delta}, \overline{\delta})$ *such that* $\begin{cases} E[T(x,\delta)] < 0,\ \frac{\partial V(\delta)}{\delta} > 0 & \textit{if}\ \delta < \delta_0 \\ E[T(x,\delta)] = 0,\ \frac{\partial V(\delta)}{\delta} = 0 & \textit{if}\ \delta = \delta_0 \\ E[T(x,\delta)] > 0,\ \frac{\partial V(\delta)}{\delta} < 0 & \textit{if}\ \delta > \delta_0 \end{cases}$

I begin by showing that such δ_0 exists for the expected RC payment. From above, I showed that $\frac{\partial E[T(x,\delta)]}{\partial \delta} > 0\ \forall\, \delta$. When $\delta = 0$, $E[T(x,0)] = -\alpha E[x] < 0$. When $\delta \to \infty$, $E[T(x,\delta)] = \infty > 0$. As a result there must be unique $\delta = \delta_0$ such that $E[T(x,\delta_0)] = 0$.

Furthermore, if the distribution of cost $\tilde{C}$ is symmetric around its mean, $\delta_0 = 1$. If the distribution is positively skewed, then $\delta_0 > 1$. Conversely, if the distribution is negatively skewed, then $\delta_0 < 1$. So if I assume that the distribution of $\tilde{C}$ is either symmetric or positively skewed around its mean, then the following lemma will hold true.

Lemma 2. *If the distribution of* $\tilde{C}$ *is symmetric or positively skewed around its mean, then under assumption 2,* $V(x,\underline{\delta}) \leq V(x,\overline{\delta})$.

Given the above set of statements, I now prove the main preposition. δ affects insurer's objective function in two ways: expected risk-corridor payments and the variance of total cost. For the expected risk-corridor payment, I showed that it is always increasing in δ, meaning the firm will want to choose $\delta = \underline{\delta}$ to minimize its expected risk-corridor payment. For the variance, I showed that it is decreasing in the absolute value of $\delta - \delta_0$ where $\delta_0 \approx 1$. That is as the firm over or underestimates its cost, its variance decreases. The minimum of such variance is achieved at either extremes. And with the assumption that the firm's lower bound and upper bound on cost over or underestimation is equal, the firm's variance of cost will also be minimized at $\delta = \underline{\delta}$. As a result, it is optimal for insurer to choose $\delta = \underline{\delta}$.

This will have an intuitive effect on insurer's optimal prices. At $\delta = 1$, insurer's expected cost remains unchanged while its variance of cost will be smaller. As a result, insurer will incur lower marginal risk-charge and hence its optimal price will be lower. At $\delta = \underline{\delta}$, insurer will incur negative expected risk-corridor payment, and its variance of cost will be even lower than at $\delta = 1$. As a result, insurer's effective marginal cost and marginal risk-charge will decrease, lowering its optimal price even further.

■

Proposition 3: Optimal δ^* to the insurer's problem in (1.9) will be $\underline{\delta} \leq \delta^* < 1$ or $1 < \delta^* \leq \overline{\delta}$ if the margin constraint does not bind or if the margin constraint strictly binds at $p^*_{rc}(1)$, respectively. And furthermore insurer's profit-maximizing price, p^*_{both} will be s.t. $p^*_{rc} \leq p^*_{both} \leq p^*_0$.

Proof: Suppose the insurer's margin constraint isn't binding at $\delta = 1$. Then insurer can underestimate its cost by setting $\delta' = 1 - \varepsilon$ for some small $\varepsilon > 0$ without violating the margin constraint. Then from the proof in earlier preposition, the expected risk-corridor payment will decrease and the variance of cost will also decrease. This will increase the insurer's objective function and hence insurer will not report truthfully. Furthermore, because of the decrease in insurer's marginal cost and marginal risk-charge, insurer will now charge a lower price.

Now suppose the insurer's margin constraint is binding at $\delta = 1$. Then it means the insurer will want to charge higher price in the absence of the constraint. Therefore the insurer can overestimate its cost and set $\delta' = 1 + \varepsilon$ for some small $\varepsilon > 0$. Insurer can then increase its price by $\varepsilon \overline{mc}$ which will be closer to the optimal price it would like to charge, increasing its expected profit. ■

A.4. Details on the Supply-Side Model

Here, I provide additional details on the supply-side model presented in (1.12). I first expand on how different parts of the insurer's objective function is constructed.

Insurer's cost is a function of individuals that it enrolls and is altered by the risk-corridor transfers ex-post. Let $\tilde{c}_{ij} = c_{ij} + \varepsilon_{ij}$ denote individual i's ex-post cost for plan j (suppressing the market index) where c_{ij} is the expected cost, and ε_{ij} is the zero-mean ex-post shock. Then the plan's total ex-post cost prior to risk-corridor transfers will be $\tilde{C}_j = \sum_i^{q_j} \tilde{c}_{ij}$ where q_j is the demand for plan j. With the risk-corridor transfers, plan's ex-post cost will be:

$$\tilde{C}_j^{rc} = \tilde{C}_j + T(\tilde{C}_j, \delta E[\tilde{C}_j]) \tag{A.6}$$

Given my model of "risk-averse" insurer, the insurer cares about both the expected value as well as the variance of the cost. The expected cost can be written as:

$$E[\tilde{C}_j^{rc}] = C_j + E[T(\tilde{C}_j, \delta C_j)] = C_j + \underbrace{\overbrace{E\left[T\left(\frac{\tilde{C}_j}{C_j}, \delta\right)\right]}^{\text{expected risk-corridor payment share}}}_{\approx \gamma_j(\delta, q_j)} C_j \tag{A.7}$$

where $C_j = E[\tilde{C}_j]$. The expected cost that insurer faces can be broken down into the expected cost, C_j component prior to any risk-corridor payments and the expected risk-corridor payments. Furthermore, the expected risk-corridor payments can be written as an expected share of expected costs.[1] This expected risk-corridor payment share will be a function of δ and q_j and so can be written as $\gamma_j(\delta, q_j)$, which is part of the insurer's main objective function in (1.12). γ_j is a function of q_j because the risk-corridor transfer, $T()$ is a non-linear function and hence the expected value will depend on higher moments of the random variable, $\tilde{C}/C_j$ which will depend on the demand.[2]

The variance of insurer's cost for the plan will be:

$$Var\left(\tilde{C}_j + T(\tilde{C}_j, \delta C_j)\right) \approx V_j(\delta, q_j) \tag{A.8}$$

which is also a function of δ and the demand, q_j and can be written as $V_j(\delta, q_j)$, which is part of the main objective function in (1.12).

I allow individuals' expected costs to vary by risk-type of the individuals. For an individual i whose risk type is t, his/her expected cost will be $c_{ij} = \kappa_t c_j$ where κ_t is risk-type t's multiplier and c_j is the baseline expected cost of plan j for an average enrollee. The multiplier κ_t is assumed to be same across different plans, meaning the ratio of cost of risk-type t to t' under the given plan is held constant regardless of which plan the risk-types are enrolled in.

[1] This hold true because the risk-corridor transfer function is homogeneous of degree one.

[2] To see this, assume that $c_{ij} = c_j$ and $Var(\varepsilon_{ij}) = \sigma_j^2$. Then $Var(\tilde{C}/C_j) = \frac{\sigma_j^2 q_j}{c_j^2 q_j^2} = \frac{\sigma_j^2}{c_j^2 q_j}$, which is a function of demand q_j.

On the revenue side, each plan will submit bids b_j's to CMS which reflects the plan's premium for an average enrollee. CMS then takes the bid and risk-adjusts the bids according to the risk profile of the individual. So the premium that the plan receives from enrolling risk-type t would be $\theta_t b_j$.[3]

Insurer's expected profit for plan j (without the risk-corridor transfers and the risk-charges), is then

$$\sum_t (\theta_t b_j - \kappa_t c_j) M_t s_j^t(b) \tag{A.9}$$

where M_t and $s_j^t(b)$ is the market-size and demand share function of consumers of risk-type t, respectively. I further make the assumption that there is perfect risk-adjustment i.e. $\theta_t = \kappa_t$.[4] The expected profit can then be re-rewritten as

$$(b_j - c_j) \underbrace{\sum_t \theta_t M_t s_j^t(b)}_{\substack{Q_j(b), \\ \text{risk-adj demand}}} \tag{A.10}$$

I allow 6 different risk-types across individuals; five health-levels for regular enrollees (the same health-level used in demand estimation) and a single type for the LIS enrollees.

[3]Here, CMS is paying the difference between $\theta_t b_j$ and b_j as enrollees are faced with the same premiums regardless of their risk profiles.

[4]As mentioned in section 1.5.2, this is mainly to help with numerical issues in the supply-side estimation

Putting the expected profit with the expected risk-corridor transfers and risk-charges, we have the following objective:

$$\sum_m \sum_{j \in J_m} \left(b_{jm} - c_{jm}\right) \underbrace{Q_{jm}(b)}_{\text{risk-adj demand}} - \underbrace{\gamma_{jm}(\delta, Q_{jm})\, c_{jm} Q_{jm}(b)}_{\text{expected rc payment}} - \underbrace{\rho V_{jm}(\delta, Q_{jm})}_{\text{risk-charge}} \tag{A.11}$$

A.4.1. Enrollee Subsidy

Given plans' bids, CMS sets the enrollee subsidy, S such that the enrollee's premium for purchasing plan j is:

$$p_j^e = \max\left\{0, b_j - \underbrace{(0.745\bar{b} - 0.255\bar{r})}_{S}\right\} \tag{A.12}$$

where $\bar{b}$ is the lagged enrollment-weighted average of all the bids across all the markets in the US and $\bar{r}$ is the average expected reinsurance subsidy per enrollee.[5] The subsidy S is set so that on average government pays for 74.5% of the benefit expenses and enrollee pays for 25.5%. To see this, enrollee's premium for purchasing an average plan would be $\bar{p}_j^e = \bar{b} - S = 0.255(\bar{b} + \bar{r})$.[6]

For the low-income subsidy (LIS) eligible population, they face an even greater subsidy rate. For a LIS enrollee, his/her premium for purchasing plan j in market

[5] In practice, insurers submit each plan's expected per-enrollee reinsurance cost to the government along with their bids. Similar to $\bar{b}$, $\bar{r}$ is the lagged-enrollment weighted average of all the plans' expected reinsurance cost across the markets.

[6] In theory, if b_j is sufficiently low enough enrollee premium for the plan could be 0. However, for the sample period of 2012-2015 regular enrollees faced no zero-premium plans.

m is:

$$p_{jm}^{LIS} = \max\left\{0, b_{jm} - \bar{b}_m\right\} \tag{A.13}$$

where $\bar{b}_m$ is the lagged enrollment-weighted average of bids within the same market, also known as the LIS benchmark premium. LIS enrollees will pay zero premium for plans below the benchmark, which by design there will always be at least one such plan.[7] For plans that are above the benchmark, LIS enrollees will pay the difference between the plan bid and the benchmark.

The above subsidy design poses several challenges in accurately modeling the supply-side due to how the demand share function looks like. For the regular enrollees, the subsidy-level S is a weighted-average of the bids and hence is a function of insurer's own bids. While I could model insurers as internalizing this effect, I assume that insurers take the subsidy-level S as given (i.e. treat it as exogenous). This seems reasonable as there are close to 1000 different plans per year that are used to construct the weighted-average bid and includes both the PDP and the MA-PD bids.

For the LIS consumers, it gets even more difficult. First, the benchmark is constructed at the market level and hence it could be more susceptible to insurers' strategic behaviors (Decarolis, 2015). While this may be problematic when insurers can offer many plans which was the case in the earlier years of Part D market,

[7] In practice, a large portion of LIS enrollees are randomly assigned to plans that are below the benchmark. However, after they are auto-enrolled in the randomly assigned plan, they are free to choose a different one.

for the years of my analysis, insurers are restricted in the number of plans they offer in a single market. More specifically, starting from 2010 CMS imposed a "meaningful difference" requirement across plans that made it harder for insurer to offer more than two plans. In the data, an insurer usually offers one or two plans and at most three plans in a single market. Second, the plans' share function will not be continuously differentiable with respect to their bids at or below the benchmark. This is because LIS premium will be zero and will not change as long as it's at or below the benchmark. As a result, I can't use a standard first-order-condition for these plans.

Similar to Decarolis (2015), I make the following assumptions. I assume that plans whose bids are sufficiently above the benchmark premium face an elastic demand and price optimally according to the demand. I refer these as regular plans. For the bids that are at or below the benchmark premium, I do not model how insurers set those bids but take them as given. So while I can still construct FOC's with respect to bids of plans whose bids are above the benchmark, I can not do the same for the plans whose bids are below the benchmark. I refer these as LIS-distorted plans.

A.4.2. First-Order Conditions

Given the insurer's objective in (1.12) and the above assumptions, we can derive the following first-order-conditions with respect to all the regular plans' bid b_{km}:

$$\begin{aligned}
& b_{km} + \sum_{j \in J_m} (b_{jm} - c_{jm}) \frac{\partial Q_{jm}(b)}{\partial b_{km}} \\
& - \sum_{j \in J_m} \left(\gamma_{jm} c_{jm} \frac{\partial Q_{jm}(b)}{\partial b_{km}} + \frac{\gamma_{jm}(\delta, Q_{jm})}{\partial Q} \frac{\partial - Q_{jm}(b)}{\partial b_{km}} c_{jm} Q_{jm}(b) \right) \\
& - \sum_{j \in J_m} \rho \frac{V_{jm}(\delta, Q_{jm})}{\partial Q} \frac{\partial Q_{jm}(b)}{\partial b_{km}} \\
& - \lambda \left(b_{km} + \sum_{j \in J_m} (b_{jm} - \overline{m} \delta c_{jm}) \frac{\partial Q_{jm}(b)}{\partial b_{km}} \right) = 0
\end{aligned}$$

and FOC's with respect to the strategic cost-reporting term δ:

$$- \sum_m \sum_{j \in J_m} \left(\frac{\partial \gamma_{jm}(\delta, Q_{jm})}{\partial \delta} c_{jm} Q_{jm} + \rho \frac{\partial V_{jm}(\delta, Q_{jm})}{\partial \delta} \right) + \lambda \overline{m} \sum_m \sum_{j \in J_m} c_{jm} Q_{jm} = 0$$

Furthermore, with the binding margin constraint we have that[8]

$$\sum_m \sum_{j \in J_m} b_{jm} Q_{jm} - \overline{m} \sum_m \sum_{j \in J_m} \delta c_{jm} Q_{jm} = 0$$

[8]I assume that the margin constraint is always binding for firms. This can be proven as long as the risk-charge term is not too big. For example, I make a reasonable economic assumption that marginal risk-charge can't be larger than the marginal costs in which case I can show that the margin constraint will be always binding.

We can rewrite the above in a vectorized form:

(A.14)

$$\frac{\partial \mathbf{Q}}{\partial \mathbf{b}}^{-1} \underbrace{\left(\mathbf{Q} + \frac{\partial \mathbf{Q}}{\partial \mathbf{b}} \mathbf{b}\right)}_{\text{MR}} = \underbrace{\frac{(1-\lambda \overline{m} \delta)}{(1-\lambda)} \mathbf{c} + \frac{1}{(1-\lambda)} \qquad \rho \overbrace{\frac{\partial \mathbf{V}}{\partial \mathbf{Q}}}^{\text{marginal risk-charge}} \qquad + \frac{1}{(1-\lambda)} \overbrace{\left(\gamma + \frac{\partial \gamma}{\partial \mathbf{Q}} \mathbf{Q}\right)}^{\text{marginal RC payment}}}_{\text{Effective MC}}$$

(A.15)

$$\text{where} \quad \lambda = \frac{\rho \frac{\partial V}{\partial \delta} - \frac{\partial \gamma}{\partial \delta}' \mathbf{Q}}{\overline{m} \mathbf{c}' \mathbf{Q}}, \quad \delta = \frac{\mathbf{b}' \mathbf{Q}}{\overline{m} \mathbf{c}' \mathbf{Q}}$$

The above FOC while similar to the standard FOC's where marginal revenue equals to the marginal costs is much more complicated. In a standard model, inverting the FOC should yield the marginal cost on the right hand side of the equation. However, here the right hand side is an "effective marginal cost" that is composed of marginal cost, marginal risk-charge as well as the marginal risk-corridor payments, some of which are non-linear due to the margin constraint.

A.5. Details on Simulating Cost Distribution

A.5.1. Computing plan-specific distribution of cost

In this section, I detail how I use the sample claims data to construct a plan-specific sample distribution of cost. The goal is to create a sample distribution of enrollees and their associated claims cost that each plan could be facing.

From the MCBS data, I observe a nationally representative sample of Medicare beneficiaries and their detailed prescription drug consumption information through the whole year. The information includes the date of the prescription drug fill, the quantity, and the specific drug or the NDC code of the drug purchased.

From the CMS Part D Prescription Drug Plan Formulary, Pharmacy Network, and Pricing Information Files, I observe detailed plan benefit design and formulary data. The information includes financial cost-sharing information like the deductible, co-insurance/co-pay rates across different tiers of drugs, drug formularly design (i.e. which set of drugs are in tier 1, tier 2 and so on) as well as plan level average monthly costs for each drug.

With the above two data sets, I can create $N \times J$ total individual-plan level cost. I follow the procedure for each market m. For a given individual i and plan j, I compute the hypothetical cost to the enrollee and the insurer if individual i were to be enrolled in plan j. This is done by feeding in individual i's prescription drug purchase information through plan j's plan formulary/benefit design information.[9] This results in an estimated object: c_{ij}^{plan}, plan j's cost of enrolling in individual i, and c_{ij}^{enrol}, individual i's out of pocket cost of enrolling in plan j with the given prescription drug consumption. After the procedure, I'm left with two matrixes that is N (number of individuals) by J (number of plans); one for plan liable cost, and the other for enrollee liable out of pocket cost. More importantly, for each plan j I'll have a sample distribution of costs of N individuals: $\{c_{ij}^{plan}\}_{i=1}^{N}$.

A.5.2. Computing Enrollee's Risk-Score

Here, I detail how I estimate the risk-adjustment factor across different enrollee types. Given the plan-individual level imputed cost data from A.5.1, I first compute the expected cost across all individual-plans i.e.

$$\bar{c} = \frac{1}{NJ} \sum_{i} \sum_{j} c_{ij}^{plan} \tag{A.16}$$

Then I compute the expected cost of each risk-type across all plans:

$$\bar{c}_t = \frac{1}{N_t J} \sum_{i, r(i)==t} \sum_{j} c_{ij}^{plan} \tag{A.17}$$

The risk-type specific adjustment factor is then computed by

$$\theta_t = \frac{\bar{c}_t}{\bar{c}} \tag{A.18}$$

[9]Here, I'm implicitly assuming there is no moral-hazard i.e. individuals' consumption of drugs do not depend on the plan's benefit generosity.

The resulting factors across different risk-types are shown in table A.3. The risk-adjustment factors follow intuitive patterns where healthier enrollees receive lower risk-adjustment factor.

Table A.3. Enrollee Risk-Type Risk-Adjustment/Cost Factor

Enrollee Health Risk-Type	θ_t
Excellent	0.45
Good	0.68
Fair	0.92
Poor	1.26
Very Poor	1.52
LIS	1.39

Notes: the table shows estimated risk-adjustment/cost factor across different risk-types of individuals.

A.5.3. Simulating $V(\delta, Q)$ and $\gamma(\delta, Q)$

Here, I detail how the variance of total cost subject to risk-corridor, $V(\delta, Q)$ and the expected risk-corridor payment share function, $\gamma(\delta, Q)$ is simulated and then estimated via a 2-dimensional spline method.

From section A.5.1, for each plan j, I have a sample distribution of individual-level cost: $\{c_{ij}^{plan}\}_{i=1}^{N}$. I then take the following steps to get a distribution of total cost that insurers may be facing and compute the associated variance and the expected risk-corridor payment share:

(1) fix a value of Q, the total demand or the number of enrollees in plan j and δ the degree of strategic cost-reporting parameter.

(2) draw Q enrolles from the distribution of cost: $\{c_{ij}^{plan}\}$ to get a vector of cost: $\{e_{ij}\}$

(3) then compute the total cost incurred to the plan: $C_j = \sum_i^Q e_{ij}$

(4) Repeat steps 2 to 3 from $k = 1$ to K times to get a distribution of total cost that the plan could be facing: $\{C_{j,k}\}_{k=1}^K$

(5) Compute the expected total cost as $\bar{C}_j = \frac{1}{K}\sum_k^K C_{j,k}$

(6) Apply the risk-corridor function to each k^{th} draw of the total cost i.e.

$$C_{j,k}^{rc} = C_{j,k} + T(C_{j,k}, \delta\bar{C}_j) \tag{A.19}$$

where $T()$ is the ex-post risk-corridor function in (1.7).

(7) compute the variance of total cost and the expected risk-corridor share function as:

$$V(\delta, Q) = \frac{1}{K-1}\sum_{k=1}^{K}(C_{j,k}^{rc} - \bar{C}_j^{rc})^2 \tag{A.20}$$

$$\gamma(\delta, Q) = \frac{1}{K}\sum_{k=1}^{K}\frac{T(C_{j,k}, \delta\bar{C}_j)}{\bar{C}_j} \tag{A.21}$$

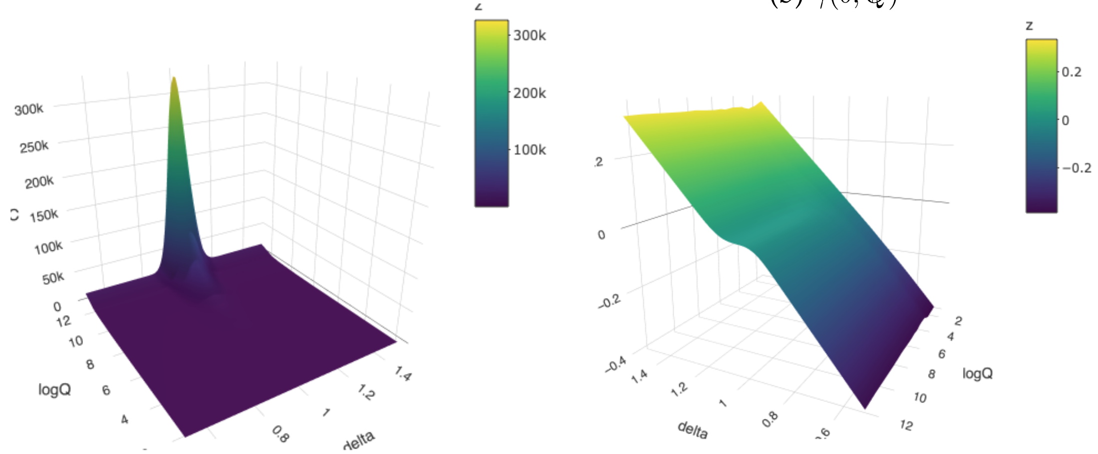
where $\bar{C}_j^{rc}$ is the average total cost after the risk-corridor function has been applied.

(8) Repeat the above steps for various values of δ and Q.

The above procedure will allow me to generate various values of the variance and the risk-corridor payment share for different values of δ and Q. Figure A.15 shows the results for a sample plan.

Figure A.15. Simulated $V(\delta, Q)$ and $\gamma(\delta, Q)$

(a) $V(\delta, Q)$ (b) $\gamma(\delta, Q)$



Panel (a) plots the simulated values of the variance of total cost as a function of δ and Q. Panel (b) plots the simulated expected risk-corridor payment share as a function of δ and Q.

While the above procedure is straightforward to implement, it can get quite computationally intensive and as such I approximate and estimate the function: $V(\delta, Q)$ and $\gamma(\delta, Q)$ using a 2-dimensional spline methods i.e. I estimate the above using a series of polynomial coefficients across different basis functions of δ, Q. The estimated function results look very similar to the simulated ones where the R^2 is close to 0.99. I do so for each plan to estimate the functions: $V_{jm}(\delta, Q)$ and $\gamma(\delta, Q)$.

A.6. Details on the Linear Risk-Sharing Rule

Here, I provide additional details on the linear risk-sharing rule of the form:

$$T(\tilde{C}, \delta C) = \alpha(\delta C - \tilde{C}) \tag{A.22}$$

where $\tilde{C}$ is the ex-post realized total cost, C is the ex-ante expected total cost and δ is the strategic cost-reporting parameter. With the linear risk-sharing rule, insurer's ex-post total cost will be

$$\begin{aligned} \tilde{C}_\alpha &= \tilde{C} + T(\tilde{C}, \delta C) \\ &= (1-\alpha)\tilde{C} + \alpha\delta C \end{aligned} \tag{A.23}$$

So the insurers' expected cost and variance of cost will be:

$$\begin{aligned} E[\tilde{C}_\alpha] &= C + E[T(\tilde{C}, \delta C)] \\ &= C \quad + \quad \underbrace{\alpha(\delta - 1)C}_{\text{expected rc payment}} \end{aligned} \tag{A.24}$$

$$Var(\tilde{C}_\alpha) = (1-\alpha)^2 Var(\tilde{C}) \tag{A.25}$$

This shows that the expected cost will still be a function of insurer's strategic cost-reporting parameter, δ. If insurer overestimates its cost ($\delta > 1$), then it will be expected to pay the government and vice versa. However, insurer's variance of cost is no longer dependent on δ as shown above. This is contrary to the existing risk-corridor function that makes both insurers' expected cost and the variance of cost be function of δ.

APPENDIX B

Additional Details for Chapter 2

B.1. Additional Tables

Table B.1. State reinsurance programs

State	Program Start Year	Program Structure
AK	2018	Covers claims costs for one or more of 33 conditions specified in state regulation.
CO	2020	Covers 15%-35% of claims costs above $30k, with the coinsurance rate depending on rating areas. Reimbursement cap is $400k per consumer.
DE	2020	Covers 20% of claims costs above $65k. Reimbursement cap is $335k per consumer.
ME	2019	Covers 10% of claims costs between $65k and $95k.
MA	2019	Covers 20% of claims costs above $20k. Reimbursement cap is $250k per consumer.
MN	2018	Covers 20% of claims costs above $50k. Reimbursement cap is $250k per consumer.
MT	2020	Covers 40% of claims costs above $40k. Reimbursement cap is $101.75k per consumer.
NH	2021	Covers 26% of claims costs above $60k. Reimbursement cap is $400k per consumer.
NJ	2019	Covers 50% of claims costs above $35k. Reimbursement cap is $245k per consumer.
ND	2020	Covers 25% of claims costs above $100k. Reimbursement cap is $1000k per consumer.
OR	2018	Covers 50% of claims costs above $83k. Reimbursement cap is $1000k per consumer.
PA	2021	Covers 40% of claims costs above $60k. Reimbursement cap is $100k per consumer.
RI	2020	Covers 50% of claims costs above $30k. Reimbursement cap is $72k per consumer.
WI	2019	Covers 53% of claims costs above $40k. Reimbursement cap is $175k per consumer.

Notes: All numbers reported in this table are from year 2021 policies. Source: Center for Medicare & Medicaid Services (2021).

B.2. Derivations for the theoretical model

B.2.1. Pass-through of reinsurance

Proof to Proposition 4:

Without loss of generality, we provide a simple example with linear demand and symmetric individual types to illustrate that pass-through rate of greater than one can be achieved to prove Proposition 4. Suppose that individual of type t's cost is identically distributed i.e. $F_\ell = F_h$, meaning $c_\ell = c_h, \sigma_\ell^2 = \sigma_h^2$. Suppose that the monopoly insurer faces an aggregate linear demand of $Q(p) = a - bp$. Then the insurer's first order condition can be re-written in the following way:

$$p^*(\theta) = \frac{1}{2}\left(c(\theta) + \rho\sigma^2(\theta) + \frac{a}{b}\right)$$

Without reinsurance, $p_0^* = c + \rho\sigma^2 + a/b$. When the government provides reinsurance of level θ, then it will decrease insurer's expected cost by $r(\theta) = c - c(\theta)$. So the corresponding pass-through rate will be

$$\frac{p^*(\theta) - p_0^*}{r(\theta)} = \frac{1}{2} + \frac{1}{2}\frac{\rho\overbrace{\left(\sigma^2(\theta) - \sigma^2\right)}^{\Delta\sigma^2(\theta)}}{r(\theta)}$$

As a result, as long as the decrease in the risk charge, $\rho\Delta\sigma^2(\theta)$ is larger than the expected reinsurance cost of $r(\theta)$, the pass-through could be greater than one.

Proof to Proposition 5:

In the absence of risk frictions, insurer will face no risk charge i.e. $\rho = 0$. Furthermore when there is no selection, individuals across different types t all are drawn from the same cost distribution i.e. $F_\ell(t) = F_h(t)\ \forall\, t$, implying $c_\ell = c_h, \sigma_\ell^2 = \sigma_h^2$.

Then the expected average reinsurance cost for given θ is

$$r(\theta) = r_\ell(\theta) = r_h(\theta)$$

The expected per-enrollee subsidy will be $s(\theta) = r(\theta)$. That is under no risk frictions and no selection, both consumer subsidy and reinsurance cost the government the same amount of expenditure.

Now if insurer is risk averse i.e. $\rho > 0$ but without selection in the market, the expected reinsurance cost will remain the same. However, the expected per-enrollee subsidy will now be

$$s(\theta) = r(\theta) + \underbrace{\rho\Delta\sigma^2(\theta)}_{>0} > r(\theta)$$

Hence, when there is just risk frictions, reinsurance which is an ex-post subsidy is more efficient in lowering the enrollee premium.

Now suppose there is adverse selection, but no risk frictions. The expected average reinsurance cost is

$$r(\theta) = \alpha(p) r_\ell(\theta) + (1 - \alpha(p))\, r_h(\theta)$$

The expected per-enrollee subsidy is

$$s(\theta) = \lambda(p) r_\ell(\theta) + (1 - \lambda(p))\, r_h(\theta)$$

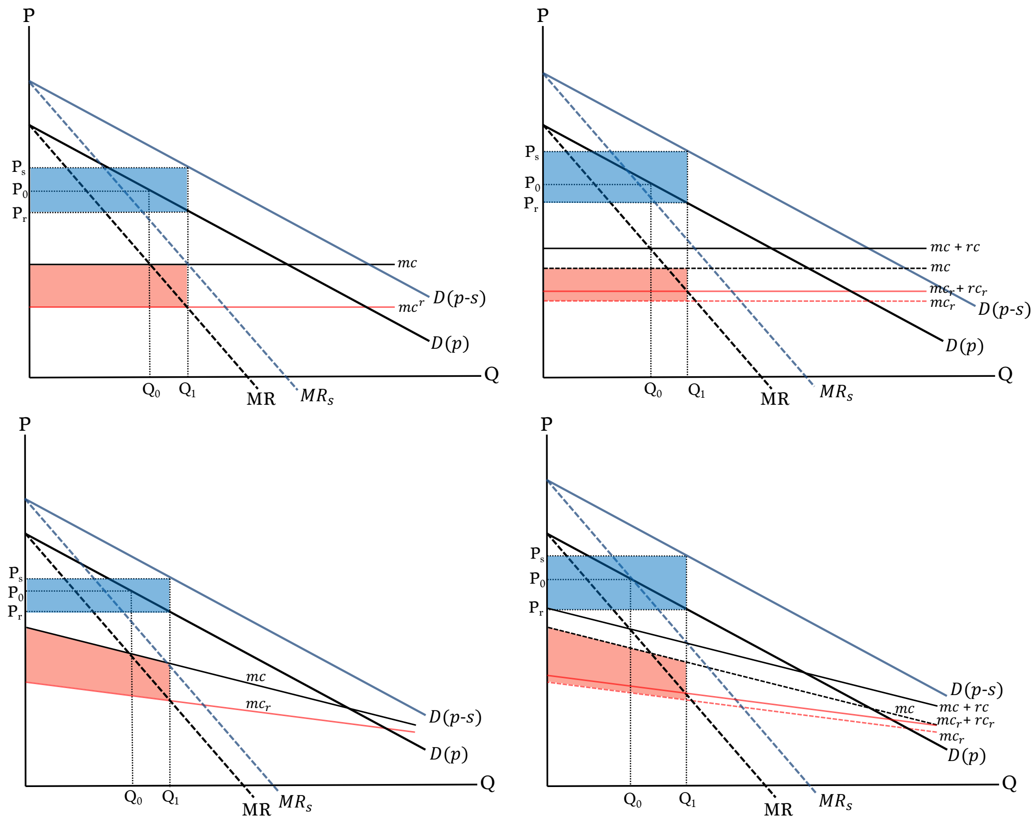
Under adverse selection, $F_h(t) < F_\ell(t)\ \forall\, t$. This directly implies that $r_\ell(\theta) < r_h(\theta)$. We now show that the marginal reinsurance cost is smaller than the average reinsurance cost. Given that $r_\ell(\theta) < r_h(\theta)$, if $\alpha(p) < \lambda(p)$ then $r(\theta) > s(\theta)$ as the average reinsurance cost uses $\alpha(p)$ as the weight for the type ℓ individual whereas the marginal reinsurance cost uses $\lambda(p)$.

$$\begin{aligned}
\lambda(p) &= \frac{\frac{\partial q_\ell(p)}{\partial p}}{\frac{\partial q_\ell(p)}{\partial p} + \frac{\partial q_h(p)}{\partial p}} \\
&= \frac{\frac{\partial q_\ell(p)}{\partial p}\frac{p}{q_\ell}}{\frac{\partial q_\ell(p)}{\partial p}\frac{p}{q_\ell} + \frac{\partial q_h(p)}{\partial p}\frac{p}{q_\ell}} \\
&= \frac{\varepsilon_\ell(p)}{\varepsilon_\ell(p) + \varepsilon_\ell(p)\frac{q_h}{q_\ell}} \\
&= \frac{q_\ell \varepsilon_\ell(p)}{q_\ell \varepsilon_\ell(p) + \varepsilon_\ell(p) q_h} \\
&= \frac{q_\ell}{q_\ell + q_h \underbrace{\frac{\varepsilon_p(p)}{\varepsilon_\ell(p)}}_{<1}} \\
&> \frac{q_\ell}{q_\ell + q_h} = \alpha(p)
\end{aligned}$$

where the last inequality comes from the assumption that type ℓ's demand is more elastic than type h's. Hence $s(\theta) < r(\theta)$ as the marginal reinsurance cost is smaller than the average reinsurance cost due to adverse selection. So when there is just adverse selection, consumer subsidy is more efficient in lowering the enrollee premium.

When there are both risk frictions and adverse selection, the efficiency will depend on which force dominates. If selection is strong in the market then consumer subsidy might be more efficient. If risk frictions dominates then reinsurance might be more efficient.

Figure B.1. Reinsurance vs. Demand Subsidy



APPENDIX C

Additional Details for Chapter 3

C.1. Proofs

Theorem 1: Under Assumptions 1–2, for each θ_i, (i) there exists a unique $f^*(\theta_i) \in [0,1]$ that solves the borrower's optimization problem, and (ii) $f^*(\theta_i)$ is weakly decreasing in θ_i.

Proof: We begin by proving part (i), the uniqueness of f^*. Throughout, we condition on T, w, and b, and let $\underline{f}' = \frac{\underline{f}}{1+\frac{T \cdot w}{b}}$. There are five cases to consider:

(1) $R(1,\theta_i) < 1$

(2) $R(1,\theta_i) \geq 1$ and $1 + \frac{1-R(\underline{f}',\theta)}{\frac{\partial R}{\partial f}(\underline{f}',\theta)} > \frac{\underline{f}}{1+\frac{T \cdot w}{b}}$

(3) $1 + \frac{1-R(\underline{f}',\theta)}{\frac{\partial R}{\partial f}} + \frac{1+r}{\gamma \cdot \underline{f} \cdot \frac{\partial R}{\partial f}} < \frac{\underline{f}}{1+\frac{T \cdot w}{b}}$

(4) $1 + \frac{1-R(\underline{f}',\theta)}{\frac{\partial R}{\partial f}} + \frac{1+r}{\gamma \cdot \underline{f} \cdot \frac{\partial R}{\partial f}} \geq \frac{\underline{f}}{1+\frac{T \cdot w}{b}}$ and $1 + \frac{1-R(\underline{f}',\theta)}{\frac{\partial R}{\partial f}} \leq \frac{\underline{f}}{1+\frac{T \cdot w}{b}}$

(5) $1 + \frac{1-R(0,\theta)}{\frac{\partial R}{\partial f}} + \frac{1+r}{\gamma \cdot \underline{f} \cdot \frac{\partial R}{\partial f}} < 0$

Consider case 1, where $R(1,\theta_i) < 1$. By Assumption 1, $R(\cdot,\theta_i)$ is increasing in its first argument. This implies that $R(f,\theta_i) < 1$ for all $f \in [0,1)$. Thus, $f^*(\theta_i) = 1$.

For the remaining cases, we must consider the borrower's first-order condition:

$$f_i^* = 1 + \frac{1 - R(f_i^*, \theta)}{\frac{\partial R}{\partial f}} + \frac{\mathbb{I}(f_i^*(1 + \frac{T \cdot w}{b}) < \underline{f})(1 + r)}{\gamma \cdot \underline{f} \cdot \frac{\partial R}{\partial f}}$$

For case 2, consider $f \in (\underline{f}', 1)$. First note that $1 + \frac{1-R(\underline{f}',\theta)}{\frac{\partial R}{\partial f}} > \frac{\underline{f}}{1+\frac{T \cdot w}{b}}$. We now must show that $G(f) = 1 + \frac{1-R(f,\theta)}{\frac{\partial R}{\partial f}}$ is strictly decreasing in f on $(\underline{f}', 1)$. We can show that

$$\frac{\partial G}{\partial f} < 0 \text{ if } (R(f, \theta) - 1) \frac{\partial^2 R}{\partial f^2} < \left(\frac{\partial R}{\partial f}\right)^2$$

which is satisfied if $R(f, \theta) - 1$ is log-concave. Because R is assumed to be concave in f, $R(f, \theta) - 1$ is also concave and, thus, log-concave in f. Therefore, by the intermediate value theorem, there exists a unique $f_i^* \in (\underline{f}', 1)$ that solves the borrower's problem. An analogous argument for case 3 can be used to prove a unique $f_i^* \in (0, \underline{f}')$.

Now, consider case 4. Monotonicity of $R(f, \theta)$ in its first argument and $1 + \frac{1-R(\underline{f}',\theta)}{\frac{\partial R}{\partial f}} + \frac{1+r}{\gamma \cdot \underline{f} \cdot \frac{\partial R}{\partial f}} \geq \frac{\underline{f}}{1+\frac{T \cdot w}{b}}$ implies that, for all $f \in [0, \underline{f}')$, $1 + \frac{1-R(f,\theta)}{\frac{\partial R}{\partial f}} + \frac{1+r}{\gamma \cdot \underline{f} \cdot \frac{\partial R}{\partial f}} > f$. Furthermore, monotonicity and $1 + \frac{1-R(\underline{f}',\theta)}{\frac{\partial R}{\partial f}} \leq \underline{f}'$ implies that, for all $f \in (\underline{f}', 1]$, $1 + \frac{1-R(f,\theta)}{\frac{\partial R}{\partial f}} < f$. It follows that the borrower's problem is solved at $f^* = \frac{\underline{f}}{1+\frac{T \cdot w}{b}}$.

Finally, consider case 5. $1 + \frac{1-R(0,\theta)}{\frac{\partial R}{\partial f}} + \frac{1+r}{\gamma \cdot \underline{f} \cdot \frac{\partial R}{\partial f}} < 0$ and monotonicity of $R(\cdot, \theta)$ in its first argument implies that, for all $f \in (0, 1]$, $1 + \frac{1-R(f,\theta)}{\frac{\partial R}{\partial f}} < 0$. Therefore, the borrower's problem is solved at $f^* = 0$.

We conclude by proving part (ii) and show that $f^*(\theta)$ is weakly decreasing in θ. Denote the borrower's objective function as $u(f, \theta)$. It suffices to show that this objective function satisfies decreasing differences for values of θ such that an interior solution is optimal:

$$\frac{\partial^2 u}{\partial f \partial \theta} = b_i \gamma \left[-\frac{\partial R}{\partial \theta} + (1 - f) \frac{\partial^2 R}{\partial f \partial \theta} \right] < 0,$$

where the inequality follows from $\frac{\partial R}{\partial \theta} < 0$ and $\frac{\partial^2 R}{\partial f \partial \theta} < 0$. ■

Theorem 2: Under Assumption 3, there exists a minimum loan amount, $\underline{b}$, such that lender j approves all loans with $b_i > \underline{b}$. The minimum loan amount is: (i) decreasing in S, (ii) decreasing in r, and (iii) increasing in $\underline{f}$. Specifically, $\underline{b}$ is given by:

$$\underline{b}(\tilde{f}_i^*) = \frac{c_{ij}}{S + (r - \delta_j) s(\tilde{f}_i^*)}$$

Proof: The lender earns the following profit from issuing a loan:

$$\pi_{ij} = S b_i - c_{ij} + s(\tilde{f}_i^*)(r - \delta_j) b_i,$$

We normalize the lender's payoff of not issuing a loan to zero, and a loan is issued if $\pi_{ij} \geq 0$, which is satisfied if

$$b_i \geq \frac{c_{ij}}{S + s(\tilde{f}_i^*)(r - \delta_j)},$$

under the assumption that $S + s(\tilde{f}_i^*)(r - \delta_j) > 0$.

We now prove the comparative statics results:

$$\frac{\partial \underline{b}}{\partial S} = \frac{-c_{ij}}{(S + s(\tilde{f}_i^*)(r - \delta_j))^2} < 0$$

$$\frac{\partial \underline{b}}{\partial c} = \frac{1}{S + s(\tilde{f}_i^*)(r - \delta_j)} > 0$$

Consider a decline in the forgiveness threshold. For $\underline{f}'' < \underline{f}'$,

$$\begin{aligned}
\underline{b}(\underline{f}'') - \underline{b}(\underline{f}') &= \frac{c}{S + (r - \delta)s(\tilde{f}(\underline{f}''))} - \frac{c}{S + (r - \delta)s(\tilde{f}(\underline{f}'))} \\
&= \frac{c(r - \delta)(s(\tilde{f}(\underline{f}')) - s(\tilde{f}(\underline{f}'')))}{(S + (r - \delta)s(\tilde{f}(\underline{f}'')))(S + (r - \delta)s(\tilde{f}(\underline{f}'))} \\
&< 0
\end{aligned}$$

■

C.2. Window Definition

In this appendix, we conduct analysis to support the window definition used in the main text. For our main sample, we restrict to loans issued within the twelve week window encompassing the four weeks prior to the passage of the PPP Flexibility Act and the eight weeks following the policy change.

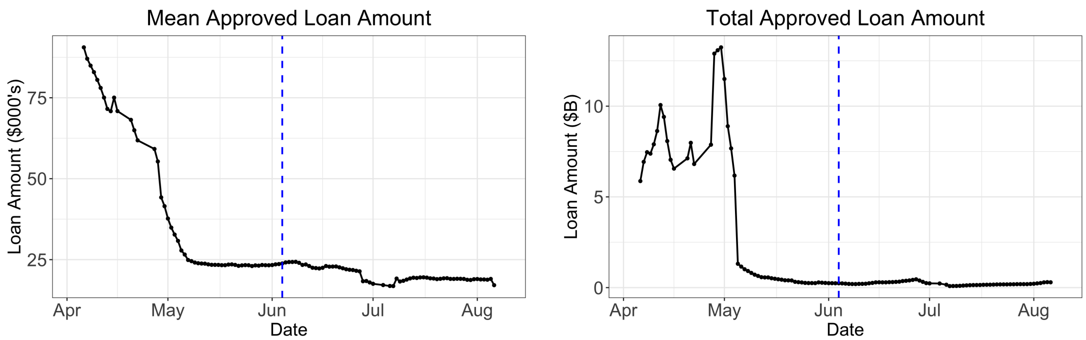
We use this sample for two main reasons. First, we seek to examine the lenders' response to the policy change and therefore seek to keep the distribution of borrowers approximately fixed. Our sample begins after the first month of the PPP. During this first month, discussions in the media and elsewhere highlighted the issues with the rollout of the program. Large businesses, such as Shake Shack, Potbelly Sandwich Shop, and Ruth's Chris Steakhouse, received loans[1], and borrowers faced numerous delays in receiving funds.[2] This difference in borrower types is apparent in the time series plot of loan amounts. Figure C.1 displays the 7-day moving average loan amount and total loan amount over time, and we see a large difference in the first four weeks of the PPP. To standardize the sample over time, we remove these loans issued at the beginning of the program.

Second, we restrict to loans issued up to eight weeks following the program because of a potential lag in policy take-up. In the plots below, we display the share of loans issued with an amount allocated to payroll of 75%, which was the threshold in the prior to the legislation change, and the share issued with an amount

[1] https://www.nytimes.com/2020/04/20/business/shake-shack-returning-loan-ppp-coronavirus.html

[2] See, for example, https://www.usatoday.com/story/money/usaandmain/2020/04/07/ppp-loan-plan-rollout-disaster-small-businesses/2963901001/.

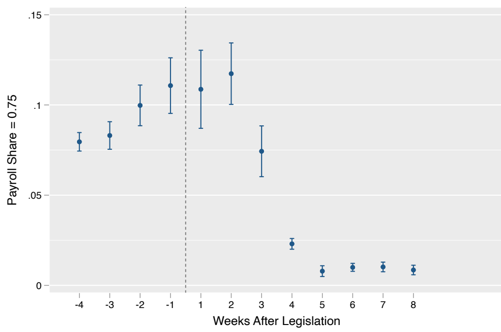
Figure C.1. 7-Day Moving Average of Mean and Total Loan Amount



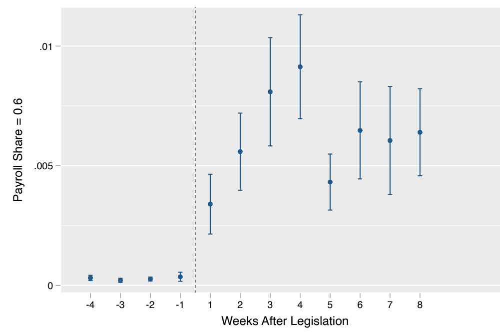
Note: This figure displays two plots. On the lefthand side, we show the 7-day moving average of mean loan amount (in thousands of dollars) and, on the righthand side, we show the 7-day moving average of total loan amount (in billions of dollars).

allocated to payroll of 60%, the threshold after the policy change. There is a transition period of about four weeks during which the share of loans issued at 75% decreases and the share issued at 60% increases. Because of this lag, we include an additional four weeks after which the policy take-up appears to stabilize. In Appendix C.3, we show the main empirical results are robust to using a donut-hole specification in which we remove the loans issued in the first four weeks following the policy change (i.e., the transition period).

Figure C.2. Share of Loans Issued at Threshold Payroll Amounts



(a) Share = 0.75



(b) Share = 0.60

Note: This figure the coefficient of a regression of an indicator of the at-origination payroll share being equal to 0.75 (lefthand side) or 0.6 (righthand size) on week fixed effects. The bars represent 95% conficence intervals computed using standard errors clustered by state.

C.3. Robustness Checks

Our main sample considers all loans issued between 4 weeks before and 8 weeks after the policy change. As described in Appendix , the take-up of the policy change occurred over time and took approximately four weeks to stabilize. For this reason, we examine robustness of our main event study results (i.e., in loan amount and in payroll share) to removing loans in the transition period. To do this, we restrict to loans issued up to four weeks before and between 5 and 8 weeks after the policy change.

Table C.1 displays the results of the event studies with loan amount on the lefthand side, and Table C.2 displays results for the payroll-share specifications. In both cases, the results are stronger for this specification than they are in the main text. However, the qualitative takeaways are identical.

Table C.1. Aggregate Loan Amount Changes - Donut-Hole Specification

	(1) Log(Loan Amount)	(2) Log(Loan Amount)	(3) Log(Loan Amount)
Post-Legislation	-0.128***	-0.139***	-0.0672***
	(0.0281)	(0.0217)	(0.0192)
Observations	661,062	660,808	660,805
Lender FE	No	Yes	Yes
Borrower Controls	No	No	Yes

* p<0.1, ** p<0.05, *** p<0.01

Note: This table presents results for the aggregate event-study specifications defined by Equation (2), restricting to loans issued up to four weeks before and between five and eight weeks after the policy change. Standard errors clustered by borrower state are shown in parentheses. Lender FEs are defined as a combination of a lender name, lender city, and lender state. Borrower controls include fixed effects for the two-digit NAICS code, business type (corporation, LLC, sole proprietorship, other), urban/rural, business age (> 2 years, ≤ 2 years, unanswered), and borrower state.

Table C.2. Aggregate Payroll Share Response - Donut-Hole Specification

	(1) Payroll Share (At Orig.)	(2) Payroll Share (At Orig.)
Post-Legislation	0.0184***	0.0174***
	(0.0020)	(0.0020)
Observations	661,062	661,055
Borrower Controls	No	Yes

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Note: This table presents results for the event-study specifications with payroll share on the lefthand side, defined by Equation (3), restricting to loans issued up to four weeks before and between five and eight weeks after the policy change. Standard errors clustered by borrower state are shown in parentheses. Borrower controls include fixed effects for the two-digit NAICS code, business type (corporation, LLC, sole proprietorship, other), urban/rural, business age ($>$ 2 years, $\leq$ 2 years, unanswered), and borrower state.