Essays on Industrial Organization

Essays on Industrial Organization

A dissertation presented

by

Shoshana Vasserman

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2019

Dissertation Advisors:
Professor Robin Lee
Professor Ariel Pakes

Author:
Shoshana Vasserman

Essays on Industrial Organization

Abstract

This dissertation consists of three independent chapters. Chapter 1 examines the role of auction design in managing risk exposure for bidders in public procurement and minimizing costs for the government. I study the mechanism used to procure bridge construction by the Massachusetts Department of Transportation (DOT). I estimate a structural model of equilibrium bidding by risk-averse firms facing uncertainty about the final specification for each project at the time of bidding. I show that the existing mechanism insures risk-averse contractors against large changes in project scope, thereby reducing risk-premiums in bidding and ultimately lowering costs incurred by the DOT. Furthermore, while reducing uncertainty could reduce costs, I find that incentivizing additional competition is a more promising direction for policy intervention.

Chapter 2 examines the extent to which a reference pricing policy could reduce prices for pharmaceutical drugs in the United States. I estimate a structural model of supply and demand for pharmaceuticals in the U.S. and in Canada. I then simulate counterfactuals in which American and Canadian regulators negotiate over prices with pharmaceutical representatives, but the United States constrains the prices that it is willing to accept by the prices set in Canada. I find that reference pricing results in a slight decrease in U.S. prices and a substantial increase in Canadian prices. Consequently, I find modest consumer welfare gains in the U.S., substantial consumer welfare losses in Canada, and a net increase in pharmaceutical firms' profits.

Chapter 3 examines the impact of a monitoring program in which a large US auto-insurer offered drivers a discount for good driving performance, as measured by a telematics device for their first contracted period. I demonstrate reduced form evidence of both adverse selection and moral hazard: drivers who choose to enter the monitoring program have lower average liability claims, but average claims rates are 23% lower during the monitoring period than after. I estimate a structural model of driver claims and insurance plan choice and simulate several counterfactual regimes with respect to the availability of monitoring. While consumers face a disutility from being monitored, monitoring induces a net social surplus: both firm profits and net consumer welfare improve.

Contents

List of Tables

List of Figures

Acknowledgments

I am indebted to my advisors Robin Lee, Ariel Pakes and Elie Tamer, whose teaching, guidance and tireless support enabled not just this thesis, but also my development as a scholar.

I am further indebted to my generous co-authors—Valentin Bolotnyy, Yizhou Jin, Ashvin Gandhi, Pierre Dubois, S. Nageeb Ali and Greg Lewis. Our collaborations made my PhD experience lively, social and productive—and I learned a tremendous amount from each one. I am indebted to Muhamet Yildiz in particular, who in addition to our co-authored work, taught me how to think like a game theorist and write like a communicator. I will never forget his encouragement to seek inspiration in Dostoevsky, relevance in history and tools in measure theory, when I was trying to figure out what it might mean to be an economist.

I am grateful to my family, and in particular to my father, who raised me with enough space and trust that I could always find my way uninhibited. I am grateful to Luis Voloch, who envisioned my path before I could conceive of it, and convinced me that the world of science and math had room for me.

I am grateful to my friends—Avital Chizhik-Goldschmidt, Alumba Dembo, Iulia Neagu, Yonadav Shavit and Cory Smith—more aptly dubbed, family by virtue of the soul. I am grateful to my friend, collaborator and friendly competitor, Jim Savage, who habitually insists on doubling the stakes on what can be learned and contributed to the world.

Finally, I am grateful to my partner and my love, Hershel Macaulay, who shows me how to find joy in every experience and endeavor—and who keeps me true to myself.

I dedicate this thesis to Joel Spencer, who saw potential in me when I was a clueless 17 year old, and taught me to see the beauty in numbers.

Introduction

This dissertation consists of three independent chapters, all related to empirical industrial organization. Chapter 1 examines the role of auction design in managing risk exposure for contractors and minimizing costs for the government in the context of public procurement. I study the mechanism used to procure bridge construction by the Massachusetts Department of Transportation (MassDOT). As in many other public procurement settings, the DOT allocates bridge construction projects via a “scaling auction". Interested contractors submit unit price bids for each task and material on an itemized design specification prepared by DOT engineers beforehand. The winner is determined by the lowest total cost—given DOT estimates of the amount of each item needed—but, critically, they are paid based on the realized quantities used. This creates an incentive for contractors to skew their bids—bidding high when they believe the DOT is underestimating an item’s quantity and vice versa—and raises concerns of rent-extraction among policymakers. For risk averse bidders, however, scaling auctions provide a distinctive way to generate surplus: they enable firms to limit their risk exposure by placing lower unit bids on items with greater uncertainty.

I demonstrate reduced form evidence that contractors participating in MassDOT bridge auctions are knowledgeable—they are able to predict which items will overrun the DOT engineers’ estimates and vice versa—but are also risk averse. Holding everything else fixed, the contractors tend to bid lower on items that are historically more variable so as to mitigate the possibility of a large shock in their ultimate compensation. Motivated by these empirical facts, I estimate a structural model of equilibrium bidding by risk-averse firms facing uncertainty about the final specification for each project at the time of bidding. I show that the existing mechanism insures risk-averse contractors against large changes in project scope, thereby reducing risk-premiums in bidding and ultimate costs to the DOT. Using counterfactual simulations, I show that while reducing uncertainty could reduce costs, incentivizing additional competition is a more promising direction for policy intervention.

Chapter 2 examines the extent to which a reference pricing policy could reduce prices for pharmaceutical drugs in the United States. I estimate a structural model of supply and demand for pharmaceuticals in the U.S. and in Canada. I then simulate counterfactuals in which American and Canadian regulators negotiate over prices with pharmaceutical representatives and the United States constrains the prices that it is willing to accept by the prices set in Canada. I find that reference pricing results in a slight decrease in U.S. prices and a substantial increase in Canadian prices. The magnitude of these effects depends on the particular structure of the policy. I consider variants of reference pricing policies in which the US is allowed a premium above the Canadian price, and in which the US references a larger country with a pharmaceutical market that is comparable to the US market in size. While referencing a larger market increases savings to US consumers, I find that the overall relative magnitudes are unchanged. Consequently, I find modest consumer welfare gains in the U.S., substantial consumer welfare losses in Canada, and a net increase in pharmaceutical firms' profits.

Chapter 3 examines the impact of a monitoring program in which a large US auto-insurer offered drivers a discount for good driving performance as measured by a telematics device for their first contracted period. I demonstrate reduced form evidence of both adverse selection and moral hazard: drivers who choose to enter the monitoring program have lower average liability claims, but average claims rates are 23% lower during the monitoring period than after. I estimate a structural model of driver claims and insurance plan choice and simulate several counterfactual regimes with respect to the availability of monitoring. While consumers face a disutility from being monitored, monitoring induces a net social surplus: both firm profits and net consumer welfare improve. However, proprietary ownership of the monitoring data is key to firm profits: If monitoring results had to be shared with competitors, price competition would significantly lower the insurer's private benefit from offering the monitoring program. Consequently, in equilibrium, the insurer would offer a significantly lower baseline (opt-in) discount to encourage the uptake of monitoring. Uptake would drop, and so the social surplus from reducing moral hazard would be drop as well.

Chapter 1

Scaling Auctions as Insurance: A Case Study in Infrastructure Procurement[1]

1.1 Introduction

Infrastructure investment underlies nearly every part of the American economy and constitutes hundreds of billions of dollars in public spending each year.[2] Infrastructure is also politically popular: voters and policy-makers alike support increasing spending on infrastructure projects by as much as 100% over the next decade.[3] However, infrastructure projects are often complex and subject to unexpected changes. Uncertainty can be costly to the firms implementing construction—many of whose businesses are centered on public works. The mechanisms used to procure construction work can play a key role in mitigating firms' exposure to risk. Limiting risk makes prospective contracts more lucrative to firms

[1]Co-authored with Valentin Bolotnyy

[2]According to the CBO, infrastructure spending accounts for roughly $416B or 2.4% of GDP annually across federal, state and local levels. Of this, $165B—40%—is spent on highways and bridges alone.

[3]Recent polls have consistently shown around 70% of voters in support of increased infrastructure spending along the lines of the $1.5 trillion plan outlined by the Trump administration. See YouGov and Gallop for example. A major infrastructure bill is expected to entertain bi-partisan support following the 2018 election (Nilson (2018)). This is in addition to a 2015 bill passed with bi-partisan support to increase infrastructure spending by $305 billion over five years. ((Berman (2015)).

and increases competition, thereby reducing tax payer expenditures.

In this paper we study the mechanism by which contracts for construction work are allocated by the Highway and Bridge Division in the Massachusetts Department of Transportation (MassDOT). As in 36 other states, MassDOT uses a *scaling auction*, whereby bidders submit unit price bids for each item in a comprehensive list of tasks and materials required to complete a project. The winning bidder is determined by the lowest sum of unit bids multiplied by item quantity estimates produced by DOT project designers. The winner is then paid based on the quantities ultimately used in completing the project.

A common concern among policy-makers is that bidders may extract rents from the DOT by “skewing" their bids: placing high unit bids on items that will over-run the DOT estimates and low unit bids on items that will under-run. Bid-skewing has been documented as far back as 1935, and referred to as commonplace as recently as 2009 (Skitmore and Cattell (2013)). Previous work on timber auctions (Athey and Levin (2001)) and highway construction (Bajari, Houghton and Tadelis (2014)) has demonstrated evidence that bidders skew correctly on average and that the most competitive bidders skew in a similar way. This suggests that competitive bidders are similarly able to predict which items will over/under-run.

As we demonstrate, the markup charged to the DOT depends not only on the level of competition in the auction, but also on the uncertainty about the ultimate needs of a project—conditional on the DOT's specification—as well as the degree of risk aversion that contending bidders face. If bidders are risk neutral and equally informed, bid-skewing produces no additional cost to the DOT in equilibrium. Contractors choose their bids using refined quantity estimates, and any information rent is competed away. Risk averse bidders, however, use bid skewing to balance the uncertainty in a project across the different items involved. As in the risk neutral case, bidders generally submit higher bids for items they believe will over-run the DOT quantity estimates. However, the incentive to raise bids on items predicted to over-run is dampened by the level of noise in this prediction. Moreover, the risk lowers the value of a project to bidders, causing them to bid less aggressively and consequently extract higher payments from the DOT.

Notably, risk averse bidders will generally submit *interior bids*—unit bids that are above zero—whereas risk neutral bidders will submit "penny" bids—unit bids of essentially zero—on all but the items that are predicted to overrun by the largest amount, absent an external force to prevent this.[4] This matches the observations in our data, in which the vast majority of unit bids are interior, but no significant penalty for penny bidding has ever been exercised.[5]

Moreover, taking uncertainty and risk aversion into account has significant implications for comparisons across auctions. Risk neutral bidders would profit identically under a scaling auction, a lump sum auction—in which bidders bid a total project price and are responsible for all realized costs—or anything in between. Risk averse bidders, however, are sensitive to the differences in risk exposure under each of these mechanisms. Scaling auctions compensate bidders for every unit that is ultimately used. As such, the only risk that bidders are exposed to (upon winning the auction) is the risk that they "mis-optimized" in selecting their bid spread across items given the ex-post quantity realizations. Under a lump sum auction, however, bidders bear the entirety of the cost risk involved in the project. If the realized quantities are substantially larger than the predicted values used during bidding, the winning bidder is liable for the differences, with no further compensation.[6] In equilibrium, bidders will insure themselves against the risk that they face by submitting higher overall bids. Thus, scaling auctions, in which the level of risk from uncertainty about the ex-post quantities in a project can be minimized by the bidders, are predicted to produce substantially lower overall costs to the DOT.

Our contributions are three-fold. First, we construct a parsimonious model of competitive

---

[4]See section 1.2 for a discussion of the model predictions under risk neutral and risk averse bidders.

[5]In a few rare instances, the DOT responded to suspicious bids by scrapping the auction all together and revising the specification for the project before putting it up for auction again. In these instances, the same bidders were able to participate, and so any cost incurred was minimal.

[6]This analysis precludes ex-post hold up problems and the like, in which the bidder might sue the DOT for additional compensation. Considerations of this sort would further increase the costs to the DOT, and so our analysis serves as a conservative estimate of the total effect.

bidding in a scaling auction with risk averse bidders who shoulder uncertainty over the quantities that will ultimately be used. We show that risk aversion and uncertainty are sufficient to explain interior bids, in contrast to previous work, which relies on heuristic penalties on penny bids. We then provide reduced form evidence that the bidding behavior observed in our data is consistent with the predictions of our model. Furthermore, we demonstrate how our model can be used to evaluate the cost that the DOT incurs due to uncertainty in its project specifications. Second, we estimate a structural model for uncertainty and optimal bidding in our data. We employ a two-stage procedure to estimate the level of risk in each project, the degree of risk aversion, and the distribution of bidders' private costs. In the first stage, we estimate a model of bidder uncertainty using the history of predicted and realized item quantities. In the second stage, we use the specification of equilibrium unit bids implied by our model to construct a GMM estimator for risk aversion and bidder cost types. Third, we use our structural estimates to simulate counterfactual auction equilibria in which: (1) the DOT eliminates all uncertainty about item quantities; (2) the DOT employs a μ-risk-sharing auction in which it compensates bidders for μ times the prespecified estimated quantity and $1 - \mu$ times the realized quantity of each item used. Finally, we calculate bounds on the cost of entry for an additional bidder to each auction, as well as the cost savings to the DOT from an additional entry.

Using the first counterfactual results, we assess the DOT's cost from uncertainty by taking the difference in the expected amount paid to the winning bidder in the baseline auction (the auction used in the status quo) and in the counterfactual setting with all uncertainty removed. We find that the DOT's cost in the baseline auction is only \$2,145—or 0.70%—higher, on average, than in the counterfactual auction with no uncertainty. However, this estimate reflects the sum of two opposing forces that are shifted by the counterfactual: risk and prediction. In the baseline, bidders use a best prediction (given available information) of the ultimate item quantities. These predictions may be inaccurate in-sample, and so the bids submitted may not be optimal (from the bidders' perspective, after observing the ex-post quantities). By contrast, in the counterfactual setting with all uncertainty eliminated, bidders know the exact quantities that will be used and optimize accordingly. Consequently, the DOT winds up paying more than in the baseline for some projects. To isolate the effect of risk itself, we repeat the counterfactual exercise under the assumption that bidders' quantity predictions are correct (but bidders still interpret these predictions as coming from noisy signals as before) in the baseline. In this case, there is no bidder mis-optimization in the baseline, and so the DOT strictly saves money from eliminating risk: $172,513 (13.74%) on average.

Using the second set of counterfactual results, we assess the extent and direction to which DOT costs would change if the DOT switched from the scaling auction to an alternative in which part or all of the amount paid to the winning bidder is fixed at the time of bidding.[7] A mechanism of this sort curbs bidders' ability to skew their bids: in the limiting case of a lump sum auction, bidders are paid the amount that they bid and so, there is no advantage to spreading unit bids across items in any particular way. It may also offer benefits to the DOT by reducing its burden in project specification and budgeting flexibility.[8] However, mechanisms of this sort effectively shift risk from the DOT to the bidders. As such, they lower the expected value of winning each auction and induce higher, less aggressive bids. We estimate that switching to a lump sum auction would increase DOT costs by 128% on average (85% on median). The losses do not scale linearly with the amount of risk sharing, however. We estimate that if the DOT were to pay the winning bidder her unit bid multiplied by a 50-50 split of the ex-ante and ex-post quantities for each item, costs would increase by 6.84% on average (3.47% on median).

Finally, while major improvements to quantity estimation may be difficult to achieve across the board, efforts to increase competition may offer an additional channel to improve

[7]To highlight the effects of the counterfactual policies themselves, we report the results of all counterfactuals assuming that bidders have correct quantity estimates (but still interpret these estimates with uncertainty) in both the baseline and the counterfactual. Results in the case that bidders use the quantity predictions estimated in our first stage are similar. We report them in the appendix for robustness.

[8]Neither moral hazard nor hold up problems are considered in our model. Moral hazard might make lump sum auctions more attractive as imposing more risk on bidders would induce more thrifty uses of material. However, the extent of moral hazard is limited by the contractors' ability to influence quantities given DOT restrictions and supervision. Hold up problems would strengthen our results.

DOT cost efficiency. We estimate that adding an additional bidder to each auction results in an average DOT savings of $82,583 (8.90%). Furthermore, our estimates of lower bounds on bidders' cost of entry suggest that an increased (guaranteed) payment of as little as $2,316 (on average) could incentivize an additional entry.

Our analysis is enabled by a rich and detailed data set, provided to us by the Highway and Bridge division of the Massachusetts Department of Transportation. For each auction in our study, we observe the full set of items involved in construction, along with ex-ante estimates and ex-post realizations of the quantity of each item, a blue book DOT estimate of the market unit rate for the item, and the unit price bid that each bidder who participated in the auction submitted. Furthermore, our setting is particularly conducive to the study of risk aversion. Bridge maintenance projects are highly standardized, and so heterogeneity across projects is well captured by the characteristics observed in our data. The winner of each auction is determined entirely by the expected cost of the project given the bidder's unit bids. Participating bidders are all pre-qualified by the DOT and neither historical performance, nor external quality considerations play a role in the allocation of contracts. In addition, while there is substantial variation between the ex-ante DOT estimates and the ex-post realizations of item quantities, all changes to the original project specification must be approved by an on-site DOT project manager or engineer, limiting the scope of moral hazard. Finally, while previous work on highway procurement auctions has discussed the role of ex-post renegotiation of unit-prices and a disincentive for bid skewing due to a possibility of having a winning bid rejected by the DOT, neither of these forces is applicable in our setting. Unit price renegotiation occurs in a negligible number of cases in our data, and MassDOT does not reject the winning bidder as a matter of policy.[9]

Connections and Contributions to the Related Literature

[9] In a handful of cases, MassDOT has withdrawn the auction all together after receiving bids, citing internal mis-estimation in the project specification, and has re-posted the auction anew after making adjustments. The same bidders were eligible to participate in the revised auction.

Our paper follows a rich literature on strategic manipulation in scoring auctions, and is closest in spirit to Athey and Levin (2001) and Bajari, Houghton and Tadelis (2014).[10] Athey and Levin (2001) first established the theoretical framework demonstrating that bid-skewing arises in equilibrium when bidders are better informed about ex-post quantities than the auctioneer. Using a general modeling framework, Athey and Levin establish a number of empirical predictions and test them in the context of US timber auctions. Notably, they test the hypothesis that bidders have superior information (beyond what is given to them by the auctioneer) by comparing the direction of bid skews: profitable skews are indicative of superior information. They find significant evidence of superior information, as well as evidence that there is little informational differentiation between the top two bidders. We discuss analogous exercises in our reduced form section and find similar results in our setting as well. Furthermore, Athey and Levin note that the absence of total skewing (e.g. penny bidding) in their setting is inconsistent with risk neutral bidders in their model, and suggest risk aversion as a more fitting explanation for what they observe. Using the Athey and Levin framework, we construct a structural model that allows us to quantify the costs—realized and hypothetical—of scaling auctions in practice.

Bajari, Houghton and Tadelis (2014) ("BHT") studies a setting similar to ours: the auctions used to procure highway construction contracts in California. As in our setting, BHT observe item-level unit bids submitted in a scaling auction in which awards are allocated based on engineers' quantity estimates, but compensated based on realized quantities.[11] However, the study's main focus is on adaptation costs—costs incurred from disruptions in work-flow due to inadequate preliminary planning. BHT propose a structural model for bidding in

[10]More recently, De Silva, Dunne, Kosmopoulou and Lamarche (2016) apply a framework similar to Bajari, Houghton and Tadelis (2014) to assess the effects of a DOT's commitment to reducing the scope of project changes.

[11]There are several notable differences between the setting in Bajari, Houghton and Tadelis (2014) and ours. Unlike MassDOT, the California DOT imposes tighter limits on quantity overruns, and does occasionally reject bidders with mathematically unbalanced bids. Furthermore, while the overall level of bid skewing, as evidenced by the relationship between quantity overruns and price overruns (as in Figure1.7) across all highway and bridge projects in Massachusetts is similar to that in California, this relationship is particularly pronounced among the bridge maintenance projects that our analysis focuses on.

which bidders are risk-neutral and have correct (on average) expectations over what the final quantity of each item used will be. They then use conditions derived from this model in conjunction with data on ex-post negotiated change orders, adjustments to unit prices, extra work-orders, and deductions (due to failures on the part of the contractor) to identify the expected cost of these adjustments that is paid by the California DOT.

Our paper differs from Bajari, Houghton and Tadelis (2014) in several significant ways. First, because BHT is primarily concerned with measuring adaptation costs, it does not aim to predict bids at the item level. By contrast, our paper is focused on predicting bids for auctions in counterfactual settings. Our approach incorporates risk and risk aversion to rationalize interior bids, allowing us to capture substitution patterns between items with a micro-founded generative model of unit bid setting.[12] Our model characterizes equilibrium bids at the auction-bidder-item level as a function of the item's historical quantity variance, the bidder's private cost type and distribution of opponent types, and the level of risk aversion in the auction. Our identification strategy leverages variation in unit bids across auctions that each bidder participated in, as well as variation across auctions that items identified by the DOT as "highly skewed" appeared in.[13]

Our approach allows us to estimate the distribution of bidder cost types in each auction, as well as the coefficient of bidder risk aversion. These parameters, along with those governing the item quantity distributions, jointly characterize the equilibrium bid distribution in each setting.[14] Using our estimates, we are able to predict the equilibrium bids that would arise in each counterfactual. We can thus assess policy-relevant outcomes: the expected

[12]Bajari et al. model bidders as risk neutral, but subject to a heuristic penalty function in bidders' utility that convexly penalizes deviations of unit bids from the DOT's cost estimates for each item. They estimate that the penalty coefficient is small and negative (suggesting bid-skewing is encouraged, rather than penalized), but not statistically significant. As part of preparing our paper, we replicated their methodology on our data set, and found our model substantially better in back-predicting item bid spreads.

[13]By contrast, Bajari et al. use aggregate optimality conditions such that each observation entering their moment condition is at the bidder-auction level. They then estimate a mean cost type across all bidders and auctions, as well as mean coefficients on adaptation costs, etc.

[14]Note that one cannot evaluate counterfactual outcomes by extrapolating from the empirical score distribution. Changes to the auction setting will change the equilibrium score distribution, and so it is necessary to compute the equilibrium from primitives in each counterfactual.

cost to the DOT, as well as the utility to prospective bidders (which may impact entry). To our knowledge, no counterfactual analysis of scaling auctions, nor any assessment of their performance in the context of mechanism design has been done before.

More generally, our paper relates to the literature on multi-dimensional auctions, and scoring auctions in particular (auctions in which bids on different dimensions of interest are aggregated into a single-dimensional score to determine the winner). Che (1993) characterizes the equilibria of auctions that employ a two-dimensional scoring rule (quality and price) with single-dimensional bidder types. Asker and Cantillon (2008) extend this to a more general setting, allowing for multi-dimensional bidder types and general quasi-linear scoring rules, by showing that a mapping of multi-dimensional attributes onto a single dimensional "pseudo-type" is sufficient to characterize equilibria up to payoff equivalence. Both of these papers assume that bidders are risk neutral, and the result on "pseudo-types" does not extend directly to the risk averse case. In our paper, we model single-dimensional bidder types, as this is the most parsimonious way to ensure a unique monotonic equilibrium.[15] However, we plan to extend our approach to a more general type space in future work. Furthermore, while our identification strategy leverages the particular properties of scaling auctions, our work may provide methodological insights for estimation and prediction in more general multi-dimensional auctions as well.

Our paper also relates to a rich literature on the theory and estimation of equilibrium bidding in auctions with risk averse bidders. Maskin and Riley (1984) and Matthews (1987) first characterized the optimal auction in the presence of risk averse bidders with independent private values (IPV). While we do not relate our results to the optimal mechanism in this version of the paper, an evaluation of the DOT cost savings under the optimal mechanism, following Matthews (1987), is under preparation for a future draft. Campo, Guerre, Perrigne

[15] To see why a multi-dimensional bidder type model is substantially more complicated, note that a monotonic equilibrium in our setting requires a single-dimensional ranking of bidder types: a bidder with a better type should have a higher chance of winning (and therefore a lower score). Whereas in the risk neutral case (as in Bajari *et al.* (2014)), better bidders are those with lower expected costs for completing a project, risk averse bidders are compared by the certainty equivalent of their profits from completing a project. As demonstrated in section 1.2, the certainty equivalent entails an interaction between bidders' item costs and item bids, making straight-forward comparisons in a general case difficult.

and Vuong (2011) first established semi-parametric identification results for estimating risk aversion parameters in single-dimensional first price auctions in an IPV setting. As in their approach, we exploit the heterogeneity across items being auctioned and a parameterization of the bidders' utility function for identification. However, as our identification leverages the optimal spread of unit bids across items at each bidder's equilibrium score, we do not require any restrictions on the distribution of the bidders' private value distribution for estimation.

Our paper is structured as follows. In Section 1.2, we give an overview of how the players involved with procurement auctions—contractors and DOT managers—have treated bid skewing in practice. We then present an illustrative example of equilibrium bidding in our setting to demonstrate how uncertainty, risk aversion, and competition influence the interpretation of bids that we may see in practice. In section 1.3, we discuss our dataset and present reduced form evidence that the bids we observe in our data support our model. In section 1.4, we present a full theoretical model of equilibrium bidding. In section 1.5, we present a structural model for estimating the auction primitives that underlie the bids in our data. In section 1.6, we present our structural estimates. Finally, in section 1.7, we present our counterfactual predictions and discuss their implications for policy.

1.2 Bid Skewing and Material Loss to the DOT

1.2.1 Scaling Auctions in Highway and Bridge Procurement

Like most other states, Massachusetts manages the construction and maintenance for its highways and bridges through its Department of Transportation (DOT). In order to develop a new project, DOT engineers assemble a detailed specification of what the project will entail. This specification includes an itemized list of every task and material (item) that is necessary to complete the project, as well as the engineers' estimates of the quantity with which each item will be needed, and a market unit rate for its cost. The itemized list of quantities is then advertised to prospective bidders.[16]

Any contractor who has been pre-qualified for a given project can submit a bid for the contract to implement it. Pre-qualification entails that the contractor is able to complete the work required, given their staff and equipment. Notably, it does not depend on past performance in any way. In order to submit a bid, a contractor posts a per-unit price for each of the items specified by the DOT. Since April 2011, all bids have been processed through an online platform, Bid Express, which is also used by 36 other state DOTs.[17] All bids are private until the completion of the auction.

Once the auction is complete, each contractor is given a score, computed by the sum of the product of each item's estimated quantity and the contractor's unit-price bid for it. The bidder with the lowest score is then awarded the rights to implement the project. In the process of construction, it is common for items to be used in quantities that deviate from the DOT engineer's specification. All changes, however, must be approved by an on-site DOT manager. The winning contractor is ultimately paid the sum of her unit price bid multiplied by the *actual* quantity of each item used.

While contractors' ability to influence the item quantities that are ultimately used is limited, bidders may be able to predict which items will over/under-run the DOT's estimates. Consequently, DOT officials have expressed concerns that bidders may manipulate unit prices to take advantage of government inaccuracies and extract rents from the taxpayer till.

1.2.2 Views of Bid Skewing by Contractors and DOT Managers

Bid Skewing Among Contractors

The practice of *unbalanced bidding*—or *bid skewing*—in scaling auctions appears, in the words of one review, "to be ubiquitous" (Skitmore and Cattell (2013)). References to bid skewing in operations research and construction management journals date as far back as 1935

[16]The DOT's estimate of market rates are not advertised to prospective bidders, and are used primarily for internal budgeting purposes.

[17]Scaling auctions using paper-bids were used for over a decade prior to the introduction of Bid Express.

and as recently as 2010. A key component of skewing is the bidders' ability to predict quantity over/under-runs and optimize accordingly. Stark (1974), for instance, characterizes contemporary accounts of bidding:

> Knowledgeable contractors independently assess quantities searching for items apt to seriously underrun. By setting modest unit bids for these items they can considerably enhance the competitiveness of their total bid.

Uncertainty regarding the quantities that will ultimately be used presents a challenge to optimal bid-skewing, however. In an overview of "modern" highway construction planning, Tait (1971) writes:

> ...there is a risk in manipulating rates independently of true cost, for the quantities schedule in the bill of quantities are only estimates and significant differences may be found in the actual quantities measured in the works and on which payment would be based.

In order to manage the complexities of bid selection, contractors often employ experts and software geared for statistical prediction and optimization. Discussing the use of his algorithm for optimal bidding in consulting for a large construction firm, Stark (1974) notes a manager's prediction that such software would soon become widespread—reducing asymmetries between bidders and increasing allocative efficiency in the industry.

> ...since the model was public and others might find it useful as well, it had the longer term promise of eroding some uncertainties and irrelevancies in the tendering process. Their elimination...increased the likelihood that fewer contracts would be awarded by chance and that his firm would be a beneficiary.

Since then, an assortment of decision support tools for estimating item quantities and optimizing bids has become widely available. A search on Capterra, a web platform that facilitates research for business software buyers, yields 181 distinct results. In a survey on construction management software trends, Capterra estimates that contractors spend an average $2,700 annually on software. The top 3 platforms command a market share of 36% and surveyed firms report having used their current software for about 2 years—suggesting a competitive environment. Asked what was most improved by the software, a leading 21%

of respondents said, “estimating accuracy", while 14% (in third place) said “bidding".

DOT Challenges to Bid Skewing

Concerns that sophisticated bidding strategies may allow contractors to extract excessively large payments have led to a number of lawsuits about the DOT’s right to reject suspicious bids. The Federal Highway Administration (FHWA) has explicit policies that allow officials to reject bids that are deemed manipulative. However, the legal burden of proof for a manipulative bid is quite high. In order for a bid to be legally rejected, it must be proven to be *materially unbalanced*.[18]

> A bid is materially unbalanced if there is a reasonable doubt that award to the bidder ... will result in the lowest ultimate cost to the Government. Consequently, a materially unbalanced bid may not be accepted.[19]

However, as the definition for material unbalancedness is very broad, FHWA statute requires that a bid be *mathematically* unbalanced as a precondition. A *mathematically unbalanced* bid is defined as one, “structured on the basis of nominal prices for some work and inflated prices for other work."[20] In other words, it is a bid that appears to be strategically skewed. In order to discourage bid skewing, many regional DOTs use concrete criteria to define mathematically unbalanced bids. In Massachusetts, a bid is considered mathematically unbalanced if it contains any line-item for which the unit bid is (1) over (under) the office cost estimate and (2) over (under) the average unit bid of bidders ranked 2-5 by more than 25%.

In principle, a mathematically unbalanced bid elicits a flag for DOT officials to examine the possibility of material unbalancedness. However, in practice, such bids are ubiquitous, and substantial challenges by the DOT are very rare. In our data, only about 20% of projects do not have a single item breaking MassDOT’s overbidding rule, and only about 10% of

[18]See Federal Acquisition Regulations, Sec. 14.201-6(e)(2) for sealed bids in general and Sec. 36.205(d) for construction specifically (Cohen Seglias Pallas Greenhall and Furman PC (2018)).

[19]Matter of: Crown Laundry and Dry Cleaners, Comp. Gen. B-208795.2, April 22, 1983.

[20]Matter of: Howell Construction, Comp. Gen. B-225766 (1987)

projects do not have a single item breaking the underbidding rule. Indeed, most projects have a substantial portion of unit bids that should trigger a mathematical unbalancedness flag.[21] However, only 2.5% of projects have seen bidders rejected across all justifications, a handful of which were due to unbalanced bids.[22]

The Difficulty of Determining 'Materially Unbalanced' Bids

A primary reason that so few mathematically unbalanced bids are penalized is that material unbalancedness is very hard to prove. In a precedent-setting 1984 case, the Boston Water and Sewer Commission was sued by the second-lowest bidder for awarding a contract to R.J. Longo Construction Co., Inc., a contractor who had the lowest total bid along with a penny bid. The Massachusetts Superior Court ruled that the Commission acted correctly, since the Commission saw no evidence that the penny bid would generate losses for the state. More specifically, no convincing evidence was presented that if the penny bid did generate losses, the losses would exceed the premium on construction that the second-lowest bidder wanted to charge (Mass Superior Court, 1984).[23] In January 2017, MassDOT attempted to require a minimum bid for every unit price item in a various locations contract due to bid skewing concerns. SPS New England, Inc. protested, arguing that such rules preclude the project from being awarded to the lowest responsible bidder. The Massachusetts Assistant Attorney General ruled in favor of the contractor on August 1, 2017.

In fact, there is a theoretical basis to question the relationship between mathematical and material unbalancedness. As we demonstrate, bid skewing plays dual roles in bidders'

[21]See figures A.5a and A.5b in the appendix for more details.

[22]Note that MassDOT does not reject individual bidders, but rather withdraws the project from auction and possibly resubmits it for auction after a revision of the project spec.

[23]In response to this case, MassDOT inserted the following clause into Subsection 4.06 of the MassDOT Standard Specifications for Highways and Bridges: "No adjustment will be made for any item of work identified as having an unrealistic unit price as described in Subsection 4.04." This clause, inserted in the Supplemental Specifications dated December 11, 2002, made it difficult for contractors to renegotiate the unit price of penny bid items during the course of construction. An internal MassDOT memo from the time shows that Construction Industries of Massachusetts (CIM) requested that this clause be removed. One MassDOT engineer disagreed, writing that "if it is determined that MHD should modify Subsection 4.06 as requested by CIM it should be noted that the Department may not necessarily be awarding the contract to the lowest responsible bidder as required." The clause was removed from Subsection 4.06 in the June 15, 2012 Supplemental Specifications.

strategic behavior. On the one hand, bidders extract higher ex-post profits by placing higher bids on items that they predict will overrun in quantity. On the other hand, bidders reduce ex-ante risk by placing lower bids on items, regarding which they are particularly uncertain. Moreover, when bidders are similarly informed regarding ex-post quantities, the profits from predicting overruns are largely competed away in equilibrium, but the reduction in ex-ante risk is passed on to the DOT in the form of cost-savings.

1.2.3 An Illustrative Example

Consider the following simple example of infrastructure procurement bidding. Two bidders compete for a project that requires two types of inputs to complete: concrete and traffic cones. The DOT estimates that 10 tons of concrete and 20 traffic cones will be necessary to complete the project. Upon inspection, the bidders determine that the actual quantities of each item that will be used – random variables that we will denote q_c^a and q_r^a for concrete and traffic cones, respectively – are normally distributed with means $\mathbb{E}[q_c^a] = 12$ and $\mathbb{E}[q_r^a] = 16$ and variances $\sigma_c^2 = 2$ and $\sigma_r^2 = 1$.[24] We assume that the actual quantities are exogenous to the bidding process, and do not depend on who wins the auction in any way. Furthermore, we will assume that the bidders' expectations are identical across both bidders.[25]

The bidders differ in their private costs for implementing the project. They have access to the same vendors for the raw materials, but differ in the cost of storing and transporting the materials to the site of construction as well as the cost of labor, depending on the site's location, the state of their caseload at the time and firm-level idiosyncrasies. We therefore describe each bidder's cost as a multiplicative factor α of market-rate cost estimate for each item: $c_c = \$8$/ton for each ton of Concrete and $c_r = \$12$/pack for each pack of 100

[24]As we discuss in section 1.4, we assume that the distributions of q_c^a and q_r^a are independent conditional on available information regarding the auction. This assumption, as well as the assumption that the quantity distributions are not truncated at 0 (so that quantities cannot be negative) are made for the purpose of computational traceability in our structural model. Note that if item quantities are correlated, bidders' risk exposure is higher, and so our results can be seen as a conservative estimate of this case.

[25]These assumptions align with the characterization of highway and bridge projects in practice: the projects are highly standardized and all decisions regarding quantity changes must be approved by an on-site DOT official, thereby limiting contractors' ability to influence ex-post quantities.

traffic cones. Each bidder i knows her own type α^i at the time of bidding, as well as the distribution (but not realization) of her opponent's type.

To participate in the auction, each bidder i submits a unit bid for each of the items: b^i_c and b^i_r. The winner of the auction is then chosen on the basis of her *score*: the sum of her unit bids multiplied by the DOT's quantity estimates:

$$s^i = 10b^i_c + 20b^i_r.$$

Once a winner is selected, she will implement the project and earn net profits of her unit bids, less the unit costs of each item, multiplied by the *realized* quantities of each item that are ultimately used. At the time of bidding, these quantities are unrealized samples of random variables.

Bidders are endowed with a standard CARA utility function over their earnings from the project with a common constant coefficient of absolute risk aversion γ:

$$u(\pi) = 1 - \exp(-\gamma\pi).$$

Note that bidders are exposed to two sources of risk: (1) uncertainty over winning the auction; (2) uncertainty over the profits that they would earn at the realized ex-post quantity of each item.

The profit π that bidder i earns is either 0, if she loses the auction, or

$$\pi(\mathbf{b}^i, \alpha^i, \mathbf{c}, \mathbf{q}^a) = q^a_c \cdot (b^i_c - \alpha^i c_c) + q^a_r \cdot (b^i_r - \alpha^i c_r),$$

if she wins the auction. Bidder i's expected utility at the time of the auction is therefore given by:

$$\mathbb{E}[u(\pi(\mathbf{b}^i, \alpha^i, \mathbf{c}, \mathbf{q}^a))] = \left(\underbrace{1 - \mathbb{E}_{\mathbf{q}^a}\left[\exp\left(-\gamma \cdot \pi(\mathbf{b}^i, \alpha^i, \mathbf{c}, \mathbf{q}^a)\right)\right]}_{\text{Expected utility conditional on winning}} \right) \times \underbrace{(\Pr\{s^i < s^j\})}_{\text{Probability of winning with } s^i = 10b^i_c + 20b^i_r}.$$

That is, bidder i's expected utility from submitting a set of bids b^i_c and b^i_r is the product of the utility that she expects to get (given those bids) if she were to win the auction, and the probability that she will win the auction at those bids. Note that the expectation of utility conditional on winning is with respect to the realizations of the item quantities q_c^a and q_r^a, entirely.

As the ex-post quantities are distributed as independent Gaussians, the expected utility term above can be rewritten in terms of the certainty equivalent of bidder i's profits conditional on winning:[26]

$$1 - \exp\left(-\gamma \cdot \text{CE}(\mathbf{b}^i, \alpha^i, \mathbf{c}, \mathbf{q}^a)\right),$$

where the certainty equivalent of profits $\text{CE}(\mathbf{b}^i, \alpha^i, \mathbf{c}, \mathbf{q}^a)$ is given by:

$$\underbrace{\mathbb{E}[q_c^a] \cdot (b_c^i - \alpha^i c_c) + \mathbb{E}[q_r^a] \cdot (b_r^i - \alpha^i c_r)}_{\text{Expection of Profits}} - \underbrace{\left[\frac{\gamma\sigma_c^2}{2} \cdot (b_c^i - \alpha^i c_c)^2 + \frac{\gamma\sigma_r^2}{2} \cdot (b_r^i - \alpha^i c_r)^2\right]}_{\text{Variance of Profits}}. \quad (1.1)$$

Furthermore, as we discuss in section 1.4, the optimal selection of bids for each bidder i can be described as the solution to a two-stage problem:

Inner: For each possible score s, choose the bids b_c and b_r that maximize $\text{CE}(\{b_c, b_r\}, \alpha^i, \mathbf{c}, \mathbf{q}^a)$, subject to the score constraint: $10b_c + 20b_r = s$.

Outer: Choose the score $s^*(\alpha^i)$ that maximizes expected utility $\mathbb{E}[u(\pi(\mathbf{b}^i(s), \alpha^i))]$, where $\mathbf{b}^i(s)$ is the solution to the inner step, evaluated at s.

That is, at every possible score that bidder i might consider, she chooses the bids that sum to s for the purpose of the DOT's evaluation of who will win the auction, and maximize her certainty equivalent of profits conditional on winning. She then chooses the score that maximizes her total expected utility.

To see how this decision process can generate bids that appear mathematically unbalanced, suppose, for example, that the common CARA coefficient is $\gamma = 0.05$, and consider a bidder in this auction who has type $\alpha^i = 1.5$.[27] Suppose, furthermore, that the bidder has

[26] See section 1.4 and the appendix for a detailed derivation.

[27] That is, for each ton of concrete that will be used will cost, the bidder incur a cost of $\alpha^i \times c_c = 1.5 \times \$8 = \$12$, and for each pack of traffic cones that will be used, she will incur a cost of $\alpha^i \times c_r = 1.5 \times \$12 = \$18$.

decided to submit a total score of \$500. There are a number of ways in which the bidder could construct a score of \$500. For instance, she could bid her cost on concrete, $b_c^i = \$12$, and a dollar mark-up on traffic cones: $b_r^i = (\$500 - \$12 \times 10)/20 = \$19$. Alternatively, she could bid her cost on traffic cones, $b_r^i = \$18$, and a two-dollar mark-up on traffic cones: $b_c^i = (\$500 - \$18 \times 20)/10 = \$14$. Both of these bids would result in the same score, and so give the bidder the same chances of winning the auction. However, they yield very different expected utilities to the bidder. Plugging each set of bids into equation (1.1), we find that the first set of bids produces a certainty equivalent of:

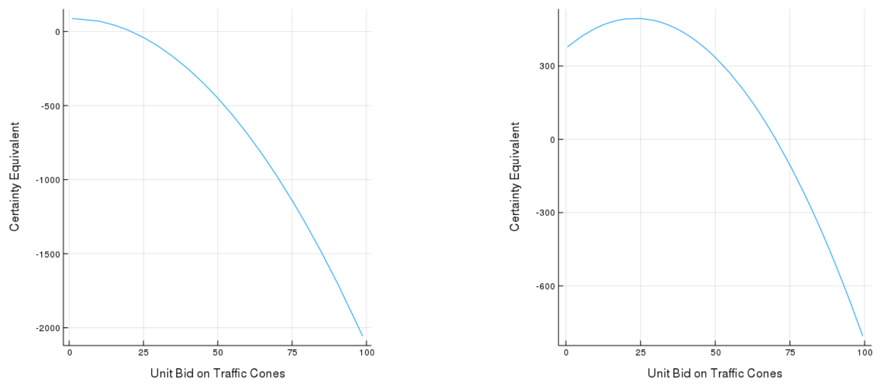
$$12 \times (\$0) + 16 \times (\$1) - \frac{0.05 \times 2}{2} \times (\$0)^2 - \frac{0.05 \times 1}{2} \times (\$1)^2 = \$15.98,$$

whereas the second set of bids produces a certainty equivalent of

$$12 \times (\$2) + 16 \times (\$0) - \frac{0.05 \times 2}{2} \times (\$2)^2 - \frac{0.05 \times 1}{2} \times (\$0)^2 = \$23.80.$$

In fact, further inspection shows that the optimal bids giving a score of \$500 are $b_c^i = \$47.78$ and $b_r^i = \$1.12$, yielding a certainty equivalent of \$87.98. The intuition for this is precisely that described by Athey and Levin (2001), and the contractors cited by Stark (1974): the bidder predicts that concrete will overrun in quantity – she predicts that 12 tons will be used, whereas the DOT estimated only 10 – and that traffic cones will underrun – she predicts that 16 will be used, rather than the DOT's estimate of 20. When the variance terms aren't too large (relatively), the interpretation is quite simple: every additional dollar bid on concrete is worth approximately 12/10 in expectation, whereas every additional dollar bid on traffic cones is worth only 16/20.

However, the incentive to bid higher on items projected to overrun is dampened when the variance term is relatively large. This can arise when the coefficient of risk aversion is relatively high or when the variance of an item's ex-post quantity distribution is high. More generally, as demonstrated in equation (1.1), the certainty equivalent of profits is increasing in the expected quantity of each item, $\mathbb{E}[q_c^a]$ and $\mathbb{E}[q_r^a]$, but decreasing in the variance of each item σ_c^2 and σ_r^2.



(a) *Score = $500*

(b) *Score = $1000*

Figure 1.1: *Certainty equivalent as a function of her unit bid on traffic cones, for the example bidder submitting a score of $500 or $1,000*

Moreover, the extent of bid skewing can depend on the level of competition in the auction. Figure 1.1 plots the bidder's certainty equivalent as a function of her unit bid on traffic cones when she chooses to submit a total score of (a) $500, and when she chooses to submit a score of (b) $1,000. In the first case, the bid that optimizes the certainty equivalent is very small, $b_r^i = \$1.12$. In the second case, however, the optimal bid is much higher at $b_r^i = \$23.33$. The reason for this is that a low bid on traffic cones implies a high bid on concrete. A high markup on concrete decreases the bidder's certainty equivalent at a quadratic rate. Thus, as the score gets higher, there is more of an incentive to spread markups across items, rather than bidding very high on select items, and very low on others.

1.2.4 Bid Skewing in Equilibrium

As we discuss in section 1.4, the auction game described above has a unique Bayes Nash Equilibrium. This equilibrium is characterized following the two-stage procedure described on in section 1.4.2: (1) given an equilibrium score $s(\alpha)$, each bidder of type α submits the vector of unit bids that maximizes her certainty equivalent conditional on winning, and sums to $s(\alpha)$; (2) The equilibrium score is chosen optimally, such that there does not exist a type α and an alternative score $\tilde{s}$, so that a bidder of type α can attain a higher expected utility with the score $\tilde{s}$ than with $s(\alpha)$.

The optimal selection of bids given an equilibrium score depends on the bidders' expectations over ex-post quantities and the DOT's posted estimates, as well as on the coefficient of risk aversion and the level of uncertainty in the bidders' expectations. High overruns cause bidders to produce more heavily skewed bids, whereas high risk aversion and high levels of uncertainty push bidders to produce more balanced bids.

In addition to influencing the relative skewness of bids, these factors also have a level effect on bidder utility. Higher expectations of ex-post quantities raise the certainty equivalent conditional on winning for every bidder. Higher levels of uncertainty (and a higher degree of risk aversion), however, induce a cost for bidders that lowers the certainty equivalent. Consequently, higher levels of uncertainty lower the value of participating for every bidder and result in less aggressive bidding behavior, and higher costs to the DOT in equilibrium.

To demonstrate this, we plot the equilibrium score, unit-bid distribution and ex-post revenue for every bidder type α in our example. To illustrate the effects of risk and risk aversion on bidder behavior and DOT costs, we compare the equilibria in four cases. First, we compute the equilibrium in our example laid out on page 11 when bidders are risk averse with CARA coefficient $\gamma = 0.05$, and when bidders are risk neutral (e.g. $\gamma = 0$). To hone in on the effects of risk in particular, and not mis-estimation, we will assume that the bidders' expectations of ex-post quantities are perfectly correct (e.g. the realization of q_c^a is equal to $\mathbb{E}[q_c^a]$, although the bidders do not know this ex-ante, and still assume their estimates are noisy with Gaussian error).

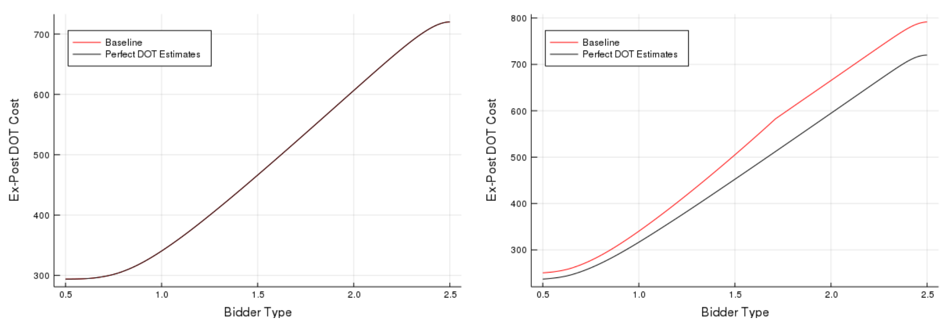
Next, we compute the equilibrium in each case under the counterfactual in which uncertainty regarding quantities is eliminated. In particular, we consider a setting in which the DOT is able to discern the precise quantities that will be used, and advertise the project with the ex-post quantities, rather than imprecise estimates. The DOT's accuracy is common knowledge, and so upon seeing the DOT numbers in this counterfactual, the bidders are certain of what the ex-post quantities will be (e.g. $\sigma_c^2 = \sigma_r^2 = 0$).

In Table 1.1, we present the expected (ex-post) DOT cost in each case. This is the

	Risk Neutral Bidders	Risk Averse Bidders
Noisy Quantity Estimates	\$326.76	\$317.32
Perfect Quantity Estimates	\$326.76	\$296.26

Table 1.1: *Comparison of Expected DOT Costs*

expectation of the amount that the DOT would pay the winning bidder $q_c^a b_c^w + q_r^a b_r^w$ at the equilibrium bidding strategy in each setting, taken with respect to the distribution of the type of the lowest (winning) bidder.[28] When bidders are risk neutral ($\gamma = 0$), the equilibrium cost to the DOT does not change when the DOT improves its quantity estimates. The reason for this is that since $\gamma = 0$, the variance term in equation (1.1) is zero regardless of the level of the noise in quantity predictions. As the bidders' quantity expectations $\mathbb{E}[q_c^a]$ and $\mathbb{E}[q_r^a]$ are unchanged, the expected revenue of the winning bidder (corresponding to the expected cost to the DOT) is unchanged as well.

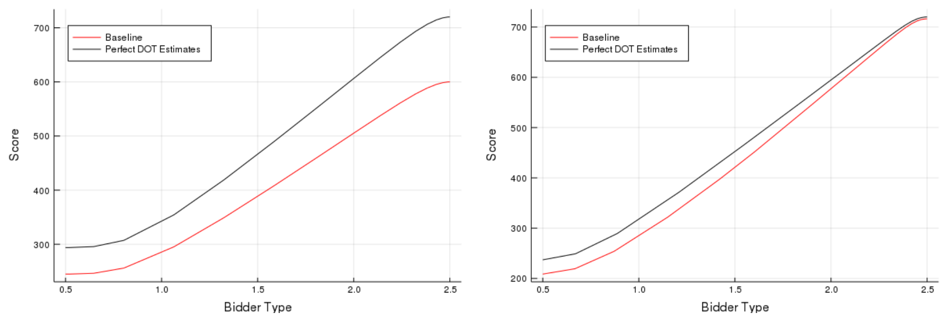


(a) *Risk Neutral Bidders* **(b)** *Risk Averse Bidders:* $\gamma = 0.05$

Figure 1.2: *Equilibrium DOT Cost/Bidder Revenue by Bidder Type*

In Figure 1.2a, we plot the revenue that each type of bidder expects to get in equilibrium when bidders are risk neutral. The red line corresponds to the baseline setting, in which the DOT underestimates the ex-post quantity of concrete, and overestimates the ex-post quantity of traffic cones. The black line corresponds to the counterfactual in which both

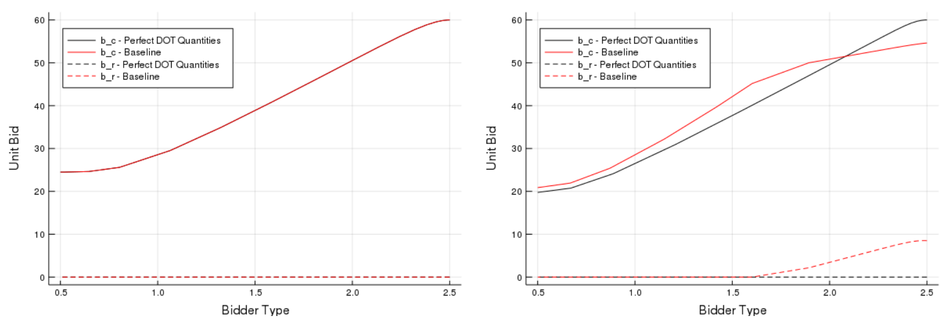
[28]In order to simulate equilibria, we need to assume a distribution of bidder types. For this example, we assume that bidder types are distributed according to a truncated lognormal distribution, $\alpha \sim \text{LogNormal}(0, 0.2)$ that is bounded from above by 2.5. There is nothing special about this particular choice, and we could easily have made others with similar results.



(a) *Risk Neutral Bidders* **(b)** *Risk Averse Bidders:* $\gamma = 0.05$

Figure 1.3: *Equilibrium Score Functions by Bidder Type*

quantities are precisely estimated, and bidders have no residual uncertainty about what the quantities will be. Note that while the ex-post cost to the DOT is the same whether or not the DOT quantity estimates are correct, the unit bids and resulting scores that bidders submit are different. In Figure 1.3a, we plot the equilibrium score for each bidder type when bidders are risk neutral. The score at every bidder type is smaller under the baseline than under the counterfactual in which the DOT discerns ex-post quantities. This is because the scores in the counterfactual correspond to the bidders' expected revenues, while the scores in the baseline multiply bids that are skewed to up-weight overrunning items by their under-estimated DOT quantities. See the appendix for a full derivation and discussion of the risk neutral case.



(a) *Risk Neutral Bidders* **(b)** *Risk Averse Bidders:* $\gamma = 0.05$

Figure 1.4: *Equilibrium Unit Bids by Bidder Type*

Figure 1.4a plots the unit bid that each type of bidder submits in equilibrium when bidders are risk neutral. As before, the red lines correspond to the baseline setting in which the DOT mis-estimates quantities, whereas the black lines correspond to the counterfactual setting in which the DOT discerns ex-post quantities perfectly. The solid line in each case corresponds to the unit bid for concrete $b_c(\alpha)$ that each α type of bidder submits in equilibrium. The dashed line corresponds to the equilibrium unit bid for traffic cones $b_r(\alpha)$ for each bidder type. Notably, in every case, the optimal bid for each bidder puts the maximum possible amount (conditional on the bidder's equilibrium score) on the item that is predicted to overrun the most, and \$0 on the other item. This is a direct implication of optimal bidding by risk neutral bidders, absent an external impetus to do otherwise. As noted by Athey and Levin (2001), this suggests that the observations of *interior* or *intermediately-skewed* bids in our data, as well as in Athey and Levin's, are inconsistent with a model of risk neutral bidders. Other work, such as Bajari, Houghton and Tadelis (2014) have rationalized interior bids by modeling a heuristic penalty for extreme skewing that represents a fear of regulatory rebuke. However, no significant regulatory enforcement against bid skewing has ever been exercised by MassDOT, and discussions of bidding incentives in related papers as well as in Athey and Levin (2001) suggest that risk avoidance is a more likely dominant motive.

In figures 1.2b, 1.3b and 1.4b, we plot the equilibrium revenue, score and bid for every bidder type, when bidders are risk averse with the CARA coefficient $\gamma = 0.05$. Unlike the risk-neutral case, the DOT's elimination of uncertainty regarding quantities has a tangible impact on DOT costs. When the DOT eliminates quantity risk for the bidders, it substantially increases the value of the project for all of the bidders, causing more competitive bidding behavior. Seen another way, uncertainty regarding ex-post quantities imposes a cost to the bidders, on top of the cost of implementing the project upon winning. In equilibrium, bidders submit bids that allow them to recover all of their costs (plus a mark-up). When uncertainty is eliminated, the cost of the project decreases, and so the total revenue needed to recover each bidder's costs decreases as well. Note, also, that the elimination of uncertainty results in different levels of skewing across the unit bids of different items. Whereas under the baseline, bidders with types $\alpha > 1.6$ place increasing interior bids on traffic cones, when risk is eliminated, this is no longer the case. However, this is subject to a tie breaking rule – when the DOT perfectly predicts ex-post quantities, there are no overruns, and so there is no meaningful different to over-bid on one item over the other. The analysis of the optimal bid (conditional on a score) here is analogous to that under risk neutrality, and so we defer details to the appendix.

CARA Coeff	Baseline	No Quantity Risk	Pct Diff
0	\$326.76	\$326.76	0%
0.001	\$326.04	\$325.62	0.13%
0.005	\$323.49	\$321.41	0.64%
0.01	\$321.01	\$316.88	1.29%
0.05	\$**317.32**	\$**296.26**	**6.64**%
0.10	\$319.83	\$285.57	10.71%

Table 1.2: *Comparison of expected DOT costs under different levels of bidder risk aversion*

While the general observation that reducing uncertainty may result in meaningful cost savings to the DOT, the degree of these savings depends on the baseline level of uncertainty in each project, as well as the degree of bidders' risk aversion and the level of competition in each auction (constituted by the distribution of cost types and the number of participating bidders). To illustrate this, we repeat the exercise summarized in Table 1.1 over different degrees of risk aversion and different levels of uncertainty. In Table 1.2, we present the expected DOT cost under the baseline example and under the counterfactual in which the DOT eliminates quantity risk, as well as the percent difference between the two, for a range of CARA coefficients.[29] The bolded row with a CARA coefficient of 0.05 corresponds to the right hand column of Table 1.1. We repeat this exercise across different magnitudes of prediction noise in Table A.10, in the appendix.

[29]That is, in the baseline, the DOT posts quantity estimates $q_c^e = 10$ and $q_r^e = 20$, while bidders predict that $\mathbb{E}[q_c^a] = 12$ and $\mathbb{E}[q_r^a] = 18$ with $\sigma_c^2 = 2$ and $\sigma_r^2 = 1$. In the No Quantity Risk counterfactual, the DOT discerns that $q_c^e = q_c^a = 12$ and $q_r^e = q_r^a = 18$, so that $\sigma_c^2 = \sigma_r^2 = 0$.

1.3 Data and Reduced Form Results

1.3.1 Data

Our data comes from MassDOT and covers highway and bridge construction and maintenance projects undertaken by the state from 1998 to 2015. We are limited by the extent of MassDOT's collection and storage of data on its projects. 4,294 construction and maintenance projects are in the DOT's digital records, although the coverage is sparse prior to the early 2000s. If we keep only the projects for which MassDOT has digital records on 1) identities of the winning and losing bidders; 2) bids for the winning and losing bidders; and 3) data on the actual quantities used for each item, we are left with 2,513 projects, 440 of which are related to bridge maintenance. We focus on bridge projects alone for this paper, as these projects are particularly prone to item quantity adjustments. Coverage is especially poor in the first few years of the available data and is especially good since 2008, when MassHighway became MassDOT and a push to improve digital records went into effect.[30]

MassDOT began using an online procurement service, called Bid Express, in April 2011. Prior to Bid Express, each contractor submitted his bids in paper form and MassDOT personnel then manually entered the bid data into an internal data set. The shift from a paper process to an online process thus likely helped data collection efforts and improved data accuracy.

The rules of the procurement process were the same, however, before and after April 2011. All bidders who participate in an auction have been able to see, ex-post, how everyone bid on each item. And all contractors have had access to summary statistics on past bids for each item, across time and location. Officially, all interested bidders find out about the specifications and expectations of each project at the same time, when the project is advertised (a short while before it opens up for bidding). Only those contractors who have been pre-qualified at the beginning of the year to do the work required by the project can bid on the project. Thus, contractors do not have a say in project designs, which are furnished

[30]See Table A.8 for a breakdown of the number of projects in our data, by year.

either in-house by MassDOT or by an outside consultant.

Once a winning bidder is selected, project management moves under the purview of an engineer working in one of 6 MassDOT districts around the state. The Project Manager assigns a Resident Engineer to monitor work on a particular project out in the field and to be the first to decide whether to approve or reject underruns, overruns, and Extra Work Orders (EWOs).[31] Underruns and overruns, as the DOT defines them and as we will define them here, apply to the items specified in the initial project design and refer to the difference between actual item quantities used and the estimated item quantities. EWOs refer to work done outside of the scope of the initial contract design and are most often negotiated as lump sum payments from the DOT to the contractor. For the purposes of our discussion and analyses, we will focus on underruns and overruns in projects relating to bridge construction and maintenance, as this is a focal point of interest to the DOT, as well as an area with a fair amount of uncertainty for the bidders.

Statistic	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Project Length (Estimated)	1.53 years	0.89 years	0.88 years	1.48 years	2.01 years
Project Value (DOT Estimate)	$2.72 million	$3.89 million	$981,281	$1.79 million	$3.3 million
# Bidders	6.55	3.04	4	6	9
# Types of Items	67.80	36.64	37	67	92
Net Over-Cost (DOT Quantities)	−$286,245	$2.12 million	−$480,487	−$119,950	$167,933
Net Over-Cost (Ex-Post Quantities)	−$26,990	$1.36 million	−$208,554	$15,653	$275,219
Extra Work Orders	$298,796	$295,173	$78,775	$195,068	$431,188

Table 1.3: *Summary Statistics*

Table 1.3 provides summary statistics for the bridge projects in our data set. We measure the extent to which MassDOT overpays the projected project cost in two ways. First, we consider the difference between what the DOT ultimately pays the winning bidder (the sum of the actual quantities used, multiplied the winning bidder's unit bids) and the DOT's initial estimate (the sum of the DOT's quantity estimates, multiplied by the DOT's estimate for each item's unit cost). Summary statistics for this measure are presented in the "Net Over-Cost (DOT Quantities)" row of Table 1.3. While it seems as though the DOT is saving

[31]The full approval process of changes in the initial project design involves N layers of review.

money on net, this is a misrepresentation of the costs of bid skewing. As we demonstrated in section 1.2, the DOT's estimate, which can be thought of the *score* evaluated using the DOT's unit costs as bids, is not representative of the ex-post amount to be paid at those bids. Rather, a more appropriate metric is to compare the amount ultimately spent against the dot product of the the DOT's unit cost estimates and the actual quantities used. This is presented in the "Net Over-Cost (Ex-Post Quantities)" row of Table 1.3. The median over-payment by this metric is about $15,000, but the 25th and 75th percentiles are about -$210,000 and $275,000. Figure 1.5 shows the spread of over-payment across projects. As we will show in our counterfactual section, the distribution of over-payment corresponds to the potential savings from the elimination of risk.

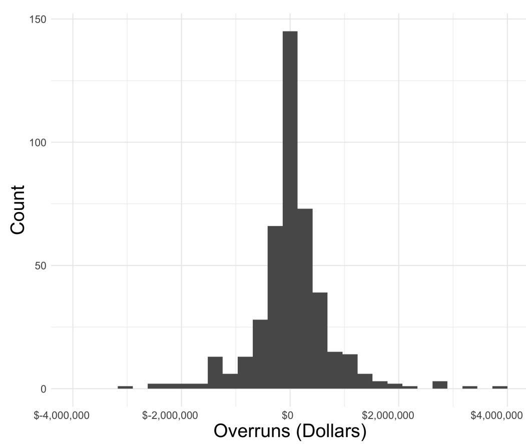


Figure 1.5: *Net Over-Cost (Ex-Post Quantities) Across Bridge Projects*

Description of Bidders

Across our data set, there are 2,883 unique project-bidder pairs (e.g. total bids submitted) across the 440 projects that were auctioned off. There are 116 unique firms that participate, albeit to different degrees. We distinguish firms that are rare participants by dividing firms into two groups: 'common' firms, which participate in at least 30 auctions within our data set, and 'rare firms', which participate in less than 30 auctions. We retain the individual identifiers for each of the 24 common firms, but group the 92 rare firms together for purposes of estimation. Common firms constitute 2,263 (78%) of total bids submitted, and 351 (80%) of auction victories.

Bidder Name	No. Employees	No. Auctions Participated
MIG Corporation	80	297
Northern Constr Services LLC	80	286
SPS New England Inc	75	210
ET&L Corp	1	201
B&E Construction Corp	9	118
NEL Corporation	68	116
Construction Dynamics Inc	22	113
S&R Corporation	20	111
New England Infrastructure	35	95
James A Gross Inc	7	78

Table 1.4: *Number of employees is drawn from estimates on LinkedIn and Manta*

Common Firm	Common Firm	Rare Firm
Number of Firms	24	92
Total Number of Bid Submitted	2263	620
Mean Number of Bid Submitted Per Firm	94.29	6.74
Median Number of Bid Submitted Per Firm	63.0	2.5
Total Number of Wins	351	89
Mean Number of Wins Per Firm	14.62	0.97
Median Number of Wins Per Firm	10	0
Mean Bid Submitted	$2,774,941	$4,535,310
Mean Ex-Post Cost of Bid	$2,608,921	$4,159,949
Mean Ex-Post Overrun of Bid	9.7%	21.97%
Proportion of Bids on Projects in the Same District	28.19	15.95
Proportion of Bids by Revenue Dominant Firms	51.67	11.80
Mean Specialization	24.44	2.51
Mean Capacity	10.38	2.75
Mean Utilization Ratio	53.05	25.50

Table 1.5: *Comparison of Firms Participating in <30 vs 30+ Auctions*

Table 1.4 presents the number of auctions participated in by each of the top 10 most common firms, as well as estimates of the number of full time employees on their payrolls. While the employee count numbers presented here are estimates, and may not include additional labor hired on a project-by-project basis, these firms are all relatively small, private, family-owned businesses.[32] Table 1.5 presents summary statistics of the two firm groups. The mean (median) common firm submitted bids to 94.29 (63) auctions and won 14.62 (10) of them. The mean total bid (e.g. the score) submitted is about $2.8 million, while the mean ex-post DOT cost implied by the firm's unit bids is $2.6 million. The mean ex-post cost overrun (the percent difference of the sum of unit bids multiplied by the ex-post quantities and the sum of blue book costs multiplied by the ex-post quantities) is 9.73%. By contrast, the mean (median) rare firm submitted bids to 6.74 (2.5) auctions and won 0.97 (0) of them. The mean total bid and ex-post scores are quite a bit larger than the common firms – $4.5 million and $4.2 million respectively, and this is reflected in a substantially larger ex-post overrun: 21.97% on average.

In addition to the firm's identity, there are a number of factors which may influence its competitiveness in a given auction. One such factor is the firm's distance from the project. Although we do not observe precise locations for each project in our data, we observe which of the 6 geographic districts that MassDOT jurisdiction is broken into each project belongs to. We then geocode the headquarters of each firm by district, and compare districts for each project-bidder pair. Among common firms, 28.19% of bids were on projects that were located in the same district as the bidding firm's headquarters. By contrast, only 15.95% of bids among rare firms were in matching districts.

Another measure of competitiveness is specialization—firms with extensive experience bidding on and implementing a certain type of project may find it cheaper to implement an additional project of the same sort. Our data involves three distinct project types, according to the DOT taxonomy: Bridge Reconstruction/Rehabilitation projects, Bridge Replacement projects, and Structures Maintenance projects. We calculate the specialization of a project-

[32]All 24 most common firms in our sample are privately owned, and so there is no publicly available, verifiable information on their revenues or expenses. The numbers of employees presented here are drawn from Manta, an online directory of small businesses, and cross-referenced with LinkedIn, on which a subset of these firms list a range of their employee counts. Note that there is some ambiguity as to who "counts" as an employee, as such firms often hire additional construction laborers on a project-by-project basis. The "family owned" label is drawn from the firms' self-descriptions on their websites.

bidder pair as the share of auctions of the same project type that the bidding firm has placed a bid on within our dataset. The mean specialization of a common firm is 24.44%, while the mean specialization of a rare firm is 2.51%. As projects have varying sizes, we compute a measure of specialization in terms of project revenue as well. We define a revenue-dominant firm (within a project-type) as a firm that has been awarded more than 1% of the total money spent by the DOT across projects of that project type. Among common firms, 51.67% of bids submitted were by firms that were revenue dominant in the relevant project type; among rare firms, the proportion of bids by revenue dominant firms is 11.8%.

A third factor of competitiveness is each firm's capacity – the maximum number of DOT projects that the firm has ever had open while bidding on another project – and its utilization – the share of the firm's capacity that is filled when she is bidding on the current project.[33] The mean capacity is 10.38 projects among common firms and 2.75 projects among rare firms. This suggests that rare firms generally have less business with the DOT (either because they are smaller in size, or because the DOT constitutes a smaller portion of their operations). The mean utilization ratio, however, is 53.05% for common firms and 25.5% for rare firms. This suggests that firms in our data are likely to have ongoing business with the DOT at the time of bidding, and are likely to have spare capacity during adjacent auctions that they did not participate in.[34]

Description of Quantity Estimates and Uncertainty

As we discuss in section 1.2, scaling auctions improve social welfare by enabling risk-averse bidders to insure themselves against uncertainty about the item quantities that will ultimately be used for each project. The welfare benefit is particularly strong if the uncertainty regarding item ex-post quantities varies across items within a project, and especially so if there are a few items that have particularly high variance. When this is

[33] We measure capacity and utilization with respect to all projects with MassDOT recorded in our data – not just bridge projects.

[34] Note that while we do not take dynamic considerations of capacity constraints into consideration, we find our measure of capacity to be a useful metric of the extent of a firm's dealings with the DOT, as well as its size.

the case, bidders in a scaling auction can greatly reduce the risk that they face by placing minimal bids on the highly uncertain items (and higher bids on more predictable items).[35]

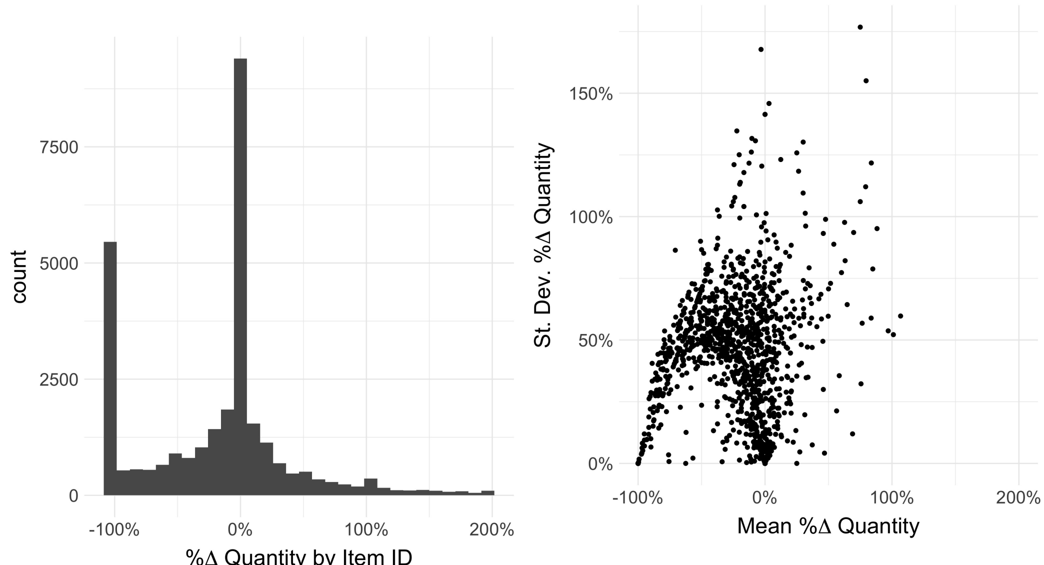
Our data set includes records of 2,985 unique items, as per MassDOT's internal taxonomy. Spread across 440 projects, these items constitute 29,834 unique item-project pairs. Of the 2,985 unique items, 50% appear in only one project. The 75th, 90th and 95th percentiles of unique items by number of appearances in our data are 4, 16 and 45 auctions, respectively.[36]

For each item, in every auction, we observe the quantity with which MassDOT predicted it would be used at the time of the auction – q_t^e in our model – the quantity with which the item was ultimately used – q_t^a – and a *blue book* DOT estimate for the market rate for the unit cost of the item. The DOT quantities are typically inaccurate: 76.7% of item observations in our data had ex-post quantities that deviated from the DOT estimates. Figure 1.6a presents a histogram of the percent quantity overrun across observations of items. The percent quantity overrun is defined as the difference of the ex-post quantity of an item observation and its DOT quantity estimates, normalized by the DOT estimate: $\frac{q_t^a - q_t^e}{q_t^e}$. In addition to the 23.3% item-project observations in which quantity overruns are 0%, another 18% involve items that aren't used at all (so that the overrun is equal to -100%). The remaining overruns are distributed, more or less symmetrically, around 0%. Furthermore, quantity overruns vary across observations of the same item in different auctions. Figure 1.6b plots the mean percent quantity overrun for each unique item with at least 2 observations against its standard deviation . While a few items have standard deviations close to 0, the majority of items have overrun standard deviations that are as large or larger than the absolute value of their

[35]A number of different factors may influence the extent of item over/under-runs in a given project: the type of maintenance needed, underlying state of the structure, time since assessment and skill of the project designer, chief among them. While our dataset is insufficient to robustly estimate the causal effects of these features on overruns, we present a brief discussion of the variation observed across DOT designers and project managers in the appendix.

[36]Part of the reason that so many unique items appear so rarely in our data is that the DOT taxonomy is very specific. For instance, item 866100 – also known as "100 Mm Reflect. White Line (Thermoplastic)" – is distinct from item 867100 – "100 Mm Reflect. Yellow Line (Thermoplastic)," although clearly there is a relationship between them. In order to take these similarities into account, we project item-project pairs onto characteristic space constructed by natural language parsing of the item descriptions, as well as a number of numerical item-project features. We discuss this at greater length in the estimation section.

means. That is, the percent overrun of the majority of unique items varies substantially across observations.[37] While this is a coarse approximation of the uncertainty that bidders face with regard to each item—it does not take item or project characteristics into account, for example—it is suggestive of the scope of risk in each auction.



(a) *Histogram of the percent quantity overrun across item-project pairs.*

(b) *Plot of the mean against the standard deviation of percent quantity overruns within each unique item.*

Figure 1.6: *Descriptive plots for item quantity overruns*

1.3.2 Reduced Form Evidence for Risk Averse Bid Skewing

As in Athey and Levin (2001) and Bajari, Houghton and Tadelis (2014), the bids in our dataset are consistent with a model of similarly informed bidders who bid strategically to maximize expected utility. In Figure 1.7, we plot the relationship between quantity overruns and the percent by which each item was overbid above the blue book cost estimate by the winning bidder.[38] The binscatter is residualized. In order to obtain it, we first regress

[37]The statement of majority here is with respect to items that appear multiple times.

[38]The percent overbid of an item is defined as $\frac{b_t - c_t}{c_t} \times 100$ where b_t is the bid on item t and c_t is the blue book unit cost estimate of item t. The percent quantity overrun is similarly defined as $\frac{q_t^a - q_t^e}{q_t^e} \times 100$ where q_t^a is the amount of item t that was ultimately used and q_t^e is the DOT quantity estimate for item t that is used to calculate bidder scores.

percent overbid on a range of controls and obtain residuals. We then regress percent overrun on the same controls and obtain residuals. Finally, to obtain the slope in red, we regress the residuals from the first regression on the residuals from the second. Controls include the DOT estimate of total project cost, initially stated project length in days, number of bidders, and fixed effects for the year in which the project was opened for bidding, project type, resident engineer, project manager, and project designer, as well as item fixed effects. Specifications that exclude item fixed effects or include an array of additional controls produce a very similar slope.[39] We use a similar procedure for all residualized bin-scatters in this section.

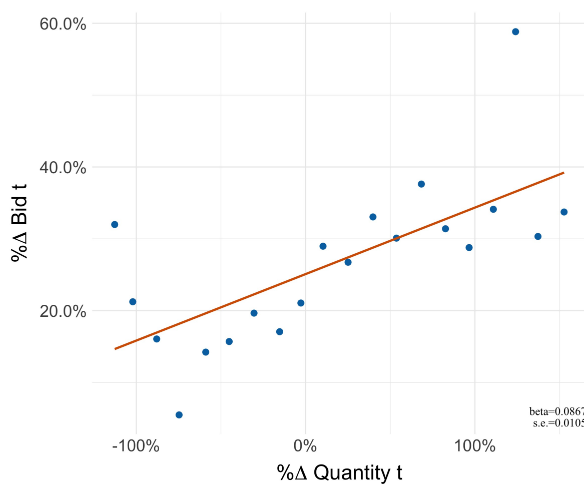


Figure 1.7: *Residualized bin-scatter of item-level percent winner overbid against percent quantity overrun*

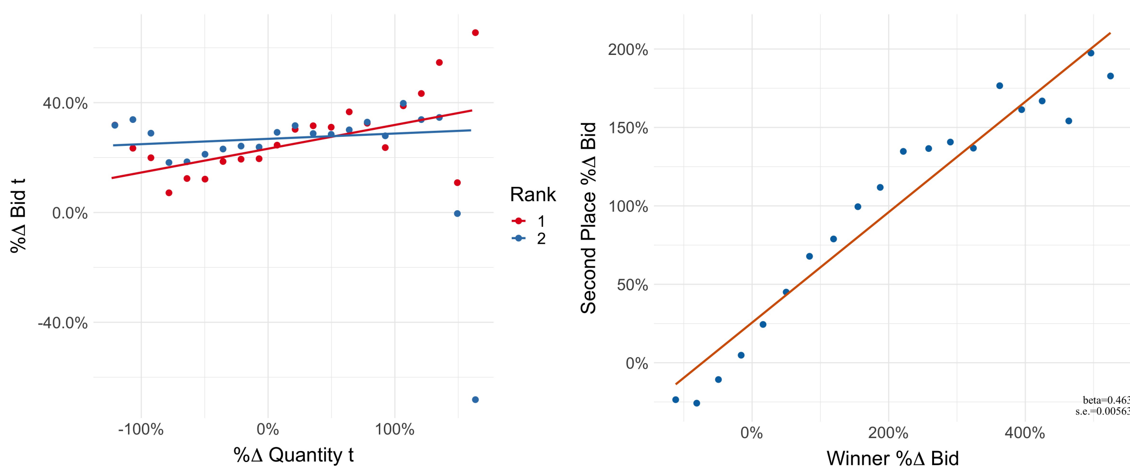
As Figure 1.7 demonstrates, there is a significant positive relationship between percent quantity overruns and percent overbids by the winning bidder. A 1% increase in quantity overruns corresponds to a 0.085% increase in overbids on average.[40] This suggests that the winning bidder is able to correctly predict which items will overrun on average. As in the example in Section 1.2, items predicted to overrun generally recieve higher bids. Thus, as higher bids correspond to items that overran in our data, we conclude that bidders are

[39]For each graph, we truncate observations at the top and bottom 1%. This is done for the purposes of clarity as outliers can distort the visibility of the general trends. We include untruncated versions in an online appendix for robustness.

[40]See the appendix for a full regression report.

informed beyond the DOT quantity estimates and skewing strategically.

Furthermore, the bid skewing relationship is similar across bidders beside the winner. Figure 1.8a plots the residualized bin-scatter of percent overbids against percent quantity overruns for the winning bidder and the second-place bidder in each auction. With the exception of a few outlying points, the relationship between overbids and overruns is very similar between the top two bidders. In the appendix, we show that this relationship is even stronger when we restrict the comparison to projects in which the first two bidder submit similar total scores. Figure 1.8b plots a residualized bin-scatter of the winning bidder's unit bid for each item against the second place bidder's bid for the same item. Overall, the direction of skewing corresponds strongly between the top two bidders – a higher overbid by the winning bidder corresponds to a higher overbid by the second place bidder as well.[41] Together, these figures suggest that bidders have access to the same information regarding quantity overruns.



(a) *Residualized bin-scatter of item-level percent overbid by the rank 1 (winning) and rank 2 bidder, against percent quantity overrun.*

(b) *Bin-scatter of item-level percent overbids by the rank 2 bidder against the rank 1 (winning) bidder.*

Figure 1.8: *Reduced form evidence that bidders have access to the same information regarding quantity overruns*

[41]Note that the percent overbids in Figure 1.8b appear to be substantially larger than those in Figure 1.8a. This is because while large overbids occur in the data, they are relatively rare and so are averaged down in the percent quantity overrun binning of Figure 1.8a.

While our data suggests that bidders do engage in bid skewing, there is no evidence of *complete* bid skewing, in which a few items are given very high unit bids and the rest are given “penny bids". The average number of unit bids worth $0.10 or less by the winning bidder is 0.51—or 0.7% of the items in the auction. The average number for unit bids worth $0.50, $1.00, and $10.00, respectively is 1.68, 2.85 and 13.91, corresponding to 2.8%, 4.73%, and 23.29% of the items in the auction. This observation is consistent with previous studies of bidding in scaling auctions. Athey and Levin (2001) argue that the interior bids observed in their data are suggestive of risk aversion among the bidders. While they acknowledge that other forces, such as fear of regulatory rebuke, may provide an alternative explanation for the lack of total bid skewing, they note that risk avoidance was the primary explanation given to them in interviews with professionals.

In addition to interior bids, risk aversion has several testable theoretical implications. First, risk averse bidders are predicted to bid more aggressively on projects that are worth more. A true reduced form test for this would require a ceteris perebis comparison of bid outcomes on identical auctions that only vary on project size. However, a suggestive proxy for aggressive bidding is the percent net over-cost: the percent by which the total amount paid to the winner in each project exceeds the total project value given by the DOT’s blue book unit cost estimates. As shown in Figure 1.9, this relationship is generally negative in our data. Interpreting percent net over-cost as a proxy for markups, this suggests that bidders extract less rents (percentage-wise) in auctions with higher stakes, as risk averse behavior would imply.

Furthermore, as we discuss in section 1.2, risk averse bidders balance the incentive to bid high on items that are projected to overrun with an incentive to bid lower on items that are uncertain. As such, we would expect bidders to bid lower on items that – everything else held fixed – have higher uncertainty. While we do not see observations of the same item in the same context with identifiably different uncertainty, we present the following suggestive evidence. In figures 1.10a and 1.10b, we plot the relationship between the unit bid for each item in each auction by the winning bidder, and an estimate of the level of

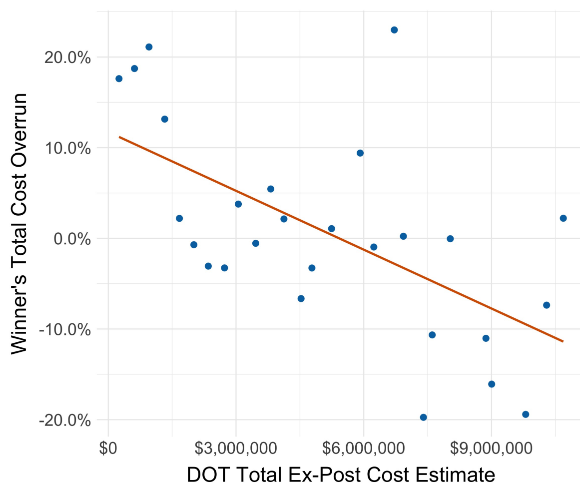


Figure 1.9: *Bin-scatter of the percent net over-cost against the total DOT estimate for the project cost*

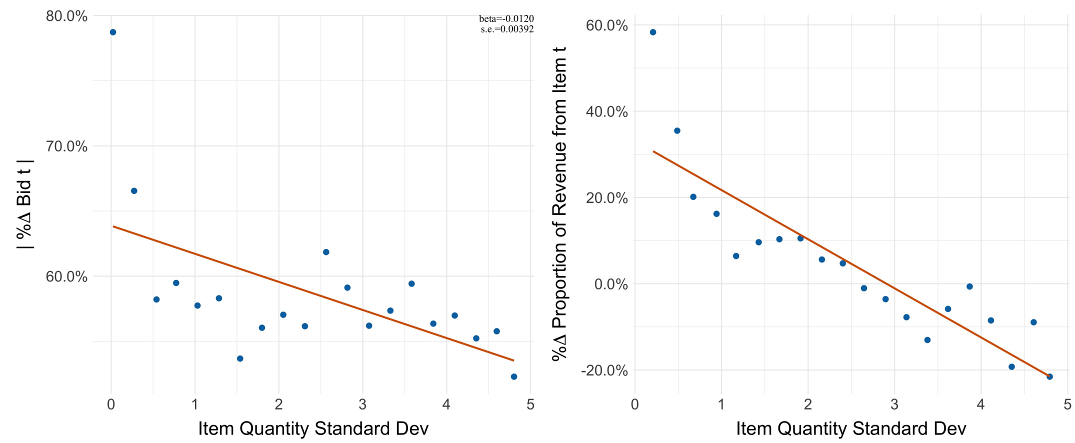
Note: DOT estimates are calculated with blue book cost estimates and ex-post quantity realizations.

uncertainty regarding the ex-post quantity of that item (in the context of the particular auction). To calculate the level of uncertainty for each item, we use the results of our first stage estimation, discussed in section 1.5.[42] For every item, in every auction, our first stage gives us an estimate of the variance of the error on the best prediction of what the ex-post quantity of that item would be, given the information available at the time of bidding.

In Figure 1.10a, we plot a residualized binscatter of the winning bidder's absolute percent overbid on each item against the item's standard deviation – the square root of the estimated prediction variance. The relationship is negative, suggesting that holding all else fixed, bidders bid closer to cost on items with higher variance, limiting their risk exposure.[43]

[42]As we discuss in section 1.5, we fit a model for the distribution of the ex-post quantity of each item in each auction. The model has two parts: first, we model the ex-post quantity of each item observation as a linear function of the DOT quantity estimate for that item and a vector of item-auction specific features, given a Gaussian error. Second, we model the variance of the Gaussian error in each observation to a lognormal distribution, the mean of which is also a linear function of the DOT quantity estimate and item-auction features. We fit this model jointly just Hamiltonian Monte Carlo using the full history of item-auction observations in our data set. Intuitively this is akin to projecting the ex-post quantity of each item observation onto its DOT estimate and feature vector, and then parametrically fitting the resulting residuals to a lognormal distribution.

[43]To account for the impact of quantity expectations, we include $\%\Delta q_t$ as a control in the specification when residualizing. However, the qualitative negative relationship persists even if we exclude it. We present this in section A.6.4 of the appendix for completeness.



(a) *Residualized bin-scatter of item-level percent absolute overbid against the square root of estimated item quantity variance.*

(b) *Residualized bin-scatter of item-level percent difference in cost contribution, against the square root of estimated item quantity variance*

Figure 1.10: *Reduced form evidence of risk aversion in observed bids*

Note, however, that this analysis does not directly account for the trade-off between quantity overruns and uncertainty. As in equation (1.1), a bidder's certainty equivalent increases in the predicted quantity of each item, but decreases in the item's quantity variance. To account for this trade-off, we consider the following alternative metric for bidding high on an item:

$$\%\Delta \text{ Cost Contribution from } t \;=\; \frac{\frac{b_t q_t^a}{\sum_p b_p q_p^a} - \frac{c_t q_t^e}{\sum_p c_t q_p^e}}{\frac{c_t q_t^e}{\sum_p c_t q_p^e}} \times 100$$

This is the percentage difference in the proportion of the total revenue that the winning bidder earned that was due to item t, and the proportion of the DOT's initial cost estimate that item t constituted. In Figure 1.10b, we plot the residualized bin scatter of the %Δ Cost Contribution due to each item against the item's quantity standard deviation. The negative relationship here is particularly pronounced, providing further evidence that bidders allocate proportionally less weight in their expected revenue to items with high variance, as our model of risk averse bidding predicts.

1.4 A Structural Model for Bidding With Risk Aversion

1.4.1 Setup

A procurement project is characterized by T items, each of which is needed in a different quantity. MassDOT (henceforth, "the buyer" or 'the DOT") initiates an auction for the project by posting a list of the T items, along with a vector of *estimated* quantities $\mathbf{q}^e = \{q_1^e, \dots, q_T^e\}$, with which it expects each item to be used. Once the auction is complete, the project is implemented in full by the winning bidder using the *actual* (ex-post) quantity q_t^a for each item t. The actual quantities $\mathbf{q}^a = \{q_1^a, \dots, q_T^a\}$ are assumed to be fixed but unknown at the time of the auction. That is, from the perspective of the buyer and the bidders, the vector of actual quantities $\mathbf{q}^a$ is an exogenous random variable. The realization of $\mathbf{q}^a$ is independent of which bidder wins the auction, and at what price.[44]

The auction is simultaneous with sealed bids, but both the set of $m > 1$ participating bidders and the buyer's quantity estimates $\mathbf{q}^e$ are fixed and common knowledge to all participants at the start of the auction. In addition, prior to the auction, the bidders receive a symmetric noisy signal $\mathbf{q}^b = \{q_1^b, \dots, q_T^b\}$ of what the ex post quantities for the project will be:

$$q_t^b = q_t^a + \epsilon_t \text{ where } \epsilon_t \sim \mathcal{N}(0, \sigma_t^2). \tag{1.2}$$

For simplicity, we assume that the signals are common across bidders. Thus, all bidders have the same expected value q_t^b for the actual quantity of item t, and the same variance σ_t^2, with which this estimate is off.[45]

Bidders differ in their private cost of production along a single dimensional *efficiency multiplier* α. At the time of the auction, every item t has a commonly-known market unit cost

[44]This assumption, which follows Bajari, Houghton and Tadelis (2014) and Athey and Levin (2001), precludes the possibility of asymmetric moral hazard. In our reduced form section, we argue that the similarity in projected overruns by the winning bidder and the runner-up suggests that if moral hazard affects bidding, its effects are anticipated symmetrically by bidders so that this assumption, too, will not harm our estimates greatly. It also precludes substitutability between items. While we cannot rule substitutions out, we argue that their scope is limited as only items on the DOT designer's project specification may be used for construction.

[45]It is not without loss of generality to assume that signals are common across bidders. However, we make this assumption for the sake of tractability.

c_t. This cost represents that market price of the materials—generally things like concrete, traffic cones, etc., which are standard and competitive—at the scale necessary for the project. However, the bidders vary in their labor and transportation costs, storage capacity, etc., yielding a multiplicative (dis)advantage over competitors. In particular, for every item t in the project, bidder i faces a unit cost of $\alpha^i c_t$ where α^i is the bidder's efficiency (multiplier) type. The efficiency type of each bidder i is drawn independently from a common, publicly known distribution with a well behaved density $f(\alpha^i)$ over a compact subset $[\underline{\alpha}, \overline{\alpha}]$ of $\mathbb{R}_+$.[46] Each bidder privately observes only her own efficiency type prior to the auction, but the distribution of competitor types is common knowledge.

To participate, each bidder i submits a vector of unit prices $\mathbf{b}^i = \{b_1^i, \ldots, b_T^i\}$, setting the amount per unit that she will be paid for each item if she wins. The winner of the auction is determined according to a first-price scoring rule. Each bidder i is given a *score* based on her unit bids and the DOT quantity estimates:

$$s^i = \sum_{t=1}^{T} b_t^i q_t^e.$$

The bidder with the lowest score wins the contract and implements the project in full. Upon the completion of the project, the actual (ex-post) quantities $\mathbf{q}^a$ of the items are realized, and the winning bidder is paid her unit bid b_t^i multiplied by the ex-post quantity q_t^a for each item. The winning bidder is responsible for securing all of the materials and labor for the project privately, and so she also incurs a cost of $\alpha^i c_t$ multiplied by q_t^a for each item.[47]

Finally, we model the bidders as risk averse, with a standard CARA utility function over their earnings from the project and a common constant coefficient of absolute risk aversion

[46]The assumption that the distribution of efficiency types is common (e.g. not specific to individual bidders) is not critical to our analysis, and relaxing it would not substantially change our estimation method or results, although it might impact the counterfactuals.

[47]Note that only the winner of the auction incurs any costs. All losing bidders receive no further cost nor revenue from the project at hand, once the auction is complete.

γ:[48]

$$u(\pi) = 1 - \exp(-\gamma\pi). \tag{1.3}$$

The profit π that bidder i earns is either 0, if she loses the auction, or

$$\pi(\mathbf{b}^i, \alpha^i, \mathbf{c}, \mathbf{q}^a) = \sum_t^T q_t^a \cdot (b_t^i - \alpha^i c_t),$$

if she wins the auction. Note that as $\mathbf{q}^a$ is a random variable from the bidder's perspective at the time of bidding, her profit from winning is stochastic as well.

Bidder i choose her bids so as to maximize her expected utility at the time of the auction:

$$\left(\underbrace{1 - \mathbb{E}_{\mathbf{q}^a}\left[\exp\left(-\gamma \sum_{t=1}^T q_t^a \cdot (b_t^i - \alpha^i c_t)\right)\right]}_{\text{Expected utility conditional on winning}}\right) \cdot \underbrace{(\Pr\{s^i < s^j \text{ for all } j \neq i\})}_{\text{Probability of winning with } s^i = \mathbf{b}^i \cdot \mathbf{q}^e} \tag{1.4}$$

where we suppress the common auction characteristics $\mathbf{c}, \mathbf{q}^e, \mathbf{q}^a$ as arguments in the utility and profit functions for ease of exposition. This is bidder i's expected utility over her profit if she were to win the auction, multiplied by the probability that her score – at the chosen unit bids – will be the lowest one offered, so that she will win. Note that the expectation in the first term is with respect to $\mathbf{q}^a$.

Bidders form their expectations based on the posterior distribution of each q_t^a given by equation (1.2) at their signals q_t^b and σ_t^2. The expected utility of bidder i can therefore be rewritten:

$$\begin{aligned}&\left(1 - \mathbb{E}_\epsilon\left[\exp\left(-\gamma \sum_{t=1}^T (q_t^b - \epsilon_t) \cdot (b_t^i - \alpha^i c_t)\right)\right]\right) \cdot (\Pr\{s^i < s^j \text{ for all } j \neq i\}) \\ &= \left(1 - \exp\left(-\gamma \sum_{t=1}^T q_t^b (b_t^i - \alpha^i c_t) - \frac{\gamma\sigma_t^2}{2}(b_t^i - \alpha^i c_t)^2\right)\right) \cdot (\Pr\{s^i < s^j \text{ for all } j \neq i\}).\end{aligned}$$

where the first equality is given by rewriting $q_t^a = q_t^b - \epsilon_t$, so that the expectation operator

[48] Note that equation (1.3) can be thought of as a normalization of the CARA utility function $u_i(\pi) = \exp(-\gamma_w) - \exp(-\gamma_w(w + \pi))$ where w is the bidder's wealth independently of the auction, and $\gamma_w = \frac{\gamma}{w}$ is the unnormalized CARA coefficient. When w is the same across all of the bidders in the auctions, this normalization is without loss of generality. While this is a strong assumption, we will maintain it throughout the main part of this paper for the purpose of tractability in this draft

in the profit term is with respect to the distribution of $\boldsymbol{\epsilon}$. The second equality follows from the closed form solution to this expectation.[49]

1.4.2 Equilibrium Bidding Behavior

We now characterize the Bayesian Nash Equilibrium of the static first-price sealed bid scoring auction described in the previous section. Our setting is similar to Bajari, Houghton and Tadelis (2014), which uses a special case of the Asker and Cantillon (2008) model, in which the project and its value to the buyer are fixed and independent of the winning bidder. We consider a linear scoring auction game with independent private values that can be characterized by a uni-dimensional "pseudo-type"— each bidder's *efficiency multiplier type* α.[50] As in Bajari, Houghton and Tadelis (2014) and Asker and Cantillon (2008), the optimal bidding problem in our setting can be decomposed into two parts: (1) given an efficiency type α, choose the optimal score s; (2) given a score s, choose the optimal bid vector $\mathbf{b}$ subject to the constraint that $\mathbf{b} \cdot \mathbf{q}^e = s$. As we describe below, the optimal choice of $\mathbf{b}$ conditional on a choice of s, a type α, and the auction characteristics, is deterministic and independent of competitive considerations. Therefore, at the optimum, the value of winning the auction to a bidder of type α, submitting a score s – that is, the bidder's expected utility from winning the auction using the optimal vector of bids $\mathbf{b}$ that yield s – is determined entirely by her choice of s, and is monotonically increasing in s. Following a sub-case of Lebrun (2006), this game has a unique monotonic equilibrium in pure strategies.

We derive the equilibrium as follows for an arbitrary bidder i with efficiency type α^i:

1. Given a (winning) score s, we find the optimal bid vector $\mathbf{b}^i(s)$ s.t. $\sum_{t=1}^{T} b_t^i(s) q_t^e = s$.

[49] $\mathbb{E}[\exp(-\gamma c \epsilon)] = \exp(-\gamma \mu_\epsilon c + \frac{\gamma^2 \sigma_\epsilon^2}{2} c^2))$ when c is a constant and $\epsilon \sim \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$.

[50] Another related reference is Che (1993), which employs a uni-dimensional bidder type, referred to as the bidders' "productive potential".

To do this, we solve the convex optimization program:

$$\max_{\mathbf{b}^i(s)} \left[1 - \exp\left(-\gamma \sum_{t=1}^{T} q_t^b (b_t^i(s) - \alpha^i c_t) - \frac{\gamma \sigma_t^2}{2} (b_t^i(s) - \alpha^i c_t)^2 \right) \right] \tag{1.5}$$

$$\text{s.t. } \sum_{t=1}^{T} b_t^i(s) q_t^e = s$$

Note that the objective function is separable in t and concave, and so this optimization problem will have a unique global maximum. Moreover, applying the monotone transformation $T(f(x)) = -log(-f(x) - 1)$, we can characterize the solution to (1.5) by the constrained quadratic program:

$$\max_{\mathbf{b}^i(s)} \left[\gamma \sum_{t=1}^{T} q_t^b (b_t^i(s) - \alpha^i c_t) - \frac{\gamma \sigma_t^2}{2} (b_t^i(s) - \alpha^i c_t)^2 \right] \tag{1.6}$$

$$\text{s.t. } \sum_{t=1}^{T} b_t^i(s) q_t^e = s.$$

The solution to this program is given by:[51]

$$b_{i,t}^*(s) = \alpha^i c_t + \frac{q_t^b}{\gamma \sigma_t^2} + \frac{q_t^e}{\sigma_t^2 \sum_{p=1}^{T} \left[\frac{(q_p^e)^2}{\sigma_p^2} \right]} \left(s - \sum_{p=1}^{T} \left[\alpha^i c_p q_p^e + \frac{q_p^b q_p^e}{\gamma \sigma_p^2} \right] \right). \tag{1.7}$$

2. Let $\mathbf{b}_i^*(s)$ be the optimal mapping from score to bid distribution for bidder i, as in equation (1.7). We find the optimal score for bidder i by maximizing her expected utility given the equilibrium distribution of opponent scores.

Let $H^j(\cdot)$ be the CDF of contractor j's score. Then by bidding a score of s, bidder i

[51] Note that this formulation of the optimal bid program does not explicitly constrain unit bids to be non-negative. This is not with loss of generality, and we apply the additional non-negativity constraint when computing counterfactual bids. However, as all observed bids are positive (meaning that the non-negativity constraint did not bind), this 'unconstrained' program serves as a very useful approximation to the solution of the fully constrained program. In particular, while the fully constrained program does not have a closed form solution and must be solved with interior point algorithms or the like, the 'unconstrained' version has a closed form solution that is linear in our parameters of interest. As we show in section 1.6, the bids predicted by our estimated model do quite well at matching the data.

obtains an expected profit of:

$$\mathbb{E}[u_i(\pi_i(s)] = \underbrace{\left(1 - \exp\left(-\gamma \sum_{t=1}^{T} q_t^b (b_{i,t}^*(s) - \alpha^i c_t) - \frac{\gamma \sigma_t^2}{2} (b_{i,t}^*(s) - \alpha^i c_t)^2\right)\right)}_{\text{Expected utility conditional on winning}} \cdot \underbrace{\left(\prod_{k \neq i} (1 - H^k(s))\right)}_{\text{Prob of win w/ } s = \mathbf{b}_i^* \cdot \mathbf{q}^e}$$

where the first phrase in parentheses is i's expected utility from the total profit that she stands to make from winning the auction, and the second phrase is the probability that s is the lowest score given the equilibrium score distributions $H^j(\cdot)$ for competing contractors $j \neq i$.

As is standard in auction theoretic analysis (see Milgrom and Segal (2002), for example), the optimal strategy is described by the first order condition:

$$\gamma \sum_{t=1}^{T} \left(q_t^b - \gamma \sigma^2 (b_{i,t}^*(s_i^*) - \alpha^i c_t)\right) \frac{\partial b_{i,t}^*(s_i^*)}{\partial s} =$$
$$\sum_{k \neq i} \frac{h^j(s^*)}{1 - H^j(s^*)} \left[\exp\left(\gamma \sum_{t=1}^{T} q_t^b (b_{i,t}^*(s_i^*) - \alpha^i c_t) - \frac{\gamma \sigma_t^2}{2} (b_{i,t}^*(s_i^*) - \alpha^i c_t)^2\right) - 1\right], \quad (1.8)$$

where $h^j(\cdot)$ is the pdf of contractor j's score distribution. Note that the exponent on the RHS is one over the certainty equivalent of the profit from winning - we will denote this as $\exp(\gamma \bar{\pi})$ as shorthand for expositional purposes. Substituting $\frac{\partial b_{i,t}^*(s_i^*)}{\partial s}$ by taking the derivative of equation (1.7) with respect to s and evaluating it at s_i^*, we obtain a global optimality first order condition:

$$\frac{\gamma^2}{\sum_{p=1}^{T} \left[\frac{(q_p^e)^2}{\sigma_p^2}\right]} \left(\sum_{p=1}^{T} \left[\alpha^i c_p q_p^e + \frac{q_p^b q_p^e}{\gamma \sigma_p^2}\right] - s_i^*\right) = \sum_{k \neq i} \frac{h^j(s_i^*)}{1 - H^j(s_i^*)} \left[\exp(\gamma \bar{\pi}) - 1\right]. \quad (1.9)$$

Note, however, that while equation 1.9 characterizes the equilibrium score s_i^* for bidder i, the equilibrium vector of bids *conditional* on s_i^* is defined entirely by the optimality of the bids with respect to bidder i's expected utility from winning the auction using s_i^*. That is, conditional on an equilibrium choice of score, the optimal bids for bidder i are given by equation 1.7, evaluated at the equilibrium score.

1.5 Econometric Model

We now present a multi-step estimation procedure to estimate the model described in the previous section. We split our parameters into two categories: (1) statistical/historical parameters, which we estimate in the first stage and (2) economic parameters, which we estimate in the second stage. The first set of parameters characterizes the bidders' beliefs over the distribution of actual quantities. The estimation procedure for this stage will use the full history of auctions in our data to build a statistical model of bidder expectations using publicly available project characteristics. However, it will not take into account bidding incentives in any particular auction. By contrast, the second stage will estimate the coefficient of risk aversion γ for each project type, and each bidder's efficiency type α in each auction that she participates in. In this stage, we take the first stage estimates as fixed and construct moments for GMM estimation using the optimality of observed bids submitted by each bidder i in auction n, given our model, as described in equations (1.7) and (1.9).

Stage 1a: Estimating the Posterior Distribution of q_t^a

In the model presented in section 1.4, we did not take a stance on what the signals in equation (1.2) are based on. The reason for this was to emphasize the flexibility of our model with respect to possible signal structures: the only required assumption is that conditional on all of the information held at the time of bidding, the posterior distribution of each q_t^a can be approximated by a normal distribution with a commonly known mean and variance. In particular, it allows for correlations between items, as well as complicated forms of correlation between the bidders' beliefs and the DOT's expectations.

For the purpose of estimation, however, we make an additional assumption. We assume that the posterior distribution of each q_t^a is given by a statistical model that conditions on q_t^e, item characteristics (e.g. the item's type classification), observable project characteristics (e.g. the project's location, project manager, designer, etc.), and the history of DOT projects. This assumption can be thought of in several ways. It can be interpreted as an additional component of the structural model: the bidders use a statistical estimation procedure to assess the likelihood of item quantities, and consequently, the value of the project, prior to bidding. The DOT quantities, item and project characteristics are indeed all publicly known at the time of bidding, as are historical records of DOT projections and ex-post quantities. Furthermore, it is likely that firms do precisely this when forming their bids. There is a competitive industry of software for procurement bid management that touts sophisticated estimation of project input quantities and costs. Alternatively, this assumption could be thought of as the econometrician's model of the signal mean q_t^b and variance σ_t^2 for each item t.

In particular, denote an auction by n and the items involved in auction n by $t \in \mathcal{T}(n)$. We model the realization of the actual quantity of item t in auction n by:

$$q_{t,n}^a = \widehat{q_{t,n}^b} + \eta_{t,n} \text{ where } \eta_{t,n} \sim \mathcal{N}(0, \hat{\sigma}_{t,n}^2) \tag{1.10}$$

such that

$$\widehat{q_{t,n}^b} = \beta_{0,q} q_{t,n}^e + \vec{\beta}_q X_{t,n} \text{ and } \hat{\sigma}_{t,n} = \exp(\beta_{0,\sigma} q_{t,n}^e + \vec{\beta}_\sigma X_{t,n}). \tag{1.11}$$

Here, $\widehat{q_{t,n}^b}$ is the posterior mean of $q_{t,n}^a$ and $\hat{\sigma}_{t,n}$ is the square root of its posterior variance—linear and log-linear functions of the DOT estimate for item t, $q_{t,n}^e$, and a matrix of item-project characteristics $X_{t,n}$. We estimate this model with Hamiltonian Monte Carlo as an efficient implementation of a likelihood method optimized for a GLM and use the posterior mode as a point estimate for the second stage of estimation.[52] We demonstrate the goodness of fit in section 1.6.

Stage 2: Estimating Cost Types and the CARA Coefficient

We now discuss our econometric model for the estimation of the CARA coefficient of risk aversion γ and bidder-auction efficiency types α_n^i. The key to our identification strategy

[52]Note that it is possible to estimate our first and second stage jointly using Hamiltonian Monte Carlo, adding further fidelity to the effect of the first stage estimates on the second stage moments along the entire posterior distribution. However, as we prefer GMM for the second stage for this version, we make do with the posterior mode. We could also simply run the second stage GMM along the posterior distribution and compute a full second stage posterior this way, but this would be very computationally burdensome, and so we do not do so at this time.

lies in the heterogeneity of unit bids that we observe in our data. Our data set contains a unit bid for every item, submitted by every participating bidder in every auction that we see. In particular, we have three main sources of heterogeneity: (1) bids submitted by different bidders in an auction with the same project characteristics, item, etc.; (2) bids submitted by the same bidders across different items and different auctions with different project characteristics, etc.; (3) bids submitted for the same items by bidders across different auctions with different project characteristics, quantity projections and participating bidders.

Denote auctions by n, the bidders participating in the auction by i and the items involved in the auction by t. The model of optimal bidding described in section 1.4 predicts that the optimal unit bid for item t for a bidder of type α_n^i in auction n is given by:

$$b^*_{t,i,n}(s^*_{i,n}) = \alpha_n^i c_{t,n} + \frac{q^b_{t,n}}{\gamma\sigma^2_{t,n}} + \frac{q^e_{t,n}}{\sigma^2_{t,n}\sum\limits_{p=1}^{T_n}\left[\frac{(q^e_{p,n})^2}{\sigma^2_{p,n}}\right]}\left(s^*_{i,n} - \sum_{p=1}^{T_n}\left[\alpha^i c_{p,n}q^e_{p,n} + \frac{q^b_{p,n}q^e_{p,n}}{\gamma\sigma^2_{p,n}}\right]\right), \quad (1.12)$$

where $s^*_{i,n}$ is the optimal score for this bidder, such that $s^*_{i,n} = \sum\limits_{t}^{T_n} q^e_{t,n} b^*_{t,i,n}(s^*_{i,n})$, and

$$\frac{\gamma^2}{\sum\limits_{p=1}^{T_n}\left[\frac{(q^e_{p,n})^2}{\sigma^2_{p,n}}\right]}\left(\sum_{p=1}^{T_n}\left[\alpha_n^i c_{p,n}q^e_{p,n} + \frac{q^b_{p,n}q^e_{p,n}}{\gamma\sigma^2_{p,n}}\right] - s^*_{i,n}\right) = \sum_{k\neq i}\frac{h_n(s^*_{i,n})}{1-H_n(s^*_{i,n})}\left[\exp(\gamma\bar{\pi})-1\right], \quad (1.13)$$

where $\bar{\pi} = \sum_{t=1}^{T_n} q^b_{t,n}(b^*_{t,i,n}(s^*_{i,n}) - \alpha_n^i c_{t,n}) - \frac{\gamma\sigma^2_{t,n}}{2}(b^*_{t,i,n}(s^*_{i,n}) - \alpha^i c_{t,n})^2$.

As discussed above, we identify $q^b_{t,n}$ and $\sigma^2_{t,n}$ with a statistical model of ex-post quantities conditional on item-project characteristics using the full history of auctions in our data. To reduce the dimensionality of our parameter space, we model the bidder-auction efficiency type α_n^i onto a bidder-specific fixed effect and a regression model of bidder-auction characteristics:

$$\alpha_n^i = \alpha^i + \beta_\alpha X_{i,n}.$$

Finally, We make the following assumption to connect our first stage estimates to our bid data and close our model:

Assumption 1. *Let $b^d_{t,i,n}$ denote the unit bid for item t submitted by bidder i in auction n, as* *observed in our data. Each observed unit bid is equal to the optimal bid $b^*_{t,i,n}$, subject to an IID, mean-zero measurement error $\nu_{t,i,n}$:*

$$b^d_{t,i,n} = b^*_{t,i,n} + \nu_{t,i,n}$$

where

$$\mathbb{E}[\nu_{t,i,n}] = 0 \text{ and } \nu_{t,i,n} \perp X_{t,n}, X_{i,n}$$

Assumption 1 states that each unit bid observed in our data is given by the optimal bid implied by our model – at the true underlying parameters – subject to an idiosyncratic error that is independent across draws, and orthogonal to auction-item and auction-bidder characteristics. Such an error might come about because of rounding/smudging in the translation between the bidder's optimal bidding choice and the record that appears to the DOT (and consequently, to the econometrician). One might alternatively frame this error as an optimization error: the optimal choice of bids is a numerical solution to a constrained quadratic program that may not produce numbers that are convenient to report in currency. To see the need for Assumption 1, note that an auction with T items and I bidders has $T \times I$ unit bids, our model allows for only T quantity predictions, T item variance terms, I bidder efficiency types, and 1 coefficient of risk aversion as free parameters to explain these bids. Absent an additional assumption, a model in which all $T \times I$ bids must match the bids in our data would be rejected in most cases. It is not, however, strictly necessary for our model to assert independence in error within bidder or project. We will therefore examine relaxations of the independence assumption in an upcoming revision.

Note that Assumption 1 implies that the optimal score $s^*_{i,n}$ is also observed with error:

$$s^*_{i,n} = \sum_{t}^{T_n} b^d_{i,t,n} q^e_{t,n} + \bar{\nu}_{i,n} = s^d_{i,n} + \bar{\nu}_{i,n},$$

where $\bar{\nu}_{i,n} = -\sum_{t=1}^{T_n} \nu_{t,i,n} q^e_{t,n}$ is also mean-zero, conditional on the project characteristics of auction n. Write $\theta_2 = (\gamma, \{\alpha^i\}, \vec{\beta}_\alpha | \beta_{0,q}, \vec{\beta}_q, \beta_{0,\sigma}, \vec{\beta}_\sigma, \vec{\beta}_s, \vec{\sigma}_s)$. By definition, the bidder-item-auction level error on each unit bid is given by:

$$\nu_{t,i,n} = b^d_{t,i,n} - \boldsymbol{\alpha^i_n}\left(c_{t,n} - \frac{q^e_{t,n}}{\sigma^2_{t,n}\sum_{p\in\mathcal{T}(n)}\left[\frac{(q^e_{p,n})^2}{\sigma^2_{p,n}}\right]}\left[\sum_{p\in\mathcal{T}(n)} c_{p,n}q^e_{p,n}\right]\right) - \frac{1}{\boldsymbol{\gamma}}\left(\frac{q^b_{t,n}}{\sigma^2_{t,n}} - \frac{q^e_{t,n}}{\sigma^2_{t,n}\sum_{p\in\mathcal{T}(n)}\left[\frac{(q^e_{p,n})^2}{\sigma^2_{p,n}}\right]}\left[\sum_{p\in\mathcal{T}(n)}\frac{q^b_{p,n}}{\sigma^2_{p,n}}q^e_{p,n}\right]\right) - \frac{q^e_{t,n}}{\sigma^2_{t,n}\sum_{p\in\mathcal{T}(n)}\left[\frac{(q^e_{p,n})^2}{\sigma^2_{p,n}}\right]}\left[s^d_{i,n} + \boldsymbol{\bar{\nu}_{i,n}}\right]. \tag{1.14}$$

where

$$\alpha^i_n = \alpha^i + \beta_\alpha X_{i,n}. \tag{1.15}$$

Note that $\nu_{i,t,n}$ is linear in α^i_n and $\frac{1}{\gamma}$, as well as in $\bar{\nu}_{i,n}$. Furthermore, under Assumption 1, since $\mathbb{E}[\nu_{t,i,n}] = 0$ and is orthogonal to the matrix of item-auction features $X_{t,n}$ and bidder-auction features $X_{i,n}$, we have that $\mathbb{E}[\bar{\nu}_{i,n}] = 0$ as well.

We therefore define a demeaned bid error

$$\tilde{\nu}_{t,i,n} = \nu_{i,t,n} - \frac{q^e_{t,n}}{\sigma^2_{t,n}\sum_{p\in\mathcal{T}(n)}\left[\frac{(q^e_{p,n})^2}{\sigma^2_{p,n}}\right]}\bar{\nu}_{i,n}, \tag{1.16}$$

and form the following moment conditions, under Assumption 1:

$$\mathbb{E}\left[\tilde{\nu}_{t,i,n}\cdot Z_{t,i,n}|X_{t,n}, X_{i,n}\right] = 0,$$

where Z is each of the following instruments:

- Indicator for unique firm IDs[53]
- Indicator for being a "top skewed item"
- The bidder-auction feature vectors that comprise $X_{i,n}$.

[53] We include a unique ID for for all firms involved in at least 10 auctions, and a grouped ID for all firms involved in 9 or less auctions. These correspond to unique α^i parameters.

Identification

The three types of instruments above correspond to three types of moments.

The first type of moment, constructed by interactions with firm ID dummies, can be interpreted as follows: the average bid error that a bidder with unique firm ID i submitted, across all auctions that i participated in, is asymptotically zero. There are 25 such moments, one for each unique bidder id i. These moments inform the fixed effects α^i, correspondingly.

The second moment focuses on items that were deemed as "top skew items" according to the DOT Engineering Office. These items are flagged as frequently being given noticeably high or low bids. According to our model, the variation in these bids is reflective of level of bidders' responses to the uncertainty regarding the quantities of these items (in absolute terms and relative to the remainder of the project). As such, we focus on this set of items to identify the coefficient of risk aversion, γ. The moment can be interpreted as follows: the average bid errors submitted on "top skew items" is asymptotically zero in the number of auctions in which these items are involved.

The third type of moment, which interacts bid errors with bidder-auction characteristics, can be interpreted as follows: the average bid error submitted in an auction n is orthogonal to each of the 14 bidder-auction features $X^j_{i,n}$, and asymptotically zero in the number of auctions. There are 14 such moments, one for each column of the feature matrix $X_{i,n}$. Each of these moments can be thought of as informing the identification of the coefficient β^j_α.

For complete details on the moment construction, see section **??**. Note that our moment conditions use only the optimality of bidders' unit price bids (given the scores that are observed in equilibrium).[54]

[54] This is conceptually similar to the classical GPV use of the empirical bid distribution to model bidders' beliefs over the distribution of opponent bids that they face.

1.5.1 Bayesian Sampling with Hamiltonian Monte Carlo

In addition to our main GMM approach, we estimate a (fully) parametric version of our structural model using Hamiltonian Monte Carlo.[55] Bayesian analysis facilitates the modeling of hierarchical relationships in bidders' efficiency types – across auctions for the same bidders, and across bidders in similar auctions. In our GMM approach, we account for these relationships in the form of bidder fixed effects, and a regression function of auction-bidder characteristics. However, a more sophisticated GMM treatment would be difficult, given the high dimensionality of the parameter space and the amount of data available. As such, we consider both approaches in our paper. We present the details of a preliminary Bayesian specification for the second stage of our structural estimation along with results from an HMC fit of the model in section A.5 of the appendix.

1.6 Estimation Results

Our structural estimation procedure consists of two parts. In the first stage, we estimate the distribution of the ex-post quantity of each item conditional on its item-auction characteristics using Hamiltonian Monte Carlo. We present parameter estimates for the regression coefficients on the predicted quantity term $\hat{q}^b_{t,n}$ as well as the variance term $\hat{\sigma}^2_{t,n}$ in Table A.2 in the appendix. A histogram of the resulting variance terms themselves are plotted in Figure 1.11, below. Prior to estimation, all item quantities were scaled so as to be of comparable value between 0 and 10. As demonstrated in the histogram, the majority of variance terms are between 0 and 3, with a trailing number of higher values.[56] In addition, we demonstrate the model fit of our first stage in Figure A.1 and Table A.1 in the appendix.

In the second stage, we estimate a common CARA coefficient γ, as well as a bidder-auction specific efficiency type $\alpha^i_n = \alpha^i + \beta_\alpha X_{i,n}$ for every bidder-auction pair in our data

[55]Hamiltonian Monte Carlo is an efficient algorithm for sampling the posterior distribution of a statistical model. See Betancourt (2017) for an accessible complete explanation.

[56]Although we do not plot it here, in general, higher variances correspond to higher quantity predictions as well.

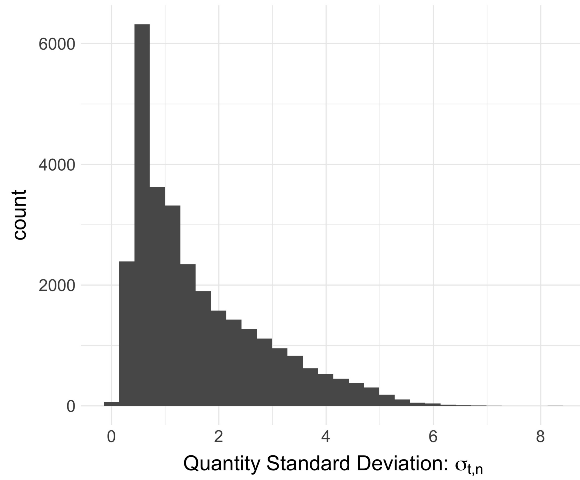


Figure 1.11: *Histogram of standard deviation estimates for each item t in each project n*

using the GMM estimator presented in section 1.5. We summarize the results in tables **??**, 1.7 and 1.8. The full parameter estimates are presented in Table A.3 in the appendix. The coefficient of risk aversion γ in our data is estimated to be about 0.046. An individual with this level of risk aversion would require a certain payment of $23 to accept a 50-50 lottery to either win or lose $1,000 with indifference, and $2,223 to accept a 50-50 lottery to win or lose $10,000.[57] As we report in Table **??**, the 95% confidence interval around our estimate is $(0.032, 0.264)$. This interval is generated by a bootstrap, in which the data set of auctions is sampled (at the auction level) with replacement in each iteration.[58]

	Parameter Estimate	95Pct CI
$\widehat{\gamma}$	0.046	(0.032,0.264)

Table 1.6: *Estimate of the coefficient of risk aversion, γ*

In Table 1.7, we present summary statistics of our estimates of bidder-auction efficiency

[57]Note that the CARA coefficient we estimate here is only identified up to a dollar scaling. For numerical efficiency, we scaled all dollar values by $1,000 in estimation and counterfactual simulation. Our results do not depend on the scaling, however. As we have verified, if we scale by an order of magnitude more (or less), the estimated CARA coefficient scales down (or up) by an order of magnitude correspondingly.

[58]At the moment, the bootstrap is only over the second stage, holding the first stage estimates fixed. A full two-stage bootstrap, which requires substantially more computation time, will be presented in a future draft.

Project Type	Mean	$\widehat{\alpha}_n^i$ St Dev	25%	Median	75%
All	0.975	0.261	0.822	0.949	1.139
Bridge Reconstruction/Rehab	1.019	0.25	0.85	1.005	1.225
Bridge Replacement	0.996	0.219	0.855	1.009	1.159
Structures Maintenance	0.919	0.312	0.782	0.873	0.978

Table 1.7: *Summary statistics of α_n^i estimates by project type*

types.[59] We break down the results by project type to highlight the differences between different types of construction. An efficiency of 1 would suggest that the bidder faces costs exactly at the rates represented by MassDOT's blue book. Our results show that the median bidder overall has an efficiency type of 0.949, consistent with estimates of bidder costs by previous papers.[60] There is heterogeneity across project types, however. We estimate that the median bidder in a bridge rehabilitation project has an efficiency type of about 1.005, suggesting that she is about 0.5% less efficient than the DOT estimates. The median bidder in structures maintenance projects, however, has an efficiency type of about 0.873, suggesting that she is about 12.7% more efficient than the DOT estimates.

In Table 1.8, we present the ex-post markups for each winning bidder given their efficiency type:

$$\text{Markup} = \frac{\sum_t q^a_{t,n} \cdot (b_{t,i,n} - \alpha^i_n c_{t,n})}{\sum_t q^a_{t,n} \cdot (\alpha^i_n c_{t,n})}.$$

This is the bidder's total ex-post profit from the project, normalized by her total cost. The numerator is given by the sum of the quantity of each item that was ultimately used $q^a_{t,n}$, multiplied by the bidder's profit from that item – her unit bid $b_{t,i,n}$ minus her private cost for that item, given by her efficiency type α^i_n multiplied by the blue book market rate estimate

[59]There are a few of decisions made by the econometrician in estimation. We considered different thresholds on the number of auctions in which a firm must have participated in order to have a separate firm fixed effect. We also identified several outlying items: items that constituted large fractions of the project cost and were always estimated and used in unit quantities. These items might better be represented as lump sum items, over which uncertainty is poorly captured in our quantity model. The substance of our results is robust to these considerations, however. We will present the results under different thresholds and when large lump items are excluded in the appendix as a robustness check.

[60]See Bajari, Houghton and Tadelis (2014) and Bhattacharya, Roberts and Sweeting (2014), for example.

$c_{t,n}$. The denominator is calculated similarly, summing over the bidder's private costs only.

Project Type	Mean	Bidder Markups St Dev	25%	Median	75%
All	17.03%	60.88%	-12.84%	5.74%	27.53%
Bridge Reconstruction/Rehab	11.39%	35.88%	-15.61%	7.34%	23.07%
Bridge Replacement	12.8%	67.43%	-12.34%	1.43%	23.67%
Structures Maintenance	23.9%	62.12%	-9.66%	10.56%	39.13%

Table 1.8: *Summary statistics of estimated winning bidders' markups given* $\hat{\alpha}_n^i$

The median markup for a winning bidder in our data set, overall, is about 5.74%. There is heterogeneity across project types: the median within bridge replacement projects is 1.43%, for instance, while it is 10.56% for structures maintenance projects. Moreover, there is substantial variation within project types as well. The mean winner markups for bridge replacement and structures maintenance projects are 12.8% and 23.9% respectively. This may be due to the heterogeneity in projects as well as the ex-post accuracy of bidders' quantity predictions. Furthermore, the 25th percentile of markups is negative for each of the projects as well. This may be due, in part, to inaccurate prediction of the ex-post quantities. However, note that the ex-post markup calculation does not take into account extra work orders. While we do not estimate profits on the extra work orders in our paper, and so cannot evaluate exactly how extra work orders would affect ex-post profits, this is a key component of BHT's estimation and likely make up the difference in mark-ups.

Finally, we demonstrate the fit of our structural model in figures A.3 and A.4, and Table A.4 in the appendix. Figure A.3 plots the unit bids predicted by our model on the x-axis, and the unit bids observed in our data on the y-axis. Figure A.4 plots a quantile-quantile plot of our model predicted bids against the data bids. While bid predictions are not perfect, the correspondence between predictions and data is quite good. Table A.4 presents a regression analysis of the predictiveness of our model fit on the observed data. Our model fit predicts data bids with an R-squared of 0.879.

1.7 Counterfactual

1.7.1 Perfectly Predicted DOT Quantities

In order to draw conclusions from our results, we return to the discussion in section 1.2. How much money *would* the DOT save if it were able to perfectly predict the actual quantities that will be required for each project?

To answer this question, we solve for the equilibrium in each of the auctions in our bridge projects dataset, under the counterfactual setting in which the DOT perfectly predicts the actual quantities. We assume that the DOT's accuracy is common knowledge and so the bidders believe that the actual quantities will be equal to the DOT's projections with variance approaching zero when making their bidding decisions.[61]

Note that it is not sufficient to simply invert the econometric model of bidding described in section 1.5 using our parameter estimates and the counterfactual conditions. The reason for this is that the distribution of competitors' scores is defined in equilibrium. As we demonstrated in section 1.2, the score that a bidder with efficiency type α will submit in equilibrium depends on the DOT quantity estimates (as well as the bidders' beliefs and all other auction characteristics). It follows that the equilibrium score distribution itself depends on the DOT quantities, and so we need to solve for the equilibrium from auction primitives afresh in each setting.

An equilibrium of an auction in our setting is determined by the following primitives: the vector of DOT quantity estimates $\boldsymbol{q}^e$, the vector of bidder quantity model predictions, $\boldsymbol{q}^b$, the vector of bidder model variances, $\boldsymbol{\sigma}^2$, the vector of DOT cost estimates $\boldsymbol{c}$, the coefficient of risk aversion γ, and the distribution of the efficiency types of bidders participating in the auction. To evaluate our counterfactuals, we compute the equilibrium bids twice: first in the baseline setting and second in the counterfactual setting. For the baseline setting, we use the DOT estimates $\boldsymbol{q}^e$ and $\boldsymbol{c}$ from the data, and the bidder quantity model parameters $\widehat{\boldsymbol{q}}^b$

[61] In particular, we exclude considerations of short term gains that the DOT might make by accurately predicting actual quantities while the bidders use noisy signals. As we assume that the bidders form their beliefs over actual quantities using statistics over historical data, any such gains would be short lived as the bidders would eventually realize that the DOT's quantities are accurate.

and $\widehat{\boldsymbol{\sigma}}^2$ from the first stage of our estimation. For the coefficient of risk aversion, we use the estimate $\hat{\gamma} = 0.046$ from the second stage of our estimation. For the distribution of bidder efficiency types, we use a parametric projection of the empirical distribution of the efficiency type estimates $\widehat{\alpha}_n^i$ from our second stage onto auction characteristics.[62] The details of the equilibrium construction are presented in section A.1 of the appendix.

In Figure 1.12, we plot a histogram of the (a) percentage and (b) dollar savings to the DOT from the perfect quantity prediction counterfactual. To calculate these savings, we compute the equilibrium bids for every efficiency type α twice: first under the baseline setting, and second under the counterfactual setting in which the DOT and bidder quantity estimates are equal to the true ex-post quantities, $\boldsymbol{q}^e = \boldsymbol{q}^b = \boldsymbol{q}^a$, and the bidders face no uncertainty, $\boldsymbol{\sigma}^2 \approx 0$.[63] In each case, we calculate the expected total amount that the DOT would pay the winning bidder in equilibrium: the expected value of the sum of the lowest efficiency type's unit bids multiplied by the ex-post item quantities $\boldsymbol{q}^a$.[64] The dollar gains in Figure 1.12b are computed by taking the difference between the expected DOT cost under the baseline setting, and under the counterfactual setting for each auction. The percent gains in Figure 1.12a are given by dividing the dollar saving amount in each auction by the expected DOT cost under the baseline. Finally, we present the bidder utility gains from the counterfactual setting in Figure 1.12c. We calculate bidder utility gains by taking the difference between the (ex-ante) certainty equivalent of a bidder participating in each auction under the baseline and the analogous certainty equivalent under the counterfactual setting.[65] We present summary statistics for all three metrics in Table 1.9.

[62]Our model assumes that bidders in a given auction are ex-ante IID, and so the distribution of bidder types must be auction, rather than bidder-auction, specific.

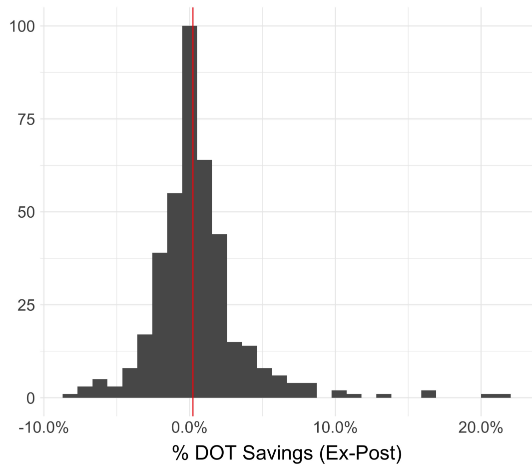
[63]We use $\boldsymbol{\sigma}^2 \approx 0$ rather than $\boldsymbol{\sigma}^2 = 0$ in order to avoid numerical overflow issues.

[64]More concretely, let $g(\alpha)$ and $G(\alpha)$by the density and cumulative probability functions of bidders' efficiency types in a given auction. Let $g^1(\alpha) = Ng(\alpha)(1 - G(\alpha))^{N-1}$ be the density of the first order statistic of g—the density of the lowest type bidder, when there are N bidders in the auction. Denote $b_t^*(\alpha)$ as the equilibrium bid for item t for a bidder with efficiency type α in that auction. The expected DOT cost is given by $\int_{\underline{\alpha}}^{\bar{\alpha}} g^1(\tilde{\alpha}) \sum_t q_t^a b_t^*(\tilde{\alpha}) d\tilde{\alpha}$.

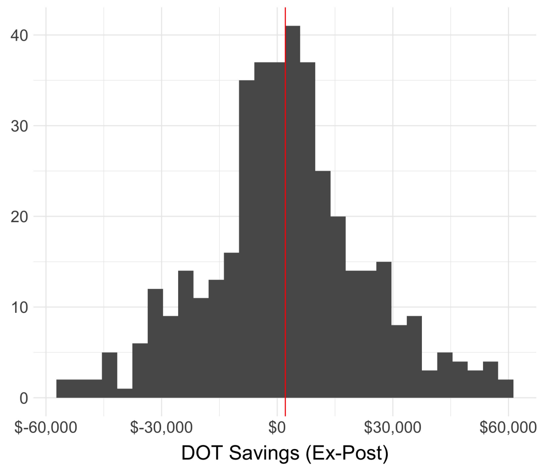
[65]The certainty equivalent is defined as the amount of money that would make a bidder indifferent between participating in the auction or forgoing the auction to accept that amount with no uncertainty.

We predict that the median expected saving to the DOT from eliminating uncertainty about ex-post quantities is about $2,203 or 0.23% of the baseline expected project cost. However, the standard deviation of savings is about $24,704 (4.25%) and the 25th and 75th percentiles are -$9,355 (-1.02%) and $13,987 (1.60%) respectively. This is reflective of the two opposing forces in effect when the DOT eliminates uncertainty. On the one hand, eliminating uncertainty drives bidder risk down, thereby increasing the value of the project to all of the bidders and causing them to bid more aggressively. On the other hand, the counterfactual allows bidders to optimize their bid choices with regard to the true quantities $\boldsymbol{q}^a$ that will be used in the project, whereas in the baseline, bidders optimize on the basis of quantity projections $\boldsymbol{q}^b$, which often differ from the true quantities. That is, whereas in the baseline, bidders optimize unit bids with regards to quantity predictions that may be inaccurate (and so, the bids may not be optimal with respect to the realized quantities, which the winner is ultimately paid for), in the counterfactual with no uncertainty, the bidders always optimize unit bids with respect to the actual quantities that will be used. As a result, in the auctions where bidders "mis-optimized" under the baseline, the DOT bears a higher cost under the counterfactual. Notably, the ex-ante value of the auction to bidders does not change very much between the baseline and the counterfactual. The median increase in bidders' ex-ante certainty equivalents under the counterfactual is a mere $17.61, and the 25th and 75th percentiles are $3.76 and $43.35, respectively. This reflects the degree to which optimal bid selection in equilibrium allows bidders to insure themselves against risk. The value of the project rises in equilibrium, adding to the certainty equivalent, but this is offset by competition and an inability to profitably skew. Consequently, the certainty equivalent rises for some auctions, falls for others, but all in all stays much the same.

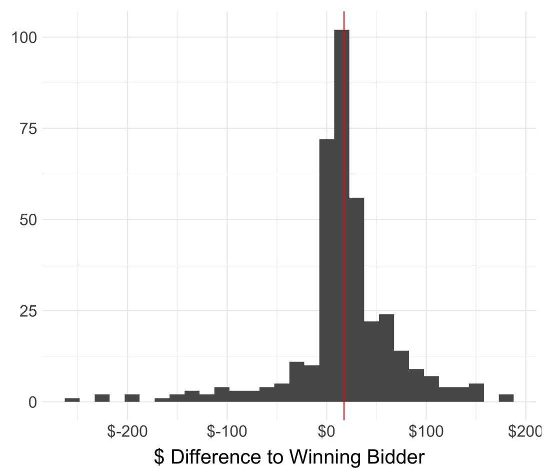
The projected expected DOT savings from eliminating risks detailed in Table 1.9 and Figure 1.12 reflect the two channels by which eliminating uncertainty changes the bidders' problem: (1) it eliminates risk, raising the value of the project and encouraging more aggressive bids; (2) it gives bidders access to the accurate ex-post quantities, allowing bidders to perfectly optimize their unit bids with respect to ex-post profits. In order to



(a) *DOT % Savings*



(b) *DOT Dollar Savings*



(c) *Winning Bidder Utility Gains*

Figure 1.12: *Percent and dollar expected DOT ex-post Savings, and bidder utility gains from a counterfactual in which risk is eliminated*

Note: The median is highlighted in red in each case.

Table 1.9: *Summary of expected DOT percent and dollar savings and bidder utility gains (in dollars) from the counterfactual setting in which the DOT reports perfectly accurate actual quantity estimates*

Statistic	Mean	St. Dev.	25%	Median	75%
Net DOT Savings	\$2,145.37	\$24,704.09	− \$9,354.61	\$2,203.49	\$13,987.89
% DOT Savings	0.70%	4.25%	−1.02%	0.23%	1.60%
Bidder Gains	\$6.64	\$145.87	\$3.76	\$17.61	\$43.35

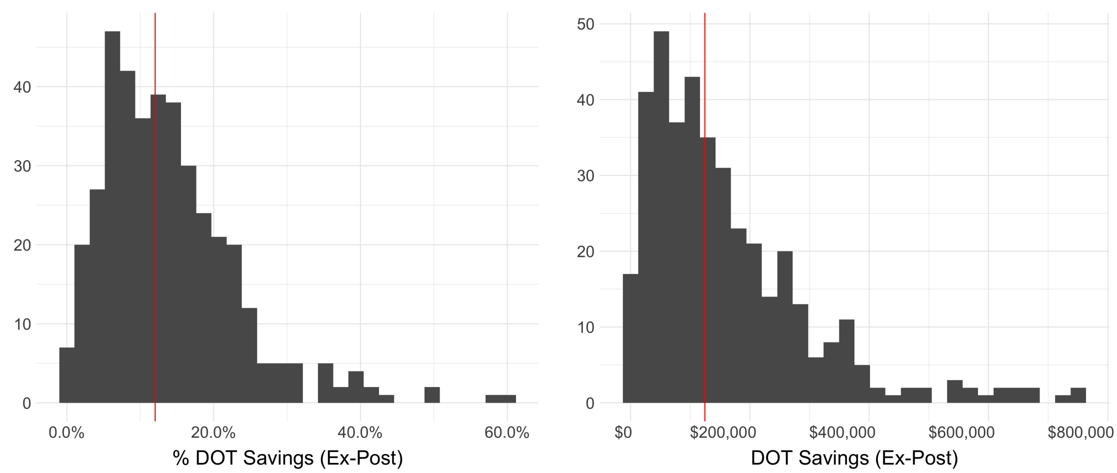
Note: Results are truncated at the top and bottom 1% to exclude extreme outliers from the mean/SD calculations.

disentangle these two effects, we repeat the counterfactual exercise under the assumption that in the baseline, bidders' quantity projections $\boldsymbol{q}^b$ are equal to the ex-post quantities $\boldsymbol{q}^a$ (but that bidders still perceive the projections to be noisy with variance $\hat{\boldsymbol{\sigma}}^2$). In this case, bidders always optimize correctly with respect to ex-post quantities, and so the second channel, by which eliminating risk can hurt DOT savings, is shut down. The resulting expected DOT savings and bidder utility gains are reported in Table 1.10 and Figure 1.13. Absent bidder mis-optimization due to inaccuracies in their quantity projections, the median expected saving to the DOT is \$125,187 or 11.98% of the (adjusted) baseline expected cost. This can be thought of as an aggressive estimate of the potential savings from eliminating risk, whereas the previous estimate is a conservative estimate. Notably, the bidder ex-ante utility gains remain modest with a median certainty equivalent gain of \$4.81 from the counterfactual. This is because ex-ante utility is evaluated with respect to bidder beliefs – according to which equilibrium bids are optimized – rather than ex-post quantities. As such, the difference in baseline quantity predictions has little effect on the ex-ante total certainty equivalent of each auction (although it does change the particular choices of optimal bids across items).

Table 1.10: *Summary of expected DOT percent and dollar savings and bidder utility gains (in dollars) from the counterfactual setting in which the DOT reports perfectly accurate actual quantity estimates, relative to a baseline in which bidders accurately predict ex-post quantities, but believe their predictions to be noisy*

Statistic	Mean	St. Dev.	25%	Median	75%
Net DOT Savings	$172,513.80	$165,129.50	$61,569.34	$125,187.10	$226,318.90
% DOT Savings	13.74%	9.05%	7.18%	11.98%	18.25%
Bidder Gains	$19.16	$124.55	−$8.48	$4.81	$37.64

Note: Results are truncated at the top and bottom 1% to exclude extreme outliers from the mean/SD calculations.



(a) *DOT % Savings* **(b)** *DOT Dollar Savings*

Figure 1.13: *Percent and dollar expected DOT ex-post Savings, and bidder utility gains from a counterfactual in which risk is eliminated, relative to a baseline in which bidders accurately predict ex-post quantities, but believe their predictions to be noisy*

Note: The median is highlighted in red in each case.

1.7.2 Alternative Risk Sharing Mechanisms: Lump Sum and μ-sharing Auctions

While highway and bridge procurement around the United States is predominately done through scaling auctions, public procurement in other departments of American DOTs, as well as in DOTs around the world, often employs auction mechanisms that place significantly more risk on contractors. The simplest example of this is a *lump sum* auction in which contractors submit a single total bid for completing the project. Subsequently, the winning contractor is responsible for all project costs incurred, independently of whether or not they exceed initial projections. Lump sum auctions have several properties that make them attractive to DOT officials. First, they require less detailed specifications from DOT engineers as bidding does not require a comprehensive itemized list of tasks and materials.[66] Second, they incentivize the winning bidder to minimize costs (as all costs are privately incurred and not directly compensated), thereby reducing the scope for moral hazard. However, lump sum auctions have worrisome incentive properties as well. First, because compensation is fixed at the time of bidding, projects that greatly exceed their scope are more likely to suffer from hold-up problems in which the winning contractor insists on negotiating additional payments before completing the project. Moreover, as we note in section 1.2, lump sum auctions greatly increase contractors' exposure to risk. The increased risk exposure reduces the value of winning the auction, and causes risk averse bidders to bid less aggressively, resulting in substantially higher costs to the DOT.

In this section, we evaluate the extent to which shifting risk exposure onto contractors, as in a lump sum auction, may be costly to the DOT. To hone in on the effect of risk exposure in particular, we maintain the main assumptions of our baseline model. Bidders are identical apart from a private, independently drawn, efficiency type α. The DOT advertises each

[66] Around 2007, the MBTA – the segment of MassDOT responsible for construction and maintenance of the public transportation system in Massachusetts – switched from scaling auctions to lump sum auctions for the majority of its procurement. We spoke to the chief engineer about the decision for this transition in 2017. Chief among his reasons was the assertion that the scope of MBTA projects is much more difficult to define (and therefore spec out ex-ante) than of highway and bridge projects. We interpreted this to mean that the difficulty/costs of producing a comprehensive list of items for MBTA projects was high. We sought data to compare costs after the switch, but were unable to obtain bidding or quantity records from before the switch.

project with a comprehensive list of items and (often inaccurate) quantity estimates $\boldsymbol{q}^e$. Bidders receive a common signal of what the ex-post quantities will be, which provides them with a vector of quantity projections $\boldsymbol{q}^b$ and a vector of variances of the projection noise $\boldsymbol{\sigma}^2$.

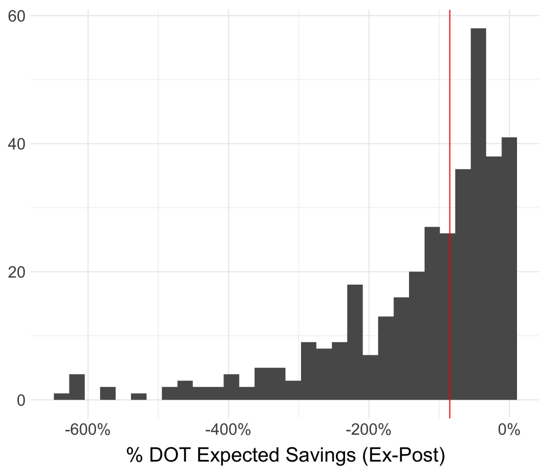
We define a μ-sharing auction for $\mu \in [0,1]$, as a scaling auction in which the winning bidder is paid

$$\sum_t (\mu q_t^a + (1-\mu) q_t^e) \cdot b_t,$$

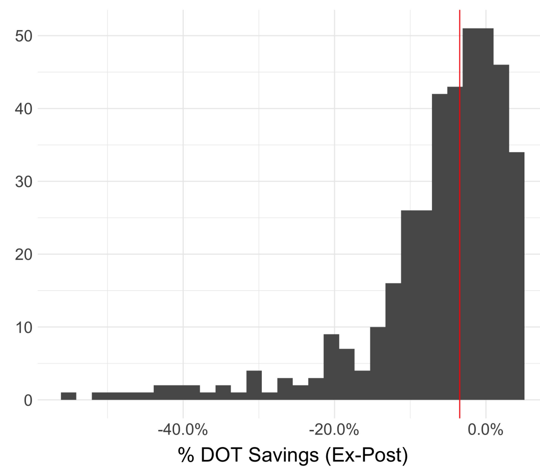
upon completion of the project. That is, for every item t involved in the project, the winning bidder is paid her bid b_t multiplied by μ times the actual quantity of t used, plus $(1-\mu)$ times the ex-ante DOT estimate for the quantity of t. When $\mu = 0$, this is equivalent to a lump-sum auction, as the bidder is paid entirely based on her score, $\boldsymbol{b} \cdot \boldsymbol{q}^e$. When $\mu = 1$, this is a standard scaling auction as in the baseline model. In general, the equilibrium bids for a bidder i with efficiency type α^i is characterized as in section 1.4.2 with the following adjustment. The certainty equivalent in the constrained quadratic program to determine the optimal distribution of bids, conditional on a candidate score (as in in equation 1.6) is replaced by its μ-sharing analog:

$$\gamma \sum_t \underbrace{(1-\mu) b_t q_t^e + (\mu b_t - \alpha c_t) q_t^b}_{\text{Expected Profits}} - \underbrace{\frac{\gamma \sigma_t^2}{2} (\mu b_t - \alpha c_t)^2}_{\text{Risk Term}}.$$

We defer a detailed derivation of the equilibrium to the appendix. As in the previous section, we calculate the change in expected DOT costs between a baseline auction in which bidders are paid according to the ex-post quantities $\boldsymbol{q}^a$ alone (e.g. $\mu = 1$) and a μ-sharing auction for $\mu \in (0,1]$. In each case, we use the DOT estimates $\boldsymbol{q}^e$, ex-post quantities $\boldsymbol{q}^a$ and blue book costs $\boldsymbol{c}$ from the data – as before – as well as our structural estimates for the CARA coefficient $\hat{\gamma}$ and the distribution of efficiency types conditional on auction characteristics. To focus in on the effect of the risk shifting alone, we shut down the bidder mis-optimization channel and assume that bidders' quantity projections $\boldsymbol{q}^b$ are equal to the actual ex-post quantities $\boldsymbol{q}^a$, but that bidders still perceive the projections to be noisy with variance $\widehat{\boldsymbol{\sigma}}^2$, from our first stage estimation. We present the percent change in expected DOT costs under a lump sum auction in Figure 1.14a, and under a $\frac{1}{2}$-sharing auction in Figure 1.14b. The median expected loss from moving to a lump sum auction is 84.84%, while the median expected loss from a $\frac{1}{2}$-sharing auction is 3.47% – both with fat tails. Summary statistics for each case are presented in Table 1.11.[67]



(a) *DOT % Cost Change*
Lump Sum ($\mu = 0$)



(b) *DOT % Cost Change*
$\mu = 1/2$

Figure 1.14: *Histogram of expected DOT percent cost change from switching to a μ-sharing auction with $\mu = 0$ (lump sum) and $\mu = 1/2$.*

Note: The median is highlighted in red in each case.

1.8 Entry

It is well known that an increase in competition benefits an auctioneer. In this section we evaluate the entry of an additional contractor to each auction in our data. First, we estimate the expected amount that the DOT would save if an additional contractor were to enter. We do this by computing the equilibrium bid function in each auction under the baseline (as in

[67]Note that small increases in risk may in fact reduce the DOT spending ex-post, as they may cause bidders to place larger bids on items with lower expected overruns (and lower risk) at a competitive score (even if the score itself rises).

Table 1.11: *Summary of expected DOT percent cost change from switching to a μ-sharing auction with $\mu = 0$ (lump sum) and $\mu = 1/2$*

DOT % Cost Change	Mean	St. Dev.	25%	Median	75%
Lump Sum	−127.93%	129.70%	−175.11%	−84.84%	−38.08%
$\mu = 1/2$	−6.84%	12.00%	−9.49%	−3.47%	0.39%

Note: Results are truncated at the top and bottom 1% to exclude extreme outliers from the mean/SD calculations.

the counterfactuals described in Figure 1.12), and then under an extension of the baseline in which the number of bidders is increased by one. We calculate the expected cost savings in each auction by taking the difference between the expected amount paid by the DOT to the winning bidder in the baseline, and in the counterfactual with an additional bidder participating. Next, we estimate bounds on the cost of entry for a prospective bidder in a procedure akin to Pakes, Porter, Ho and Ishii (2015), using the assumption that bidders enter if they anticipate to profit more than the cost of entry and total entry is set in equilibrium.

1.8.1 An Equilibrium Model of Entry

Each auction is advertised to a set of prospective (pre-approved) contractors. Upon receiving an advertisement, each prospective bidder observes the common auction characteristics: the location of the project, identity of involved DOT employees, the vector of DOT quantity estimates $\boldsymbol{q}^e$ and the blue book cost estimates, $\boldsymbol{c}$, as well as the refined quantity signals components $\boldsymbol{q}^b$ and $\boldsymbol{\sigma}^2$. Given this information, each bidder is also able to infer the distribution of efficiency types of the prospective contractors.[68] However, in order to discover her own (private) efficiency type, each bidder must invest a fixed amount K. For simplicity, we assume that K is common across bidders. The timeline of each prospective bidder's interaction with the auction is therefore as follows:

[68] As before, we assume that this distribution is the same for all prospective bidders conditional on auction characteristics.

1. Bidder observes project characteristics and the entry cost
2. Bidder calculates the expected utility of entering and determines whether or not to participate
3. If she participates:
   - Bidder observes her private efficiency type α
   - Bidder chooses optimal unit bids given α, according to the equilibrium strategy

The expected utility of entry is as follows:

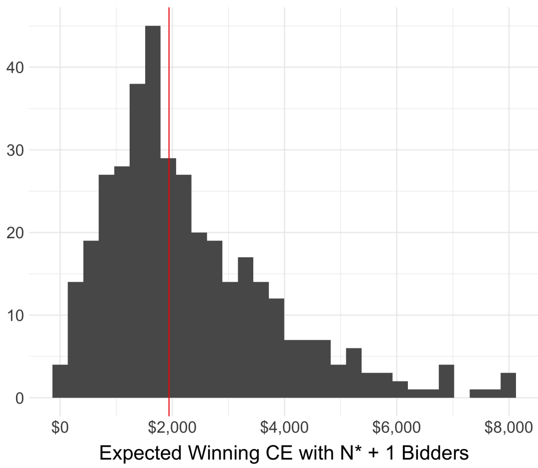
$$E[u(\pi)|N^*] = \int_{\underline{\alpha}}^{\overline{\alpha}} [E[u(\pi(s(\tilde{\alpha}), \tilde{\alpha})|N^*)] \cdot f(\tilde{\alpha})] \, d\tilde{\alpha}$$

where N^* is the equilibrium number of bidders participating in the auction, and $E[u(\pi(s(\tilde{\alpha}), \tilde{\alpha})|N^*)]$ is the expected utility from participating in the auction (and paying K) given efficiency type $\tilde{\alpha}$. In order for N^* to be the equilibrium number of bidders, it must be that the N^*th bidder found it profitable to enter, whereas the $N^* + 1$st bidder did not. That is:

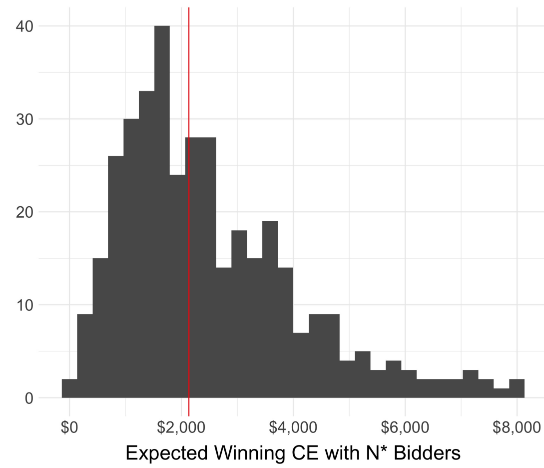
$$E[u(\pi)|N^*] \geq 0 \geq E[u(\pi)|N^* + 1].$$

As such, the certainty equivalent of $E[u(\pi)|N^* + 1]$ (absent an entry cost) provides a lower bound on K, and the certainty equivalent of $E[u(\pi)|N^*]$ provides an upper bound on K.[69] We plot the distribution of upper and lower bounds on the cost of entry K in each auction in figures 1.15a and 1.15b, respectively. In Figure 1.15c, we plot the expected savings to the DOT from the entry of an additional bidder. Summary statistics are presented in Table 1.12. The median lower (upper) bound on entry costs is \$1,959 (\$2,147), while the median DOT savings amount to \$49,335. The distribution of DOT savings is quite fat tailed, however. While the mean lower (upper) bound on entry costs is \$2,316 (\$2,567), the mean DOT saving is \$82,583. This suggests that there is substantial potential value to encouraging entry with a relatively modest guaranteed bonus payment to the winning bidder.

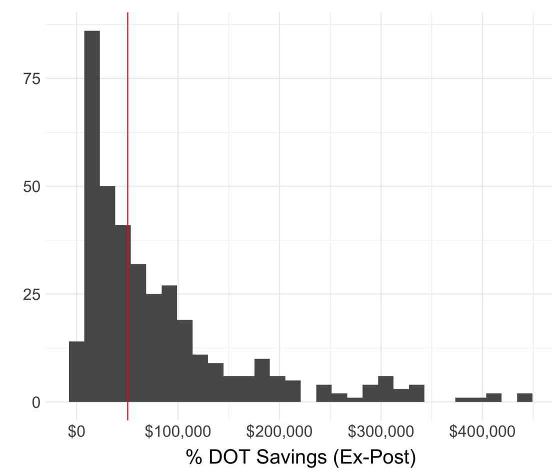
[69] See Lemma 1 in the appendix for a formal proof.



(a) *Distribution of lower bounds on the cost of entry*



(b) *Distribution of upper bounds on the cost of entry*



(c) *Distribution of the expected dollar savings to the DOT from the entry of an additional bidder to each auction*

Figure 1.15: *Welfare impacts of an additional entry to each auction*

Note: The median is highlighted in red in each case.

Table 1.12: *Summary of the welfare impacts of an additional bidder (not paying an entry cost) to each auction*

Statistic	Mean	St. Dev.	25%	Median	75%
Net DOT Savings	$82,583.25	$87,568.51	$22,296.89	$49,335.35	$103,379.50
% DOT Savings	8.90%	8.45%	2.06%	5.65%	13.47%
Entry Cost Lower Bound	$2,315.80	$1,524.88	$1,264.95	$1,959.42	$3,135.44
Entry Cost Upper Bound	$2,567.53	$1,683.76	$1,369.73	$2,147.20	$3,445.23

Note: Results are truncated at the top and bottom 1% to exclude extreme outliers from the mean/SD calculations.

1.9 Conclusion

This paper studies the bidding behavior of construction firms that participate in scaling procurement auctions hosted by the Massachusetts Department of Transportation. In particular, we analyze the incentives for bidders to strategically *skew* their bids. We show that while bidders do skew, placing high bids on items they predict will overrun the DOT's quantity estimates and low bids on items they predict will underrun, this is not necessarily indicative of rent extraction. For risk averse bidders, skewing facilitates diversifying bidders' exposure to the risk of items being used in quantities far outside their expectations. In a competitive environment, such as the one in MassDOT's bridge maintenance auctions, skewing generates substantial savings to the DOT. If bidders were compensated entirely based on the DOT's quantity estimates (or equivalently, using a lump sum auction), they would not be able to skew their bids. However, in this case, bidders would be responsible for all unanticipated modifications to the project specification and raise their bids on the whole to account for the added risk. Our estimates suggest that the DOT would subsequently pay nearly 85% more for the median project.

While switching to a lump sum auction would increase DOT expenditures, increasing bidders' exposure to risk a little bit may not be as harmful. A mixed compensation auction in which bidders are paid half on the DOT's estimates and half on the realized quantities only increases the median project's cost to the DOT by 3.5%. For a few projects, this mixed auction may even save the DOT a little. This suggests that policies that limit bidders' ability to fully optimize their bids—such as the minimum bid requirement considered by MassDOT—may be helpful in reducing DOT expenditures. Examining a counterfactual with a minimum bid requirement, and other potential mechanism design interventions is left for future work.

Chapter 2

Bargaining and International Reference Pricing in the Pharmaceutical Industry[1]

2.1 Introduction

The pharmaceutical industry represents a significant part of the global economy: global pharmaceutical sales amounted to $1.1 trillion in 2016, one third of which came from the US.[2] Policymakers around the world face the challenge of balancing the long-term benefits of pharmaceutical R&D incentives against the more immediate benefits of regulating or negotiating lower drug prices (Lakdawalla (2018), Lakdawalla *et al.* (2009)). Innovating new drugs is expensive: the Pharmaceutical Research and Manufacturers of America (PhRMA) estimates that the average cost to develop a drug (including the cost of failure) has increased from $140 million in the 1970s to $1.2 billion in the early 2000s (both in adjusted 2000 dollars), and only 2 out of 10 drugs ever achieve sufficient revenue to cover these R&D costs.[3] DiMasi

[1] Co-authored with Ashvin Gandhi and Pierre Dubois

[2] QuintilesIMS Global Pharma Outlook 2016.

[3] PhRMA 2014 profile.

et al. (1991, 2003, 2016) document a steady evolution in the cost of innovation—figures that rise from $230 million (1987) to $500 million (2000) to $1.4 billion (2013).[4] Given the substantial cost of R&D, the profits that a pharmaceutical firm expects to make off of a drug play a large role in the firm's decision to invest in developing it. New drugs are generally protected from competition by patents in order to ensure adequate profitability, and breakthrough drug prices often greatly exceed their marginal costs of production. For example, Gilead Sciences recently priced its breakthrough hepatitis C drug, Sovaldi, at $1,000 per pill—a price that almost certainly exceeds its marginal cost.[5] Even in less extreme cases, margins can be substantial: Dubois and Lasio (2018) find margins in the range of 10-50% in the French anti-ulcer industry—in spite of French price constraints—and Linnosmaa *et al.* (2004) estimate Finnish drug margins to be in the range of 59-67%.

The social planner's problem is further complicated by the fact that the benefits to pharmaceutical R&D may spill over to other countries. While there exists a theoretical literature on this topic—see, for example, Helpman (1993) and Grossman and Lai (2004)—there is very limited empirical work. One notable exception is **?**, which examines quinolone sales data to determine the effect of TRIPS global patent protection on welfare. Chaudhuri et. al. find substantial welfare losses to the Indian economy, resulting from the enforcement of foreign pharmaceutical intellectual property rights in India. Moreover, it has been shown that pharmaceutical industry profits as a whole affect R&D. Acemoglu and Linn (2004) and Dubois *et al.* (2015a) demonstrate a positive elasticity of innovation in relation to market size. Acemoglu *et al.* (2006) examines whether the introduction of Medicare affected pharmaceutical innovation and shows a positive effect, as well. Filson (2012) defines a dynamic-stochastic equilibrium model of innovation and fits it to industry facts in order to assess counterfactuals in which either the US adopts price controls or other countries drop theirs. Dynamic models of R&D have also been employed to study other industries, such as high- and low-tech manufacturing Peters *et al.* (2017).

[4] To obtain these numbers, we adjusted the figures reported in the papers for inflation.

[5] "Sales of Sovaldi, New Gilead Hepatitis C Drug, Soar to $10.3 Billion." Feb. 3, 2015. *New York Times*.

However, as the US spends twice as much as European countries per inhabitant in pharmaceuticals—not only because of larger consumption but also because of substantially higher prices—price controls in the US are increasingly being called for in policy circles Salter (2015); OECD (2017), as well as, recently, by the US administration.[6] For example, Salter (2015) discusses international reference pricing for the US as a way to reduce pharmaceutical spending, using experience in other developed countries as evidence of price reduction effects. Weiss *et al.* (2016) say that the US government may reduce the differential pricing that exists with respect to other markets by using an international reference pricing policy (though price controls may only be achieved following re-referencing as the US is typically a first-launched market). Such a policy was implemented on a small scale in the 1990s when the US Federal Government included a Most Favored Customer clause on pharmaceutical product prices supplied to Medicaid. Scott-Morton (1997) shows that, while firms had to provide Medicaid at their lowest price, the rule resulted in higher prices to some non-Medicaid consumers of pharmaceuticals. Most price control policies base price negotiations on external reference pricing—pricing of the same drugs in other countries. In the case of the US, and unlike Canada or most European countries, drug pricing is not currently negotiated by a centralized regulatory authority that can adopt more or less aggressive negotiating standards. The advantage of an international reference pricing policy is then that it only requires an ex post control that US prices should not be higher than prices for the same drugs in referenced countries.

In this paper, we develop a model that allows us to simulate a counterfactual international reference pricing policy in which price controls are introduced in the US, in reference to other countries' prices. Such a policy may imply changes in equilibrium prices, both in the US and the reference country. Using data from the US and Canada, our paper develops and estimates a structural model of supply and demand that allows us to assess how prices are set both in Canada and the US. In Canada, this amounts to estimating the marginal costs

[6]See New York Times, October 25, 2018: "Trump Proposes to Lower Drug Prices by Basing Them on Other Countries' Costs".

of products and the bargaining weights of firms that negotiate prices with regulators. In the US, it entails a Bertrand-Nash equilibrium in prices across competing firms. This gives us a setting in which we can evaluate counterfactual prices, demand, and welfare given different international pricing regimes. In particular, we simulate a policy in which the US constrains prices offered in its markets by the prices offered in Canada. In equilibrium, firms internalize the restrictions imposed by US reference pricing when negotiating with Canada. They also internalize the effects Canadian price setting when negotiating with the US.[7] Our approach is novel in that we study the equilibrium price setting that results due to reference pricing—both on prices in the country adopting a price control and in the reference countries. As such, we determine welfare and profit effects in the global pharmaceutical market equilibrium.[8]

We use detailed data on drug quantities and prices from IMS Health to estimate a random coefficient logit model of demand with estimated drug class-specific market sizes. We then model the price setting in a country with regulated prices (such as Canada) as the result of negotiation between pharmaceutical manufacturers and a centralized regulator under a Nash bargaining equilibrium Horn and Wolinsky (1988); Crawford and Yurukoglu (2012); Grennan (2013); Gowrisankaran *et al.* (2015). With these supply side assumptions, we are able to separately identify costs and bargaining parameters.[9] Since Nash bargaining involves maximizing the weighted log-sum of both parties' transaction utility, we can interpret the bargaining parameters as the degree to which countries' policymakers choose to trade off between firm profits and immediate consumer welfare.

Given our estimates of preferences, marginal costs, and bargaining parameters, we then assess counterfactual policy simulations in which pharmaceutical prices in the United States are subject to international reference pricing. Under the assumption that cost and demand

[7]In counterfactuals in which the US imposes reference pricing, we assume that price setting is set via negotiations with regulators as is the case in other countries that use reference pricing schemes.

[8]Danzon and Chao (2000) and Danzon *et al.* (2005) also study the equilibrium effects of international reference pricing, examining its effects on delayed entries of new drugs in reference countries.

[9]Dubois and Lasio (2018) instead chooses to model price setting in France as setting price ceilings that constrain firms.

parameters would not change, we simulate the counterfactual prices that result. In our counterfactual equilibrium, firms internalize the constraint that US prices must be lower than prices in Canada, while simultaneously price negotiations in Canada internalize the impact of the their result on price setting in the US.

Our results show that such a policy results in a slight decrease in US prices and a substantial increase in Canadian prices. The magnitude of these effects depends on the particular structure of the policy. The effect appears to be asymmetric because of the size differences in pharmaceutical markets across countries, the bargaining parameter value in Canada, firms' marginal costs and the shape of demand in each country. Overall, we find modest consumer welfare gains in the US, but substantial consumer welfare losses in Canada. Moreover, we find that pharmaceutical profits increase in net, suggesting that reference pricing of this form would constitute a net transfer from consumers to firms. Our analysis sheds new light on the price effects of reference pricing and shows the costs and benefits of a most favored nation policy in the US.

The effects demonstrated by our analysis are in addition to the negative impacts that previous work has shown reference pricing to have on entry in referenced countries (**?**, Danzon *et al.* (2005), Maini and Pammoli (2017)). Our analysis holds entry/exit fixed and so it does not internalize such an effect. Moreover, while our analysis shows the effects on consumer welfare and manufacturing profits, it likely underestimates the long-term welfare impact as revealed preferences from current consumers and regulators' behavior probably do not fully internalize the trade-off between current and future generations.

Our paper is structured as follows. Section 2.2 presents the data used for Canada and the US. Section 2.3 presents the demand model that we use for each market and country, as well as its identification method. Section 2.4 introduces the supply side models, both for regulated and unregulated pharmaceutical markets, that we estimate in order to identify structural supply side parameters. It then presents the supply side identification method and estimation results. Finally, section **??** develops a counterfactual model of international reference pricing. Section 2.6 concludes.

2.2 Data and Descriptive Statistics

We use data from IMS Health on revenues and quantities of drugs at the quarter level from 2002 to 2013. Our data spans the United States and Canada—the main markets in our study—as well as France, Germany, the UK, Italy, and Spain, which we use to construct instrumental variables for our identification strategy. Observations in our data are at the product-dosage level by country and quarter, and by hospital, retail or other channel of use. The data also includes product characteristics and the manufacturer name. We aggregate drugs across multiple dosage forms and administering methods (e.g., tablets and injections) using "standard units", the minimal dosage of a given drug. We use the international drug name in the data to match drug names across countries. We aggregate sales to the molecule-corporation-market level and aggregate generics for each molecule. We focus on prescription drugs and do not study the OTC market. We leave the question of the consequences of having country-specific definitions of OTC versus prescription drugs for future research. We compute quarterly drug prices as the ratio of total revenue and total quantity in standard units.

Our data details each drug's Anatomical Therapeutic Chemical (ATC) Class. In the ATC system, all drugs are classified into groups at five different (nested) levels. Our data contains the fifth ATC classification level (ATC-5) for each drug. For example, the classification of metformin (brand names: Glumetza, Fortamet, Glucophage, Riomet) is at the 1st Level (Anatomical Main Group): (A) Alimentary tract and metabolism; at the 2nd Level (Therapeutic Subgroup): (A10) Drugs used in diabetes; at the 3rd Level (Pharmacological Subgroup): (A10B) Blood glucose lowering drugs; at the 4th Level (Chemical Subgroup): (A10BA) Biguanides; and at the 5th Level (Chemical Substance): (A10BA02) Metformin.

We define markets at the ATC-4 class level. We restrict our focus to the 31 ATC-4 classes for which we have at least one on-patent molecule both in Canada and in the US.[10] These 31 ATC-4 classes are drawn from a set of 24 ATC-3 classes that covers 93% of total hospital

[10]That is, we exclude ATC-4 classes in which Canada does not have any on-patent molecules, while the US does. This typically happens because of the delayed entry of new molecules in Canada.

drug expenses in the US and 72% in Canada. We describe the treatment types by ATC-3 classes covered in our analysis in table 2.1 below.

ATC 3 Class	Treatment Type
A2B	Antiulcerants
B1B	Heparins
C2A	Antihyper- Tensives
C7A	Beta-Blocking Agents
C8A	Calcium Antagonists
C9A	Ace Inhibitors
C10A	Cholesterol- And Triglyceride-Regulating Preparations
L1A	Alkylating Agents
ATC-2 L1	Antineoplastic Alkylating Agents
L1B	Antimetabolites
L1C	Plant-Based Antineoplastics
L1D	Antineoplastic Antibiotics
L1X	Other Antineoplastics
L2B	Cytostatic Hormone Antagonists
L4X	Other Immunosuppressants
M1A	Nonsteroidal Antirheumatics
M5B	Bone Calcium Regulators
N1A	General Anesthetics
N1B	Local Anesthetics
N2A	Narcotics
N2B	Nonnarcotics And Antipyretics
N3A	Antiepileptics
N5A	Antipsychotics
N6A	Antidepressants And Mood Stabilizers

Table 2.1: *ATC-3 Description and Summary Statistics*

Table 2.2 shows descriptive statistics on the number of molecules by on-patent/off-patent branded and generic status within each ATC-4 class, in the US and in Canada. In addition, Table 2.2 displays the share of expenditures of US and Canadian hospital sector pharmaceutical spending that each ATC-4 class represents. There is variation across ATC-4 classes in the proportion of drugs with enforceable patents. ATC-4 classes in which most molecules' patents are expired typically have most drugs available in generic—and so, inexpensive—form. In these cases, lowering prices in the US is of less interest.

There is also variation in the share of expenditures that different ATC-4 classes represent between Canada and the US. In Canada, anti-cancer drugs (L1 class) represent a relatively larger share of total expenses (around 35%) than the 20% that they represent in the US. By contrast, the share of US spending on anti-thrombotic agents is much larger (16.8%) than in Canada (7.9%). The distribution of relative expenses across drug classes is thus different between the two countries, even though the US spends more in absolute value in every ATC-4 class and pays higher prices on almost all drugs, as shown in Table B.1 in Appendix B.1.2. Although the composition of drugs sold within each class in each country is different, the ATC-4 level average price is much higher in the US in almost every class and quarter. In fact, there is likely to be a negative correlation between prices and quantities within each class that makes the average price by ATC-4 class potentially less different across countries, in addition to the fact that some expensive drugs are sometimes not even sold in the US.

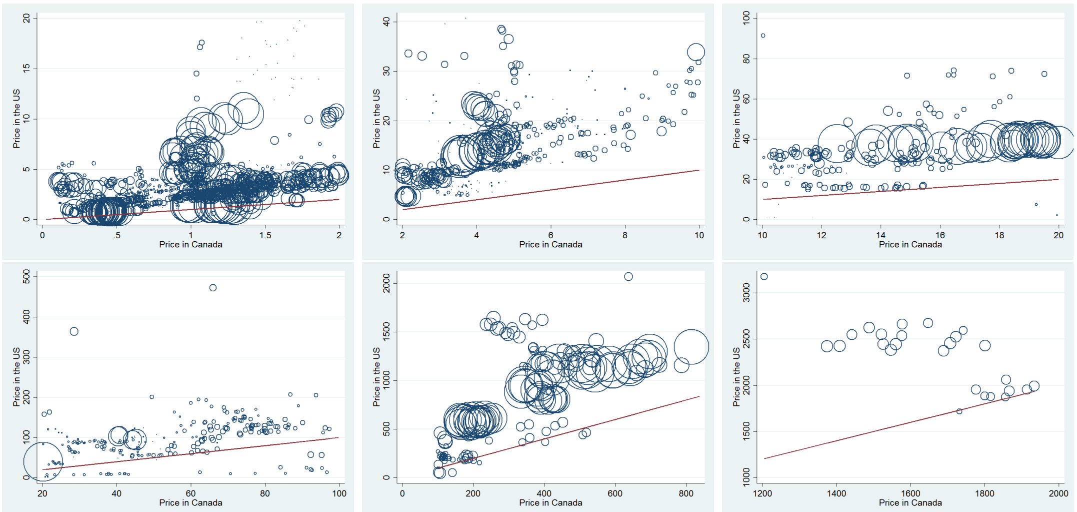
For drugs that are sold in both the US and Canada, it is interesting to verify that prices are indeed higher in the US than in Canada, as this is one of the motivation for policymakers to propose price control policies. Figure 2.1 shows a scatter plot of prices in the US against prices in Canada for the on-patent drugs present in both countries and across different ranges of prices in $US per standard unit. As shown in the figure, most drugs are more expensive in the US than in Canada by a large amount that is increasing in absolute value with the price of the drug in Canada. The ratio of prices between the US and Canada slightly decreases, however, so that the most expensive drugs are priced similarly across the two countries.

Table 2.2: *Number of molecules and expenditure shares by ATC-4*

		Canada				US			
		Number				Number			
ATC4	Label	On Patent	Branded Off	Generics	Exp Share	On Patent	Branded Off	Generics	Exp Share
A10C1	H INSUL+ANG FAST ACT	3	0	0	0.66	4	0	0	1.16
A2B2	ACID PUMP INHIBITORS	1	4	4	3.36	4	4	4	4.12
B1B2	FRACTIONATED HEPARINS	4	0	0	7.98	3	0	0	16.81
C10A1	STATINS (HMG-COA RED	3	1	3	3.19	3	3	3	2.39
C2A2	ANTIHYPER.PL MAINLY PERI	1	2	4	0.32	2	1	4	0.51
C7A0	B-BLOCKING AGENTS,PLAIN	2	3	10	1.22	2	8	12	2.18
C8A0	CALCIUM ANTAGONIST PLAIN	1	4	4	1.90	1	6	7	2.50
C9A0	ACE INHIBITORS PLAIN	9	1	2	1.55	6	3	5	0.58
C9C0	ANGIOTEN-II ANTAG, PLAIN	2	4	5	1.10	5	3	3	0.96
L1A0	ALKYLATING AGENTS	6	2	3	1.75	9	4	5	2.06
L1B0	ANTIMETABOLITES	7	1	3	7.90	5	3	7	6.84
L1C0	VINCA ALKALOIDS	3	3	5	10.84	4	3	5	4.79
L1D0	ANTINEOPLAS. ANTIBIOTICS	3	3	5	4.07	4	4	5	2.17
L1X4	A-NEO PROTEIN KINASE INH	8	0	0	9.31	10	0	0	0.96
L1X9	ALL OTH. ANTINEOPLASTICS	2	1	2	2.67	7	0	3	1.26
L2B2	CYTO ANTI-ANDROGENS	1	2	3	0.91	1	1	2	0.11
L2B3	CYTOSTAT AROMATASE INHIB	3	0	0	1.87	4	0	0	0.14
L4X0	OTHER IMMUNOSUPPRESSANTS	5	1	2	3.72	9	3	4	1.75
M1A1	ANTIRHEUMATICS NON-S PLN	1	4	7	0.38	2	8	12	0.40
M5B3	BISPHOSPH OSTEOPOROSIS	2	2	3	0.59	4	2	2	0.47
N1A1	INHAL GEN ANAESTHETICS	1	2	2	3.68	1	2	2	8.26
N1A2	INJECT GEN ANAESTHETICS	2	4	6	2.27	3	6	8	6.36
N1B1	ANAESTH LOCAL MEDIC INJ	2	3	3	0.98	1	2	5	1.12
N1B3	ANAESTH LOCAL TOPICAL	1	1	2	1.73	3	2	3	1.16
N2A0	NARCOTIC ANALGESICS	1	5	10	5.19	1	4	17	7.06
N2B0	NON-NARCOTIC ANALGESICS	1	6	8	0.56	2	5	15	0.93
N3A0	ANTI-EPILEPTICS	4	5	8	2.71	12	3	7	6.67
N5A1	ATYPICAL ANTIPSYCHOTICS	3	1	1	14.81	5	1	1	13.16
N5A9	CONVNTL ANTIPSYCHOTICS	7	4	8	1.13	3	4	8	0.71
N6A4	SSRI ANTIDEPRESSANTS	1	4	5	1.11	1	3	5	1.70
N6A9	ANTIDEPRESSANTS ALL OTH	3	3	12	0.54	5	9	12	0.71

Note: Average number of molecules (rounded to closest integer) and expenditure shares (in %) within country over 2002-2013, by ATC-4 classes. Some ATC-4 abbreviated labels have been revised and are not used anymore. See details of classification in European Pharmaceutical Market Research Association (2018).

Figure 2.1: *Comparisons of Prices of On-Patent Drugs present in both the US and Canada*



Note: Graphs for different scales of the price in Canada because of the enormous variation of prices of drugs in $US per standard unit. Within each graph, the circle size is proportional to the sales value of this drug in the US.

2.3 Demand Model

Pharmaceutical bargaining depends, in large part, on consumers' substitution between competing drugs at different price levels. Regulators consider how each proposed price change will impact total consumption (and subsequently welfare), while manufacturers consider how it will impact profits. In order to take this into account, we estimate a flexible model of aggregate consumer demand for drugs within each market. We use a standard random utility discrete choice model in which consumers' utility is a function of prices and available drug characteristics. We cannot observe data on the behavior of insurers, healthcare providers and other intermediaries between patients and drug manufacturers, and so we abstract away from modeling them and do not disentangle their role in aggregate revealed preferences. We focus on purchases made in the hospital sector. Hospitals typically fully internalize the prices of drugs that they purchase on behalf on patients, who compensate the hospitals at a per-diem basis. By contrast, consumers making purchases in the retail sector often defer to doctors' prescriptions and pay co-pays that do not fully reflect the differences in prices. As such, consumers in the retail sector may not fully internalize differences in drugs, and so the revealed preference expressed in their observed purchase decisions is more difficult to interpret for the purpose of welfare analysis. Thus, while we do observe retail sales data, we focus on the hospital sector for our analysis.

2.3.1 Demand Specification

We model the drug choice problem of a representative consumer as follows. A drug market is defined by a level 4 Anatomical Therapeutic Chemical (ATC-4) class, a country (e.g. Canada and the US), and a fiscal quarter. We denote fiscal quarters by t, countries by c and ATC-4 classes by m. Consumer preferences for each drug in a market are defined according to a random coefficient logit framework for differentiated products, following Berry *et al.* (1995) and Nevo (2001).

Within each country c, a representative individual i chooses to purchase a drug j from the set of choices $j = 0, 1, .., J_{m(j)}$ available in j's market, according to the indirect utility:[11]

$$U_{ijt} = u_{ijt} + \varepsilon_{ijt}$$

where

$$u_{ijt} = \alpha_i \ln p_{jt} + \beta_{im(j)} g_j + \gamma_i + \lambda_{m(j)} x_j + \phi_j + \mu_{m(j)t} + \xi_{jt}.$$

We normalize the utility for the outside good (choosing not to purchase anything), u_{i0t}, to zero. We denote p_{jt} for the price of drug j at t. Drug characteristics are captured by the drug's molecule identifier, patent status and generic status. In our utility specification, g_j is a binary variable indicating whether drug j is generic, x_j is a binary variable indicating whether j's molecule patent has expired by quarter t and ϕ_j is a molecule fixed effect. An unobserved shock at the drug-quarter level is denoted by ξ_{jt}.

Consumer preferences are captured by three types of random effects. Individual value

[11] All parameters and variables in the utility function, as well as the choice set within an ATC-4 class, are country-specific. We suppress the country index c for ease of exposition. Since each drug is only available in one ATC-4 class, we also suppress the m subscript in market denotations. That is, we consider the demand model country by country, and each unique market that a drug j is available in is denoted by t.

for purchasing an inside good is captured by the random effect γ_i. Individual disutility from higher prices is captured by by the random coefficient α_i on log prices.[12] Individual preference for branded drugs is captured by the random coefficient β_{im} on the branded indicator variable. We assume that random coefficients are independently normally distributed with $\alpha_i \sim \mathcal{N}(\alpha, \sigma_\alpha)$, $\beta_{im} \sim \mathcal{N}(\beta_m, \sigma_\beta)$, $\gamma_i \sim \mathcal{N}(0, \sigma_\gamma)$, and denote the vectors of parameters $\theta = (\sigma_\alpha, \sigma_\beta, \sigma_\gamma)$. The mean utility for drug j in quarter j is thus given by

$$\delta_{jt} = \alpha \ln p_{jt} + \beta_{m(j)} g_j + \lambda_{m(j)} x_j + \phi_j + \mu_{m(j)t} + \xi_{jt}.$$

Assuming that ε_{ijt} is i.i.d. extreme value distributed, the expected market share of product j in market mt where $m = m(j)$ is given by the aggregate probability that j will be chosen from the choice set in m:

$$s_{jt}\left(\delta_{jt}, \xi_{jt}, \theta\right) = \int \frac{\exp\left(u_{ijt}\right)}{1 + \sum_{k=1}^{J_m} \exp\left(u_{ikt}\right)} dF(\nu_{im}; \theta) \tag{2.1}$$

where ν_{im} denotes the vector of random coefficients $\{(\alpha_i - \alpha), (\beta_{im} - \beta_m), \gamma_i\}$ and $F\left(.; \theta\right)$ denotes their joint c.d.f.

2.3.2 Demand Identification

We estimate our demand model according to the standard BLP method with instrumental variables for prices Berry *et al.* (1995). We construct drug-quarter demand shocks $\xi_{jt}(\delta_{jt}, s_{jt}, \theta)$ by inverting a system that matches the theoretical market shares in equation 2.1 to observed market shares. We then form moment conditions by interacting the inverted demand shocks with a set of orthogonal instruments Z_{jt} so that

$$\mathbb{E}\left[Z_{jt}\xi_{jt}(\delta_{jt}, s_{jt}, \theta)\right] = 0.$$

The key challenge to estimation is the consistent estimation of the price coefficient distri-

[12]We use a log price specification that fits better the data because we have very heterogeneous prices across different ATC-4 markets. While widely used in the literature Bjornerstedt and Verboven (2016); Gowrisankaran and Rysman (2012); Berry *et al.* (1995), it is known that this specification does not correspond to a closed form solution for its direct utility function.

bution. We expect the process of Price-setting to be affected by unobserved demand shocks ξ_{jt}, and so observed prices are likely to be correlated with $\xi_{jt}(\delta_{jt}, s_{jt}, \theta)$. Our identification thus depends on the use of instruments that affect prices but are orthogonal to ξ_{jt}. While the gold standard would be to collect direct cost-shifters for each drug, this is impractical for our exercise. In order to assess the effect of an international reference pricing policy on total hospital drug spending, we examine a large number of drugs across a large number of therapeutic classes. As such, it is unlikely that we would be able to find detailed cost-shifters that are relevant to all of the classes of drugs that we cover. Similarly, it would be unfeasible to collect specific cost-shifters for each drug or therapeutic class. One possibility would be to restrict our analysis to a few therapeutic classes, find class-specific cost shifters and identify the price coefficient only off of those therapeutic classes. However, this would limit the scope of our empirical assessment.

In this version of the paper, we leverage observed differences and changes in consumers' choice sets from quarter to quarter as our primary source of identification. In particular, we form instruments by collecting, for each drug j in each quarter t, the number of products in j's ATC-4 class, its (broader) containing ATC-3 class, the numbers of generics and off-patent branded drugs, both for j's molecule and in general within j's therapeutic class, and the number of countries (out of France, Germany, Canada, Spain, Italy, the UK and the US) in which j is offered in the hospital sector. These variables capture variation in the composition of drug j's competition that is largely driven by the entry of new drugs, the expiration of patents, and the exit of outdated drugs. Similarly to BLP instruments, identification is premised on the assumption that isolation in the product space predicts prices through the competitive channel. Similar logic may still hold even if prices are set through bargaining: products that are innovative and without clear substitutes may be able to extract more rent when bargaining. Moreover, while changes in the competitive landscape for drug j is thus likely to impact its price, the changes themselves are largely driven by the ascendance of time and technological progress. Drugs often face delays in entering markets outside the US due to additional regulatory hurdles. Furthermore, patent protection is determined long in advance and entry decisions can take years. Even generic entries often face delays from regulations, start-up costs, etc. and so they provide an additional source of choice set variation. As such, it is unlikely that any of these instruments will correlate with the idiosyncratic demand shocks ξ_{jt}.

We might still be concerned that such instruments (as well as direct cost shifters and Hausman instruments) only weakly correlate with marginal cost, yielding a weak instrument problem. There are a number of reasons why marginal costs and prices may not covary in the pharmaceutical industry as much as they do in other industries. First, markups can be very high in the pharmaceutical industry, as marginal costs are typically very small relative to the fixed cost of R&D. As such, changes in the level of marginal costs may not correspond to meaningful changes in prices. Second, since drug prices are usually set through some form of bargaining—either between pharmaceutical companies and the government or between pharmaceutical companies and insurers—prices may be less responsive or slower to respond to marginal cost changes. First, given the discrete schedule of negotiations, drug prices may not be renegotiated sufficiently frequently to respond to marginal cost variation. Second, negotiations may constrain the scope of price changes, making small changes in prices costly and difficult. For instance, price increases may be explicitly prohibited by negotiated contracts, or prices may be tied to benchmarks (other countries' prices, value contribution, etc.) that lag or weaken prices' correlation with marginal costs.

In addition to checking the power of instrumental variables in a first stage regression, we consider using Hausman style instruments, as in Dubois and Lasio (2018). Identification using such instruments relies on the correlation between prices across markets due to common cost shocks rather than common demand shifters. To construct such instrumental variables, we perform country-level regressions of price on active ingredient dummies and quarter fixed effects, and we use the residuals as instruments for price. The instruments for the price of product j in market $m(j)$ are the contemporaneous residuals for the price of product j in other countries. As an example, we instrument for the price of the drug Sovaldi in the United States using the price residuals of Sovaldi in France, Germany, Canada, Spain, Italy, and the UK. The reason we use residuals as instruments is that these allow us to control for temporal, regional, and quality components that may contribute to contemporaneous demand-based variation in prices. We also allow for different relationships across countries for brand name drugs and generic drugs. We take additional care for producers with multiple drugs or for the fact that some drugs are available in only a subset of countries. When a product is not available in all other countries, we use residuals from available countries. When a product is available in only one country, we use the average residuals of other products within the same ATC in other countries as instruments. The main possible concern is that there is insufficient variation in these instruments to precisely identify price sensitivity, but this is again an empirical question of the power of instrumental variables, and we investigate this in our empirical estimates, to be detailed in an upcoming revision.

Finally, it is important to note that the estimation of BLP-type demand models requires the definition of market shares for products within each market. Quantities of drugs sold and normalized by standard units allow us to construct market shares but require the definition of a market size. Market sizes across many ATC-4 markets and across countries for the hospital sector are not obviously defined and can change over time and be very different. However, we do not observe an external estimate of market sizes, nor of the outside share (which would be equivalent). Instead, we approximate the aggregate yearly market size denoted by M_{mt} for each ATC-4 market using a nonlinear least squares calibration procedure similar to that in Huang and Rojas (2013); Huang, Dongling and Rojas, Christian (2014). We describe this procedure in detail in Appendix B.1.1. On average, we find that the estimated outside market share is 29% in Canada and 24% in the US with some variation across ATC-4 classes (see detailed estimates in Appendix B.1.2).

2.3.3 Empirical Results on Demand Estimation

We present key estimated demand parameters for the US and Canada in table 2.4. We find that the random coefficients on log prices in Canada and the US have similarly negative means. The standard deviation of the price coefficient in Canada shows substantial heterogeneity. There are a number of reasons that might underlie this. For instance, hospitals may choose to stock drugs at discrete intervals, and so lag in responding to price changes. Our estimates aggregate over many hospitals, which may vary in their frequency of decision making and stickiness of brand preferences.

We also find differences in the dimension of preference heterogeneity between Canada and the US. In the US, our estimate of the random coefficient on the generic indicator suggests that there is substantial heterogeneity in preferences for branded drugs. By contrast, in Canada, much of the heterogeneity in demand is captured in the constant term and is thus common to all drugs. We account for molecule fixed effects, ATC-4 specific year effects, and ATC-4 specific off-patent and generic effects as well, but do not report this in our main paper for the sake of exposition.

Table 2.4: *Demand Estimates for US and Canada*

Country		US		Canada	
Log Price	α	-2.254	(0.146)	-2.241	(0.206)
	σ^{α}	0.024	(0.246)	0.892	(0.224)
Generic Dummy	σ^{β}	1.628	(0.169)	0.357	(1.195)
Constant	σ^{γ}	0.042	(1.103)	1.562	(0.312)
Molecule dummies		Yes		Yes	
Off patent * ATC-4 dummies		Yes		Yes	
Generic * ATC-4 dummies		Yes		Yes	
Year * ATC-4 dummies		Yes		Yes	
Quarter dummies		Yes		Yes	

Note: Standard error in parenthesis. All dummy coefficients are not reported.

We present the average own- and cross-price elasticities for hospitals in the US and Canada in table 2.5. These elasticities are computed using our estimated demand function in every country, ATC-4 market and quarter. We present the average elasticities across ATC-4 classes and quarters within each country, in aggregate and by branded status. Overall, average price elasticities are similar between the US and Canada. However, own-price elasticities are slightly higher for generics than branded drugs in Canada, suggesting that hospitals in Canada are more responsive to price changes in generic drugs.

Table 2.5: *Average Price Elasticities for Canada and US*

	US		Canada	
	Own	Cross	Own	Cross
All	-2.033	0.124	-2.017	0.158
Branded	-2.044	0.155	-1.809	0.185
Generic	-2.021	0.147	-2.262	0.163

Note: Average own price elasticities across all products of ATC-4 markets and over quarters.

2.4 Supply Side Modeling and Estimates

2.4.1 Price setting with Bargaining

We model price setting for pharmaceuticals in Canada with a Nash Bargaining model in which firms maximize profits, while government regulators maximize consumer welfare. Nash Bargaining models of this sort (see for instance, Crawford and Yurukoglu (2012); Grennan (2013); Gowrisankaran *et al.* (2015); Ho and Lee (2017)) provide a parsimonious way to characterize the trade-offs facing policy-makers, who must balance producer profits against consumer welfare. In Canada, this bargaining may be interpreted literally, as the Canadian Patented Medicine Prices Review Board negotiates prices with drug manufacturers to ensure that they are not "excessive". Moreover, this model applies more generally to price-regulated pharmaceutical markets such as those in most European countries, absent international reference pricing. We assume that there is no international reference pricing (in the baseline), and so pricing is determined independently within each country. We thus exclude a country-specific index for exposition.

Firm profits are defined as follows. Within a market m at time t, firm f selling products $j \in F_{fm}$ receives flow profits:

$$\Pi_{fmt} = \sum_{j \in F_{fm}} \Pi_{jmt} = \sum_{j \in F_{fm}} (p_{jt} - c_{jt}) q_{jt}(\mathbf{p}_{mt}).$$

Here, c_{jt} and p_{jmt} are the marginal cost and price of drug j, respectively. Their difference (the firm's markup) multiplies q_{jt}, the total quantity of drug j demanded in market m, given the vector of prices $\mathbf{p}_{mt} = (p_{1t}, .., p_{J_{mt}})$ of drugs available in the market.[13] Firm f's total profit is the sum of its profits across markets:

$$\Pi_{ft} = \sum_m \Pi_{fmt}.$$

Government regulators maximize aggregate consumer welfare as revealed by the demand model in their country. We denote the welfare for consumers in market m at period t by: Small and Rosen (1981):

$$\begin{aligned} W_{mt}(\mathbf{p}_{mt}) &= M_{mt} \int W_{imt}(\mathbf{p}_{mt}) dF(\nu_{im}; \theta) = M_{mt} \int \ln \left[1 + \sum_j \exp\left(u_{ijt}\right) \right] dF(\nu_{im}; \theta) \\ &= M_{mt} \int \ln \left[1 + \sum_j \exp\left(\alpha_i \ln p_{jt} + \beta_{im} g_j + \gamma_i + \lambda_m x_j + \phi_j + \mu_{mt} + \xi_{jt}\right) \right] dF(\nu_{im}; \theta). \end{aligned}$$

That is, consumer welfare is given by the sum of the expected utility produced by each drug available in market m. We assume that bargaining takes place product-by-product, so that neither firms nor regulators are able to bargain jointly over their portfolio of pharmaceutical drugs. This excludes the possibility of cross-product considerations such as might occur in bundling arrangements.[14] Thus, at each market m and quarter t, prices are set product-by-product via Nash bargaining between the producer and the market m regulator, in order to maximize the Nash product of firm profits and consumer welfare:

$$\underbrace{\left(\Delta_{jm}\Pi_{ft}\left(p_{jt}, \mathbf{p}_{-jmt}\right)\right)^{\rho_{jm}}}_{\text{Profit from } j \text{ in } m} \underbrace{\left(\Delta_j W_{mt}(p_{jt}, \mathbf{p}_{-jmt})\right)^{1-\rho_{jm}}}_{\text{Welfare gain from } j \text{ in } m}.$$

Here, $\rho_{jm} \in [0, 1]$ is the bargaining parameter that determines the relative weight of the firm's (profit) objective in determining the Nash bargaining solution. In order to account for heterogeneity in the bargaining process across drug types, we allow ρ_{jm} to vary across ATC-4 markets and by each drug's status as on-patent, branded off-patent or generic. The

[13] Note that the quantity demanded is given by the size of the market multiplied by drug j's market share: $q_{jt} = M_{mt} s_{jt}$.

[14] This assumption is made in order to simplify modeling, as joint optimization of prices is a notoriously difficult problem in itself, and different approaches may yield to different equilibria. In the absence of data that would enable us to distinguish between possible bundling arrangements, we defer to the simplest setting as an approximation of reality.

firm's objective is defined as the equilibrium profit generated by offering drug j at price p_{jt}:

$$\Delta_{jm}\Pi_{ft}(p_{jt}, \mathbf{p}_{-jmt}) \equiv \Pi_{ft} - \sum\nolimits_{j' \neq j, j' \in F_f} \Pi_{j'm(j')t} = \Pi_{jmt}(p_{jt}, \mathbf{p}_{-jmt}),$$

where $\mathbf{p}_{-jmt}$ denotes the vector of prices for all drugs other than j in market m and quarter t. Note that this is just the profit directly accrued from the sale of drug j, as we have assumed that firms do not take into account substitution across different drugs in their portfolios when setting prices. Similarly, $\Delta_j W_{mt}(p_{jt}, \mathbf{p}_{-jmt})$ denotes the change in consumer welfare generated by the presence of drug j in market m and quarter t:

$$\Delta_j W_{mt}(p_{jt}, \mathbf{p}_{-jmt}) \equiv W_{mt}(p_{jt}, \mathbf{p}_{-jmt}) - W_{mt}(\infty, \mathbf{p}_{-jmt}).$$

We assume a Nash-in-Nash equilibrium. That is, the vector of competitor prices in the vector $\mathbf{p}_{-jmt}$ in the case of disagreement are assumed to be equal to the equilibrium prices. Thus, for each drug $j = 1, .., J_m$, the equilibrium price is set according to:

$$p_{jt} = \arg\max_{p_{jt}} \quad \left\{ \Pi_{jmt}(p_{jt}, \mathbf{p}_{-jmt})^{\rho_{jm}} (\Delta_j W_{mt}(p_{jt}, \mathbf{p}_{-jmt}))^{1-\rho_{jm}} \right\}. \tag{2.2}$$

The necessary first-order conditions of the Nash bargaining equilibrium definition in equation (2.2) imply that for all $j = 1, .., J_m$:

$$c_{jt} = p_{jt} + \frac{1}{\underbrace{\frac{\partial \ln q_{jt}(\mathbf{p}_{mt})}{\partial p_{jt}}}_{\text{Demand semi-elasticity}} + \frac{1-\rho_{jm}}{\rho_{jm}} \underbrace{\frac{\partial \ln \Delta_j W_{mt}(\mathbf{p}_{mt})}{\partial p_{jt}}}_{\text{Welfare semi-elasticity}}}$$

where

$$\frac{\partial \Delta_j W_{mt}\left(\mathbf{p}_{mt}\right)}{\partial p_{jt}} = \frac{\partial W_{mt}\left(\mathbf{p}_{mt}\right)}{\partial p_{jt}} = M_{mt} \int \frac{\partial W_{imt}\left(\mathbf{p}_{mt}\right)}{\partial p_{jt}} dF(\nu_{im};\theta) = M_{mt} \int s_{ijt} \frac{\partial u_{ijt}}{\partial p_{jt}} dF(\nu_{im};\theta) \tag{2.3}$$

Note that when $\rho_{jm} = 1$, pricing is set according to an unrestricted Bertrand-Nash equilibrium in prices where firms maximize profits and (2.4.1) simplifies to the usual condition:

$$c_{jt} = p_{jt} + \frac{q_{jt}\left(\mathbf{p}_{mt}\right)}{\partial q_{jt}\left(\mathbf{p}_{mt}\right) / \partial p_{jt}} \tag{2.4}$$

In such a case, an estimate of c_{jt} is straightforward to compute given demand parameter estimates. In the case of the US, we will use this special case to identify marginal costs, as we know that there is no central regulation of hospital prices akin to a bargaining game as in Canada. When $\rho_{jm} = 0$, we have price equal to marginal cost $p_{jt} = c_{jt}$.

2.4.2 Supply Side Parameters Identification and Estimation

The set of first-order conditions (2.4.1) relates marginal costs to the shape of demand, drug prices, and the bargaining parameters ρ_{jm}. With known bargaining parameters, these first-order conditions allow us to identify the vector of marginal costs c_{jmt} as functions of ρ_{jm}.

As we noted before, in the US, we assume that $\rho_{jm} = 1$ because prices are freely chosen and not regulated for the hospital sector.[15] In that case, the first-order conditions simplify to the usual Bertrand-Nash first-order conditions (2.4) and allow identifying all marginal costs, which we denote c_{jUSt} for a product j in a market belonging to the US as in Nevo (2001). For generics in the US, we impose that prices equal to marginal costs and do not estimate margins, which is consistent with the typical fact that once many generics have entered, prices are low and close to marginal costs.

In Canada, prices are set through bargaining and so we must identify the bargaining parameters ρ_{jm} in addition to marginal costs using equations (2.4.1). Without any restriction on marginal costs, we cannot identify marginal costs and bargaining parameters. We could use sign restrictions on marginal costs and markups in order to obtain lower and upper bounds on the bargaining parameter. However, it is natural to add restrictions based on parameterization to marginal costs functions as in Berry *et al.* (1995). One way to identify costs and bargaining parameters is to let marginal costs be constant over time, constant across countries, or both. We assume that marginal costs can be parameterized as additively

[15] Notable exceptions to unconstrained pricing include pharmaceutical sales to the "Big Four:" Department of Veteran Affairs ($3.4 billion in 2003), Department of Defense ($4 billion in 2003), Public Health Service, and the Coast Guard, which receive discounted drug prices negotiated with manufacturers. Medicaid also receives effective discounts, but these are in the form of ex post rebates paid directly to the state rather than lower prices paid at the register. Medicare, on the other hand, is prohibited from negotiating prices.

separable functions of supply-side covariates and an orthogonal error term as follows:

$$c_{jt}\left(\rho_{jm(j)}\right) = z'_{jt}\lambda + \omega_{jt} \tag{2.5}$$

with

$$\mathbb{E}\left[z_{jt}\omega_{jt}\right] = 0 \qquad \forall j, t \tag{2.6}$$

and where $c_{jt}\left(\rho_{jm(j)}\right)$ is solution of (2.4.1). In our application, z_{jt} include a molecule-specific and country-time-specific effect as well as the estimated US marginal cost c_{jUSt} from (2.4). We thus have further identification power by leveraging our assumption that pricing is known to be set through an unconstrained Bertrand-Nash pricing game for all products sold in the US (excluding Federal sales).

The orthogonality conditions (2.6) allow to define for any market m in Canada and all j such that $m(j) = m$:

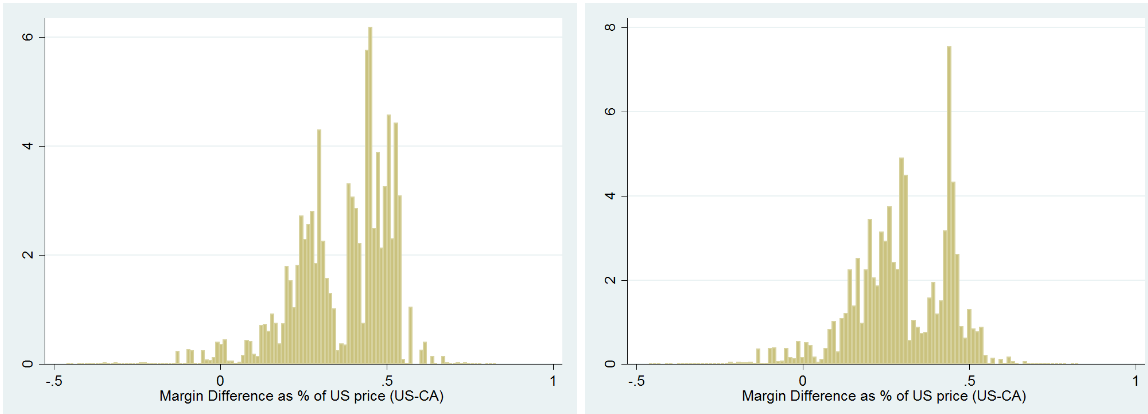
$$\omega_{jt}\left(\rho_{jm}\right) = \left[1 - z'_{jt}\left(z'_{jt}z_{jt}\right)^{-1} z'_{jt}\right] c_{jt}\left(\rho_{jm}\right)$$

Thus, we solve for any ATC-4 class m in Canada:

$$\{\rho_{jm}\}_{\{j=1,..,J\}} = \arg \min_{\{\rho_{jm}\}_{\{j=1,..,J\}}} \sum_{j,t} \omega_{jt}^2\left(\rho_{jm}\right) \tag{2.7}$$

Table B.3 in Appendix B.1.3 shows the estimated average margins in percentage of the maximum average price of US and Canada (which is almost always the US) by ATC-4 class so that we can compare them across countries. The results show relatively large margins—which is not surprising in the case of pharmaceuticals. We also find that the margins are larger in the US than in Canada for most drugs. Figure 2.2 draws the distribution of the differences of margins between US and Canada as a percentage of the US price, weighting the distribution either by quantity sold in the US or in Canada. The difference is most often positive as very few drugs have higher margins in Canada than in the US. The graph shows that many of products have margins in the US that are larger than in Canada by an amount that is more than 25% of the US price and up to 50%, which can mean extremely large differences in absolute dollars according to the US price level.

Figure 2.2: *Estimated Margins Differences between US and Canada for on Patent Drugs*



Note: The left panel shows the distribution of margins differences weighted by the US quantities of the drug. The right panel shows the distribution of margins differences weighted by the Canadian quantities of the drug. These distribution are for the sample of on-patent drugs present in both the US and Canada.

The supply model estimates also provide bargaining parameters estimates for Canada, as shown in Table B.4 in Appendix B.1.3. The parameters vary between 0 and 1.

2.5 Counterfactual Policies

In this section, we use our structural model to evaluate the impact of several counterfactual reference pricing policies. The primary reference pricing rule we consider prohibits pharmaceutical companies from setting higher prices for on-patent drugs in the United States than in Canada. In other words, this rule requires that for any on-patent drug j sold in both the United States (US) and Canada (CA):[16]

$$p_j^{US} \leq p_j^{CA}. \tag{2.8}$$

This type of policy is often referred to as an "international reference pricing" policy, or a "most favored nation" clause. The stated objective of such a rule is typically to reduce prices in the referencing country since they ensure that prices paid in the referencing

[16] To simplify notation, we exclude the time and drug-class subscripts in this section.

country (United States) are as least as low as those in the reference country (Canada). In equilibrium, however, reference pricing rules can also affect the price in the referenced country. In particular, profit-maximizing pharmaceutical companies may set or negotiate rates in the referenced country taking into account the impact on the price they can set in the referencing country. We incorporate this interdependence by allowing negotiations between pharmaceutical companies and Canada to account for the impact of the Canadian price on potential profit in the United States.

We present our primary counterfactual specification in Section 2.5.1. In this counterfactual we allow the Canadian price to act as a *price ceiling* in the United States. Implicitly, this assumes that pharmaceutical company cannot commit to a price in the United States prior to negotiation with Canada. This can be equivalently be seen as a timing assumption that prices are set in Canada prior to being set in the United States.

Because the implications of reference pricing policy may depend on the specific details of implementation,we also study alternative counterfactual specifications. The first of these is to allow the firm to commit to a price in the United States prior to negotiating with Canada. In this case, the price in the United States behaves as a *price floor* in the firm's negotiations with Canada. We include simulations for this counterfactual in Section 2.5.2. In future iterations of this paper, we hope to include two additional counterfactual specifications that are described in Appendix B.1.6. In the first of these additional specification, the United States regulator requires that the pharmaceutical company provide a comparison price in Canada. In particular, this rules out the possibility that the firm can serve only the United States and thereby weakens its negotiating power against Canada. Lastly, we will simulate the impact of a counterfactual in which the United States regulator references the average of an index of prices in other countries. This index reference rule resembles Health and Human Services' "International Pricing Index" proposal.[17]

[17] See https://www.cms.gov/sites/drupal/files/2018-10/10-25-2018%20CMS-5528-ANPRM.PDF

2.5.1 Counterfactual: Price Ceiling

When the negotiated price in Canada acts as a price-ceiling, pharmaceutical companies set prices in the United States to maximize profits subject to that ceiling:[18]

$$p_j^{US}(p_j^{CA}, \mathbf{p}_{-j}^{US}) \equiv \arg \max_{p \in [0, p_j^{CA}] \cup \{\infty\}} \Pi_j^{US}\left(p, \mathbf{p}_{-j}^{US}\right) \mathbf{1}_{\{p \le p_j^{CA}\}}. \tag{2.9}$$

This defines a correspondence giving the firm's optimal price in the United States given its price in Canada and other products' prices in the United states.

We allow negotiations between pharmaceutical companies and the Canadian regulator to account for the impact of the Canadian price on profitability in the United States. Given negotiated price p_j^{CA} in Canada, the pharmaceutical company expects to earn $\Pi_j^{CA}(p_j^{CA}, \mathbf{p}_{-j}^{CA})$ in Canada and $\Pi_j^{US}(p_j^{US}(p_j^{CA}, \mathbf{p}_{-j}^{US}), \mathbf{p}_{-j}^{US})$ in the United States, where $p_j^{US}(p_j^{CA}, \mathbf{p}_{-j}^{US})$ is from (2.9). The agreement surplus for the firm in negotiation is therefore:

$$\Delta\Pi_j(p_j^{CA}, \mathbf{p}_{-j}^{US}, \mathbf{p}_{-j}^{CA}) = \underbrace{\Pi_j^{US}(p_j^{US}(p_j^{CA}, \mathbf{p}_{-j}^{US}), \mathbf{p}_{-j}^{US}) + \Pi_j^{CA}(p_j^{CA}, \mathbf{p}_{-j}^{CA})}_{\text{global profit under agreement}} - \underbrace{\Pi_j^{US}(p_j^{US}(\infty, \mathbf{p}_{-j}^{US}), \mathbf{p}_{-j}^{US})}_{\text{profit if in US only}}. \tag{2.10}$$

Following Horn and Wolinsky (1988), the negotiated price in Canada maximizes the Nash product:

$$p_j^{CA}(\mathbf{p}_{-j}^{US}, \mathbf{p}_{-j}^{CA}) \equiv \arg\max_p \left(\underbrace{\Delta\Pi_j(p, \mathbf{p}_{-j}^{US}, \mathbf{p}_{-j}^{CA})}_{\text{profit gain from agreement}}\right)^{\rho_j} \left(\underbrace{\Delta_j W_{CA}(p, \mathbf{p}_{-j}^{CA})}_{\text{welfare gain in CA from agreement}}\right)^{1-\rho_j}. \tag{2.11}$$

In equilibrium, the prices for on-patent drugs sold in both the United States and Canada satisfy (2.9) and (2.11), respectively.[19] In other words, equilibrium prices $\{(p_j^{US*}, p_j^{CA*})\}_j$ are

[18] We again use $p_j^{US} = \infty$ to denote exit from the United States market. This occurs when $p_j^{CA} < c_j^{US}$. This is most plausible as an equilibrium outcome when the Canadian market is large, Canadian consumers are price sensitive, and marginal cost is very low, while the US market is small, US consumers are price sensitive, and marginal cost is very high in the United States.

[19] The usual profit maximization and Nash bargaining conditions must also be satisfied for all other products

characterized by:

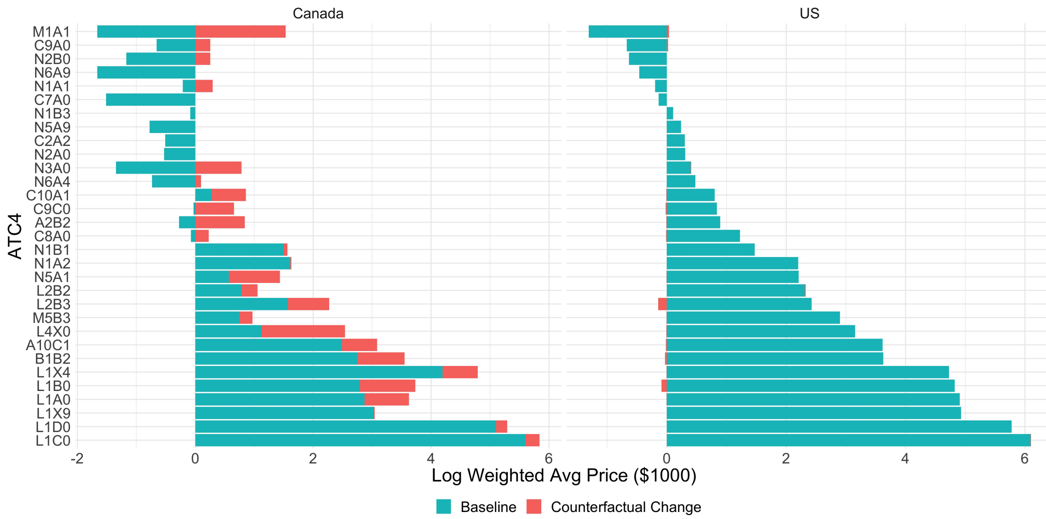
$$\begin{aligned} p_j^{US*} &= p_j^{US}(\mathbf{p}_j^{CA*}, \mathbf{p}_{-j}^{US*}), \\ p_j^{CA*} &= p_j^{CA}(\mathbf{p}_{-j}^{US*}, \mathbf{p}_{-j}^{CA*}), \end{aligned} \tag{2.12}$$

for all j. We expect that under this counterfactual simulation that prices in the United States will decrease while those in Canada will increase. We prove this in Appendix B.1.5 given simplifying assumptions.

Using our estimates for the parameters governing supply and demand from sections **??** and **??**, we simulate price setting in the US and Canada without reference pricing. In each therapeutic class, we simulate equilibrium bargaining in Canada, drug by drug, and Bertrand price-setting in the US, subject to the reference pricing policy detailed in Section 2.5.1. Although the reference constraint applies only to patented drugs, we also simulate the pricing decisions for generic and branded off-patent drugs since their optimal pricing decisions are likely to change as their on-patent competitors change prices. The graphs in this section therefore include off-patent drugs in addition to on-patent drugs unless otherwise stated.

in the US and Canada.

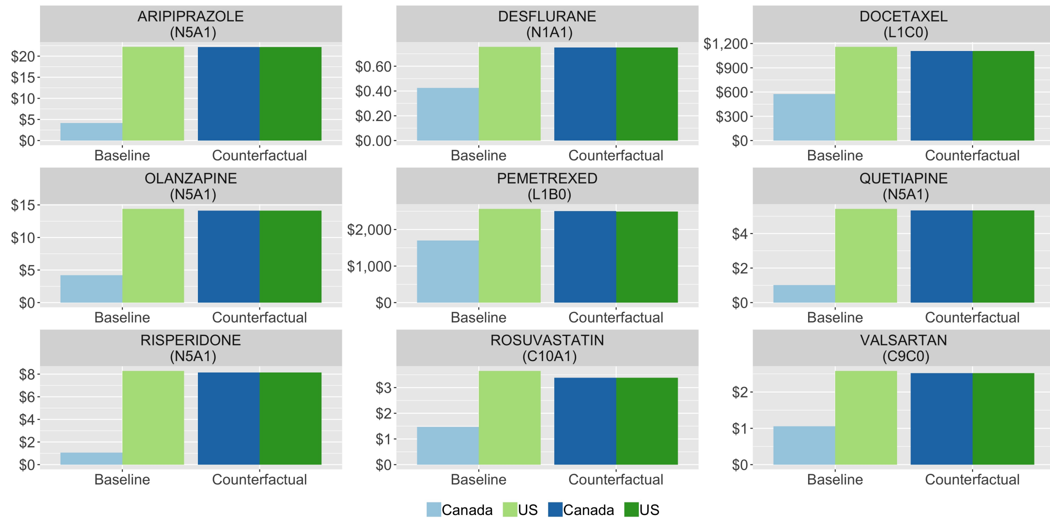
Figure 2.3: *Counterfactual Prices under Price Ceiling (Log Scale)*



Note: Each blue bar indicates the log of average prices in each ATC-4 class in the baseline without reference pricing. The red bar indicates the change in log average prices resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that prices increased by the length of the red bar. A red bar to the left of the blue bar indicates that prices decreased by the length of the red bar.

To evaluate the effects of our counterfactual, we consider several measures of impact. First, we consider the impact of counterfactuals on the prices of pharmaceuticals in both the United States and Canada. Figure 2.3 compares average prices by therapeutic class in the baseline to those that result from the introduction of reference pricing. We find that while the reference pricing policy leads only to a slight reduction in US prices, it leads to significant increases in Canadian prices. Figure 2.4 gives the price changes for nine drugs with large revenues in the United States. In all of these cases, the reference pricing rule results in a binding price constraint (i.e., $p_j^{US} = p_j^{CA}$) and equilibrium prices very close to the baseline United States prices. As such, they illustrate the dominant mechanism through which prices change when reference pricing is imposed: US prices decrease slightly, while Canadian prices rise to match them. As most Canadian prices are substantially lower than US prices in the baseline, rising to near baseline US levels constitutes a large increase. As we show next, this increase in Canadian prices dominates the quantity response, so that expenditure and profits both increase.

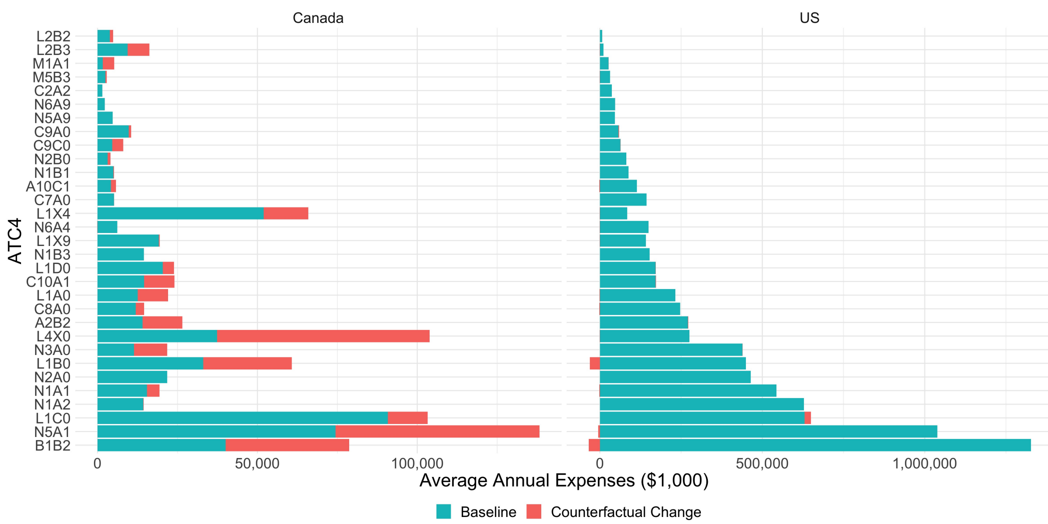
Figure 2.4: *Example Counterfactual Price Changes*



To illustrate the impact of reference pricing, we first present our counterfactual predictions for a few significant therapeutic classes. As a first case, we examine the Atypical Antipsychotics class (N5A1), which constitutes 14.8% of expenditures in the United States and 13.1% of expenditures in Canada, among the therapeutic classes we analyze. The impact of introducing the reference pricing rule on Canada is substantial: expenditures and profits increase by 85.8% and 91.1%, respectively. By comparison, these figures decrease by just 0.5% and 1.2% in the United States. Similarly, welfare decreases by 24.1% in Canada, and increases by only .7% in the US. Driving these results, the equilibrium prices in the both markets under the reference pricing rule are much closer to the baseline prices in the United States. Figure 2.4 shows that the prices of on-patent drugs Apriprazole, Olanzapine, Quetiapine, and Risperidone—all part of the N5A1 ATC-4 class—increase by hundreds of percent in Canada to near-parity with the baseline United States price.

Another key example is the Vinca Aklaloid class (L1C0), a class of chemotherapy drugs which constitutes 4.79% of expenditures in the US and 10.8% of expenditures in Canada. This class has a large concentration of patented drugs—4 out of 12 in the US and 3 out of 11 in Canada. Figure 2.4 shows the price changes of one such drug, Docetaxel, that result from the introduction of reference pricing. While the Canadian price is approximately half of the US price in the baseline, the prices are equal, at a level slightly below the baseline US price in the counterfactual. This is true on aggregate for patented drugs in L1C0: average drug prices increase in Canada by 66.1% but decrease in the US by 4.6%.[20] However, total expenses increase both in Canada (13.7%) and in the US (3.2%), reflecting a slight increase in US prices for branded off-patent drugs. Aggregate average profits and welfare follow the general pattern found across therapeutic classes: profits increase substantially (44.7%) in Canada, but decrease slightly (1.5%) in the US, while welfare decreases substantially (9.6%) and increases slightly (1.4%) in the US.

Figure 2.5: *Counterfactual Expenditure under Price Ceiling*



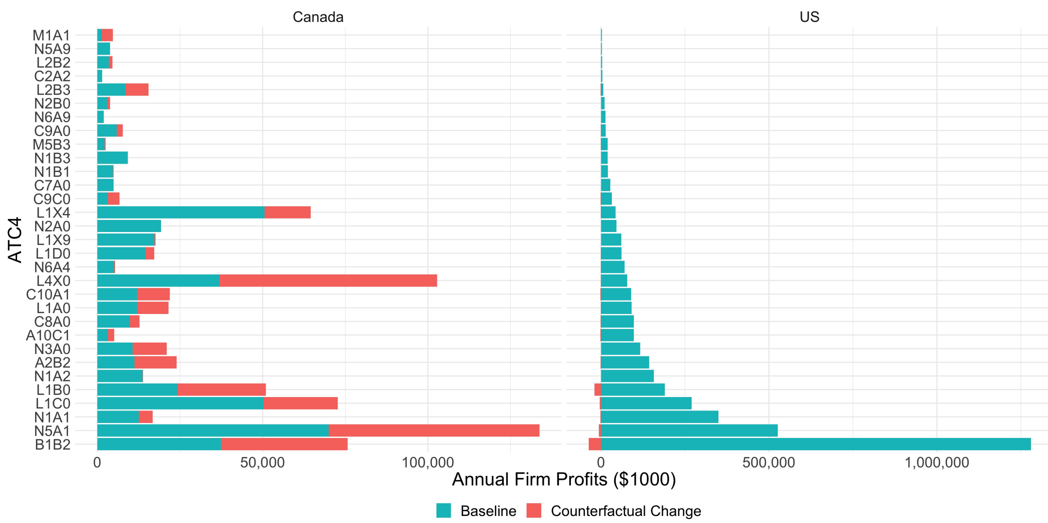
Note: Each blue bar indicates the average annual expenditure in each ATC-4 class in the baseline without reference pricing. The red bar indicates the change in average annual expenditure resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that expenditure increased by the length of the red bar. A red bar to the left of the blue bar indicates that expenditure decreased by the length of the red bar.

Figures 2.5, 2.6, and 2.7 depict the effect of the reference pricing rule on expenditures, profits, and welfare by therapeutic class. Tables of numerical values are given in Appendix B.1.7. While there is significant variation across ATC-4 classes, our results show that expenditures and profits overwhelmingly increase (60.6% and 65.6%, respectively) in Canada while Canadian consumer welfare decreases (12.3%). The impacts are significantly smaller

[20]See table B.9 for a breakdown of price changes by ATC-4 and drug status.

but generally reversed for the United States.[21] (Expenditure and profits in the United States decrease by 0.75% and 1.9%, respectively, while welfare increases .2%.) In our simulations, four on-patent drugs from four ATC-4 classes choose to exit from the United States market.[22] However, these products represent small market shares and small expenditure, suggesting that reference pricing does not generate large market distortions by incentivizing drug exits.[23] The results in the class of statins (C10A1) that Rosuvastatin belongs to, and the class of beta-blockers (C9C0) that Valsartan belongs to are similar.

Figure 2.6: *Counterfactual Profit Changes on All Drugs with Canada as Price Ceiling for the US*



Note: Each blue bar indicates the average annual profits in it ATC-4 class in the baseline without reference pricing. The red bar indicates the change in average annual profits resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that profits increased by the length of the red bar. A red bar to the left of the blue bar indicates that profits decreased by the length of the red bar.

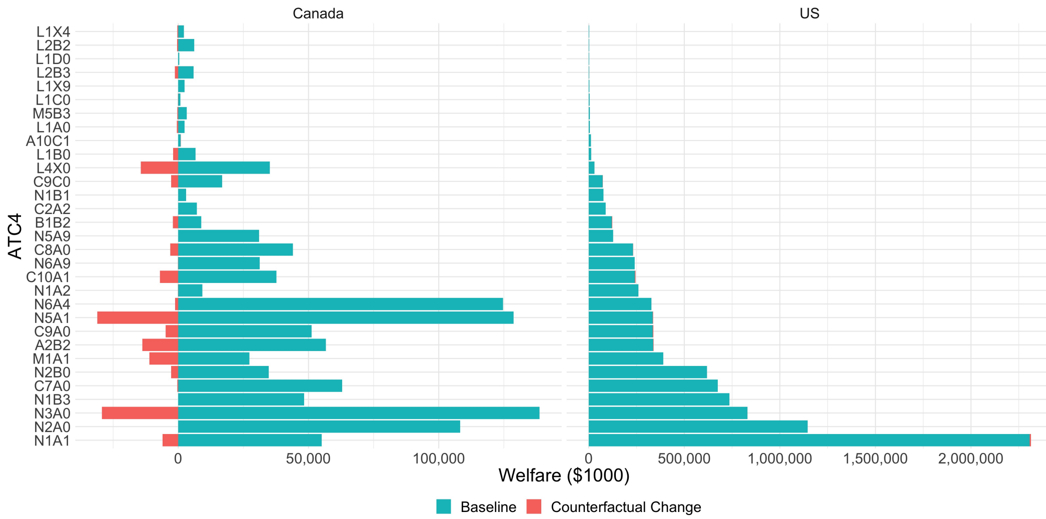
---

[21]It is worth noting that expenditure increases in the United States for a few ATC-4s, suggesting that substitution patterns may dominate small price decreases in these markets.

[22]Firms never choose to exit the Canadian market since given the firm's US price, a sufficiently high Canadian price always exists that generates positive profit and allows the reference price constraint to be satisfied.

[23]The largest is in the anti-rheumatics therapeutic class (M1A1) and combines Diclofenac with Misoprostol. This product represents 4.8% of average expenses in the class, and exits between 2008 and 2011. The other drugs that exit are the ATC-4 classes A10C1, B1B2, and C9A0, and have expenditure shares between 0.1% and 1.5%.

Figure 2.7: *Counterfactual Welfare Changes on All Drugs with Canada as Price Ceiling for the US*

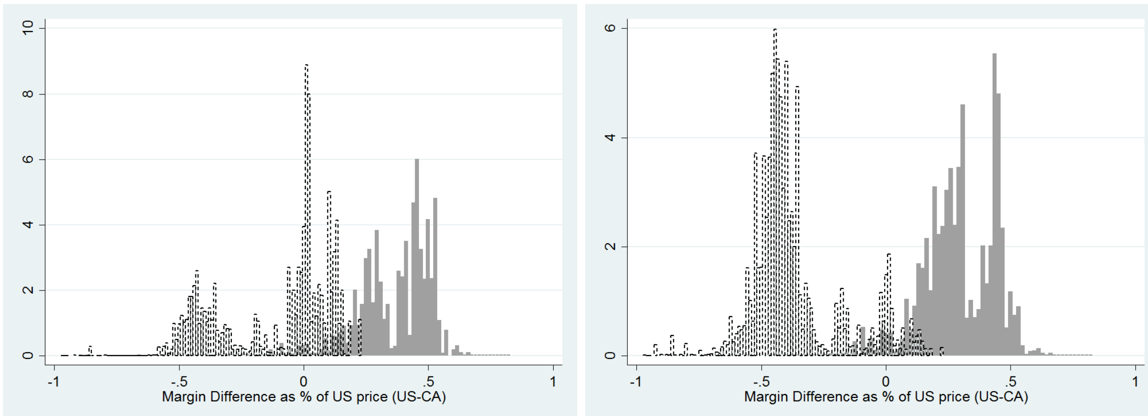


Note: Each blue bar indicates the average annual welfare in it ATC-4 class in the baseline without reference pricing. The red bar indicates the change in average annual welfare resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that welfare increased by the length of the red bar. A red bar to the left of the blue bar indicates that welfare decreased by the length of the red bar.

Analogously to Figure 2.2, Figure 2.8 shows the difference in patented drug margins in the United States and Canada both in the baseline and under the reference pricing counterfactual. This figure shows that the international reference pricing policy results in generally higher margins in Canada than in the United States, the reverse of what we find in the baseline without reference pricing.[24]

[24]The left graph of Figure 2.8 shows that when weighting the distribution by the US quantities of each drug, a significant number on-patent drugs will exhibit higher margins in Canada by an amount around 40% of the price of the drug. The right graph of Figure 2.8 shows that the share of drugs with substantially higher Canadian margins is amplified when weighting by Canadian quantities.

Figure 2.8: *Counterfactual Margins Differences between US and Canada for on Patent Drugs*

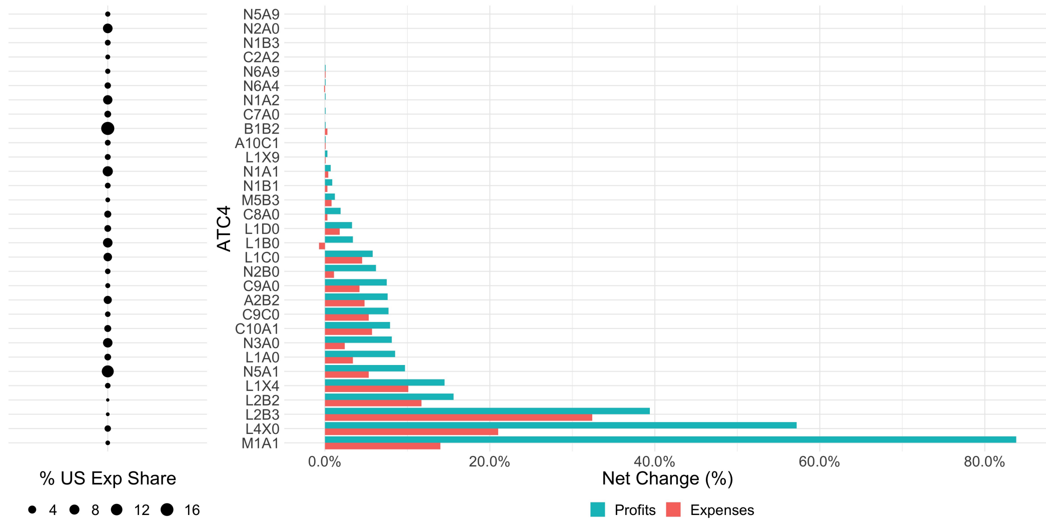


Note: The empirical distribution of the difference between margins in Canada and the US, $(p^{CA} - c^{CA}) - (p^{US} - c^{US})$*, normalized by each drug's US price and weighted by the quantity of the drug sold in the US (left) and in Canada (right).*

These graphs show that if the status quo margins are larger in the US, such that the distribution of differences is largely on the positive, then international reference pricing will not make the distribution of differences centered on zero,. Rather, margins will be higher in Canada for a substantial quantity of on-patent drugs. This occurs despite the fact that prices become close because marginal costs are typically higher in the US than in Canada. That is, international reference pricing policy makes prices more equal across countries but makes margins lower in the US and thus makes the US contribute less than Canada to pharmaceutical profits by unit of consumption.

Finally, figure 2.9 shows the net effect of imposing international reference pricing on global (that is, the US and Canada combined) expenses and profits. Overall, total expenses in the US and Canada increase by 2.7%, while total profits increase by 5.1%. Most of the changes occur in Canada, whose scale is much smaller than the US. However, the price increases in Canada are so large that on net expenses increase. In summary, an international reference pricing policy in the US has globally negative effects on the referenced country, but is not able to substantially decrease either prices or expenses on drugs in the US. Detailed results are presented in table B.11 in appendix B.1.7.

Figure 2.9: *Net Global Percent Changes in Expenses, Profits and Welfare from Reference Pricing*



Note: Each bar represents the net percent change in global expenses (red), profits (green) and welfare (blue) in each ATC-4 that results from moving from the baseline without reference pricing to our main counterfactual. For scale, on the left, we present the percent of baseline US expenditures that each ATC-4 represents.

Variations on Market Size and MFN Rule

As demonstrated in our counterfactual simulations, the international reference pricing policy is likely to have small effects in most ATC-4 classes in the US, with a few notable exceptions. However, it would have generally very large effects in the reference country—in our case, Canada. Our simulation results show that, in general, it would be too costly for pharmaceutical firms to decrease prices in the US. Rather, firms would respond to the policy by increasing prices in Canada—even if regulations in Canada can impose some downward pressure on price-setting in Canada. In the 31 ATC-4 classes of drugs, total spending in Canada is $472 million on average annually, while it is $ 6,946 million in the US. This difference is explained by both the fact that prices are much higher in the US than in Canada, and also because Canada is a much smaller country than the US in population.

Given these results, we investigate several variants of the international reference pricing. The first variant that we consider is a Most Favored Nation clause in which the United States allows pharmaceutical companies to set prices in the US so long as the price in the United

States does not exceed a maximum allowed premium above the price in Canada. In other words, in order for an on-patent drug to be sold in both the United States and Canada:

$$p_j^{US} \leq (1+\eta) p_j^{CA}, \tag{2.13}$$

where η is the maximum allowable premium. Given that Canadian prices are typically lower than in the US, we consider premiums set at 33% and 50%.[25] A policy of this sort would be a priori less stringent on price-setting in the US and Canada, and could result in smaller price changes than when the allowed premiumis 0%.

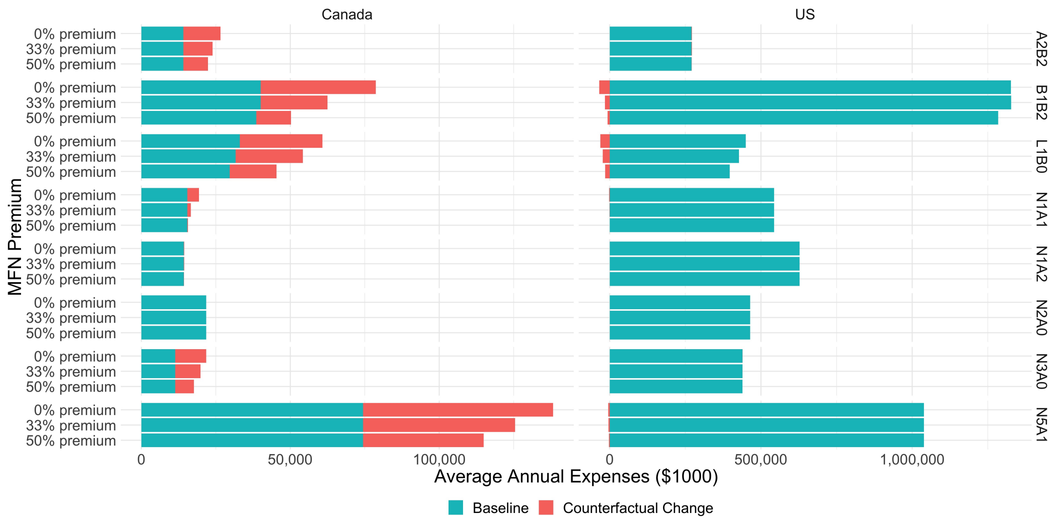
The second policy variant that we consider is reference pricing with respect to a different country, especially one with a larger market. To approximate the implications of referencing a larger country without re-estimating our model for other countries, we simulate the international reference pricing policy in our main counterfactual section with a scaled up market size for Canada, such that it represents half of the US market, or is of the same size as the US market.

Figures 2.10 and 2.11 shows the results of this counterfactual simulation for a subset of ATC-4 classes that each represent more than 3% of pharmaceutical expenditures in the US. Table B.12 in Appendix B.1.7 presents the detailed results. We present the counterfactual expenses in the benchmark case (e.g. referencing Canadian prices), as well as for simulations with marked-up reference pricing (but keeping the Canadian market as is), and simulations with an inflated Canadian reference market (but the baseline reference pricing rule). In this table, the column "MFN", which takes values 0, 33 and 50, refers to simulations in which pricing in the US is referenced with respect to Canadian prices plus a markup of 0%, +33%, and +50% respectively. The column "Share US market", which takes values 0, 50 and 100, refers to simulations in which either the baseline Canadian market size (0), or a Canadian market size that is scaled up to represent 50% or being 100% of the US market, respectively.

We find that allowing for a +33% or +50% markup on reference prices in the US would

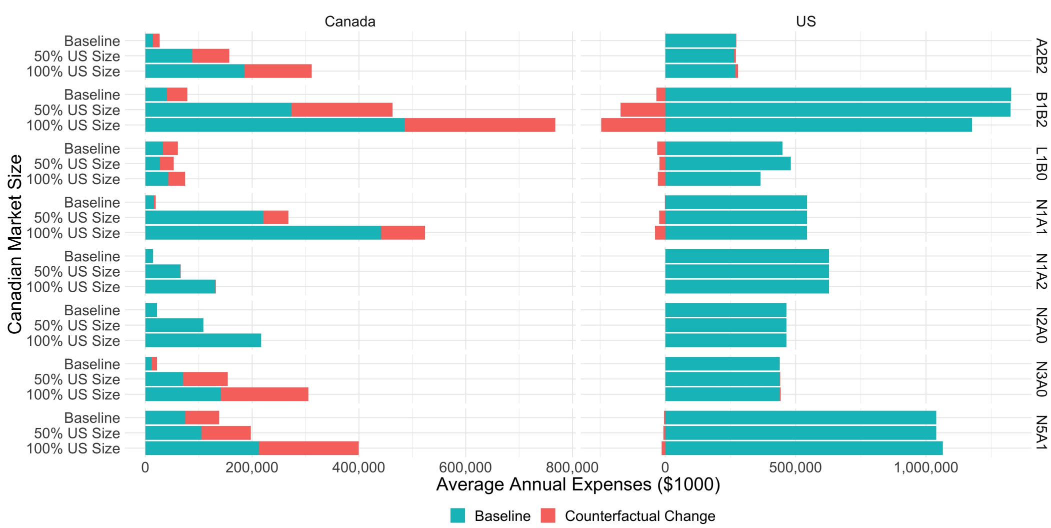
[25]While we only examine weakly positive η, many settings with most favored nation contracts involve negative η. Such contracts guarantee the referencing country (or firm) a better price than others by at least a fixed percentage. These contracts are often referred to as "MFN-plus" contracts.

Figure 2.10: *Counterfactual Expenses Changes in Large ATC-4s with A Varying MFN rule (0, +33%, +50%)*



Note: Each blue bar indicates the average annual expenditure in it ATC-4 class in the baseline without reference pricing. The red bar indicates the change in average annual expenditure resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that expenditure increased by the length of the red bar. A red bar to the left of the blue bar indicates that expenditure decreased by the length of the red bar.

Figure 2.11: *Counterfactual Expenses Changes in Large ATC-4s with a Larger Reference Market*



Note: Each blue bar indicates the average annual expenditure in it ATC-4 class in the baseline without reference pricing. The red bar indicates the change in average annual expenditure resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that expenditure increased by the length of the red bar. A red bar to the left of the blue bar indicates that expenditure decreased by the length of the red bar. There is some variation in convergence for different US sizes.

lead to smaller price increases in Canada as well as similarly small changes in expenses in the US. Again, the simulations demonstrate that when the reference country's market size is relatively small, the reference pricing policy would mostly affect the referenced country, without benefiting the US to a large degree. Our results show that increasing the market size of the referenced country to be comparable to the US—half of or comparable to the US market—implies greater reduction in the US price and a decrease in the Canadian price. In the ATC-4 class B1B2, which covers fractionated heparins (an anticoagulant), for instance, US expenses decrease by 13% and 21% when referencing a Canadian market that is scaled up to be half the size of the US market, and the same size as the US respectively. This suggests that reference pricing may be more effective when referencing larger countries. Nonetheless, expenses in the US do not decrease substantially across ATC-4s, and the large asymmetry in effect size between the US and Canada remains.

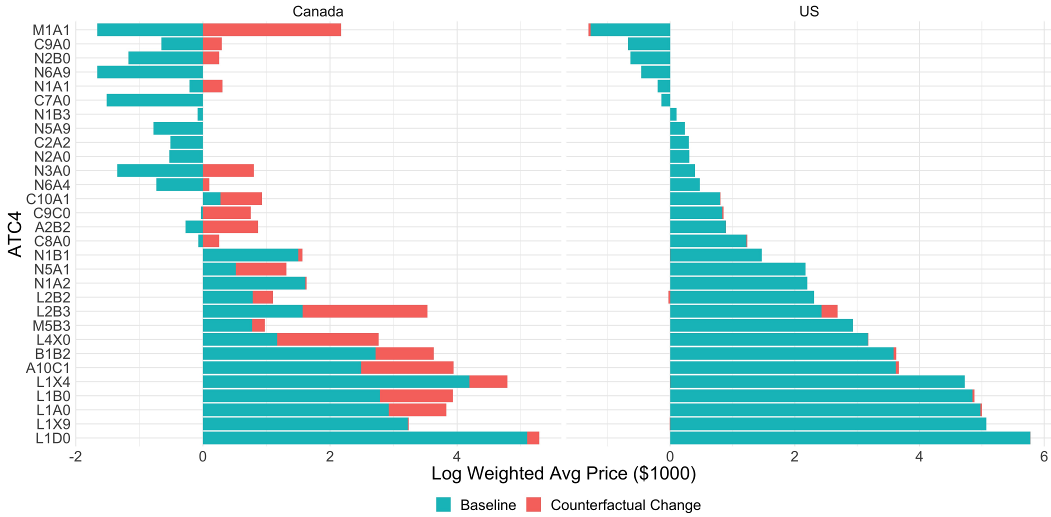
2.5.2 Alternative Specifications

Price Floor

In this section, we simulate the impacts of a reference pricing counterfactual in which the pharmaceutical company is able to commit to a price p_j^{US} in the United States prior to negotiating a price p_j^{CA} with the Canadian regulator. The firm's chosen United States price behaves effectively as a price floor in negotiations with Canada: if the negotiated rate in Canada is lower than the price floor, then the firm is forced to exit the United States market. Negotiations between the pharmaceutical company and the Canadian regulator take this into account so that the firm's agreement surplus and negotiated rate in Canada are:

$$\Delta\Pi_j(p_j^{CA}, p_j^{US}, \mathbf{p}_{-j}^{CA}, \mathbf{p}_{-j}^{US}) \equiv \underbrace{\Pi_j^{US}(p_j^{US}, \mathbf{p}_{-j}^{US})\mathbf{1}_{\{p_j^{CA} \geq p_j^{US}\}} + \Pi_j^{CA}(p_j^{CA}, \mathbf{p}_{-j}^{CA})}_{\text{total profit of } j \text{ if agrees in CA}} - \underbrace{\Pi_j^{US}(p_j^{US}, \mathbf{p}_{-j}^{US})}_{\text{profit of } j \text{ if only in the US}}, \tag{2.14}$$

Figure 2.12: *Counterfactual Prices with US as Price Floor in Canada*



Note: Each blue bar indicates the log of average prices in each ATC-4 class in the baseline without reference pricing. The red bar indicates the change in log average prices resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that prices increased by the length of the red bar. A red bar to the left of the blue bar indicates that prices decreased by the length of the red bar.

$$p_j^{CA}(p_j^{US}, \mathbf{p}_{-j}^{CA}, \mathbf{p}_{-j}^{US}) \equiv \arg\max_p \left(\underbrace{\Delta\Pi_j(p_j^{CA}, p_j^{US}, \mathbf{p}_{-j}^{CA}, \mathbf{p}_{-j}^{US})}_{\text{profit gain from agreement}} \right)^{\rho_j} \left(\underbrace{\Delta_j W_{CA}(p, \mathbf{p}_{-j}^{CA})}_{\text{welfare gain in CA from agreement}} \right)^{1-\rho_j} . \quad (2.15)$$

Strategic pharmaceutical companies will account for the correspondence in (2.15) and set their price in the United States to maximize their global profit:[26]

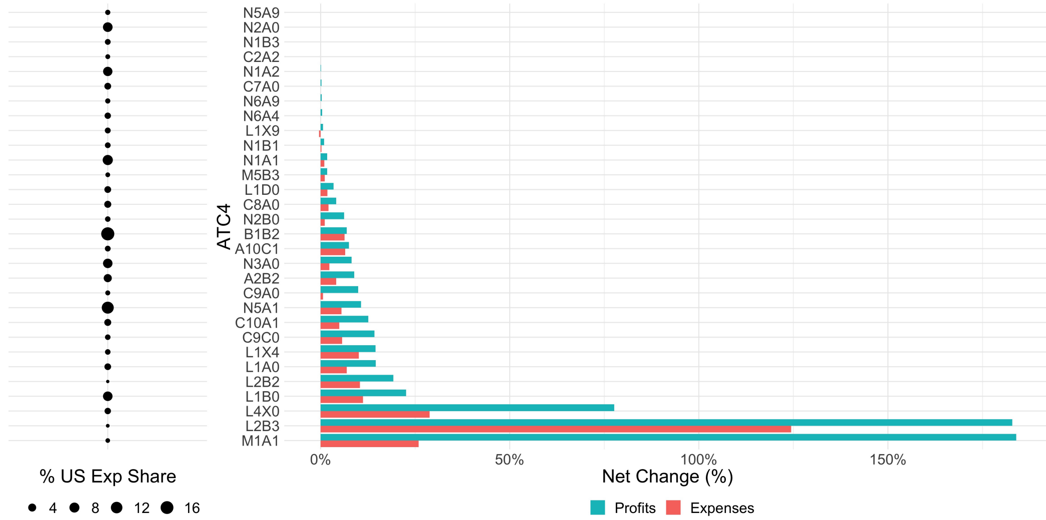
$$p_j^{US}(\mathbf{p}_{-j}^{CA}, \mathbf{p}_{-j}^{US}) = \arg\max_p \quad \Pi_j^{US}\left(p, \mathbf{p}_{-j}^{US}\right) + \Pi_j^{CA}(p_j^{CA}(p, \mathbf{p}_{-j}^{CA}, \mathbf{p}_{-j}^{US}), \mathbf{p}_{-j}^{CA}) \quad (2.16)$$

In equilibrium, the prices $\{(p_j^{US*}, p_j^{CA*})\}_j$ will satisfy (2.9) and (2.11) for each j:

$$\begin{aligned} p_j^{US*} &= p_j^{US}(\mathbf{p}_{-j}^{CA*}, \mathbf{p}_{-j}^{US*}), \\ p_j^{CA*} &= p_j^{CA}(p_j^{US*}, \mathbf{p}_{-j}^{CA*}, \mathbf{p}_{-j}^{US*}). \end{aligned} \quad (2.17)$$

[26]It is also possible that the firm prefers to exit the United States market. We additionally allow the firm to exit the United States when the profit from unrestricted sales in Canada exceed the firm's maximum global profit when serving both markets under the reference pricing rule.

Figure 2.13: *Net Global Percent Changes in Expenses, Profits and Welfare from Reference Pricing with US as Price Floor for Canada*



Note: Each bar represents the net percent change in global expenses (red) and profits (green) in each ATC-4 that results from moving from the baseline without reference pricing to our main counterfactual. For scale, on the left, we present the percent of baseline US expenditures that each ATC-4 represents.

Using our model estimates, we simulate the impact of such a reference pricing rule in which the pharmaceutical company first commits to a price in the United States and then negotiates a price in Canada. We show some of the findings from these simulations in in Figures 2.12 and 2.13. Additional figures and tables can be found in Appendix B.1.7. As in our previous counterfactuals, Figure 2.12 indicates that prices increase when the United States. However, unlike in the previous counterfactuals, it is sometimes the case that prices in the United States increase. Intuitively, this is because the profit maximizing price in Canada may be higher than in the United States. This incentivizes facilities to raise the price in the United States in order to create a price floor in Canada that moves the negotiated rate in Canada closer to the profit maximizing price in Canada. Comparing Figure 2.13 to Figure 2.9 shows that allowing firms to commit to a price in the United States leads to changes in global expenditure amd profits that are larger in magnitude (5.3% and 11.7% versus 2.7% and 5.1% without commitment).

2.6 Conclusion

We employ detailed quantity and price data from IMS Health in our analysis to estimate a random coefficients logit demand model with a structural quality metric for each drug. Under the assumption that prices are set according to Nash bargaining between the country and firm (Horn and Wolinsky, 1988; Crawford and Yurukoglu, 2012; Grennan, 2013; Gowrisankaran *et al.*, 2015) in a regulated price country such as Canada, we are able to separately identify costs and bargaining parameters. Since Nash bargaining involves maximizing the weighted log-sum of both parties' transaction utility, we can interpret the bargaining parameters as the degree to which countries' policymakers choose to trade off between firm profits and immediate consumer welfare. We then perform counterfactual simulations of a most favored nation policy in the US involving international reference pricing constraints from other markets. We develop two possible models of international reference pricing that differ according to which degree of ax ante commitment the international reference pricing can have. Without ex ante commitment of the international reference pricing rule in the US, the Canadian prices would serve as price ceilings for the same drugs sold in the US. With ex ante commitment, the US prices would serve as the price floor for the prices of the same drugs in Canada. In both cases, although with some slight and interesting variations across drug classes, we find that such policy would decrease prices slightly in the US but increase them dramatically in Canada because firms will internalize the across-country restrictions involved by the US reference pricing. We find that expenses on pharmaceuticals would increase considerably in Canada but not change significantly in the US. When comparing margins of on-patent drugs present in Canada and the US, we find that while the distribution of margins differences between the US and Canada is currently skewed towards higher margins in the US, the international reference pricing policy would skew this difference towards higher margins in Canada, while prices would be close because the US would not pay over Canada for it higher marginal costs. The effects on profit and welfare show that profits of firms would increase significantly in Canada while consumer welfare would decrease, and the effects in the US remain small. Overall, we find modest consumer welfare gains in the US, but substantial consumer welfare losses in Canada. Moreover, we find that pharmaceutical profits increase in net, suggesting that reference pricing of this form would constitute a net transfer from consumers to firms. Some variants of the simulations show that one would need a much larger reference market for this policy to have significant price reduction effects in the US.

Chapter 3

Buying Data from Consumers: The Impact of Monitoring Programs in U.S. Auto Insurance[1]

Introduction

Innovations in technology and regulation have facilitated the collection, analysis and commercialization of data consumer behavior led to a proliferation of direct transactions of consumer data. Firms directly incentivize consumers to voluntarily reveal information, while keeping the collected data as proprietary. How does this type of data collection influence social surplus and its division among firms and different types of consumers?

In this paper, we develop an empirical framework to quantify the welfare and profit impact of an auto-insurance *monitoring program* ("pay-how-you-drive") in the U.S., a prominent example of direct transactions of consumer data. New customers are invited to plug a simple device into their cars, which tracks and reports their driving behavior for up to six

[1]Co-authored with Yizhou Jin

months.[2,3] In exchange, the insurer uses the data to better assess accident risk and adjusts future premiums accordingly. Unlike most traditional pricing factors such as age or claim history, monitoring data is not shared with other firms.

In 2017, insurers serving over 60% of the $267 billion U.S. auto insurance industry offered monitoring programs.[4] Similar programs have been introduced in other industries, such as life insurance and lending.[5] Despite this growing relevance, empirical evidence on the economic impact of monitoring programs or other types of direct transactions of consumer data is sparse.

We acquire proprietary datasets from a major U.S. auto insurer that detail drivers' characteristics, the price menu they face, insurance contracts purchased, and realized insurance claims. A monitoring program is introduced during our research window. For each driver who opts in, we observe a monitoring score that the firm uses in determining premium adjustments. To understand competition, we further match each observation with price menus of the firm's main competitors in each state. Taken together, our analysis uses a panel dataset of over 1 million drivers and 50 million insurance quotes.

Our analysis overcomes three main challenges. First, we quantify the degree to which monitoring can both incentivize safer driving and reveal drivers' risk types, and in doing so, the ability for the monitoring program to improve social surplus Akerlof (1970); Fudenberg and Villas-Boas (2006); Einav, Liran and Finkelstein, Amy and Schrimpf, Paul (2010). Second, we estimate a model of consumer demand to capture complex correlations between drivers' choices of monitoring, insurance coverage, and insurer, as well as the cost to insure them. Third, as the data is proprietary, firms can raise markups and their share of the social surplus. But they also incur costs to "produce the data in the first place" Posner (1978). We

[2]See Figure C.1 in the appendix.

[3]See Bordhoff and Noel (2008); Reimers and Shiller (2018) for analyses of monitoring using aggregate data.

[4]According to 2017 data published by the National Association of Insurance Commissioners.

[5]The Vitality program from life insurer John Hancock tracks and rewards exercise and health-related behaviors. Ant Financial incentivizes users to conduct more personal finance transactions in exchange for borrowing discounts.

model dynamic and multi-product considerations in firm pricing that underpin both factors. This allows us to simulate counterfactual equilibria that endogenize the firm's information set and offer new insights into regulatory proposals that curb proprietary data.[6]

We find three main results: (i) Data collection changes consumer behavior. Drivers become 30% safer when monitored, which boosts total surplus and alters the informativeness of the data. (ii) Safer drivers are more likely to opt in. But monitoring take-up is low due to both consumers' innate preference against being monitored and attractive outside options from other insurers. Nonetheless, compared to a counterfactual with no monitoring, consumer welfare and total surplus both increase. (iii) Proprietary data facilitate higher markups and raise the firm's share of the surplus created by monitoring. But they also protect the firm's incentives to produce the data in the first place. A counterfactual equilibrium in which the firm must share monitoring data with competitors harms both profit and consumer welfare. This is because the firm offers smaller upfront incentives for monitoring opt-in, so that fewer drivers are monitored in equilibrium.

Our empirical analysis starts with a pair of reduced-form facts. The first one shows that drivers become safer when monitored – an incentive effect. The monitoring program is only offered to new customers and ends within the first six-month period. We therefore directly compare claim rates of the same monitored drivers during and after monitoring. A difference-in-differences estimator is used in which the control group consists of unmonitored drivers. Taking into account additional variation in monitoring duration, we find that the average opt-in driver *becomes* 30% safer when monitored. Our estimates are robust to various control specifications. We also conduct a test for parallel trends in periods after monitoring ends.

Despite the behavioral distortion, we document that monitoring data still captures substantial differences in drivers' risk types that are previously unobserved. This may

[6]The General Data Protection Regulation (2016) in the EU aims to curb the accumulation of proprietary data by allowing consumers to rescind consent and take their data to other firms, and by requiring firms to be transparent about how consumer data is used in pricing (see here). The National Telecommunications and Information Administration in the U.S. is considering similar regulatory proposals (see press release here).

lead to adverse selection into higher insurance coverage and advantageous selection into monitoring. We look at cost differences across monitoring groups conditional on observables. Monitored drivers who score one standard deviation above the mean are 29% riskier in the subsequent (unmonitored) period. Further, the within-driver risk reduction we measured above only explains 64% of the risk difference across monitored and unmonitored groups in the first period.[7]

With the reduced-form facts in mind, we develop a cost model for consumer (claim) risk and the monitoring program. Each driver has a latent risk type that partially depends on his or her observables. This risk type can change when the driver is monitored. Meanwhile, new customers can choose to be monitored during the first period. Doing so sends an informative signal of their risk types exclusively to the monitoring firm.

Consumers' monitoring opt-in choice is more complex and captures the following intuition. First, drivers may anticipate risk reduction during monitoring. Second, drivers form expectations over potential renewal discount (from monitoring) based on their risk types. Safer drivers may therefore be more likely to opt in. But the monitoring signal is noisy, which adds to drivers' uncertainty over their future premiums and deter risk-averse drivers from opting in (reclassification risk). Lastly, drivers need to actively opt into monitoring and may incur privacy or effort costs. They therefore suffer disutility from being monitored.

We develop a demand model that features key parameters that drive the intuition above and link consumers' monitoring opt-in decision with their choices of insurer and insurance coverage as well as the cost of insuring them. We start from an insurance framework Einav *et al.* (2010) that features risk preference, heterogeneous inertia costs, expected renewal premium, as well as the latent risk type from the cost model. We then parameterize consumers' disutility from being monitored as a random effect that varies based on both observables and unobserved latent risk type. Parameters in the cost and monitoring models are identified based on variance and covariance of claims and monitoring scores

[7] Opt-in drivers are only monitored for fractions of the first six-month period, so the incentive effect (within-driver risk changes across periods) is only 23% in the data as opposed to the full 30% outlined in the above paragraph.

conditional on observables. Identification of demand parameters relies on rich variation in prices and contract space conditional on observables used in the firm's pricing rules. For example, attrition rates under different competitive pricing environments allow us to estimate consumers' inertia in switching firms. Eligibility for monitoring also depends on location and time. This, in addition to variation in the monitoring opt-in discounts, helps us pin down consumers' monitoring disutility.

To facilitate estimation, we augment the demand model to admit a mixed logit structure and use a simulated maximum likelihood approach Train (2009). Our estimates produce a close fit to the empirical distribution of monitoring scores among monitored drivers, which is endogenously generated based on drivers' monitoring opt-in choices. We further cross-validate our demand model on a hold-out dataset in which the mandatory minimum coverage changed in one (U.S.) state. The model accurately predicts changes in monitoring opt-in rate, coverage share, and attrition rate from the firm.

Our demand estimates show that the average driver suffers a $93 disutility from being monitored. However, monitoring disutility is lower for safer drivers (lower risk type). This means that conditional on the objective financial rewards and risk from monitoring, safer drivers are yet more likely to opt in, which exacerbates advantageous selection into monitoring. Meanwhile, the average driver forgoes $284 financial gain per year from not exploiting outside options from competitors. Further, drivers are only modestly risk-averse in their auto insurance choices. Improving the monitoring score's signal precision therefore has little impact on monitoring demand. Our cost estimates are consistent with the reduced-form findings above.

We then conduct several counterfactual simulations. The first one compares the current regime with a counterfactual with no monitoring, holding fixed baseline prices.[8] Introducing monitoring raises both firm profit (by $7.9 per driver annually, a 23.6% increase) and consumer welfare (by $11.6, in certainty equivalent, or 1.5% of premium). Total surplus increases by $13.3 (1.7% of premium), 64% of which can be attributed to the risk reduction

[8]The firm did not raise prices for unmonitored drivers when introducing monitoring (Appendix C.0.2).

during monitoring. In contrast, although monitoring strongly mitigates information asymmetry, allocative efficiency gain is suppressed due to mandatory purchase of auto insurance and large preexisting competitive price variation.[9]

Next, we propose a pricing model that endogenizes the production of monitoring data and therefore the firm's information set. This is used to derive market equilibrium (i) when the firm optimizes prices without constraint, and (ii) when the firm must share its proprietary data with competitors. In the data, firm prices are likely to be sub-optimal due to regulatory constraints. But the pricing levers used imply that optimal pricing balances two motives: "investing" in data production and "harvesting" from the collected data. The latter receives far more attention from the literature: proprietary data facilitate higher markups and raise the firm's share of the surplus created by monitoring.[10] In our equilibrium simulation, we find that the firm reaches optimal pricing by reducing rent-sharing with consumers by 19.6%. This creates a flatter discount-surcharge schedule, representing more aggressive price discrimination. But the firm must first produce monitoring data. To do so, it can offer opt-in discounts or surcharge the unmonitored pool. Without competition, the firm can use the latter to force drivers into monitoring because auto insurance is mandatory. In contrast, the optimal pricing includes a surcharge of only 2.7% on the unmonitored pool. Price competition therefore effectively limits the firm's ability to coerce drivers into monitoring. Instead, the firm should raise the monitoring opt-in discount to 22.1% from 5%. This benefits the firm by producing monitoring data and simultaneously reducing risk. But it also represents a significant "investment" in the production of monitoring data.

Lastly, we endogenize competitor prices and explore the equilibrium implications of a regulation that requires the firm to share monitoring data with competitors. This turns monitoring into a public good. However, monitoring can still benefit the firm through risk reduction (the incentive effect) and high firm-switching costs (imperfect competition).

---

[9] A large literature focuses on the impact of risk classification on insurance allocation and consumer welfare. Examples include Crocker and Snow (1986); Finkelstein *et al.* (2009); Handel *et al.* (2015).

[10] If a driver is priced at $100 by all insurers but is revealed to be 30% safer through monitoring, then the firm can offer a discount far lower than $30 and still be confident about retaining her.

Nonetheless, we find that the firm significantly scales back investment in the program by reducing the incentives it offers for monitoring opt-in. Compared to the equilibrium without the information sharing mandate, this leads to a large drop in the opt-in rate for monitoring. Although the firm charges lower markups on monitored drivers and thus takes a lower share of the surplus created by monitoring, consumer welfare and total surplus both decrease.

Related Literature Our research contributes to several literatures. First, we extend the empirical literature on information asymmetry and selection markets. We investigate firms' strategy to acquire – and consumers' willingness to reveal – risk information. In doing so, we establish a novel channel through which firms can unilaterally improve the information environment it faces while enhancing its market power. Many studies have quantified the impact of regulations that exogenously change the availability of public information.[11] In fact, firms' information set is often assumed to be exogenous when considering firm strategy and competition. Mahoney and Weyl (2017) posit that market power may further depress quantity under adverse selection, reducing total surplus.[12] Our paper expands this point by endogenizing the production of proprietary consumer data that creates market power but mitigates adverse selection. Meanwhile, our study is among the first to link demand factors driving consumers' willingness to reveal information to those driving product demand as well as price competition.[13]

[11]Government mandates on community-rating (limits to risk categorization) are most common. See Finkelstein *et al.* (2009); Einav, Liran and Finkelstein, Amy and Schrimpf, Paul (2010); Agarwal *et al.* (2015); Cox (2017); Nelson (2018).

[12]Crawford *et al.* (2018) provide empirical evidence in the Italian small-business lending market. See also Einav, Liran and Levin, Jonathan and Jenkins, Mark (2012); Hendren (2013); Veiga and Weyl (2016) for how firms screen consumers using pricing, rejection, and contract levers. Endogenous product differentiation and vertical restraints are well studied in the U.S. health insurance setting Dafny (2010); Ho, Kate and Pakes, Ariel (2014); Shepard (2014); Ho and Lee (2017); Tebaldi (2017).

[13]See Cohen, Alma and Einav, Liran (2007); Fang *et al.* (2008); Barseghyan *et al.* (2013); Bai (2018); Handel (2013); Handel, Benjamin R and Kolstad, Jonathan T and Spinnewijn, Johannes (forthcoming); Handel, Benjamin R and Kolstad, Jonathan T (2015) for various product market demand factors that can be relevant for monitoring opt-in. We also show that consumers are forward-looking. See Hendel, Igal and Lizzeri, Alessandro (2003); Handel *et al.* (2015); Aron Dine *et al.* (2015) for studies on reclassification risk in health insurance.

Second, we contribute to the literature on dynamic contracting. Monitoring facilitates *asymmetric learning* on consumer risk type for the firm Cohen (2012); Hendel (2017),[14] in which the monitoring discount (or surcharge) is a form of voluntary renegotiation after the firm–and not its competitors–learns more about consumers' risk type Hart (1983); Dewatripont and Maskin (1990). We demonstrate that this contract structure has important implications on incentive provision (reducing consumer risk) and markups. A related theory literature focuses on price discrimination enabled by consumers' online purchase histories.[15] Most studies point out incentive distortion and ambiguous impact on market efficiency. Empirically, Hubbard (2000) studies required monitoring in labor contracts for truck drivers. He finds evidence that effort is highly modifiable. Managers can also allocate jobs more efficiently. Auto insurance is much less valuable than labor contracts. Yet drivers still respond to expected future incentives. We also account for additional forces associated with a voluntary mechanism and price competition. In general, empirical studies on dynamic contracting (without commitment) are sparse but relevant in many real-world settings. For example, Nevo *et al.* (2016) study usage-based pricing in the residential broadband market. Our framework can be easily modified to account for the three-part tariff in their setting.

Third, our research directly relates to the literature on the economics of privacy. Here, privacy concerns the efficient ownership of socially valuable information Posner, Richard A (1981); Stigler (1980); Hermalin and Katz (2006). We first estimate consumers' willingness to be monitored in an opt-in mechanism. In particular, although safer drivers are more likely to opt into monitoring, this advantageous selection does not lead to unraveling.[16] Second, we point out that product market competition strongly influences consumers' outside options when deciding whether to reveal their information. Furthermore, we emphasize and empirically validate an argument in the literature that has received little

[14]See also Rajan (1992)Nilssen2000Thadden2004DeGaridelThoron2012.

[15]See Taylor (2004); Acquisti, Alessandro and Varian, Hal R (2005); Fudenberg and Villas-Boas (2006); Stole (2007).

[16]This is similar to quality disclosure by firms as studied by Jin, Ginger Zhe and Leslie, Phillip (2003); Jin (2005); Dranove and Jin (2010).

attention: proprietary data is a form of privacy (or ownership) given to the data-collecting firm. Posner (1978), in particular, points out the importance for the government to protect firms' property right to their data and be mindful of firms' incentives to produce socially valuable information. Similarly, our study also extends a growing empirical literature on information technology and privacy regulation.[17]

The rest of the paper proceeds as follows. Section I describes the data and provides background information on auto insurance and the monitoring program we study. Section II conducts reduced-form tests that measure monitoring's ability to reduce risk and mitigate information asymmetry. Section III presents our structural model and identification arguments as well as the estimation procedure to recover key demand and cost parameters. Section VI discusses estimation results and counterfactual simulation procedures for welfare analyses. Section V proposes a model of monitoring pricing and investigates equilibrium implications of optimal pricing and information sharing. Section VI concludes.

3.1 Background and Data

In this section, we provide background information on U.S. auto insurance and the monitoring program we study. We also describe our datasets.

3.1.1 Auto Insurance

Auto insurers in the U.S. collected $267 billion dollars of premiums in 2017.[18] There are two main categories of insurance: liability and property. Property insurance covers damage to one's own car in an accident, regardless of fault. Liability insurance covers injury and property liability associated with an at-fault accident. In all states we study, liability

[17] Goldfarb and Tucker (2011) study efficiency impact. Acquisti *et al.* (2013); Acquisti, Alessandro and Taylor, Curtis and Wagman, Liad (2016a); Jin, Ginger Zhe (2018) study factors that drive consumers' willingness to reveal information.

[18] This is according to the National Association of Insurance Commissioners. This number is calculated as premiums from property annual statements plus state funds.

insurance is mandatory, the required coverage ranging from $25,000 to $100,000.[19]

Insurance prices are heavily regulated. Major insurers collect large amount of consumer information in risk-rating, most of which is public or shared across firms. Firms are required to publish filings that detail their pricing algorithms. In most states, the insurance commissioner needs to approve such filings.[20] An important focus of the regulator is deterring excessive price discrimination based on demand elasticity.[21] In general, a pricing rule can be summarized by the following structure, where price p for a (single-driver-single-vehicle) policy choosing certain liability coverage (limit) is:[22]

$$\begin{aligned} p = & \text{ base rate} \times \text{driver factor} \times \text{vehicle factor} \times \text{location factor} \\ & \times \text{tier factor} \times \text{coverage factor} + \text{markups and fees} \end{aligned} \tag{3.1}$$

Within each firm, price variation is based on observable characteristics and choices. Base rates vary only by state and calendar time. Driver, vehicle, and location factors include age, vehicle model, and zipcode-level population density, etc. This information is often verified and cross-referenced among various public or industry databases. Tier factors incorporate information from claim and credit databases, which include accident, traffic violation (DUI, speeding, etc.), or financial (delinquency, bankruptcy, etc.) records in the past[23]. Conditional on the factors above, choosing a higher coverage (liability limits) scales prices by a positive factor. Lastly, firms charge a fee that includes markups and overhead for operational and marketing expenditures.[24]

---

[19] All states that we study follow an "at-fault" tort system and mandate liability insurance. In reality, liability insurance is specified by three coverage limits. For example, 20/40/10 means that, in an accident, the insurer covers liability for bodily injuries up to $40,000 overall, but no more than $20,000 per victim; it also covers liability for property damage (cars or other infrastructure) for up to $10,000. We quote the highest number here.

[20] Some states follow a "use-and-file" system, which means that insurers can seek pricing approval ex-post as long as any price changes are reflected in public filings.

[21] "Price optimization" on top of risk rating is typically not allowed by state insurance commissioners.

[22] See Appendix C.0.4, e.g. Figure C.14.

[23] See Appendix Figures C.20 and C.21)

[24] The latter is often referred to as the loading factor in the literature.

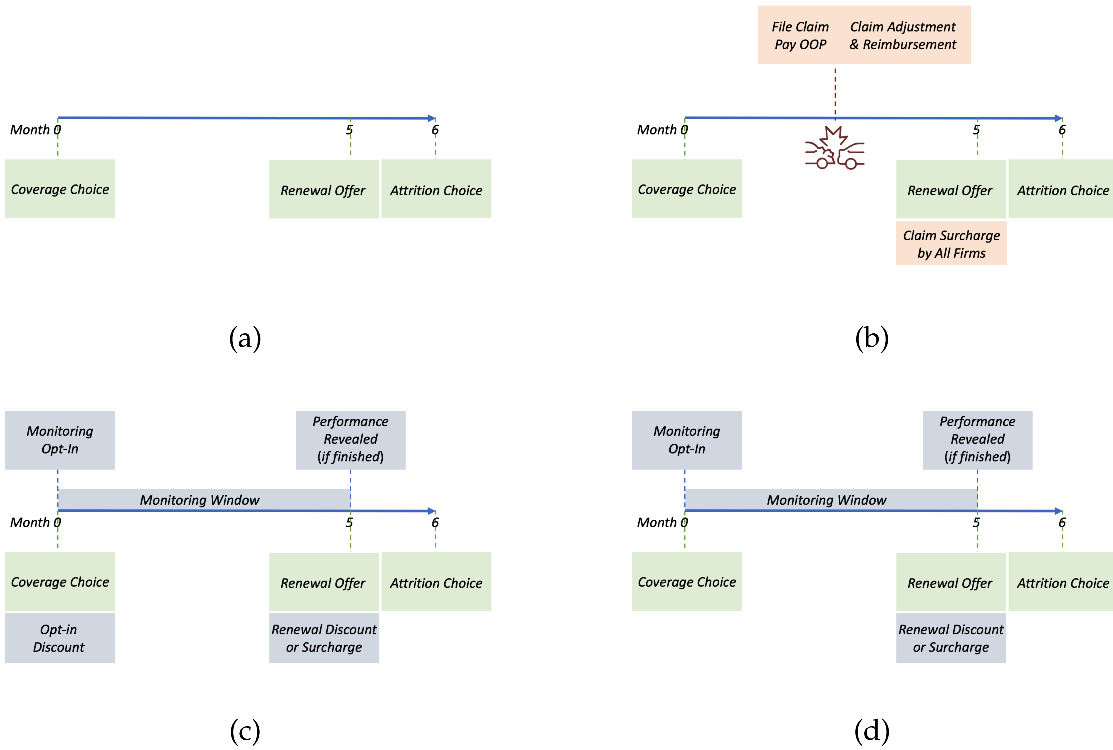


Figure 3.1: *Auto Insurance and Monitoring*

A typical period for new customers is summarized in 3.1a. At time $t = 0$, new customers arriving at the firm are required to report observable characteristics. This information facilitates risk rating, based on which the firm generates individualized price menu. Consumers can take one of the coverage options offered or go to other firms.

There is no long-term commitment in U.S. auto insurance. In our setting, each period lasts for six months. At the end of month five, firms provide their customers with renewal quotes. Drivers decide whether to renew at the end of month six. During the policy (six-month) period, if an auto accident occurs (3.1b), the insured files a claim immediately and, depending on the claim type, pays some costs out-of-pocket. Insurance adjusters will then evaluate the accident and determine reimbursement and pay-out. As soon as a claim is filed, this information is recorded in industry databases in real time. As a result, the driver will likely face a claim surcharge renewing at the current firm or switching to other firms.

Dataset 1 - Panel data from an auto insurer Our first dataset comes from a national auto insurer in the U.S. that offers a large monitoring program. It is a panel that spans 2012 to 2016, and covers 22 states. For tractability, we narrow the scope of our analyses to *single-driver-single-vehicle* insurance policies sold online or via phone. Nonetheless, we observe more than 1 million drivers, for an average duration of 1.86 years (3.73 periods)[25]. The date range spans periods pre- and post-introduction of monitoring.

At the beginning of each period, we observe each driver's observable characteristics as well as the price menu offered, which includes all available options offered by the firm and their prices. We also see the driver's coverage choice. For simplicity, we limit our attention to *liability coverage* (limits). Not only is liability the most expensive type of coverage (for the average driver), but its mandatory nature also strongly influences firms' competitive strategy and consequently, the allocative benefit that monitoring provides. Moreover, liability covers auto accidents involving two or more parties, in which the policy holder is at least partially at-fault. As such, our focus also mitigates concerns about under-reporting.[26] At renewal, drivers who have filed a claim experience a surcharge on their premium that ranges from 10% to 50%.[27] Absent a claim, however, the average driver experiences close to no price change in a typical renewal period. Overall, around 5% to 20% of drivers leave the firm after each period.[28] Table 3.1 presents summary statistics of prices, coverage levels, and claims. It also lists key observable variables. The average driver is 33 years old, drives a 2006 vehicle, lives in a zipcode area with average annual income of $142,000, and had 0.3 at-fault accidents in the past 5 years. Per six-month period, he pays $380 in liability premium and files 0.05 liability claims. We also observe his assigned risk class, which is the premium

[25]The panel is right-censored, but the censoring is plausibly uninformative.

[26]In contrast, claim filing for single-car accidents is almost entirely discretionary.

[27]See Figure C.12 in the appendix for the empirical distribution of surcharges added after the occurrence of an at-fault claim. Variation in the level of the surcharge depends only on drivers' existing claim and traffic violation records, which is summarized in points.

[28]The first renewal is somewhat different, as some one-time discounts are removed. These are mostly cost-based discounts, such as e-signature or online quoting discounts. It therefore sees a higher attrition than subsequent ones.

calculated for him before coverage factor and markups and fees.

Statistic	Mean	St. Dev.	Min	Median	Max
Total Premium (6-month)	631.50	364.02	69	548	22,544
Liability Premium	379.95	208.23	32.00	335.88	10,177
Risk Class	254.73	172.22	50.00	212.23	9,724
Liability Coverage ('000)	126.16	118.86	25	60	500
Mandatory Minimum Ind.	0.36	0.48	0	0	1
Liability Coverage Ranking	2.10	1.15	1	2	8
Renewal Count	1.76	2.01	0	1	9
Calendar Month	6.25	3.43	1	6	12
Calendar Year	2.66	1.38	0	3	5
Number of Drivers	1	0	1	1	1
Number of Vehicles	1	0	1	1	1
Claim (6-month)	323.47	2,821.78	0	0	544,814
Liability Claim	164.49	2,209.17	0	0	513,311
Claim Count	0.18	0.67	0	0	12
Liability Claim Count	0.05	0.32	0	0	7

Note: This table reports summary statistics of our main panel data. Risk class is defined as the net premium calculated for each policy's liability coverage before markups and fees. Main observable characteristics include indicators for gender, age thresholds of 18, 25 and 60, a college degree, a post-graduate degree, credit and credit report availability, home-ownership, garage verification, having an out-of-state license, having a leased vehicle, having a "Class C" vehicle, having an BAS, having a safe device installed, having a clean record, and having (some) prior insurance. In addition, observables includes driver age, years of education, years of license, years of prior insurance, credit tier, local population density, vehicle age, vehicle replacement value, length of vehicle ownership, number of accident "points", number of at-fault accidents and number of DUIs. See Table C.2 for a list of observables used in our estimation procedure.

Table 3.1: *Summary Statistics of Observables, Premium, Coverage and Claims*

Dataset 2 - Price menus of competitors based on price filings To understand competition, we need to account for drivers' outside options. To do this, we complement our main dataset with competitors' price menus that drivers face when making insurance and monitoring choices. This additional dataset includes quotes from all liability coverage options offered by the firm's top five competitors in each state. As noted above, this information is contained in publicly available price filings, and we retrieve it using Quadrant Information Services' reputable proprietary software. Each observation in our first dataset is matched with the competitive price menu the driver faces at the time of choice. We are able to achieve precise matches based on main observable characteristics, including state and calendar time.[29]

In Table 3.2, we compare the quotes for the five most common liability coverage options across competitors for all drivers from our main dataset in one large U.S. state. We report average quotes because different states have different sets of coverage options. In this state, the mandatory minimum changed from $40,000 to $50,000 within our dataset.The NA ratio calculated the portion of plan that cannot be rated. This is mostly due to the mandatory minimum increase, as well as location-based rejection. The bottom panel reports summary statistics of claim variables. While claim coverage is similar, there is substantial variation in premiums across competitors, stemming from differences in overhead costs, company strategy, and estimated loss ratios.

Looking ahead, observing competitor prices is instrumental to identifying parameters such as consumers' inertia to switch firms based on observed attrition choices in renewal periods. In counterfactual analyses, competitive prices can also help us enumerate our sample of new customers of the firm to the full market. Our ability to do so is further enhanced by prior insurance records.[30] On average, 48% new customers switched from another firm, about half from one of the top five competitors. We default the other switchers into the largest insurer of each state. 33% of new customers are previously uninsured (including new drivers), and 19% have a rewritten policy (by far the most common reason being an out-of-state move).

3.1.2 Monitoring Program

Our research focuses on the firm's one-time monitoring program for new customers.[31] The monitoring process is summarized in Figures 3.1c and 3.1d. When customers arrive, they

[29]We match based on variables in Table 3.1, as well as other traffic violation records, zip-code, vehicle make and model.

[30]This is part of tier information and is verified by the firm. It carries significant pricing weight.

[31]The firm has also offered a continuous monitoring program, but not during our research period.

Liability Coverage Limits	40	50	100	300	500
Quotes	335.14	343.43	382.03	422.13	500.48
- Competitor 1	482.68	506.11	564.34	626.81	730.56
- Competitor 2	263.14	279.15	314.46	347.69	405.22
- Competitor 3	319.42	348.97	388.48	428.64	464.36
- Competitor 4	511.24	567.58	613.74	682.87	790.83
- Competitor 5	421.84	363.96	403.64	433.17	497.79
NA Ratio	0.61	0.00	0.00	0.00	0.00
- Competitor 1	0.62	0.00	0.00	0.00	0.00
- Competitor 2	0.62	0.00	0.00	0.00	0.00
- Competitor 3	0.62	0.00	0.00	0.00	0.00
- Competitor 4	0.61	0.01	0.01	0.01	0.01
- Competitor 5	0.62	0.00	0.00	0.00	0.00
Claim	256.87	285.27	306.68	297.73	293.96
	(9.15)	(7.10)	(11.72)	(15.04)	(46.80)
Liability Claim	154.98	155.54	154.16	143.43	107.54
	(7.31)	(5.31)	(8.89)	(12.56)	(23.83)
Claim Count	0.09	0.10	0.10	0.10	0.09
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Liability Claim Count	0.05	0.05	0.04	0.03	0.03
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Share within Firm	0.19	0.39	0.20	0.19	0.03

Note: This table reports summary statistics of average quotes by liability coverage for our our firm as well as the top 5 competitors in one US state.

Table 3.2: *Summary Statistics by Coverage*

choose whether to opt into monitoring right before seeing their price menu. Before doing so, customers are provided with information on the kinds of driving behavior that are tracked and rewarded. Specifically, high mileage driven, driving at night, high speed, and hard braking are highlighted as monitored behaviors, through which good performance is promised to yield discounts. However, the exact discount schedule is opaque. Across several monitoring programs offered by large U.S. auto insurers, drivers can expect a renewal discount of up to 20-50%. They can also receive a surcharge of up to 5-20% for poor performance. In some states and calendar times, drivers are given an up-front discount for opting into monitoring, ranging from 1 to 20%.

Should a driver opt into monitoring, a monitoring device is mailed within the next week. She then has until the end of month five to accumulate around 100-150 days of monitored driving. If completed, the firm will evaluate her performance and include an appropriate renewal discount when giving out renewal quotes. In the case of an accident, monitoring data is not used in claim adjustment or reporting. Monitoring continues after any disruptions from the accident.

27% of drivers who start monitoring do not finish. Our main analysis ignores these drivers and focuses on analyzing consumers' decision to start and finish monitoring. Non-finishers likely have incorrect beliefs about the potential discounts they get or the costs they incur from monitoring. Most drop out during a two-month grace period in which they are allowed learn about the monitoring program and their own risk, without penalty.[32] Our analysis therefore does not account for the costs and benefits associated with this learning process.

During the monitoring period, monitored drivers receive real-time feedback on their performance. Different monitoring programs have different methods of communication. Insurers often post daily summaries of key statistics on recorded trips online and via mobile apps, particularly on the highlighted behaviors mentioned above. Insurers may also offer more active reminders: some send text messages or mobile app notifications, while others design their monitoring devices to beep whenever they record a hard brake.

Nevertheless, monitoring data is *proprietary* information to the firm that administers the program. Firms face both practical and regulatory hurdles in rating monitoring information from another firm. First, it is hard to verify a customer's claim that she has gotten certain monitoring results from another firm. Even if verified, each firm's monitoring program and preexisting risk algorithm are idiosyncratic. Therefore, it is very difficult for an insurer to determine and publicly file a discount for a driver who has received a monitoring score in nominal terms from a competitor's program.[33] Furthermore, according to the privacy

[32]Drivers can drop out of the monitoring program for the first two months without penalty. Afterwards, dropping out results in the maximum amount of renewal surcharge.

[33]See Figure C.22 in the appendix for an example of an insurer's rate filing, including monitoring. Discerning the monetary value of a monitoring score using this type of filing is near impossible.

policy and usage terms agreed to when opting into monitoring, no personally identifiable data can be resold. It is no surprise, therefore, that each firm only prices based on its own monitoring information according to price filings.

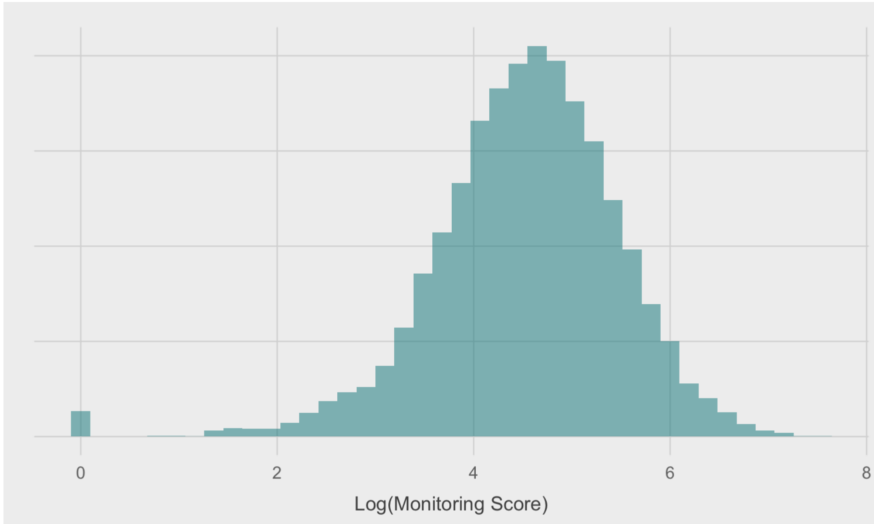
For the same reason, we are not able to empirically account for competitive monitoring programs in our analyses. Public filings contain very limited information on these programs; even monitoring start dates often do not coincide with the proposed dates in public filings. However, during our research window, monitoring in general takes up a small fraction of the market, especially around its introduction. We therefore do not consider this as a significant factor that may influence our empirical results. In addition, our firm is the only one offering monitoring in some states and time periods. We replicate our empirical results in these subsamples for robustness.

Dataset 3 - Monitoring Our data on the firm's monitoring program includes its pricing schedule, drivers' opt-in choices, and monitoring scores and renewal discounts for finishers. The firm's monitoring pricing is discussed in section 3.6 as well as in Appendix C.0.2. Across calendar time and states, the average monitoring finish rates are around $10-20\%$.

Monitored drivers' performance is summarized by a score, the distribution of which is plotted in Figure 3.2. The more punishable behavior recorded for a given driver, the *higher* her score. Drivers who received a zero score plugged in the device continuously for enough days but did not drive. We ignore these drivers in all subsequent analyses.

We treat this score as the output of the monitoring technology. It represents the firm's belief about *future* accident risk, based on a monitored driver's performance in the first (and monitoring) period. To see this, Figure 3.3 plots the average claim count in period two ($t = 1$, after monitoring) across monitoring groups. Compared to unmonitored drivers, those who finished monitoring are 22% safer. Among finishers, the quintile of their monitoring score strongly predicts their second-period risk, which ranges from 60% better to 40% worse than the opt-out pool.

Monitoring finishers face the same renewal choices as other drivers, except that their renewal quotes include appropriate monitoring discounts or surcharge. Figure 3.4 compares



Note: This graph plots the density of the (natural) log of monitoring score for all monitoring finishers. The lower the score the better. Drivers that received zero score plugged in the device continuously for enough days but did not drive. We ignore these drivers in all subsequent tests.

Figure 3.2: *Distribution of monitoring scores*

the distribution of first-renewal pricing change across monitoring groups. We benchmark the baseline price change to center around one. On average, monitored drivers received a 7% discount. Moreover, the monitoring discount is persistent after monitoring ends.[34] This is consistent with the firm's upfront communication with consumers during their opt-in decision.

3.2 Reduced-form Evidence

This section documents two reduced-form facts. First, drivers that opt into monitoring becomes safer when they are monitored. Despite this change in behavior, monitoring still reveals previously unobserved risk differences across drivers, which can lead to selection in consumer demand for monitoring and for insurance.

[34]See Figure C.11 in the appendix for a plot of the average progression of monitoring discounts across renewal periods for monitoring finishers in our sample.

Note: This is a binned-scatter plot comparing average claim count of the first renewal period ($t = 1$, after monitoring ends) across various monitoring groups. The benchmark is the unmonitored pool, which is the "opt-out" group. Group "opt-in" includes all monitored drivers that finished the program per definition in section 3.1.2. Groups "1" to "5" breaks down the "finish" group based on the quartile of the drivers' monitoring score. Lower monitoring score means better performance.

Figure 3.3: *Comparison of subsequent claim cost across monitoring groups*

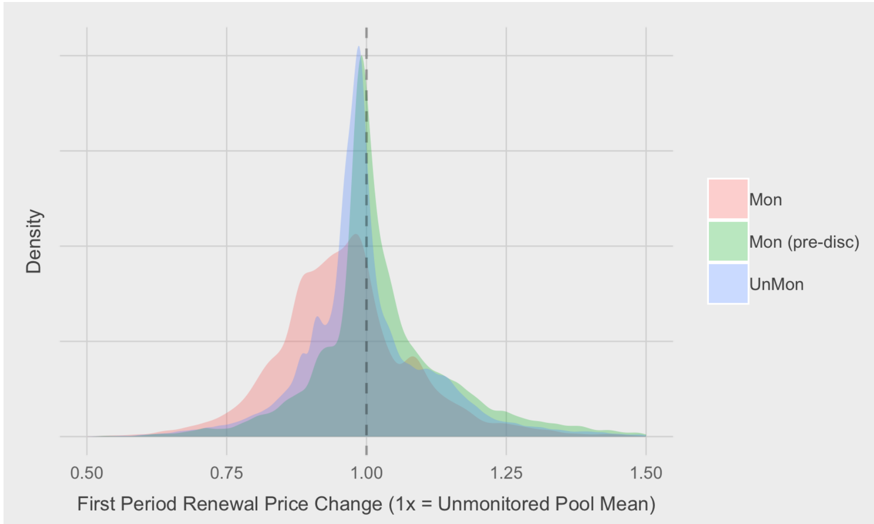
3.2.1 Risk Reduction and the Incentive Effect

If monitoring technology is effective, drivers may want to appear safer when monitored.[35] If this incentive effect is important and if drivers' risk is modifiable, then we should expect the *same* drivers to be riskier in unmonitored periods than in the monitored one.

Since monitoring is temporary, we can directly measure this effect by comparing claim outcome for the *same* monitored drivers before and after monitoring ends. This exercise requires us to balance our panel. We focus on the first three periods (18 months).[36] There may be spurious trends in claim rate across periods that are irrelevant to monitoring. We

[35]This effect is studied in Fama (1980); Holmström (1999). A similar setting is online tracking of consumers' purchase history Taylor (2004); Fudenberg and Villas-Boas (2006). If consumers know that buying expensive items online may label them as inelastic shoppers and lead to higher prices in the future, they may refrain from purchasing those items online.

[36]In our robustness check, we show results with only two periods. Attrition is about $10 - 15\%$ per period and our data is right-censored, so balancing the panel eliminates 46% of our data.



Note: This graph plots the benchmarked (per firm request) distribution of renewal price change during the first renewal, by monitoring group. 1x represents mean renewal price change factor for the unmonitored group. Initial/upfront monitoring discount is not counted towards this. So that monitoring price change is discounted monitored price divided by undiscounted new business price. "Mon" and "UnMon" are monitored and unmonitored groups, while "Mon (pre-disc)" represents the renewal price change for monitored drivers without the monitoring discount.

Figure 3.4: *First Renewal Price Factor by Monitoring Group*

account for this effect with exhaustive observable controls and a difference-in-differences approach. Among monitored drivers, we take the first difference in claim counts[37] between post-monitoring and monitored periods. This difference is then benchmarked against its counterpart among unmonitored drivers (control group).

$$C_{it} = \alpha + \tau m_i + \omega \mathbf{1}_{post,t} + \theta_{mh} m_i \cdot \mathbf{1}_{post,t} + \mathbf{x}_{it}' \mathbf{fi} + \epsilon_{it} \tag{3.2}$$

Here, i, t index driver and period in our panel dataset. C denotes claim count, and m_i is a driver-specific indicator for whether i has finished monitoring. x is a rich set of observable characteristics that the firm uses in pricing.[38]

[37] Throughout our reduced-form analyses, we use claim count as our cost proxy. This is because claim severity is extremely noisy and skewed. This is also common practice in the industry, where many risk-rating algorithms are set to predict risk occurrence only. We therefore present our estimates mostly in percentage comparison terms.

[38] See Table 3.1 for a list of main observable characteristics. We also include controls for trends and seasonality including third-order polynomials of the calendar year and the month when each driver i starts period t with

Our main specification includes only monitored drivers who finish monitoring in the first period. To test for parallel trends of the monitored and unmonitored groups, we conduct the same test in subsequent periods after monitoring. In reality, some monitored drivers do not finish monitoring until subsequent periods.[39] To make use of this plausibly exogenous variation in monitoring duration and timing across the first and subsequent periods, we introduce another specification, adding additional variation in relative monitoring duration in the pre-period, z_i. It is calculated as the fraction of days monitored in the first period minus the same fraction in post periods.[40]

Results are reported in Table 3.3. We find a large and robust incentive effect. Column (3) corresponds to the specification in equation 3.2, with the addition of insurance coverage fixed effects.[41] It shows that monitored drivers' average claim count is 0.009 or 23% lower during the monitoring period, compared to after it. Adjusting for the average monitoring duration of first-period monitoring finishers (142 days), a fully-monitored period would be 29.5% less costly to insure for the same driver. Incorporating additional variations in monitoring duration generates similar results (Column (6)). We test for parallel trends between the monitored and unmonitored groups by repeating the baseline specification in subsequent (unmonitored) periods. As shown in Columns (7-10), no differential claim change across periods can be detected between the two groups.

We discuss two important caveats of our results. First, monitoring provides a way for drivers to build a reputation for their risk (but only to the monitoring firm) Fama (1980); Holmström (1999). Moral hazard is therefore mitigated by drivers' concern over their future

---

the firm.

[39]Based on interviews with managers, among finishers, delays in finishing is predominantly caused by device malfunction or delayed start of monitoring due to mailing issues, etc.

[40]As discussed above, some drivers started monitoring but dropped out without finishing. This would bias our results if claims itself leads to non-finish. Out of more than 10,000 claims we observe among monitored drivers, only 13 occurs within 7 days before or after monitoring drop-out. In Table C.4, we further test the robustness of our results by repeating our main analyses on all drivers who started monitoring. This implies larger moral hazard effect adjusting for monitoring duration. However, if some monitored drivers drop out as they discover that they cannot change their risk, the incentive effect estimate would be contaminated by this selection effect.

[41]This soaks up any coverage adjustments between periods.

explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)	Parallel Trend/Placebo		
	dependent variable: claim count (C)								
constant	0.045***	0.002	0.003	0.046***	0.003	0.004	0.001	0.002	0.002
	(0.000)	(0.005)	(0.005)	(0.000)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)
post monitoring indicator	−0.001*	−0.003***	−0.003***	−0.001**	−0.003***	−0.003***	0.001*	−0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)
monitoring indicator (m)	−0.013***	−0.012***	−0.012***	0.008***	0.005***	0.005***	−0.005***	−0.005***	−0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
monitoring duration (z)				−0.026***	−0.020***	−0.020***			
				(0.002)	(0.002)	(0.002)			
interaction ($\mathbf{1}_{post} \times m$)	0.008***	0.009***	0.009***	−0.005**	−0.005***	−0.005***	-0.001	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
interaction ($\mathbf{1}_{post} \times M$)				0.015***	0.016***	0.016***			
				(0.002)	(0.002)	(0.002)			
observables controls (x)	N	Y	Y	N	Y	Y	Y	Y	Y
coverage fixed effects	N	N	Y	N	N	Y	Y	Y	Y
implied risk reduction (%)	28.0	29.4	29.5	27.5	29.4	29.6			
pre- / post-periods - "1st diff"	0 / 1-2						1 / 2	2 / 3	3 / 4
treatment / control - "2nd diff"	$t = 0$ finisher / unmonitored			all finishers / unmonitored			$t = 0$ finisher / unmonitored		
number of drivers per period	755,614			809,784			755,614	539,296	397,642

Note: This table reports results of (**??**). The estimate on the interaction term ($\mathbf{1}_{post} \times m$ or z) measures the "treatment effect" of monitoring ending on claim count across periods. We first balance our panel data to include all drivers who stay till the end of the third semester ($t = 3$). This gives us two renewal semesters ($t \in \{1, 2\}$) after the monitoring semester ($t = 0$). We control for a full set of observables, including driver and vehicle characteristics and tiers (past records of violations or claims). It also includes third-order polynomials of calendar year and month. Continuous observable characteristics are normalized. We report estimates with and without these controls.

Columns (3) and (6) are our main specification. Column (3) focuses on monitored drivers who finished within the first period, while Column (6) introduces additional variation in monitoring duration and timing and looks at all monitoring finishers. Columns (1,2,4,5) show robustness of our estimates to observable and coverage fixed-effect controls. The right-most columns are placebo tests for parallel trends among treatment/control groups after monitoring ends. We first try to detect a similar change from $t = 1$ to $t = 2$. We drop all observations from period 0, and roll the post-period cutoff one period forward, so that $1_{post,t} = 1 \iff t \geq 2$ (changed from $t \geq 1$). Naturally, we look at the future trends of monitored drivers who finished within the first semester and drop other monitored finishers. We find similar results by repeating this test in subsequent periods. As we need to balance panels, number of drivers drop in these tests.

Table 3.3: *Estimates From Moral Hazard Regression*

reputation as opposed to by directly contracting on effort as in a continuous monitoring setting. The magnitude of risk reduction can be different in the latter setting.[42] On the flip side, our result provides evidence that at least some drivers are forward-looking and respond greatly to future incentives. This means that uncertainty in dynamic premium (reclassification risk) may be nontrivial.

Second, our estimate measures a treatment-on-treated effect. If significant heterogeneity in the incentive effect exists across drivers and that it influences consumers' opt-in decision, then we would face external validity concerns in counterfactual simulations. In equilibrium, the firm assesses the signal monitored drivers send based on future claim records when drivers are no longer monitored, which corresponds to the renewal discount it gives. Therefore, risk reduction is compensated only to the extent to which it is correlated with drivers' future risk type. If safer drivers' risk levels are also more responsive to incentives, as suggested by a pure effort cost model for example, selection on the incentive effect can be important.[43] In this case, the effect we find will be larger than the population average (or the average treatment effect) Einav, Liran and Finkelstein, Amy and Ryan, Stephen P. and Schrimpf, Paul and Cullen, Mark R. (2013). In our counterfactual analyses, we therefore maintain the opt-in structure of the monitoring program and do not extrapolate to scenarios where the market monitoring rate is high.

3.2.2 Private Risk and the Selection Effect

Are drivers who choose monitoring safer than those who do not? Table 3.4 reports the results of regressing claim count in the first period ($t = 0$) on monitoring indicator, controlling for the same variables as in Column (3) of Table 3.3. The incentive effect only accounts for 64%

---

[42]We are also unable to disentangle the "Hawthrone effect" from drivers' responsiveness to financial incentives in our estimate. Since consumers must be aware of the data collection to be incentivized for it, we consider this effect as part of the incentive effect.

[43]Perfect revelation of a continuum of risk types is possible, as characterized in Mailath (1987), with a monotonicity condition similar to the single-crossing condition. However, consumers likely have multidimensional heterogeneity in reality, so drivers' performance during monitoring may not perfectly reveal their risk types Frankel and Kartik (2016).

of the risk differences across the two group. Had the monitored drivers not been monitored in the first semester, they would still be safer than the average unmonitored driver. It thus suggests that drivers possess private information on their own risk. Therefore, there may be strong advantageous selection into monitoring.

	Dependent variable:
	Claim Count ($t = 0$)
constant	-0.004^{**}
	(0.009)
monitoring indicator	-0.014^{***}
	(0.001)
observable controls	Y

Note: This table reports results of a regression where the dependent variable is first period claim count, and the independent variables are the monitoring indicator and observable controls. This is done within all first-period finishers of the monitoring program. This variable is consistent with the monitoring indicator in the incentive effect regression 3.2 (Table 3.3), so as to facilitate comparison and decomposition.

Table 3.4: *First Period Claim Comparison*

Selection into monitoring suggests that the technology is effective at capturing previously unobserved differences in drivers' risk types, further allowing the firm to dynamically select safer drivers. The following regression examines both factors. It shows how average costs in future (unmonitored) periods vary based on monitoring choice and score among all drivers.

$$C_{it} = \alpha_t + \theta_{m,t} m_i + \theta_{s,t} s_i + \mathbf{x}'_{it}\mathbf{fi_t} + \epsilon_{it} \tag{3.3}$$

Again, $m = 1$ for monitored drivers who finished within the first period. s denotes monitoring score, which is normalized among monitored drivers and set to 0 for others. Figure 3.6 reports $\hat{\theta}_t$ for renewal periods $t = 1$ to 5 (three years), translated into percentage difference terms.[44] Looking at the main specification (left grey series), the estimate for $\theta_{s,t}$ implies that a monitored driver who scores one standard deviation above the mean has a

[44]Regression on a balanced panel of drivers (who stayed till the end of period 5) produces similar results.

29% higher average claim count in the first renewal (after monitoring ends). However, this informativeness diminishes dynamically, and disappears after 3 years. Further, controlling for claims does not alter our estimate much. This suggests that although claim realization is a direct measure of risk, its sparsity may significantly limit how informative it is of risk in the short run. In Figure 3.5, our results also show that the monitored pool is persistently safer in periods after monitoring ends.

These results are shaped not only by the selection into monitoring, but also by selective attrition due to more accurate risk rating. We may see the monitored pool becoming safer as risky drivers receive higher prices and leave the firm at higher rates. Regardless, both effects suggest that monitoring technology is effective at capturing previously unobserved driver risk.

In reduced-form analyses, it is difficult to disentangle these two effects or to detect coverage-level adverse selection. In general, exogenous and *unilateral* variation in the pricing of policies and monitoring is rare in our setting. As shown in equation 3.1, price revisions often trigger changes in various inter-dependent prices that activate several demand margins at once. Therefore, in the next section, we propose a structural model to jointly account for several demand margins, including firm, coverage, and monitoring choices.

3.3 Cost and Demand Models of Auto Insurance and Monitoring

This section develops a structural model for consumers' monitoring opt-in choice in relations to their choices of insurer and insurance coverage as well as the cost of insuring them. We start from the canonical insurance framework, which consists of models for consumer preference and how their choices influence insurer costs. We then introduce additional features to incorporate drivers' opt-in decision and how monitoring technology can reveal driver risk. This allows us to link consumer demand in the information market with product (insurance) market fundamentals.

We describe our model in two parts. First, we outline a choice model conditional on the

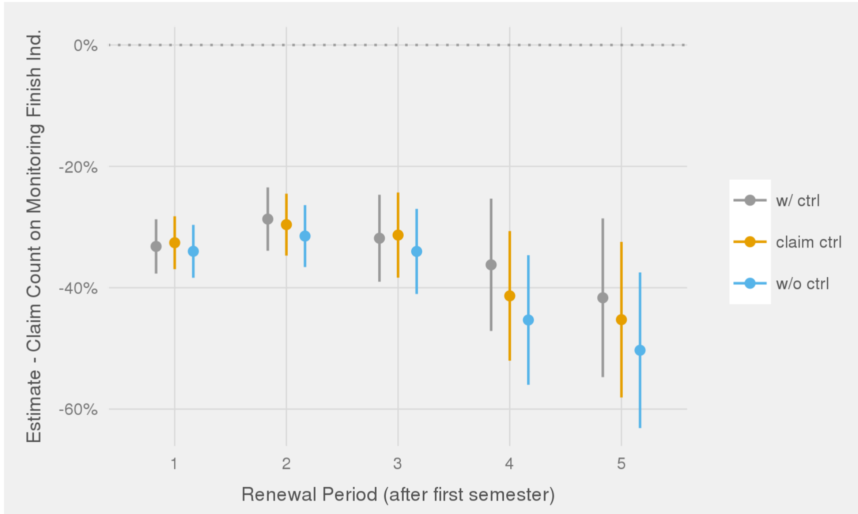


Figure 3.5: *Regression results - dynamic informativeness of monitoring participation*

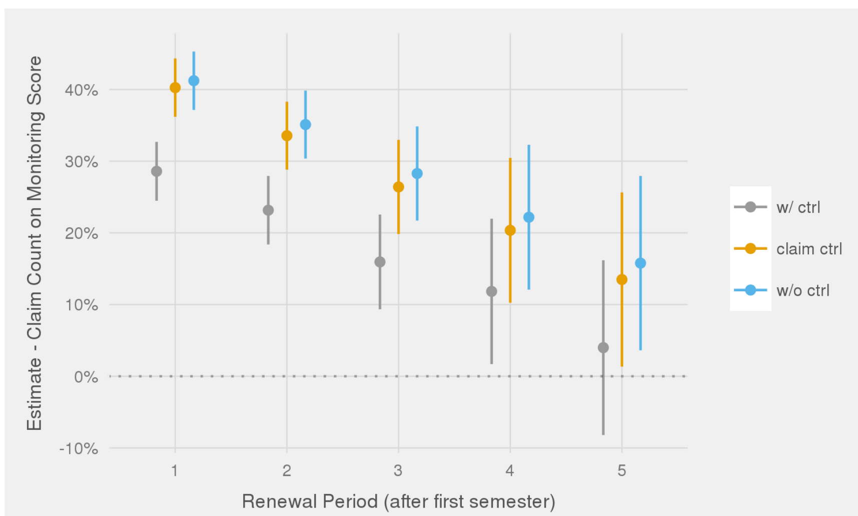


Figure 3.6: *Regression results - dynamic informativeness of monitoring score*

Note: 3.5 and 3.6 report the estimate for θ_t and γ_t from regression (3.3), translated into percent increase terms. Monitoring participation is defined as an indicator for finishing monitoring. For each $t > 0$, we take all drivers who stayed with the firm till at least the end of period t. θ_t is the coefficient of claim count of driver i in period t on monitoring score of i, and γ_t is that on monitoring finish indicator of i. Monitoring score is normalized, and defaulted as 0 for unmonitored drivers. So θ_t measures the effect of getting a score one standard deviation above the mean during the monitoring period ($t = 0$). γ_t compares unmonitored drivers with the average monitoring finisher. To further translate these effects into percent increase terms, we divide the estimate of θ_t and γ_t by the average claim count in period $t$of all *monitored* drivers. The horizontal axis represents different regressions for different renewal period $t > 0$.
Different colors and positions within each t value represent different specifications of control variables (x_{it}). The grey (left-most) series represents estimates from regressions with the full set of x_{it}; the orange (middle) one includes only claim records revealed since $t = 0$; the blue (right) series includes no control.

realization of claim and monitoring score. It captures customers' considerations regarding risk aversion, out-of-pocket expenditure, firm-switching inertia, as well as disutility from being monitored and expected price renewals. We then describe the data generating processes for claims and for monitoring scores in a cost model that features risk heterogeneity, the incentive effect, and monitoring score's signaling precision. This unifies the cost and demand factors under an rational expectation expected utility framework and introduces selection effects. Finally, we provide an informal discussion of model identification.

3.3.1 Choice model

In our setting, consumers make firm, coverage, and monitoring participation choices. Drivers, periods,[45] and choice options are indexed by i, t, and d, respectively. Conditional on the realization of claims C and monitoring score s, the choice model specifies realized utilities $u_{idt}(C, s)$.[46]

Besides consumers' risk type, our choice model highlights three factors. (i) Risk aversion governs both preference for insurance and disutility from price fluctuations. (ii) Demand frictions: firm-switching inertia leads to imperfect competition among insurers. Consumers' disutility from being monitored accounts for factors such as privacy or effort cost associated with monitoring. They also sustain partial pooling equilibrium, in which only a fraction of the population is monitored. (iii) Future prices contain most of the benefit of monitoring and depends on claims and monitoring score.

Drivers have standard von Neumann-Morgenstern preferences $u(\cdot)$. We assume that they are twice continuously differentiable and globally increasing and concave, which pin down drivers' absolute risk aversion, denoted by γ. Each driver-period i, t starts with annual income w_{it}. Different choice options are denoted by $d = \{f, y, m\}$, where f, y,

[45]Monitoring takes place in the first period ($t = 0$).

[46]We model u structurally for two reasons. First, key structural quantities outlined below are of interest. Second, the model must explain consumers' choice to be monitored consistently with their insurance choice. As monitoring introduces additional uncertainty in future prices, we need to understand the micro-structure of how consumers handle risk.

and $m \in \{0,1\}$ index firm, coverage, and monitoring choices, respectively. For the same driver-period, differentiation in choice options is purely financial and impact utility through the consumption term h_{idt}.

$$u_{idt}(C,s) = u_\gamma\left(w_{it} + h_{idt}(C,s)\right) \tag{3.4}$$

$$h_{idt}(C,s) = -p_{idt} - \underbrace{\mathbf{1}_{d,t-1} \cdot \psi_{idt}}_{\text{friction}} - \underbrace{e(C,y_d)}_{\text{oop}} - \underbrace{p_{idt} \cdot R_{idt}(C,s)}_{\text{renewal price}} \tag{3.5}$$

$$\text{where } \psi_{idt} = \underbrace{\mathbf{1}_{d,t-1} \cdot \eta_0}_{\text{baseline inertia}} + \underbrace{\mathbf{1}_{f_d,t-1} \cdot \eta(x_{it};\grave{}\,\jmath)}_{\text{firm-switching inertia}} + \underbrace{\mathbf{1}_{m_d} \cdot \mathbf{1}_{t=0} \cdot \xi(x_{it},\lambda;\grave{}\,\text{¸})}_{\text{monitoring disutility}} \tag{3.6}$$

Consumption h includes four main components. Drivers pay prices p and friction costs ψ up front. The latter is broadly defined as a cost to change choices compared to the previous period ($\mathbf{1}_{d,t-1} = 1$). Drivers also form expectations over the realization of claims and of monitoring score. These influence out-of-pocket expenditures, e, and changes in renewal prices, R. Like prices, the out-of-pocket expenditure (oop) covers two periods, so that the overall consumption term h is of a one-year horizon.[47]

Demand friction (ψ_{idt}) includes heterogeneous disutility consumers experience from being monitored, $\xi(x_{it},\lambda)$, since monitoring is only offered to new customers of the firm. We allow it to vary across both observable characteristics and risk λ. Including the latent risk type λ is important in fitting selection into monitoring well. In its absence, the differential benefit of monitoring across safe and risky drivers is deterministic conditional on expected renewal prices. This may not accurately capture both the popularity of monitoring and the (risk) selection pattern into monitoring.

Demand friction also includes consumers' inertia associated with adjusting choices. We model them as implied monetary costs. The baseline inertia η_0 prevents consumers from making any choice adjustments. Heterogeneous firm-switching inertia $\eta(x_{it})$ further

[47] We assume that consumers are myopic but have a one-year (two-period) horizon, during which they do not consider changing choices after the first period. This is the simplest model that captures different types of costs and benefits of monitoring programs to consumers. In particular, dynamic premium risk (reclassification) is incorporated, as higher uncertainty in renewal prices diminishes ex-ante utility. It is unclear whether a two-period or fully dynamic model can be separately identified. Our model can also be interpreted as approximating a two-period dynamic model with infinite adjustment costs.

deters consumers from exploiting financially lucrative outside options. These terms capture imperfect competition that supports the observed attrition rate given price dispersion in the data. They capture the effect of search and switching costs consumers face when adjusting firms across periods as well as potential brand differentiation Farrell and Klemperer (2007); Honka (2012); Handel (2013).

Renewal prices are influenced by a baseline price change factor (R_0) that can be influenced by monitoring results, as well as by claim surcharges (R_1). We separately model the two components to capture the correlation between out-of-pocket expenditures and renewal prices.[48]

$$R_{idt}(C, s) = R_{0,idt}(s) \cdot R_1^C \tag{3.7}$$

Monitored drivers gets a renewal price discount based on score s. We use a Gamma distribution to model renewal price change R_0.[49] It is influenced by observables x and, if monitored, the monitoring score.[50] Notice that monitoring only impact own firm ($f^\star$) options.

$$R_{0,idt}(s) \sim \text{Gamma}(\alpha_{R,m}(x_{it}, s; \text{`}_{\mathbf{R}}), \beta_R) \tag{3.8}$$

By definition, out-of-pocket expenditure e is non-decreasing in claim, but non-increasing in the amount of coverage y.[51] Similarly, renewal price R is *non-decreasing* in both of its arguments. The choice-specific utility v_{idt} is simply the expectation of u over C and s.

$$v_{idt} = \mathbb{E}_{C,s}\left[u_{idt}(C, s)\right] \tag{3.9}$$

---

[48]Notice that R_1 changes R_{idt} for all d, including those at other firms. In reality, it is between 1.1 and 1.5 (see Figure C.12 in the appendix for details). In contrast, monitored drivers are reclassified within the monitoring firm only. However, our myopia assumption diminishes this difference. We consider it as a realistic assumption because, as we will show later, the average switching cost is much larger than the potential surcharge that a monitored driver can receive.

[49]Figure 3.4 shows the actual distribution of the first-renewal price-change factor.

[50]In subsequent renewals, prices are very stable. We therefore assume that $\alpha_R = \beta_R$ in those periods so that, in expectation, prices do not change without claims.

[51]We abstract away from strategic reporting behavior.

Lastly, we adopt a mixed logit structure Train (2009) to model discrete choice.

$$d_{it} = \arg\max_{d \in D_{it}} \{v_{idt} + \varepsilon_{idt}\} \tag{3.10}$$

$$\text{where } D_{it} = \begin{cases} \mathcal{F}_{it} \times \Upsilon_{\mathcal{F},it} & t > 0 \text{ or } i \text{ ineligible} \\ \mathcal{F}_{it} \times \left\{\Upsilon_{-f^{\star},it}, \Upsilon_{f,it} \times \{0,1\}_m\right\} & t = 0 \text{ and } i \text{ eligible} \end{cases} \tag{3.11}$$

The choice space D can vary based on driver-period i, t, but always includes firm space $\mathcal{F}$ and the corresponding coverage space Υ. It covers all firms we observe, including the monitoring firm $f^{\star}$. As discussed in Section 2, we assume that no other firms offer monitoring, for which only new customers that come to the firm after monitoring introduction are eligible. In addition, we abstract away from monitored drivers that drop out; the opt-in indicator m therefore represents drivers' decision to finish monitoring. Lastly, ε follows a type 1 extreme value distribution with scale σ.

Our *demand parameters* include risk aversion, baseline inertia, intercept and slope parameters for heterogeneous firm-switching inertia, monitoring disutility, as well as the (expected) renewal pricing rule:

$$\Theta_d = \{\gamma, \eta_0, \grave{}_{\mathbf{J}}, \grave{}_{\text{¸}}, \grave{}_{\mathbf{R}}, \beta_R, \sigma\}.$$

3.3.2 Cost model

Let λ be defined as the expected claim count (C) per period. We model λ as follows:

$$\lambda_{imt} = \mu_\lambda(x_{it}, m; \grave{}\breve{}) + \epsilon_{\lambda,i} \tag{3.12}$$

$$\ln \epsilon_{\lambda,i} \sim \mathcal{N}(0, \sigma_\lambda) \tag{3.13}$$

$$C \sim \text{Poisson}(\lambda) \tag{3.14}$$

We interpret $\epsilon_{\lambda,i}$ as the persistent private risk of driver i that can be captured by monitoring. We further assume that it is distributed i.i.d. log-normally.[52] Let M denote the set of

[52] Risk parameters are non-negative. Cohen, Alma and Einav, Liran (2007); Barseghyan *et al.* (2013) use the same distributional assumption. We also investigate a robustness check with normally distributed λ.

monitored drivers. Then advantageous selection into monitoring implies that:

$$\mathbb{E}[\epsilon_{\lambda,i}|i \in M] < \mathbb{E}[\epsilon_{\lambda,i}|i \notin M] \tag{3.15}$$

The incentive effect may reduce monitored drivers' risk during the monitoring period. We adopt a reduced-form approach towards modeling this effect to avoid making further assumptions about the underlying structure of effort provision and risk determination. We assume that the incentive effect is homogeneous across drivers and that it enter risk in an mechanical and additive-separable fashion:[53]

$$\mu_\lambda(x_{it}, m = 1) = \mu_\lambda(x_{it}, m = 0; \grave{}_{,\mathbf{0}}) + \theta_{\lambda,m} \cdot \mathbf{1}_{t=0} \tag{3.16}$$

In order to get out-of-pocket expenditure, we need to model not only the severity of claims, but also that of accident loss conditional on occurrence. Let ℓ denote the latter quantity, which is assumed to be independent from claim count arrival and drawn from a Pareto distribution:

$$\ell_{idt} \overset{\text{i.i.d.}}{\sim} \text{Pareto}(\ell_0, \alpha_\ell) \tag{3.17}$$

α_ℓ is the main (shape) parameter. In the primary specification, we assume that α_ℓ is homogeneous across drivers. Importantly, we assume that there is no unobserved heterogeneity in the conditional loss severity.

Monitoring Technology (Score) We model monitoring score s as an informative signal of private risk ϵ_i. Monitoring score is driver-specific and is revealed once for monitored drivers after the first semester ($t = 0$).

$$\ln s_i \sim \mathcal{N}\left(\mu_s(x_i, \ln \lambda; \grave{}_{\mathbf{s}}), \sigma_s\right) \tag{3.18}$$

We assume that the signal noise has a log-normal distribution with mean μ_s and and precision σ_s, similar to the latent risk type λ that it tries to capture. When $\frac{\partial \mu_s}{\partial \lambda} \neq 0$ and σ_s

[53]For more careful treatment of moral hazard and risk determination, see Jeziorski *et al.* (2014).

is finite, the realization of s is informative of λ conditional on observable x. On the other hand, s perfectly reveals λ as $\sigma_s \to 0$.

Overall, we can define the *cost parameters* as the intercept and slope parameters for unmonitored latent risk type λ, the incentive effect parameter, the spread of latent risk type conditional on observables, intercept and slope parameters for conditional accident loss, the intercept and slope parameters for monitoring score, and montioring score precision.

$$\Theta_c = \{`˘_{,\mathbf{0}}, \theta_{\lambda,m}, \sigma_\lambda, \mathbf{ff}_\ell, `_{\mathbf{s}}, \sigma_s\}$$

3.3.3 Identification

We now provide an informal discussion of the data variation and model structure that allow us to identify cost and demand parameters.

Cost parameters All parameters contained in Θ_c can be identified with cost data alone. Variations in average claim count and monitoring scores across observable groups identify `˘ and `$_{\mathbf{s}}$ (slope parameters). $\theta_{\lambda,m}$ is identified with the same data variation outlined in the reduced-form section in equation 3.2. As in Cohen, Alma and Einav, Liran (2007), σ_λ is identified when sufficient number of drivers file for multiple claims per period, conditional on observables. In addition, the monitoring score brings additional restrictions to the distribution of private risk, conditional on signal precision σ_s. Therefore, σ_s and σ_λ are jointly identified in our setting by the variance of claim counts and monitoring score conditional on observables and on one another.

In modeling and identifying loss severity, we attempt to accurately capture both insurer cost, which we observe, and out-of-pocket expenditure in consumers' expectation, which is unobserved. The Pareto distribution does a good job balancing these two objectives. With appropriate location parameter, it fits the average claim amount well. At the same time, it is sufficiently long-tailed so that loss events significantly larger than coverage limits still have non-degenerate support in consumer's expectation. This is important in fitting the share of large coverage limits.

Demand parameters Our demand identification largely relies on price and contract space variation. Controlling for the observable characteristics used in firms' pricing rules, the remaining price variation depends on location and calendar time. We specifically model consumers' risk differences across these dimensions by including each consumers' assigned risk class in the cost model. We further include controls for yearly trend, seasonality, and zipcode characteristics including income and population density in our demand parameters. Therefore, we are left with price changes associated with the firm's and its competitors' rate revisions (back-end changes in pricing rules) as well as cross-location differences that are plausibly exogenous from consumer demand. Specifically, the firm also changed monitoring opt-in discount over time.

We also observe variation in consumers' contract space conditional on observables. Specifically, monitoring eligibility differs based on state, time, specific vehicle models, and renewal period. For instance, drivers arriving before monitoring introduction in their states or with vehicles older than 1995 are not eligible. Monitoring is also only available to new customers. Meanwhile, mandatory minimum coverage also changed in two states within our research window. We use one of these states in our demand estimation and reserve the other for cross-validation.[54]

Our primary concern is in identifying monitoring disutility (ξ) well. Given cost parameters and risk aversion, we can determine the relative attractiveness of the same coverage option with and without monitoring based on objective financial risk and rewards. However, just because a driver can financially benefit from monitoring does not mean that she will opt in. The monitoring disutility is pinned down by the observed monitoring share (under different pricing environments) given cost parameters and risk aversion. The slope parameter on risk type ($\theta_{\xi,\lambda}$) further turns the monitoring disutility term into a risk-specific shifter that flexibly controls the share of each risk type opting into monitoring. It therefore helps us fit both the share of monitoring and selection based on risk.[55]

[54]See Table 3.2 for summary statistics on quotes and coverage for the state in our main sample. Table C.1 in the appendix summarizes the cross-validation results.

[55]Simply raising baseline monitoring cost for all risk types (conditional on observables) enhances selection

Another parameter of interest is risk aversion γ. For a given i, t, different γ values imply different gradient of Δv_{idt} across the multiple coverage options we observe in the data.[56] Therefore, conditional on risk parameters, risk aversion can be identified by how the empirical coverage share changes given contract space and pricing environment.[57] In our demand estimation, the Pareto severity parameters can also affect changes in coverage attractiveness. However, we restrict the Pareto distribution to approximate the actual (truncated) claim severity that we observe.

We also need to separately identify baseline inertia (η_0) and consumers' firm-switching inertia (η). Conditional on observables, different levels of these parameters imply unique combinations of the share of drivers who adjust coverage versus leaving the firm at renewals. We also observe rich variation in competitive pricing environments conditional on observables. Under a given pricing environment, these parameters imply a corresponding threshold under which drivers would stay with the firm, and another one under which drivers would not adjust choices at all.

3.4 Estimation

In this section, we propose econometric specifications in order to take our model above to the data. We also discuss identification, our estimation procedure, the model fit, and cross-validation results.

---

but also necessarily reduces monitoring share.

[56]This is conditional on the fixed effect for the mandatory minimum plan (ψ_1). The fixed effect adds an additional degree of freedom to more flexibly fit the gradient of willingness-to-pay across coverage options.

[57]Specifically, based on the company's pricing rule in Equation 3.1, the price gradient across coverage options only depends on the actuarial risk class assigned to each consumer and the coverage factor. The latter is heavily regulated. Each state offers an official guidance on the coverage options that auto insurers should offer and the corresponding coverage factors. Firms need to provide actuarial support to deviate from the guidance in order to avoid regulatory scrutiny. Empirically, coverage factor is rarely changed in our demand estimation states based on rate revision filings.

3.4.1 Econometric Specifications

Intercept and slope parameters We parameterize heterogeneous latent parameters linearly:

$$\eta(x_{it}) = (1, x_{it})' \text{\`}_{\mathbf{J}}$$

$$\xi(x_{it}) = (1, x_{it}, \ln \lambda)' \text{\`}_{\text{¸}}$$

$$\alpha_{R,m}(x_{it}) = \begin{cases} \mathbf{x}_{it}^{R\prime} \text{\`}_{\mathbf{R,0}} & m_d = 0 \\ \left(\mathbf{x}_{it}^{R}, s\right)' \text{\`}_{\mathbf{R,1}} & m_d = 1 \end{cases}$$

$$\mu_\lambda(x_{it}, m = 0) = (1, x_{it})' \text{\`}_{\text{˘},\mathbf{0}}$$

$$\mu_s(x_i, \lambda) = (1, \ln \lambda_i, x_i^s)' \text{\`}_{\mathbf{s}}$$

Broadly consistent with actual firm pricing rules, x_{it}^R and x_i^s only include a polynomial and the log of risk class, which represents firm's risk assessment without monitoring information.

Nest structure Incorporating additional alternative-level random effects can further enrich our model. In our primary specification, we add a random coefficient, ζ, on all choices within $f^\star$. This allows us to capture correlations between choices within the firm. Here, we assume ζ is an independently normally distributed with mean zero and standard deviation σ_ζ Train (2009). This allows us to escape the Independence of Irrelevant Alternatives property of a simple logit model. The model can therefore achieve better fit on attrition rate differences across consumers facing different contract spaces across states or when mandatory minimum changes.

Taylor approximation approach for nonlinear utility Next, following the literature on auto insurance choices Cohen, Alma and Einav, Liran (2007); Barseghyan *et al.* (2013), we start with an approximation approach to model the utility function described in equation 3.4. Assuming that third- or higher-order derivatives are negligible, the utility function can be expressed by a second-order Taylor approximation of the utility function around income w. Normalizing by marginal utility evaluated at w, we get the following expression, in which γ is the absolute-risk-aversion term:

$$v_{idt}(\lambda,\zeta) = \mathbb{E}\left[h_{idt}|\,\lambda,\zeta\right] - \frac{\gamma}{2}\mathbb{E}\left[h^2_{idt}|\lambda,\zeta\right] \tag{3.19}$$

This further simplifies product differentiation into consumption bundles with different mean and variance profiles. It also allows us to interpret v in monetary values, as the second term of equation 3.19 is exactly the risk premium, while the first is expected consumption. We are currently running robustness checks for alternative utility assumptions such as CARA and CRRA, as well as to allow for richer heterogeneity in risk preference.

3.4.2 Estimation

Our model includes random coefficients that enter utility non-linearly. Private risk, in particular interacts with various observed monitoring and coverage characteristics (renewal price, out-of-pocket expenditure), as well as unobserved demand parameters (risk aversion and monitoring cost). Therefore, we use a simulated maximum likelihood approach (Train 2002; Handel 2013). In particular, the mix logit structure implies that the choice probability is numerically integrated as follows:

$$\begin{aligned}\Pr(d_{it}|\lambda) &= \Pr(\epsilon_{idt} - \epsilon_{id't} > [v_{idt}(\lambda) - v_{id't}(\lambda)]\ \ \forall d' \neq d \\ &= \frac{\exp\left[v_{idt}(\lambda)/\sigma\right]}{\sum_{d'}\exp\left[v_{id't}(\lambda)/\sigma\right]} \qquad (3.20)\\ \Pr(d_{it}) &= \int \Pr(d_{it}|\lambda) f_{\lambda}(\lambda) d\lambda \qquad (3.21)\end{aligned}$$

In general, for each parameter proposal Θ_d, we simulate 50 independent draws of private risk (ϵ_λ) and the zero-mean firm dummy (ζ).[58] Then, we compute the likelihood for observed choices, claim count and severity, monitoring score, and renewal price change.

[58] We test the effect of increasing the number of draws in estimation on a 10,000 sub-sample. The effect of going from 50 to 200 draws is minimal.

These are averaged over to get the simulated log likelihood. The estimator $\theta^{\star}$ maximizes the log likelihood. Notice that the Taylor approximation allows us to derive closed-form solutions for the first two moments of out-of-pocket expenditures and renewal prices.[59] We therefore do not simulate claim losses or monitoring scores within each draw of random coefficients.

As discussed above, our cost model is easier to estimate but requires a large amount of data to estimate precisely. Our demand model faces the opposite challenge, being computationally demanding but also making use of rich variations in choice environment and outcome. Therefore, we adopt a two-step estimation procedure. First, risk and monitoring score parameters $(\theta_\lambda, \sigma_\lambda, \theta_s, \sigma_s)$ are estimated in the full dataset (except the loss severity parameter, per the discussion above). We then feed the estimates into the demand models as truth.[60] We lose precision by doing so, but both models are identified standalone.

3.4.3 Fit and cross-validation

We demonstrate that our demand model is flexible enough to produce accurate fit for four critical moments of the data in Table 3.5 and Figure 3.7. We present two specifications: a basic one that excludes a firm dummy (ζ random coefficients) or private monitoring cost ($\theta_{\xi,\lambda}$), and a comprehensive one that includes these variables. As Table 3.5 demonstrates, we match monitoring and coverage shares within our firm well. Further, first-renewal attrition rates – the share of outside option – is also broadly consistent. More importantly, the primary specification is able to accurately fit the expected monitoring score. This demonstrates that the model is capable of capturing selection as well as the effectiveness of the monitoring

[59] Further, we restrict α_ℓ to be larger than 2 so that the mean and variance of the distribution are both finite, as both moments enter consumers' utility. The mean of the Pareto distribution is thus no more than $2\ell_0$. Therefore, to fit the average cost to the firm well, we set $\ell_0 = 3000$, roughly half the empirical mean of the claim distribution. This parameter is selected in cross-validation, om which we compare model performance in a hold-out dataset by directly calculating the likelihood. In a robustness check, we are also fitting a Gamma model for calculating the firm's cost only.

[60] Standard errors for the demand estimates are current not adjusted for two-step estimation. In a robustness check, we are correcting those standard errors and implementing a joint estimation.

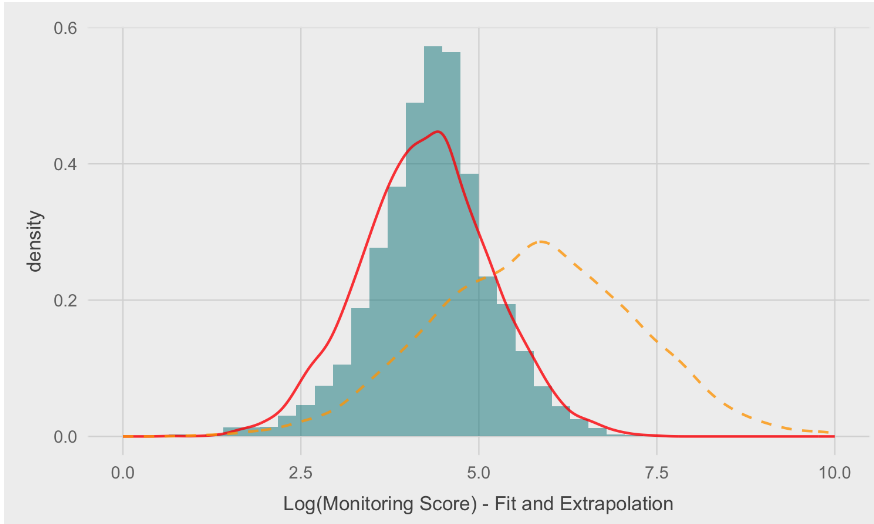
score. Figure 3.7 confirms this graphically: we calculate the expected monitoring score for each driver over all random-coefficient draws. The red line plots the simulated score weighted by the corresponding monitoring choice probability in each draw. The orange line plots the full distribution of expected monitoring scores, had everyone in the data finished monitoring.

	Basic Specification	**Primary Specification**	**Data**
Monitoring share (if eligible)	17.7%	15.6%	15.3%
Expected score	5.46	4.25	4.30
Selection effect (risk)	6.7%	21.2%	-
Coverage share			
30K	13.7%	12.5%	12.7%
40K	9.1%	8.2%	8.5%
50K	53.2%	49.8%	47.1%
100K	13.0%	15.4%	17.0%
300K	9.3%	11.9%	12.3%
500K	1.8%	2.3%	2.4%
First renewal attrition (indexed)	133.0%	102.9%	100.0%

Note: This table reports the fit of our demand model. The basic specification does not include firm-level random coefficient (ζ) or private monitoring cost ($\theta_{\xi,\lambda}$). The primary specification is outlined in our econometric model section. Monitoring share is conditional on eligibility. For coverage shares, our demand estimation data pools across three states with different mandatory minimum. One state changed mandatory minimum from 30K to 50K; estimation data is drawn from only the pre-period of that state to capture monitoring introduction. First renewal attrition rate is benchmarked to data per the firm's request (reporting percent differences, not percentage point differences). Expected score is calculated as the monitoring score weighted by monitoring probability in each simulation, normalized by overall monitoring share. Similarly, selection effect is calculated as the unmonitored risk type ($\lambda_{i,m=0,t=0}$) weighted by monitoring probability in each simulation, compared to the same quantity but weighted by the inverse of monitoring probability.

Table 3.5: *Demand Model Fit*

Using these estimates, we can calculate the expected unmonitored risk type (no incentive effect) of *monitored* drivers in the first period. Specifically, when we numerically integrate over private risk ϵ_λ, we simply weight it by the choice probability of monitoring. This gives us the expected (unmonitored) risk type in the monitored pool. Vice versa for the



Note: The green histogram is the empirical distribution of monitoring score for monitoring finishers in our demand estimation data. The red line plots the fitted distribution as outlined above. The orange dotted line plots the density of the extrapolated distribution of monitoring scores had all drivers finished monitoring.

Figure 3.7: *Monitoring Score - Fit and Extrapolation*

unmonitored pool. The selection effect is therefore a ratio between the two. The 21% ratio between the two pools is similar to the 17% back-of-the-envelope calculation we did in the reduced-form section.

The availability of un-used demand data allows us to perform cross-validation. In particular, one state in our dataset increased its mandatory minimum from $30,000 to $50,000. In our demand estimation, we draw from only the pre-change period for this state. The hold-out sample, however, contains all drivers in that state arriving in the post-period. Our model performs well out of sample, as demonstrated in Table C.1 in the appendix.

3.5 Estimation Results and Welfare Calculations

We present the raw estimates for homogeneous parameters in our models in Table 3.6 below. The heterogeneous latent parameters are reported in Table C.2 in the appendix. In this section, we highlight some key results and provide intuition. In particular, we use a simulation exercise to demonstrate the relative importance of different demand factors. We also conduct welfare calculations. Importantly, all simulation exercises in this section hold observed prices as fixed.

	Cost		Demand
$\ln \gamma$	-9.235^{***} (0.089)	η_0	134.262^{***} (2.228)
$\ln \sigma_{\lambda,\text{new driver}}$	-0.266^{***} (0.060)	$\beta_{R,\text{new}}$	66.953^{***} (0.403)
$\ln \sigma_{\lambda,\text{old driver}}$	-0.840^{***} (0.070)	$\beta_{R,\text{monitoring}}$	59.680^{***} (0.902)
$\ln \sigma_s$	-0.081^{***} (0.007)	$\beta_{R,\text{renw}}$	78.571^{***} (0.315)
$\ln \alpha_\ell$	-1.480^{***} (0.063)	σ_ζ	98.989^{***} (2.303)
		σ	39.213^{***} (0.632)
		Note:	*p<0.1; **p<0.05; ***p<0.01

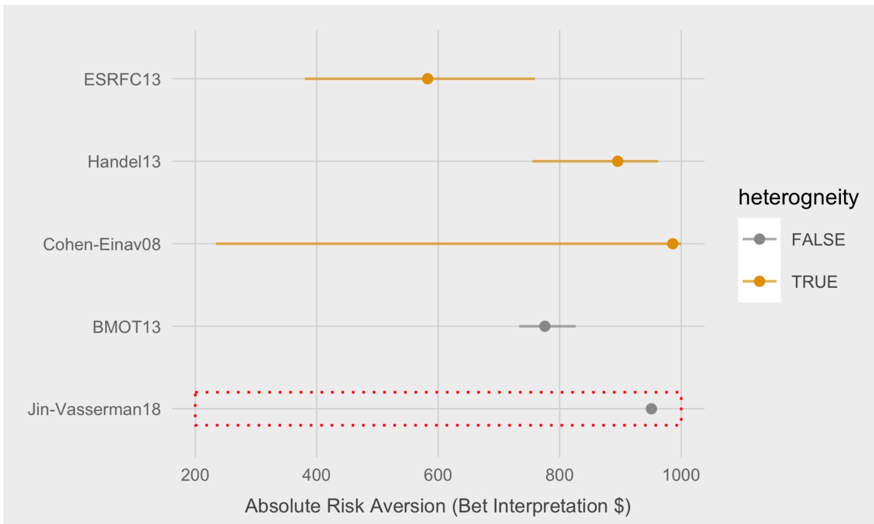
Note: This table reports estimates for homogeneous parameters in our structural model, including the absolute risk aversion coefficient γ, the spread of private risk $\sigma_{\lambda,\text{new driver}}$ and $\sigma_{\lambda,\text{old driver}}$. New drivers are defined as those licensed in the past three years. σ_s is the monitoring score's signal precision. Claim severity follows a Pareto distribution with parameters ℓ_0 and α_ℓ, where ℓ_0 is set at $3,000 per discussion in the text. On the demand side, we have the baseline inertia term η_0. β_R's are the rate parameters for the renewal price change Gamma distribution (R_0). Lastly, σ_ζ is the variance of the independent random coefficients on own firm dummy. σ is the scale of the logit error.

Table 3.6: *Homogeneous Parameters*

The magnitude of private risk and the monitoring score's signal precision are presented in the left panel of Table 3.6. Compared to Cohen, Alma and Einav, Liran (2007), we find significantly more unobserved heterogeneity in driving.[61] This can be attributed to

[61]Our private risk spread is 0.43 ($\exp(\ln \sigma_\lambda)$) for non-new drivers, compared to Cohen, Alma and Einav, Liran (2007)'s estimate of 0.15.

our ability to capture information contained in an additional signal of private risk – the monitoring score. New drivers who do not have past claim records see particularly high spread of private risk. Our estimates also capture the efficacy of the monitoring technology and the firm's renewal prices as well. In particular, we find that monitoring scores rise with driver risk, as do renewal prices for monitored drivers.[62]



Note: This figure benchmarks our risk aversion parameter estimate to the literature. Risk aversion is interpreted as the indifference value between inaction and taking a 50-50 bet on gaining $1000 versus losing that value. Heterogeneity indicator means that the author allows risk aversion to vary across people, in which case we plot the range of risk aversion parameters in the population. Otherwise we plot the 95% confidence interval of the homogeneous risk aversion parameter.

Figure 3.8: *Risk Aversion Parameter Estimates - Benchmark*

Figure 3.8 benchmarks our risk-aversion parameter against the literature. Our primary specification assumes homogeneous risk aversion and the estimate is broadly consistent with the literature.[63]As in the prior literature, we find that demand frictions are empirically important: many drivers who would benefit from monitoring choose not to participate. In Table 3.7, we present the empirical distribution of both firm-switching and monitoring costs in the population. The average driver foregoes $283—44% of annual premium—of gain by

[62]See Table C.3 in the appendix for details.

[63]Barseghyan *et al.* (2013), in particular, differentiate between probability distortion (wrong belief about one's own risk) and risk aversion.

not choosing an outside option from other firms. Disutility from monitoring is also large and heterogeneous across drivers. We estimate that the average driver needs to expect a gain of $93 to participate in monitoring, with a standard deviation of $19.21.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Firm-switching Inertia $\eta(x)$	283.63	35.39	157.71	264.85	286.46	307.11	406.56
(/ annual premium)	0.44	0.17	0.11	0.31	0.41	0.55	0.55
Monitoring Disutility $\xi(x, \lambda)$	92.83	19.21	9.52	79.97	92.54	105.21	187.20
(/ annual premium)	0.14	0.06	0.01	0.10	0.13	0.18	0.25
Claim Risk λ	0.05	0.05	0.001	0.02	0.03	0.06	1.48

Note: This table reports the distribution of heterogeneous latent parameters in our dataset. We simulate a distribution of private risk and calculate these parameters based on our demand estimates.

Table 3.7: *Latent Parameters*

Moreover, monitoring disutility increases with private risk.[64] This further accelerates advantageous selection into monitoring, while suggesting that observed renewal prices alone are not enough to explain the empirical selection pattern. At the same time, we see that older and more educated drivers tend to have lower monitoring costs, as well as those with newer cars, better prior insurance records and less traffic violation points.

As shown in the right panel of Table 3.6, the fixed inertia cost that drivers need to overcome when adjusting choices is $134. This adds to firm-switching and monitoring costs and further prevents safe drivers from being monitored. All else equal, the average driver only prefers the mandatory minimum coverage by $26, which seems low given that the plan commands almost 50% market share. This suggests that the rational amount of coverage for many drivers may be below the mandatory minimum, which restricts how monitoring can affect allocative changes across coverage.

[64]Column (2) of Table C.2 in the appendix reports the slope parameter for private risk.

3.5.1 Fixed-price Counterfactuals and Welfare Calculations

In this section, we use several simulation exercises to understand the demand and profit impact of removing different elements of the demand model as well as the welfare impact of introducing monitoring. We hold prices fixed here, and study equilibrium implication in the next section.

Simulation methodology Consistent with our demand model, we take a one-year horizon. The following procedure is used to calculate ex-ante and expected realized (ex-post) quantities.

1. For each driver i, simulate random coefficients (private risk and firm dummy) $L \in \mathbb{N}^+$ times.

2. For each draw $l \in \{1, ..., L\}$, calculate ex-ante utility directly and the corresponding certainty equivalent.[65] First-period choice probabilities are also calculated, which gives us the monitoring share. Expected cost of the first semester can be calculated directly. But we also need to form an expectation of the second-period cost (and prices) in order to calculate total surplus (and profit):

3. Simulate $K \in \mathbb{N}^+$ draws of first-period claim occurrence and monitoring score based on private risk.[66] Each draw pins down the renewal price change that driver i would face in the second period. All other prices remain constant. For each first-period choice d, we can then calculate the second period choice probability and the corresponding expected cost.

Sample enumeration Since we observe new customers' origins, as well as the competitive prices they face when coming to the firm, we can use our model to enumerate a full sample

[65]Due to our Taylor approximation, this should be the negative root of the polynomial.

[66]For simplicity, we assume that R_0 is deterministic conditional on C and s. In reality, the spread of baseline R_0 without claims and monitoring may have subtle nonlinear effects on consumer choice, which we assume away.

of potential new customers Train (2009). To do so, we first calculate the probability of each new customer arriving at the firm. We then follow the same procedure as outlined above, but weight each driver by the inverse of the calculated probability. The simulation is carried out assuming that monitoring is available for all new customers.[67] Overall, our simulated dataset is expanded by a factor of 4.03, which gives us a market share (among the top six firms for which we have data) close to the reality in the states we study.[68] This also allows us to derive a realistic proxy for competitor profit under a symmetric cost assumption; that is, the distribution of risk that we estimate in our dataset is valid when extrapolated to the simulated market.

In order to enumerate the market, we need to extrapolate the estimated attrition elasticity the firm faces to understand how the firm competes with other firms in the first period. To do so, we make a *no-brand-differentiation assumption*: liability insurance contracts offered by different firms only differ financially. This means that our firm-switching inertia estimate consists only of search and switching costs that are state-dependent (on consumers' preexisting firm choice) and that consumers have no unobserved preference for our firm, which is not state-dependent. In the context of our counterfactual simulations, this assumption essentially maintains that the price elasticity the firm's competitors face when the firm tries to poach customers away from them (in the first period) is the same as the price elasticity the firm faces when trying to retain existing customers.

This assumption follows naturally from our data limitation: we do not observe comprehensive micro-level choice or quantity data for the firm's competitors. But it is also supported by empirical evidence. Honka (2012) uses a survey dataset that includes individual consumer choices across auto insurers. She is then able to tease out switching cost from firm-specific preferences. She finds that the mean firm preferences are not significantly different from 0 for all companies.[69]

[67]Part of the estimation data is pre-monitoring introduction. We use the average opt-in discount for these drivers.

[68]We winzorize the re-weighting scaling factor to be between 1 and 20 to deal with outliers.

[69]Her estimate of search and switching cost is lower than our estimate. However, for the firm from which

Counterfactual demand models In this section, we show simulation results of removing key components of the demand model, as an illustration of their relative importance in determining monitoring share and the firm's profitability.

First, compared to the baseline model, the "No Safer" model assumes that drivers do not take into account the incentive effect of monitoring on reducing their risk. As shown in Figure 3.9, monitoring share drops by 6.3pp.[70] Drivers substitute to the unmonitored pool and to competitors, leading to a 1.3pp drop in unconditional monitoring share but only a 0.6pp drop in market share.[71]

Second, the "Perfect Sig." model assumes that the monitoring signal is perfect in consumers' expectation by setting σ_s to zero. The market share, unconditional and conditional monitoring shares increase by 0.4pp, 0.6pp, and 2.6pp, respectively. In reality, our specification is consistent with a dynamic framework in which firm-switching is infinitely costly within a year. This will likely overstate the effect of reclassification risk. Nevertheless, the impact of a perfect signal on demand is small compared to that of other forces.[72]

Demand frictions are the most important deterrent against monitoring participation. The third model removes firm-switching inertia, which dramatically lowers the barrier for drivers with good private risk to participate in monitoring. However, It also clears the way for drivers to explore attractive outside options. We find that the firm is able to gain market share by 12.6pp, while increasing its monitoring share by 12.1pp so that 5.9% of drivers in the market has monitoring. Lastly, we remove monitoring cost. This generates the biggest impact on monitoring by far. In particular, any driver with good private risk would prefer monitoring with any coverage within the firm. The monitoring share rises to 61.3%, with

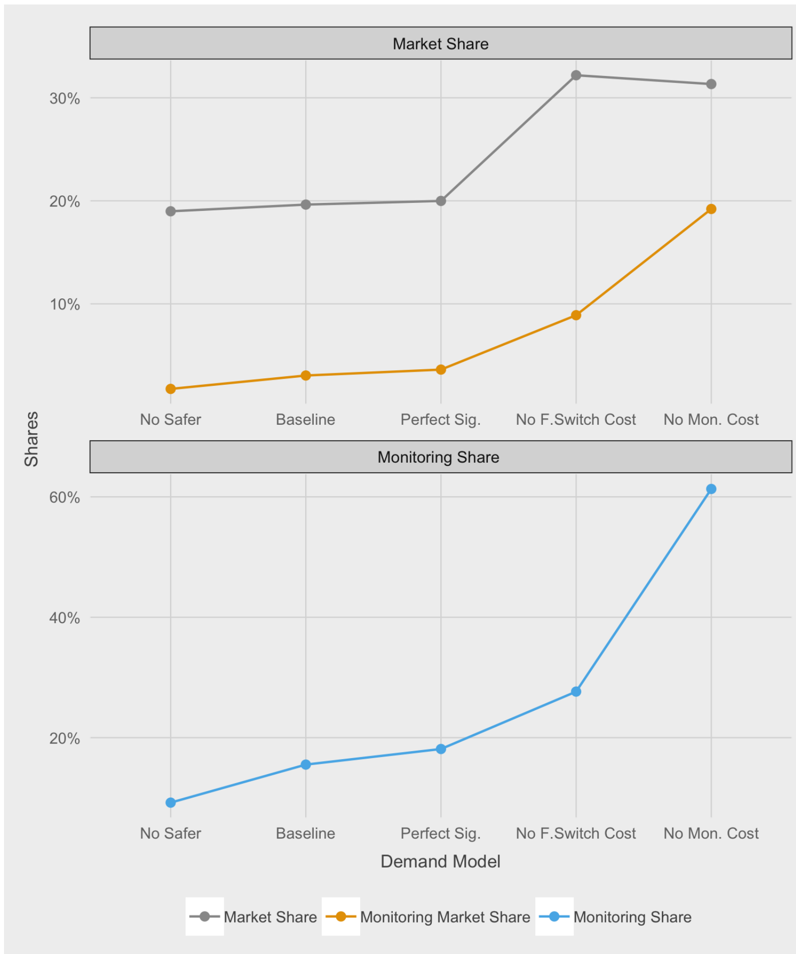
---

our administrative dataset comes from, the reported attrition rate in her dataset is more than three times as large as what we observe. Her estimate is therefore likely biased downwards.

[70] "pp" denotes percentage points.

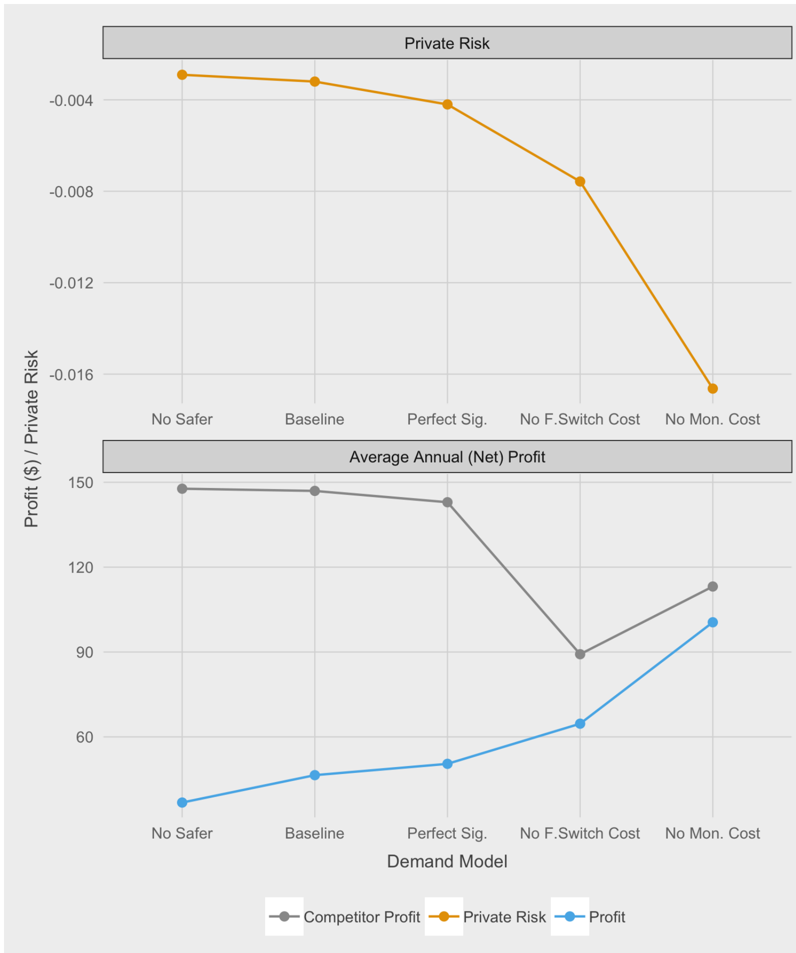
[71] Market share here is calculated as the average choice probability for the monitoring firm $f^{\star}$ in the simulation.

[72] A caveat is that we assume rational expectation in our model. This means that the effect of a systematic over- or under-estimation of the monitoring signal's noise would show up in drivers' monitoring cost instead of be attributed to reclassification risk.



Note: These figures correspond to our analyses in section 3.5.1. The top graph plots the counterfactual market share of the firm, as well as the unconditional share of monitored drivers in the market, when prices are fixed but the demand model changes. The bottom graph plots the conditional monitoring share within the firm. See main text for definitions of each model - importantly, changes in model features are *not* cumulative from left to right. We also enumerate our sample of new customers to the full market with model-predicted likelihood of each new customer being in our dataset.

Figure 3.9: *Demand Share Simulation Across Demand Model Assumptions*



Note: Corresponding to the figure above, these graphs plot firm profit and competitor profit, holding prices fixed. The top graph plots the expected private risk among the firm's customers. Notice that private risk has mean zero in the population. It is numerically integrated over in the counterfactual simulations. With each draw, we weight each person's private risk with her probability of arriving at the firm to get the number shown above. It therefore represents both the monitored and the unmonitored pools of the firm.

Figure 3.10: *Simulation - Profit Under Different Demand Model Assumptions*

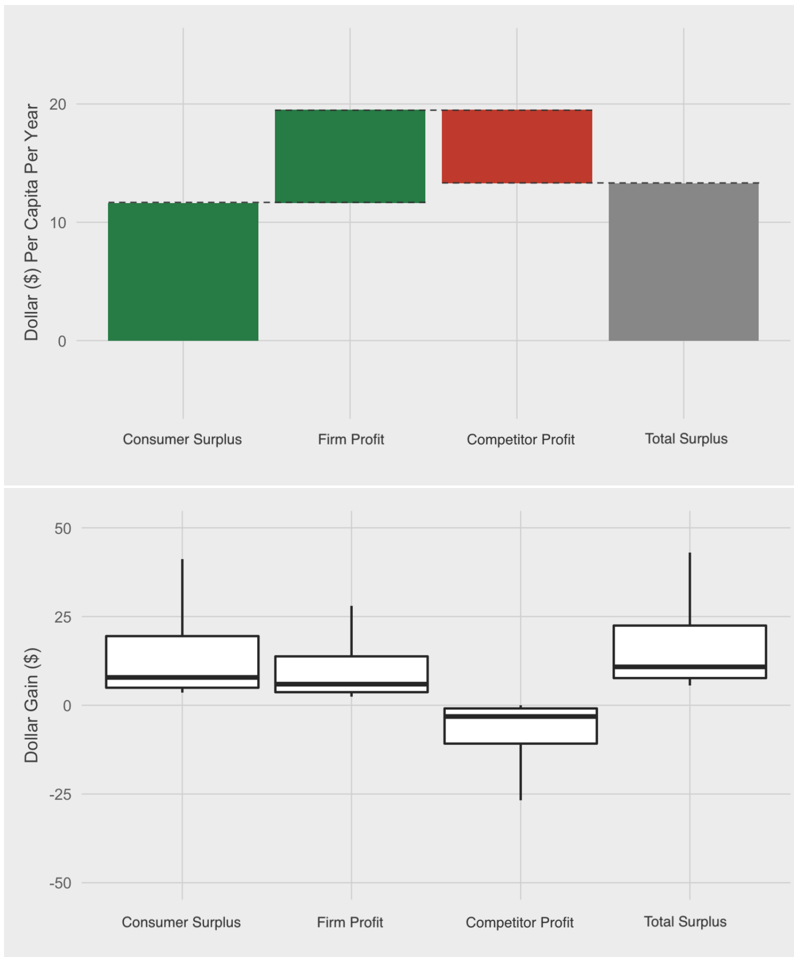
16.2% of the market opting in the firm's monitoring program.

Firm profit is influenced not only by its market share, but also by risk selection. To directly visualize this, we isolate the risk selection effect from the overall profit impact in Figure 3.10. It plots the expected private risk parameter ($\epsilon_{\lambda,i}$, mean 0) for the firm's customers, both monitored and unmonitored. This clarifies the changes in the private risk of the marginal customers that come to the firm as we relax demand factors, which is crucial in understanding competition in selection markets. As the firm cream-skims better drivers in its monitored pool, the unmonitored pool in and outside of the firm deteriorates. These pool may therefore eventually unravel as firms adjust prices.

Welfare calculation We evaluate the welfare and total surplus of introducing monitoring by comparing the current monitoring regime to a simulated counterfactual where no monitoring is offered. As mentioned above, we take a certainty equivalent approach in calculating ex-ante welfare. Total surplus is the difference between welfare and total expected cost over two periods. Profits are given by observed prices (and renewal pricing parameters) minus the same expected cost. We also take into account the resource cost for the firm to administer monitoring. It is unobserved and is difficult to estimate since actual prices may be suboptimal. In our simulations, it is set at $35 per monitored period, based on interviews with the program manager and on industry estimates. It includes manufacturing, wireless data transmission, depreciation, inventory, and mailing costs as well as R&D, marketing, and other overheads.

Figure 3.11 plots the results in per-capita per-year terms. The average consumer gains $11.6 in certainty equivalent, or 1.5% of premium. Profit increases by $7.9 per capita, a 23.6% increase. Under our symmetric cost and no-brand-preference assumptions, competitors see a profit decline of $6.2. This isolates the impact of cream skimming by the monitoring firm because the firm can offer lower prices to some monitored drivers despite charging higher markups. The combined total surplus increases by $13.3 (1.7% of premium) over the no-monitoring scenario.

To disentangle the welfare consequence of the incentive effect (risk reduction) and



Note: These figures plot results from our welfare exercise outlined in Section 3.5.1. The amount denotes the change moving from a regime where no monitoring is offered to one we observe in the data. We plot the differences in ex-ante certainty equivalent, expected profit (across two-periods) for both the monitoring firm and its competitors, as well as total surplus (welfare minus expected cost). The top graph is a waterfall graph decomposing how the components of total surplus changes. The color green indicates an increase while red indicates a decrease. The box plot show 10/25/50/75/90 percentiles.

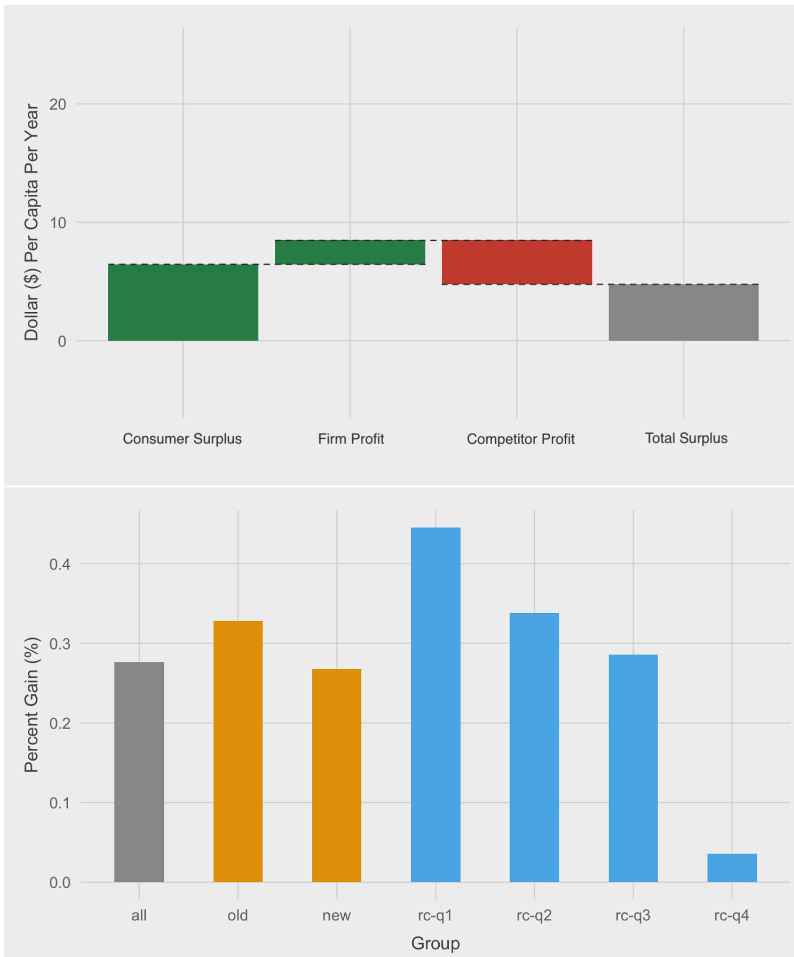
Figure 3.11: *Welfare Calculations*

allocative changes from mechanical monetary transfers across drivers, we first redo the welfare calculation without the incentive effect. Consumers' expected utility from monitoring and firms' expected cost for monitored drivers will both suffer, reducing the total surplus to \$4.8 per capita. The top panel of Figure 3.12 plots this effect. This attributes almost 64% of total surplus gain to better driving, implying small allocative efficiency gains. To investigate this further, we look at changes in the quantity of insurance purchased, comparing the observed regime with the no-monitoring one. Because liability insurance is mandatory, the result we find here is entirely due to changes in coverage levels. Overall, insurance coverage increases, but only by 0.28%. Looking across various observable pools, the safer risk classes stand out despite the fact that they already pay lower premiums. Meanwhile, without risk reduction, overall profit in the industry falls as the monitoring firm offers lower prices to good monitored drivers at the expense of its competitors' profit.

Importantly, our simulation in this section do not consider how the introduction of monitoring may have changed baseline firm prices for unmonitored drivers. This is because, as shown in Appendix C.0.2, the firm did not raise prices on the unmonitored pool during the introduction of monitoring. Therefore, any cream-skimming effect in our simulation would reduce profit in the unmonitored pool as opposed to reduce welfare of unmonitored drivers. In the next section, we propose a model for pricing where the firm can freely surcharge unmonitored drivers.

3.6 Firm Pricing Model and Equilibrium Implications

In this section, we propose a dynamic multi-product model of firm pricing that links together firm's ex-ante incentive to produce information (monitoring data) and its ex-post incentive to extract rent from the data. The model endogenizes the firm's information set and allows us to simulate two counterfactual equilibria. First, we allow the monitoring firm to optimize prices without constraints, holding competitor pricing fixed. This highlights that profit maximization implies an "invest-and-harvest" pricing dynamic. Second, we endognize competitor prices and simulate an equilibrium in which the firm is required to



Note: The top figure plots the same welfare calculation assuming away risk reduction during monitoring based on the incentive effect, per our discussion in the main text. The bottom figure plots average change in coverage amount in percentage across observable groups. “rc-q1” means risk class being in the first quartile at time of choice.

Figure 3.12: *Incentive Effect and Coverage Reallocation*

disclose monitoring data to competitors. This helps us understand the impact of regulatory proposals that aim to curb markups by restricting proprietary data.

3.6.1 Firm Pricing

In our data, the firm uses two pricing levers for the monitoring program. First, it uses upfront discounts to encourage monitoring opt-in. Second, it uses non-uniform markups in giving monitoring discounts.[73] However, actual prices for monitoring may be suboptimal for profit maximization, largely because prices are heavily regulated in the insurance industry. In order to understand the broader equilibrium implications for an unregulated market, we propose the following *two-period, two-product* model for the firm's pricing of the monitoring program.

Suppose the firm's pricing rule is driven by a vector of parameters $\vec{\kappa}$ that maximizes profit Π, which depends on aggregate demand, heterogeneous costs, and competitor prices. For illustrative simplicity, we suppress coverage choice below.[74] The firm therefore has two products: insurance with and without monitoring. Further, since all prices already takes observables x into account, we suppress that notation.

$$\begin{aligned}\Pi(\vec{\kappa}) = \sum_i \int_\lambda \Bigg\{ & \sum_m \underbrace{\Pr\left(f^\star, m \,|\, \lambda, \mathbf{p_m}, \mathbf{p}_{-f^\star}; \Theta\right)}_{\text{demand share}} \cdot \Bigg[\underbrace{\left(p_{m,0}(\vec{\kappa}|\mathbb{I}_{0,i}) - c(\lambda, m) - m \cdot c_m\right)}_{\text{markup}} \\ & + \delta \cdot \mathbb{E}_{C,s|\lambda} \Bigg[\underbrace{\Pr\left(f^\star \,|\, \lambda, p_{m,1}, \mathbf{p}_{-f^\star}; \Theta\right)}_{\text{retention rate}} \cdot \underbrace{\left(p_{m,1}(\vec{\kappa}|\mathbb{I}_{1,m,i}(C,s)) - c(\lambda, 0)\right)}_{\text{retention markup}} \Bigg] \Bigg] \Bigg\} g(\lambda) d\lambda\end{aligned} \tag{3.22}$$

The firm jointly optimizes two-period profit for all potential customers i whose latent risk types λ are distributed according to the distribution $g(\lambda)$. It forms expectations over the realization of stochastic claims C and monitoring scores s. In each period, it faces demand

[73] See Appendix C.0.3 for more details. In particular, we conduct an event study around monitoring introduction to show that the firm did not raise prices for the unmonitored pool. Meanwhile, we show that the retention elasticity drops as the firm gives more discounts.

[74] Coverage choice is incorporated in our simulation exercises and results.

and incur cost to insure drivers ($c(\lambda, m)$) and cost to monitored drivers that choose to opt in ($c_m = 35$).

Our main focus is the pricing adjustments related to monitoring, $\mathbf{p_m} = \{p_{m,0}, p_{m,1}\}$, which can change the firm's information set as they influence demand for monitoring. In the first period, the firm's information set consists of observables x and consumers' monitoring choice m. In the second period, the firm gains additional signal about drivers by observing claim realization and, for monitored drivers, the monitoring score. Competitors' information sets do not include monitoring information.

$$\mathbb{I}_{0,i} = \{x_i, m\} \qquad \mathbb{I}_{-f^\star,0,i} = \{x_i\}$$
$$\mathbb{I}_{1,m,i}(C, s) = \{x_i, C, m \cdot s\} \qquad \mathbb{I}_{-f^\star,1,i}(C) = \{x_i, C\}$$

Next, we need to specify the pricing rule $\mathbf{p_m}(\vec{\kappa}|\mathbb{I})$ given the information set. The firm's complete pricing rule is extremely complex in reality. The price filings we obstain are frequently thousands of pages long. To make the pricing problem tractable, we start from the firm's existing price rule $p(\cdot)$ observed in the data and parameterize $\vec{(}\kappa)$ as simple adjustments related to the monitoring program.

In the first period, the firm faces price competition while aiming to produce valuable information. Based on its information set, on top of the existing price schedule $p(x)$, it can surcharge the unmonitored pool by κ_0 and discount the monitored pool by κ_1. Both of which can potentially nudge drivers towards monitoring, which intuitively represent the firm's "investment" in *information production*.

$$p_{m,0}(\vec{\kappa}|\mathbb{I} = x_i, m) = \begin{cases} \kappa_0 \cdot p(x_i) & m = 0 \\ \kappa_1 \cdot p(x_i) & m = 1 \end{cases} \tag{3.23}$$

In the second period, the firm continues to face competition, but among monitored drivers, it gains an information advantageous by observing the monitoring score s. For a monitored driver that is 30% safer than previously expected, the firm may be able to offer a discount much less than 30% and still be confident that she would not leave the firm. The firm essentially solves an optimal *rent-sharing* problem with the monitored drivers in order to "harvest" the value of the collect data.

Firm's monitoring price schedule becomes continuous with the revelation of monitoring score s, which is captured by our renewal price change model $R(C, s)$. For any given score s, conditional on observables x, the wedge between this and the unmonitored price change $R(C)$ represents the firm's rent-sharing schedule observed in the data. We use a single parameter κ_s to represent linear deviations from this rent-sharing schedule.

$$p_{m,1}(\vec{\kappa}|\mathbb{I} = \{x_i, C, s\}) = \begin{cases} p(x_i) \cdot R_{m=0}(C) & m = 0 \\ p(x_i) \cdot [\kappa_s \cdot R_{m=1}(C, s) + (1 - \kappa_s) \cdot R_{m=0}(C)] & m = 1 \end{cases} \tag{3.24}$$

When $\kappa_s = 0$, the firm keeps all the rent and performance in monitoring has no impact on monitored drivers' renewal pricing. On the other hand $\kappa_s > 1$ means that the firm is sharing more rent with consumers than it does in the current regime.

3.6.2 Equilibrium Implication

Optimal pricing With the proposed model above, we can find the optimal pricing rule $\vec{\kappa}^\star$, taking demand and cost estimates as given. In particular, profit is simulated using the procedures outlined in Section 5.3.

Our results show that, in the first period, the firm should optimally surcharge the unmonitored pool by 2.7%, while offering a 22.1% upfront discount for opting into monitoring.[75]

Without competition, our model contains no outside option for consumers, given that auto insurance is mandatory. The firm can therefore arbitrarily surcharge prices for the unmonitored drivers. In contrast, our dataset only includes five competitors, yet the optimal pricing only includes a modest surcharge of 2.7% for the unmonitored pool. Price competition in the industry therefore significantly limits the firm's ability to coerce drivers into monitoring and to extract excessive rent. Instead, the large monitoring opt-in discount suggests that the firm can benefit from more investment in eliciting (producing) monitoring

[75] Consistent with our model, this discount is given for all drivers that **finish** monitoring.

data, which not only enhances ex-post competitive advantage, but it also directly reduces the cost to insure drivers in the first period.

In the renewal period, we show that optimal pricing implies 19.6% less rent-sharing than observed in the data. This means less discount for good drivers and less surcharge for bad ones, which coincidentally implies more aggressive price discrimination: good drivers receive a discount only from the monitoring firm, and are therefore less likely to leave the firm; bad ones, however, face competitive pricing without a monitoring surcharge and are therefore more price-sensitive. This pattern is documented descriptively in Appendix C.0.2.

Overall, monitoring opt-in rate increases to 4.4% (unconditional for coming to the firm). Consumer welfare and market surplus both increase. Intuitively, although the firm is taking a larger share of the surplus, it also creates more surplus in the first place by eliciting more monitoring data.

Information sharing Building on the optimal pricing regime, we now endogenize competitor prices and impose a regulation requires the firm to share its monitoring data with competitors. This turns the monitoring program into a public good. However, significant firm-switching inertia may form an effective barrier against other firms "cream-skimming" monitored drivers. In addition, the firm also directly benefits from the risk reduction during monitoring. In this section, we endogenize competitor prices and simulate an equilibrium in which competitors do not offer monitoring but can set alternative rent-sharing schedules to entice drivers who have finished monitoring.

This is similar to the patent dynamic studied in a typical innovation setting.

We make two important assumptions to facilitate this exercise. First, information sharing is complete and credible. Therefore, firms have symmetric knowledge about the expected cost of monitored drivers, given observables and monitoring score. Second, competitors do not adjust baseline prices. Instead, the focus is solely on competitors' cream-skimming motive, given the monitoring information revealed in the second period.[76] As a result, for

[76]Consumers face higher reclassification risk when their monitoring information is made public. However, due to our myopia assumption, this does not influence the attractiveness of monitoring. We see this as a fairly

this exercise, competitors need only determine a competing rent-sharing schedule $\kappa_{-f^\star,s}$.

The overall equilibrium is achieved when $\vec{\kappa}$ optimizes the firm's own profit while $\kappa_{-f^\star,s}$ optimizes competitor profit. We use a best-response algorithm to compute Nash equilibrium. We start with the optimal pricing $\vec{\kappa}^{(0)}$ we derived above and calculate the optimal competitor response $\kappa^{(0)}_{-f^\star,s}$. Taking the latter as given, we update the monitoring firm's optimal pricing to $\vec{\kappa}^{(1)}$, which is conditioned upon in calculating $\kappa^{(1)}_{-f^\star,s}$. The algorithm converges after 16 iterations with a tolerance of a total of 1-percentage-point adjustment on all four markup parameters.

Results are presented in Table 3.8. We find that competitors offer an 81% "steeper" rent-sharing schedule than what the firm offers in the current regime. The firm is then forced to share more rent with monitored drivers, by 14% compared to the current regime and by 43% compared to the optimal pricing regime. In response, the firm also significantly scales back investment in the monitoring program, offering only 8.3% opt-in discount and surcharging the unmonitored pool by 0.8%. Overall, as profit reallocates across firms, consumer welfare and total surplus decreases slightly compared to the equilibrium without the information sharing mandate (optimal pricing regime). This implies that the positive impact of information sharing on curbing ex-post markups is outweighed by the firm's adjustments in investment level, which lowers monitoring participation. This suggests that existing price competition and consumer demand frictions already significantly limit the firm's pricing power. Data regulation on proprietary data should jointly consider their markup implications and firms' incentive to produce information in the first place.

Limitations There are several important limitations to our equilibrium simulations. First, our simplistic pricing framework may not fully capture the firm's pricing structure for the monitoring program. The latter can vary nonlinearly and interact with baseline prices in complex ways. Moreover, we maintain our assumption of symmetric cost across firms for monitored drivers. In reality, however, competitors have different preexisting belief about

innocuous omission given large firm-switching inertia and our demand simulations in section 3.5.1.

	Current Regime	Optimal Pricing	Data Sharing
Firm Profit	46.5	61.2	49.3
Competitor Profit	149.2	138.2	147.1
Consumer Welfare (CE)	-	+4.7	+2.2
Total Surplus	-	+8.4	+2.9
Monitoring Market Share	3.0%	4.4%	3.4%
Invest			
Unmonitored surcharge	0.0%	2.7%	1.6%
Opt-in discount	4.6%	22.1%	8.3%
Harvest			
Rent-sharing (κ_s)	1	0.80	1.14
Competitor rent-sharing ($\kappa_{s,-f}$)	-	-	1.81

Note: This table reports results from our counterfactual equilibrium simulations in section 3.6. The simulation procedure to calculate welfare, profits, and total surplus is outlined in section 3.5.1. These quantities are reported in dollar per driver per year terms as we translate utility with a certainty equivalent approach. We further enumerate our sample of new customers to the full market by calculating driver weight as in section 3.5.1. The time frame we report is one year (two-period). The level of consumer welfare and total surplus is not identified, so we report only the change in those values in counterfactual regimes compared to the current regime. "Optimal Pricing" represents our equilibrium simulation in section 3.6.2. "Data Sharing" represents the equilibrium simulation in section 3.6.2, where the monitoring firms is required to share monitoring data to competitors. The "Current Regime" uses monitoring pricing we observe in the data. The rent-sharing parameter (κ_s) is indexed against the one observed in the "Current Regime". Empirically, it is a scalar on top of the firm's existing monitoring renewal schedule. $\kappa_s = 0$ means no rent sharing with consumers (flat pricing schedule regardless of monitoring outcome). $\kappa_s > 1$ means a steeper monitoring discount schedule than observed. This represents more rent-sharing with the consumers.

Table 3.8: *Counterfactual Equilibrium Simulations*

these drivers' risk based on their observables. Further, due to our utility assumptions, different regimes influence consumers' ex-ante welfare only by changing the prices and expected renewal prices they face at the monitoring firm. This is because they do not anticipate potential adjustments after the first period in our model; baseline competitor prices are also held fixed in the simulations. Therefore, our simulations will likely underestimate the changes in welfare and surplus across different regimes. In addition, firms' profit function do not take into account loading factor (overhead and administrative expenses unrelated to monitoring) on top of claim costs because we cannot separate loading factor from markups charged in our micro data. We therefore will overestimate the firm's profitability from attracting customers. Lastly, we restrict our simulation to two periods, as we find that the value of monitoring data diminishes dynamically (see Figure 3.6).

3.7 Conclusion

Firms are increasingly collecting consumer data in direct transactions. This influences social surplus and its division in complex ways. Beyond testing for the presence of various economic forces, it is important to quantify the underlying primitives and incentives to understand their interactions and joint effects.

In this paper, we acquire novel datasets that give us direct visibility into how valuable proprietary data are collected and used by firms. We also develop an empirical framework that links together the information market in which data transactions occur with the underlying product market. We conclude by revisiting three main results and discussing their real-world implications and caveats.

First, data collection changes consumer behavior. Drivers become 30% safer when monitored. We show that this is the primary reason why the monitoring program boosts social surplus in the short run. In other settings, consumer behavior may be distorted in a way that harms social surplus. For example, if consumers know that buying expensive items may label them as inelastic shoppers and lead to higher prices in the future, they may delay or refrain from purchasing those items. In general, firms learning about consumers can change consumer incentives and behavior, but the direction and magnitude of such distortion depends on how consumers perceive their information will be used by firms in the future.

Especially for selection markets such as insurance and lending, additional data on consumers cause differential price changes across consumers that alter allocation in the product market. In our setting, almost half of the drivers are in the state mandatory minimum plan, price adjustments therefore lead to only modest gain in allocative efficiency. This effect can be much greater in other selection markets that do not mandate participation, such as life insurance and student loans.

Second, we find that even though safer drivers are more likely to opt into monitoring, most drivers who would receive a monitoring discount (in expectation) do not. This low take-up rate is primarily driven by two factors. First, consumers suffer large disutility from being monitored. Our data does not allow us to identify the micro foundation of this disutility term. It may include "real" costs like privacy and effort costs. It can also incorporate the effect of systematic misconceptions of monitoring's benefit. In addition, it might also include the effect of salience issues related to an opt-in system. When considering a government mandate for monitoring or an opt-out mechanism, these costs will disappear. Nonetheless, our results show that in the context of direct transactions of consumer data, firms may face inelastic demand when incentivizing consumers to reveal information.

Competition in the product market also strongly influences the number of drivers choosing monitoring in equilibrium. Drivers have attractive outside options from other auto insurers due to fierce price competition. This limits the firm's ability to coerce drivers into monitoring by raising baseline prices. In many online settings, large firms hold significant market power and can afford to make their service contingent upon data collection without losing too many customers. For instance, after the EU's sweeping privacy regulation GDPR went into effect in 2018, the *Wall Street Journal* reports that large firms such as Google and Facebook achieved far higher consent rate for targeted ads than most competing online-ad services.[77] This further reinforces large firms' competitive advantage. In light of our results, the reason for their high opt-in rates is perhaps not only the value of their services but also the poor outside options consumers face.[78] More generally, our study shows that adding an additional informational demand margin can further amplify preexisting market power large firms have in the product market. Regulators should be cautious about this trade-off between consumer privacy and imperfect competition.

Lastly, the notion of privacy pertains not only to consumers' ownership of their data but also to firms' ownership of valuable proprietary data that they have collected. Our research develops a framework to jointly consider firms' incentives to "invest" in producing proprietary data and to "harvest" its value through higher markups. Our counterfactual simulation demonstrates that, in the short run, the government should protect the firm's ownership to the monitoring data in order to preserve its investment incentives to produce the data. In the long run, however, markup implications will likely dominate. The optimal regulation for proprietary data may therefore resemble a patent mechanism when the product market is sufficiently competitive and when data collection is costly but socially valuable.

[77] https://www.wsj.com/articles/eus-strict-new-privacy-law-is-sending-more-ad-money-to-google-1527759001.

[78] See this report for the opt-in process used by large multinational firms following the implementation of the GDPR.

References

Abbring, J. H., Chiappori, P.-A. and Zavadil, T. (2008). Better safe than sorry? ex ante and ex post moral hazard in dynamic insurance data.

Acemoglu, D., Cutler, D., Finkelstein, A. and Linn, J. (2006). Did Medicare Induce Pharmaceutical Innovation. *American Economic Review, Papers and Proceedings*, pp. 103–107.

— and Linn, J. (2004). Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry. *The Quarterly Journal of Economics*, **119** (3), 1049–1090.

Acquisti, A., John, L. K. and Loewenstein, G. (2013). What Is Privacy Worth? *The Journal of Legal Studies*, **42** (2), 249–274.

Acquisti, Alessandro and Taylor, Curtis and Wagman, Liad (2016a). The economics of privacy. *Journal of Economic Literature*, **54** (2), 442–92.

Acquisti, Alessandro and Taylor, Curtis and Wagman, Liad (2016b). The Economics of Privacy. *Journal of Economic Literature*, **54** (2), 442–492.

Acquisti, Alessandro and Varian, Hal R (2005). Conditioning Prices on Purchase History. *Marketing Science*, **24** (3), 367–381.

Agarwal, S., Chomsisengphet, S., Mahoney, N. and Stroebel, J. (2015). Regulating Consumer Financial Products: Evidence from Credit Cards. *The Quarterly Journal of Economics*, pp. 111–164.

—, Grigsby, J., Hortaçsu, A., Matvos, G., Seru, A. and Yao, V. (2017). Search and Screening in Credit Markets. pp. 1–77.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, **84** (3), 488.

Alesina, A. and Tabellini, G. (2007). Bureaucrats or politicians? Part I: A single policy task. *American Economic Review*, **97** (1), 169–179.

— and — (2008). Bureaucrats or politicians? Part II: Multiple policy tasks. *Journal of Public Economics*, **92** (3-4), 426–447.

Altshuler, A. and Luberoff, D. (2004). *Mega-Projects: The Changing Politics of Urban Public Investment*. Brookings Institution Press.

Andre, P., Curto, V., Duggan, M., Einav, L., Ellis, R., Gandhi, A., Gregory, J., Hendricks, K., Ho, K., Larsen, B., Levin, J., Layton, T., Polyakova, M., Riordan, M., Sorensen, A., Taber, C., Tebaldi, P. and Town, B. (2016). No Title.

Aron Dine, A., Einav, L., Finkelstein, A. and Cullen, M. (2015). Technology Diffusion and Productivity Growth in Health Care. *Review of Economics and Statistics*, **97** (4), 725–741.

Asker, J. and Cantillon, E. (2008). Properties of scoring auctions. *RAND Journal of Economics*, **39** (1), 69–85.

Athey, S. and Haile, P. (2006). Empirical models of auctions. pp. 1–48.

— and Levin, J. (2001). Information and competition in US forest service timber auctions. *Journal of Political Economy*, **109** (2), 375–417.

Attar, A., Mariotti, T. and Salanie, F. (2017). Private Information and Insurance Rejections: A Comment. *Ssrn*, **81** (5), 1713–1762.

Augenblick, N. and Bodoh-creed, A. (2017). To Reveal or Not to Reveal : Privacy Preferences and Economic Frictions.

— and Lazarus, E. (2018). Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence.

—, Niederle, M. and Sprenger, C. (2015). Working Over Time: Dynamic Inconsistency in Real Effort Tasks.

— and Rabin, M. (). An Experiment on Time Preference and Misprediction in Unpleasant Tasks. *Quarterly Journal of Economics*.

Azevedo, E. M. and Gottlieb, D. (2017). Perfect Competition in Markets With Adverse Selection. *Econometrica*, **85** (1), 67–105.

Bafumi, J. and Gelman, A. (2006). Fitting Multilevel Models When Predictors and Group Effects Correlate. *Political Science*.

Bai, J. (2018). Melons as Lemons : Asymmetric Information , Consumer Learning and Quality Provision. *Working Paper, Harvard Kennedy School*.

Bajari, P., Houghton, S. and Tadelis, S. (2014). Bidding for incomplete contracts: An empirical analysis. *American Economic Review*, **104** (April), 1288–1319(32).

—, McMillan, R. and Tadelis, S. (2009). Auctions versus negotiations in procurement: An empirical analysis. *Journal of Law, Economics, and Organization*, **25** (2), 372–399.

Barseghyan, L., Molinari, F., O'Donoghue, T. and Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, **103** (6), 2499–2529.

Benetton, M. (2018). Leverage regulation and market structure: An empirical model of the uk mortgage market. *Working Paper, University of California, Berkeley*.

Berman, R. (2015). A major infrastructure bill clears congress. *The Atlantic*.

Berry, S. T., Levinsohn, J. and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, **63** (4), 841–890.

Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*.

Bhattacharya, V., Roberts, J. W. and Sweeting, A. (2014). Regulating bidder participation in auctions. *The RAND Journal of Economics*, **45** (4), 675–704.

Bjornerstedt, J. and Verboven, F. (2016). Does merger simulation work? Evidence from the swedish analgesics market. *American Economic Journal: Applied Economics*, **8** (3), 125–164.

Bordhoff, J. E. and Noel, P. J. (2008). *Pay-as-You-Drive Auto Insurance: A Simple Way to Reduce Driving-Relayed Harms and Increase Equity. The Hamilton Project.* , Discussion paper 08-09, Brookings Institution, Washington DC.

Bulow, J. and Klemperer, P. (1996). Auctions versus negotiations. *The American Economic Review*, **86** (1), 180–194.

Bundorf, M. K., Levin, J. and Mahoney, N. (2012). Pricing and welfare in health plan choice. *American Economic Review*, **102** (7), 3214–3248.

Campo, S., Guerre, E., Perrigne, I. and Vuong, Q. (2011). Semiparametric estimation of first-price auctions with risk-averse bidders. *The Review of Economic Studies*, **78** (1), 112–147.

Cardon, J. H. and Hendel, I. (2001). Asymmetric Information in Health Insurance : Evidence from the National Medical Expenditure Survey. *The RAND Journal of Economics*, **32** (3), 408–427.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, **34** (3), 305–334.

Chaudhuri, S., Goldberg, P. K. and Jia, P. (2006). The Effects of Global Patent Protection in Estimating Pharmaceuticals : A Case Study of Quinolones in India. *American Economic Review*, **96**, 1477–1514.

Che, Y.-k. (1993). Design Competition Through Multidimensional Auctions Author. *RAND Journal of Economics*, **24** (4), 668–680.

Chiappori, P. A. and Salanie, B. (2000). Testing for Asymmetric Information in Insurance Markets. *Journal of Political Economy*, **108** (1), 56–78.

Cohen, A. (2012). Asymmetric learning in repeated contracting: An empirical study. *Review of Economics and Statistics*, **94** (2), 419–432.

Cohen, Alma and Einav, Liran (2007). Estimating Risk Preference from Deductible Choice. *American Economic Review*, **97** (1994), 745–788.

Cohen Seglias Pallas Greenhall and Furman PC (2018). Unbalanced bidding.

Cox, N. (2017). The Impact of Risk-Based Pricing in the Student Loan Market: Evidence from Borrower Repayment Decisions. *Working Paper, Princeton University*.

Cranor, L., Tsai, J. Y., Egelman, S., Cranor, L. and Acquisti, A. (2007). The Effect of Online Privacy Information on Purchasing Behavior : An Experimental Study The Effect of Online Privacy Information on Purchasing Behavior : An Experimental Study. *Information Systems Research*, **22** (January), 254–268.

Crawford, G. S., Pavanini, N. and Schivardi, F. (2018). Asymmetric information and imperfect competition in lending markets. *American Economic Review*, **108** (7), 1659–1701.

— and Yurukoglu, A. (2012). The Welfare Effects of Bundling in Multichannel Television Markets. *American Economic Review*, **102** (2), 643–685.

Crocker, K. J. and Snow, A. (1986). The Efficiency Effects of Categorical Discrimination in the Insurance Industry. *Journal of Political Economy*, **94** (2), 321–344.

Cummins, J. D. and Tennyson, S. (1996). Moral hazard in insurance claiming: Evidence from automobile insurance. *Journal of Risk and Uncertainty*, **12** (1), 29–50.

Dafny, L. (2010). Are Health Insurance Markets Competitive? *American Economic Review*, **100** (4), 1399–1431.

—, Duggan, M. and Ramanarayanan, S. (2012). Paying a premium on your premium? Consolidation in the US health insurance industry. *American Economic Review*, **102** (2), 1161–1185.

Danzon, P. and Chao, L. (2000). Prices, competition and regulation in pharmaceuticals: a cross-national comparison.

Danzon, P. M., Wang, Y. R. and Wang, L. (2005). The impact of price regulation on the launch delay of new drugs - Evidence from twenty-five major markets in the 1990s. *Health Economics*, **14** (3), 269–292.

Davis, L. W. (2016). The Distributional Effects of US Clean Energy Tax Credits.

— and Metcalf, G. E. (2016). Does Better Information Lead to Better Choices ? Evidence from Energy-Ef fi ciency Labels. **3** (3).

de Garidel-Thoron, T. (2005). Welfare Improving Asymmetric Information in Dynamic Insurance Markets. *Journal of Political Economy*, **113** (1), 121–150.

De Silva, D. G., Dunne, T., Kosmopoulou, G. and Lamarche, C. (2016). Project modifications and bidding in highway procurement auctions. *Working Paper*.

Dewatripont, M. (1989). Renegotiation and Information Revelation over Time: The Case of Optimal Labor Contracts. *The Quarterly Journal of Economics*, **104** (3), 589–619.

— and Maskin, E. (1990). Contract renegotiation in models of asymmetric information*. *European Economic Review*, **34** (2-3), 311–321.

DiMasi, J. A., Hansen, R. W. and Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of Health Economics*, **22**, 151–185.

—, — and — (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, **47**, 20–33.

—, Hansen William E Simon, R. W., Grabowski, H. G. and Lasagna, L. (1991). *Cost of innovation in the pharmaceutical industry**. Tech. rep.

Dionne, G. (2013). *Handbook of insurance: Second edition*.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, **48** (4), 935–63.

Dubois, P., de Mouzon, O., Scott Morton, F. and Seabright, P. (2015a). Market size and pharmaceutical innovation. *The RAND Journal of Economics*, **46** (4), 844–871.

—, De Mouzon, O., Scott-Morton, F. and Seabright, P. (2015b). Market size and pharmaceutical innovation. *The RAND Journal of Economics*, **46** (4), 844–871.

— and Lasio, L. (2018). Identifying Industry Margins with Price Constraints: Structural Estimation on Pharmaceuticals. *American Economic Review, forthcoming*.

Dunning, I., Huchette, J. and Lubin, M. (2017). Jump: A modeling language for mathematical optimization. *SIAM Review*, **59** (2), 295–320.

Economides, N., Hermalin, B. E., The, S., Journal, R., Winter, N. and Economides, N. (2012). The Economics of Network Neutrality. *The RAND Journal of Economics*, **43** (4), 602–629.

Einav, L., Finkelstein, A. and Levin, J. (2010). Beyond testing: Empirical models of insurance markets. *Annual Review of Economics*, **2** (1), 311–336.

Einav, Liran and Finkelstein, Amy and Ryan, Stephen P. and Schrimpf, Paul and Cullen, Mark R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, **103** (1), 178–219.

Einav, Liran and Finkelstein, Amy and Schrimpf, Paul (2010). Optimal Mandates and The Welfare Cost of Asymmetric Information: Evidence from The U.K. Annuity Market. *Econometrica*, **78** (3), 1031–1092.

Einav, Liran and Levin, Jonathan and Jenkins, Mark (2012). Contract Pricing in Consumer Credit Markets. *Econometrica*, **80** (4), 1387–1432.

European Pharmaceutical Market Research Association (2018). *Anatomical Classification Guidelines V2018*. Tech. rep.

Evdokimov, K. (2010). Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity. *Princeton University: Discussion Paper*, pp. 1–62.

Fama, E. F. (1980). Agency Problems and the Theory of the Firm. *Journal of Political Economy*, **88** (2), 288–307.

Fang, H., Keane, M. and Silverman, D. (2008). Sources of Advantageous Selection: Evidence from the Medigap Insurance Market. *Journal of Political Economy*, **116** (2), 303–350.

Farrell, J. and Klemperer, P. (2007). Chapter 31 Coordination and Lock-In: Competition with Switching Costs and Network Effects. *Handbook of Industrial Organization*, **3** (06), 1967–2072.

Farrellt, J. and Katzt, M. L. (2005). Competition or Predation? Consumer Coordination, Strategic Pricing and Price Floors in Network Markets. *The Journal of Industrial Economics*, **53** (2), 203–231.

Fershtman, C. and Pakes, A. (2012). Dynamic Games with Asymmetric Information: A Framework For Empirical Work. *The Quarterly Journal of Economics*, pp. 1611–1661.

Filson, D. (2012). A Markov-perfect equilibrium model of the impacts of price controls on the performance of the pharmaceutical industry. *RAND Journal of Economics*, **43** (1), 110–138.

Finkelstein, A. and McGarry, K. (2006). Multiple Dimensions of Private Information: Evidence from the Multiple Care Insurance Market. *American Economic Review*, **96** (4), 938–958.

—, Poterba, J. and Rothschild, C. (2009). Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities. *Journal of Financial Economics*, **91** (1), 38–58.

Flyvbjerg, B. (2009). Survival of the unfittest: Why the worst infrastructure gets built-and what we can do about it. *Oxford Review of Economic Policy*, **25** (3), 344–367.

Frankel, A. and Kartik, N. (2016). Muddled information. *Working Paper, University of Chicago, Booth School of Business and Columbia University*.

Fudenberg, D. and Villas-Boas, J. M. (2006). Behavior-based price discrimination and customer recognition. *Handbook on Economics and Information Systems*, **1**, 377–436.

— and — (2012). Price Discrimination in the Digital Economy. (June 2018), 1–22.

Goldfarb, A. and Tucker, C. E. (2011). Privacy Regulation and Online Advertising. *Management Science*, **57** (1), 57–71.

Gowrisankaran, G., Nevo, A. and Town, R. (2015). Mergers When Prices Are Negotiated: Evidence from the Hospital Industry. *American Economic Review*, **105** (1), 172–203.

— and Rysman, M. (2012). Dynamics of Consumer Demand for New Durable Goods. *Journal of Political Economy*, **120** (6), 1173–1219.

Grennan, M. (2013). Price Discrimination and Bargaining: Empirical Evidence from Medical Devices. *American Economic Review*, **103** (1), 145–177.

Grossman, G. M. and Lai, E. L. C. (2004). International Protection of Intellectual Property. **94** (5), 1635–1653.

Grossman, S. J. and Hart, O. D. (1983). An Analysis of the Principal-Agent Problem. *Econometrica*, **51** (1), 7.

— and — (1986). The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy*, **94** (4), 691–719.

Guerre, E., Perrigne, I. and Vuong, Q. (2009). Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica*, **77** (4), 1193–1227.

Hackmann, By Martin B and Kolstad, Jonathan T and Kowalski, Amanda E (2015). Adverse Selection and an Individual Mandate: When Theory Meets Practice. *American Economic Review*, **105** (3), 1030–1066.

Haile, P. A. and Tamer, E. (2003). Inference with an Incomplete Model of English Auctions. *Journal of Political Economy*, **111** (1), 1–51.

Handel, B., Hendel, I. and Whinston, M. D. (2015). Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk. *Econometrica*, **83** (4), 1261–1313.

Handel, B. R. (2013). Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts. *American Economic Review*, **No. 17459**, 1–48.

Handel, Benjamin and Kolstad, Jonathan (2017). Wearable technologies and health behaviors: New data and new methods to understand population health. *American Economic Review*, **107** (5), 481–85.

Handel, Benjamin R and Kolstad, Jonathan T (2015). Health Insurance for "Humans" - Information Frictions, Plan Choice, and Consumer Welfare. *American Economic Review*, **105** (8), 2449–2500.

Handel, Benjamin R. and Kolstad, Jonathan T and Spinnewijn, Johannes (). Information Frictions and Adverse Selection: Policy Interventions in Health Insurance Markets. *Review of Economics and Statistics*.

Handel, Benjamin R and Kolstad, Jonathan T and Spinnewijn, Johannes (forthcoming). Information frictions and adverse selection: Policy interventions in health insurance markets. *The Review of Economics and Statistics*.

Handel, By Benjamin R and Kolstad, Jonathan T (2015). Health Insurance for “ Humans ”: Information Frictions , Plan Choice , and Consumer Welfare †. **105** (8), 2449–2500.

Hart, O. (1983). Optimal labour contracts under asymmetric information: an introduction. *The Review of Economic Studies*, **50** (1), 3–35.

— and Tirole, J. (1988). Contract Renogiation and Coasian Dynamics. *Review of Economic Studies*, **55** (4), 509–540.

Helpman, E. (1993). Innovation, Imitation, and Intellectual Property Rights. *Econometrica*, **61** (6), 1247–1280.

Hendel, I. (2017). Dynamic Selection and Reclassi ... cation Risk : Theory and Empirics. In *Advances in Economics and Econometrics: Eleventh World Congress*, vol. 1, p. 99.

Hendel, Igal and Lizzeri, Alessandro (2003). The Role of Commitment in Dynamic Contracts : Evidence from Life Insurance. *The Quarterly Journal of Economics*, **118** (1), 299–327.

Hendren, N. (2013). Private Information and Insurance Rejections. *Econometrica*, **81** (5), 1713–1762.

Hermalin, B. E. and Katz, M. L. (2006). Privacy, Property Rights and Efficiency: The Economics of Privacy as Secrecy. *Quantitative Marketing and Economics*, **4**, 209–239.

— and Weisbach, M. S. (2018). American Finance Association Information Disclosure and Corporate Governance Source : The Journal of Finance , Vol . 67 , No . 1 (FEBRUARY 2012), pp . 195-233 Published by : Wiley for the American Finance Association Stable URL : https://www.jstor.org/stable/41419675 Information Disclosure and Corporate Governance. **67** (1), 195–233.

Hinze, J. (2010). *Construction Contracts*. McGraw-Hill Education.

Ho, K. and Lee, R. S. (2017). Insurer Competition in Health Care Markets. *Econometrica*, **85** (2), 379–417.

Ho, Kate and Pakes, Ariel (2014). Hospital Choices, Hospital Prices, and Financial Incentives to Physicians. *American Economic Review*, **104** (12), 3841–3884.

Holmström, B. (1999). Managerial Incentive Problems : A Dynamic Perspective. *Review of Economic Studies*, **66** (1), 169–182.

Honka, E. (2012). Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry. *The RAND Journal of Economics*, **45** (4), 847–884.

Horn, H. and Wolinsky, A. (1988). Bilateral Monopolies and Incentives for Merger. *The RAND Journal of Economics*, **19** (3), 408.

Huang, D. and Rojas, C. (2013). The Outside Good Bias in Logit Models of Demand with Aggregate Data. *Economics Bulletin*, **12** (1), 198–206.

Huang, Dongling and Rojas, Christian (2014). Eliminating the outside good bias in logit models of demand with aggregate data. *Review of Marketing Science*, **12** (1), 1–36.

Hubbard, T. (2000). The Demand for Monitoring Technologies: The Case of Trucking. *The Quarterly Journal of Economics*, **115** (2), 533–560.

Illanes, G. (2017). Switching Costs in Pension Plan Choice. *Working Paper, Northwestern University*.

Jappelli, T. (2015). Information Sharing in Credit Markets. **48** (5), 1693–1718.

Jeziorski, P., Krasnokutskaya, E. and Ceccarini, O. (2014). *Adverse Selection and Moral Hazard in the Dynamic Model of Auto Insurance*. Tech. rep., Mimeo.

Jin, G. Z. (2005). Competition and disclosure incentives: an empirical study of hmos. *RAND Journal of economics*, pp. 93–112.

Jin, Ginger Zhe (2018). *Artificial Intelligence and Consumer Privacy*. , National Bureau of Economic Research.

Jin, Ginger Zhe and Leslie, Phillip (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, **118** (2), 409–451.

Kahneman D. Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking.

Kehoe, P. J., Larsen, B. J. and Pastorino, E. (2018). Dynamic Competition in the Era of Big Data. *Working Paper, Stanford University*, (April), 1–67.

Krasnokutskaya, E. (2011). Identification and estimation of auction models with unobserved heterogeneity. *Review of Economic Studies*, **78** (1), 293–327.

Krishna, V. (2009). *Auction theory*. Academic press.

Lakdawalla, D. N. (2018). Economics of the Pharmaceutical Industry. *Journal of Economic Literature*, **56** (2), 397–449.

—, Goldman, D. P., Michaud, P. C., Sood, N., Lempert, R., Cong, Z., De Vries, H. and Gutierrez, I. (2009). U.S. pharmaceutical policy in a global marketplace. *Health Affairs*, **28** (1), 138–150.

Lebrun, B. (2006). Uniqueness of the equilibrium in first-price auctions. *Games and Economic Behavior*, **55** (1), 131–151.

Levitt, S. D. and List, J. A. (2011). Was there really a hawthorne effect at the hawthorne plant? An analysis of the original illumination experiments. *American Economic Journal: Applied Economics*, **3** (1), 224–238.

Lewis, G. and Bajari, P. (2011). Procurement contracting with time incentives: Theory and evidence. *Quarterly Journal of Economics*, **126** (3), 1173–1211.

— and — (2014). Moral hazard, incentive contracts, and risk: Evidence from procurement. *Review of Economic Studies*, **81** (3), 1201–1228.

Linnosmaa, I., Hermans, R. and Hallinen, T. (2004). Price-cost margin in the pharmaceutical industry. Empirical evidence from Finland. *The European Journal of Health Economics*, **5** (2), 122–128.

Locke, R. M. and Brause, A. (2007). Does Monitoring Improve Labor Standards? Lessons From Nike. *Industrial and Labor Relations Review*, **61** (1).

Mahoney, N. (2015). Bankruptcy as Implicit Health Insurance. *American Economic Review*, **105** (2), 710–746.

— and Weyl, E. G. (2017). Imperfect Competition in Seleciton Markets. *Review of Economics and Statistics*, **99** (4), 637–651.

Mailath, G. J. (1987). Incentive compatibility in signaling games with a continuum of types. *Econometrica*, pp. 1349–1365.

Maini, L. and Pammoli, F. (2017). Reference Pricing as a Deterrent to Entry: Evidence from the European Pharmaceutical Market.

Maskin, E. and Riley, J. (1984). Optimal auctions with risk averse buyers. *Econometrica: Journal of the Econometric Society*, pp. 1473–1518.

Masscases.com (1984). Department of labor and industries vs. boston water and sewer commission, 18 mass. app. ct. 621.

Matthews, S. (1987). Comparing auctions for risk averse buyers: A buyer's point of view. *Econometrica: Journal of the Econometric Society*, pp. 633–646.

Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, **70** (2), 583–601.

Nelson, S. T. (2018). Private information and price regulation in the us credit card market. *Working Paper, University of Chicago, Booth School of Business*, pp. 1–79.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, **69** (2), 307–342.

—, Turner, J. L. and Williams, J. W. (2016). Usage-Based Pricing and Demand for Residential Broadband. *Econometrica*, **84** (2), 411–443.

Nilson, E. (2018). Democrats still want to make infrastructure week happen. *Vox*.

Nilssen, T. (2000). Consumer lock-in with asymmetric information. *International Journal of Industrial Organization*, **18** (4), 641–666.

OECD (2017). *Health at a Glance 2017 OECD InDICatOrs*.

Padilla, A. J., Pagano, M. and Federico, N. (2018). The Society for Financial Studies Endogenous Communication Among Lenders and Entrepreneurial Incentives Published by : Oxford University Press . Sponsor : The Society for Financial Studies . Stable URL : http://www.jstor.org/stable/2962261 Endogenous Comm. **10** (1), 205–236.

Pakes, A., Porter, J., Ho, K. and Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, **83** (1), 315–334.

Peters, B., Roberts, M. J., Vuong, V. A. and Fryges, H. (2017). Estimating dynamic R&D choice: an analysis of costs and long-run benefits. *RAND Journal of Economics*, **48** (2), 409–437.

Pitsis, T. S., Kornberger, M. and Clegg, S. (2004). The art of managing relationships in interorganizational collaboration. *Management*, **7** (3), 47–67.

Porter, R. H. and Zona, J. D. (1993). Detection of Bid Rigging in Procurement Auctions. *Journal of Political Economy*, **101** (5), 255–268.

Posner, R. A. (1978). The Right of Privacy. *Georgia Law Review*, **12** (3), 393.

Posner, Richard A (1981). The Economics of Privacy. *The American Economic Review*, **71** (2).

Puelz, R. and Snow, A. (1994). Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market. *Journal of Political Economy*, **102** (2), 236.

Puentes, R. (2015). Why infrastructure matters: Rotten roads, bum economy. *Brookings*.

Rackauckas, C. and Nie, Q. (2017). Differentialequations.jl – a performant and feature-rich ecosystem for solving differential equations in julia. *Journal of Open Research Software*, **5**, 1.

Rajan, R. (1992). Insiders and outsiders: the choice between informed and arm's-lenght debt. *The Journal of Finance*, **47** (4), 1367–1400.

Reimers, I. and Shiller, B. (2018). Welfare Implications of Proprietary Data Collection: An Application to Telematics in Auto Insurance. *Working Paper*.

Salter, M. (2015). Reference Pricing: An Effective Model for the U.S. Pharmaceutical Industry? *Northwestern Journal of International Law & Business J. Int'l L. & Bus. Northwestern Journal of International Law & Business*, **35** (2).

Schmalensee, R. (1981). Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination. *The American Economic Review*, **71** (1), 242–247.

Scott-Morton, F. (1997). The Strategic Response by Pharmaceutical Firms to the Medicaid Most- Favored-Customer Rules. *The RAND Journal of Economics*, **28** (2), 269.

Sharpe, S. a. (1990). American Finance Association Asymmetric Information , Bank Lending and Implicit Contracts : A Stylized Model of Customer Relationships Author (s): Steven A . Sharpe Source : The Journal of Finance , Vol . 45 , No . 4 (Sep ., 1990), pp . 1069-1087 Publ. *Journal of Finance*, **45** (4), 1069–1087.

Shemkus, S. (2015). State's Aging Infrastucture Worries Businesses.

Shepard, M. (2014). Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange. *Working Paper, Harvard Kennedy School*, **000186**, 1–68.

Skitmore, M. and Cattell, D. (2013). On being balanced in an unbalanced world. *Journal of the Operational Research Society*, **64** (1), 138–146.

Small, K. A. and Rosen, H. S. (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica*, **49** (1), 105–130.

Somaini, P. (2015). Identification in auction models with interdependent costs.

Stark, R. M. (1974). Unbalanced highway contract tendering. *Journal of the Operational Research Society*, **25** (3), 373–388.

Stigler, G. J. . (1952). *Theory of Price*. New York: The Macmillan Company of Canada, Limited, 2nd edn.

— (1980). An Introduction to Privacy in Economics and Politics. *The Journal of Legal Studies*, **9** (4), 623–644.

Stole, L. A. (2007). Price discrimination and competition. *Handbook of industrial organization*, **3**, 2221–2299.

Tait, W. (1971). The role of the civil engineer in the planning, design and construction of a modern highway. *Proceedings of the Institution of Civil Engineers*, **49** (2), 211–220.

Taylor, C. R. (2004). Consumer Privacy and the Market for Customer Information. *The RAND Journal of Economics*, **35** (4), 631.

Taylor, F. W. (1911). the Principles of Scientific Management. *The Academy of Management Review*, **2** (1), 143.

Team, S. D. (2018a). Rstan: the r interface to stan. *R package version*, **2** (17), 3.

— (2018b). Stan modeling language users guide and reference manual. *Version*, **2**, 18.

Tebaldi, P. (2017). Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA. *Working Paper, University of Chicago*, pp. 1–78.

Thadden, E. l. V. (2004). Asymmetric Information , Bank Lending and Implicit Contracts : The Winner's Curse. *Finance Research Letters*, **1** (1), 11–23.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*, vol. 2.

US Department of Transportation (1988). *Memorandum: Bid Analysis and Unbalanced Bids*. Tech. rep., Washington DC.

Varian, H. (1985). Price Discrimination and Social Welfare. *The American Economic Review*, **75** (4), 870–875.

— (1987). Price Discrimination. *Working Paper*, (87).

— (2001). Economics of information technology. *University of California, Berkeley*, **24** (July 2001), 1–53.

Veiga, A. and Weyl, E. G. (2016). Product design in selection markets. *The Quarterly Journal of Economics*, **131** (2), 1007–1056.

Weiss, J., Hakim, P. and Degun, R. (2016). Impact of Implementing International Reference Pricing on Pharmaceutical Prices for United States Medicare. *Value in Health*, **19** (3), A89.

Wolfe, J. R. and Goddeeris, J. H. (1991). Adverse selection, moral hazard, and wealth effects in the medigap insurance market. *Journal of Health Economics*, **10** (4), 433–459.

Appendix A

Appendix to Chapter 1

A.1 Scaling Equilibrium Construction

We construct the unique pure-strategy, monotonic equilibrium of a DOT procurement auction with DOT quantities $\boldsymbol{q}^e$, bidder quantity signals $\boldsymbol{q}^b$ and variances $\boldsymbol{\sigma}^2$, DOT cost estimates $\boldsymbol{c}$, and I participating bidders. Each bidder has a privately observed efficiency type α^i, that is publicly known to have been drawn from a well-behaved probability distribution over a bounded domain $[\underline{\alpha}, \overline{\alpha}]$. We denote the CDF and pdf of this distribution by $F(\alpha)$ and $f(\alpha)$, respectively.

In particular, for our counterfactual simulations, we assume that α^i is distributed according to a bounded log-normal distribution with a mean that depends on project characteristics, and a project-type-specific variance:

$$\alpha_n^i \sim \text{LogNormal}(\mu_n^\alpha, {\sigma_n^\alpha}^2) \tag{A.1}$$

where $\mu_n^\alpha = X_n \beta_\alpha$ and σ_n^α is project-type specific. We estimate $\vec{\beta}_\alpha$ and $\vec{\sigma}_n^\alpha$ from the estimated distribution of α types, using Hamiltonian Monte Carlo with MC Stan. We continue to use $F(\cdot)$ and $f(\cdot)$ to refer to the CDF/PDF of this distribution for the remainder the derivation for notational convenience.

The equilibrium assigns a unique equilibrium score $s(\alpha)$ to each efficiency type α. It is monotonic in the sense that $s(\cdot)$ is strictly increasing in α:

$$\alpha > \alpha' \iff s(\alpha) > s(\alpha'), \text{ for each pair } \alpha, \alpha' \in [\underline{\alpha}, \overline{\alpha}] .$$

Under this condition, the probability that $s(\alpha^i)$ is smaller than $s(\alpha^j)$ in equilibrium is equal to the probability that α^i is smaller than α^j, for any α^i and α^j. We can therefore write the equilibrium expected utility of an arbitrary bidder i, using the distribution of α:

$$E[u(\pi(s(\alpha),\alpha))] = \underbrace{\left(1 - \exp\left(-\gamma \sum_{t=1}^{T} q_t^b (b_t^*(s(\alpha)) - \alpha c_t) - \frac{\gamma \sigma_t^2}{2}(b_t^*(s(\alpha)) - \alpha c_t)^2\right)\right)}_{\text{Expected utility conditional on winning}} \cdot \underbrace{(1 - F(\alpha))^{N-1}}_{\text{Prob of win w/ } s(\alpha) = \mathbf{b}^*(s(\alpha)) \cdot \mathbf{q}^e},$$

where N is the number of bidders participating in the auction. In order for $s(\cdot)$ to hold in equilibrium, it must be optimal for every bidder of efficiency type α to submit $s(\alpha)$ as her score. By the envelope theorem, this is ensured when the first order condition of expected utility with respect to $s(\alpha)$ holds:

$$\frac{\partial \mathbb{E}[u(\pi(\tilde{s}, \alpha))]}{\partial \tilde{s}}|_{\tilde{s}=s(\alpha)} = 0.$$

Evaluating the derivative and rearranging, we characterize the equilibrium score function by the solution to the Ordinary Differential Equation:

$$s'(\alpha) \sum_{t=1}^{T} \left[\left(\gamma q_t^b - \gamma^2 \sigma^2 (b_t^*(s(\alpha)) - \alpha c_t) \right) \frac{\partial b_t^*(s(\alpha))}{\partial s} \right] = [\exp(\gamma \bar{\pi}(\alpha)) - 1] \sum_{k=1}^{N-1} \frac{f(\alpha)}{1 - F(\alpha)}, \tag{A.2}$$

where $\bar{\pi}(\alpha) = \sum_{t=1}^{T} q_t^b (b_t^*(s(\alpha)) - \alpha c_t) - \frac{\gamma \sigma_t^2}{2} (b_t^*(s(\alpha)) - \alpha^i c_t)^2$ and the bidding function $\mathbf{b}(s(\alpha))$ is optimal (given α). That is, given an equilibrium score $s(\alpha)$, the bidding function solves:

$$\max_{\mathbf{b}(s(\alpha))} \left[1 - \exp\left(-\gamma \sum_{t=1}^{T} q_t^b (b_t(s(\alpha)) - \alpha c_t) - \frac{\gamma \sigma_t^2}{2} (b_t(s(\alpha)) - \alpha c_t)^2 \right) \right] \tag{A.3}$$

$$\text{s.t. } \sum_{t=1}^{T} b_t(s(\alpha)) q_t^e = s$$

$$b_t(s(\alpha)) \geq 0 \text{ for each item } t.$$

Note that for the counterfactual, we add the further restriction that the optimal bid vector be non-negative. In principle, this restriction should always hold, but we ignored it for the purpose of estimation as all observed bids are positive. For the counterfactual however, it is possible that the optimal unrestricted bids would be negative, and so it is important to include the restriction explicitly. With the additional non-negativity constraint, the convex programming problem in A.3 has no closed form solution and must be solved numerically. However, given a solution that determines which of the items have interior bids (rather than zero bids) at the optimum, the solution can be characterized as follows:

$$b_t^*(\cdot) = \max \begin{cases} \alpha c_t + \frac{q_t^b}{\gamma \sigma_t^2} + \frac{q_t^e}{\sigma_t^2 \sum_{t:b_t^*(\cdot)>0} \left[\frac{(q_t^e)^2}{\sigma_t^2} \right]} \left(s(\alpha) - \sum_{t:b_t^*(\cdot)>0} \left[\alpha c_t q_t^e + \frac{q_t^b}{\gamma \sigma_t^2} q_t^e \right] \right) \\ 0 \end{cases} \tag{A.4}$$

Note that when all items have interior bids, this is equivalent to equation 1.7. We solve the ODE in (A.2) numerically using a state-of-the-art stiff ODE solver using the DifferentialEquations library in Julia.[1] At every evaluation of equation (A.4) in the ODE solver, we compute the optimal bid vector at every score by numerically solving the program in (A.3) using the IPOPT optimization suite through JuMP framework. We then compute the partial derivative $\frac{db_t^*(\cdot)}{s}$ using the (analytical) derivative of equation (A.4), evaluated at the optimal bids found with the numerical solver.

Note that this ODE is unique up to a boundary condition. As such, to ensure that this indeed characterizes an equilibrium, we require that the highest possible efficiency type $\overline{\alpha}$ submits a score $s(\overline{\alpha})$, that provides zero profit at the optimal bidding strategy. We compute $s(\overline{\alpha})$ numerically using this critereon directly, and use this to initialize the ODE solver.

[1]We would like to particularly thank the lead developer of DifferentialEquations.jl for helping us work through numerical issues in getting this to work.

A.2 Entry Cost Proofs

Lemma 1. *Consider an auction in which N^* bidders enter in equilibrium given an entry cost K. The cost of entry K is bounded from below by the certainty equivalent of participating in the auction, absent an entry cost, when N^*+1 bidders participate. K is bounded from above by the certainty equivalent of participating in the auction absent an entry cost when N^* bidders participate.*

Proof. We break our proof into two steps. First, we argue that if a bidder of type α prefers to enter an auction at a cost of K, then:

$$(1-\exp(-\gamma\bar{\pi}(\alpha)))\cdot(1-F(\alpha))^{N^*-1} \geq 1-\exp(-\gamma K) \tag{A.5}$$

where

$$\bar{\pi}(\alpha) = \sum_{t=1}^{T} q_t^b(b_t^*(s(\alpha)) - \alpha c_t) - \frac{\gamma\sigma_t^2}{2}(b_t^*(s(\alpha)) - \alpha^i c_t)^2$$

is the bidder's certainty equivalent of profits conditional on winning the auction. This condition states that the bidder's expected utility of participating in the auction absent the entry cost K is at least as large as her utility of "keeping" K and not participating.

To see this, consider a bidder of type α and knows her type, but must still pay an entry fee of K in order to enter a given scaling auction, in which there are N^*-1 opposing bidders. In order for the bidder to prefer to enter the auction, she must expect that her utility upon entering will be higher than her utility otherwise:

$$\mathbb{E}[u(\pi(s(\alpha),\alpha))] \geq 0, \tag{A.6}$$

$$\mathbb{E}[u(\pi(s(\alpha),\alpha))] = \underbrace{(1-\exp(\gamma K))}_{\text{Utility on entering and losing}} \cdot \underbrace{\left[1-(1-F(\alpha))^{N^*-1}\right]}_{\text{Prob of losing}} +$$
$$\underbrace{\left(1-\exp\left(\gamma K - \gamma\sum_{t=1}^{T} q_t^b(b_t^*(s(\alpha)) - \alpha c_t) - \frac{\gamma\sigma_t^2}{2}(b_t^*(s(\alpha)) - \alpha^i c_t)^2\right)\right)}_{\text{Expected utility conditional on entering and winning}} \cdot \underbrace{(1-F(\alpha))^{N^*-1}}_{\text{Prob of win w/ } s(\alpha)}.$$

Substituting and rearranging inequality A.6, we obtain that the bidder prefers to enter if and only if:

$$[1-\exp(\gamma K)\cdot\exp(-\gamma\bar{\pi}(\alpha))]\cdot(1-F(\alpha))^{N^*-1}+[1-\exp(\gamma K)]\geq[1-\exp(\gamma K)]\cdot(1-F(\alpha))^{N^*-1}$$

Rearranging once more, we obtain:

$$1-\exp(\gamma K)\left[1-[1-\exp(-\gamma\bar{\pi}(\alpha))]\cdot(1-F(\alpha))^{N^*-1}\right]\geq 0$$

and so,

$$\exp(-\gamma K)\geq 1-[1-\exp(-\gamma\bar{\pi}(\alpha))]\cdot(1-F(\alpha))^{N^*-1}$$

from which we obtain

$$[1-\exp(-\gamma\bar{\pi}(\alpha))]\cdot(1-F(\alpha))^{N^*-1}\geq 1-\exp(-\gamma K).$$

as in equation A.5.

Lower Bound

We now derive a lower bound on K by considering the entry of the N^*+1st bidder, where N^* is the equilibrium number of entrants to the auction given the entry cost K. By definition of N^*, it is unprofitable (in expectation) for the N^*+1st bidder to enter. That is,

$$\int_{\underline{\alpha}}^{\overline{\alpha}}[\mathbb{E}[u(\pi(s(\tilde{\alpha}),\tilde{\alpha}))|N^*+1]\cdot f(\tilde{\alpha})]\,d\tilde{\alpha}\leq 0,$$

where $\mathbb{E}[u(\pi(s(\tilde{\alpha}),\tilde{\alpha}))|N]$ is the bidder's expected utility from entering given N total entrants (including her) if she turns out to have type $\tilde{\alpha}$, as defined above.

We proceed as follows. Let $\mathbb{E}_\alpha[\cdot]$ denote the integral over α: $\int_{\underline{\alpha}}^{\overline{\alpha}}[\cdot]\,f(\tilde{\alpha})d\tilde{\alpha}$.

$$\mathbb{E}_\alpha[\mathbb{E}[u(\pi(s(\tilde{\alpha}),\tilde{\alpha}))|N^*+1]]=$$

$$\mathbb{E}_\alpha\left[(1-\exp(\gamma K))\cdot\left(1-(1-F(\alpha))^{N^*}\right)\right]+\mathbb{E}_\alpha\left[(1-\exp(\gamma K)\exp(-\gamma\bar{\pi}(\alpha)))\cdot(1-F(\alpha))^{N^*}\right]$$

$$=1-\exp(\gamma K)\cdot\left(1-\mathbb{E}_\alpha\left[(1-F(\alpha))^{N^*}\right]+\mathbb{E}_\alpha\left[\exp(-\gamma\bar{\pi}(\alpha))\cdot(1-F(\alpha))^{N^*}\right]\right). \quad \text{(A.7)}$$

Rearranging equation (A.7), we have that if $\mathbb{E}_\alpha[\mathbb{E}[u(\pi(s(\tilde{\alpha}),\tilde{\alpha}))]|N^*+1] \leq 0$, then:

$$1 - \exp(-\gamma K) \geq \mathbb{E}_\alpha \left[(1 - \exp(-\gamma \bar{\pi}(\alpha))) \cdot (1 - F(\alpha))^{N^*} \right]. \quad \text{(A.8)}$$

That is, the utility of having K dollars is greater than a bidder's expected utility of entering the auction at zero cost when there are $N^* + 1$ total entrants. Solving inequality (A.8) for K, we obtain that the certainty equivalent of entering the auction at zero cost given $N + 1$ bidders provides a lower bound on the cost of entry.

Upper Bound

We now derive an upper bound on K by considering the entry of the N^*th bidder. By definition of N^* as the equilibrium number of entrants, it is profitable in expectation for this bidder to enter.

$$\int_{\underline{\alpha}}^{\overline{\alpha}} [\mathbb{E}[u(\pi(s(\tilde{\alpha}),\tilde{\alpha}))|N^*] \cdot f(\tilde{\alpha})] \, d\tilde{\alpha} \geq 0.$$

Writing $\mathbb{E}_\alpha[\cdot]$ for the integral over α: $\int_{\underline{\alpha}}^{\overline{\alpha}} [\cdot] f(\tilde{\alpha}) d\tilde{\alpha}$ as before, and rearranging as before, we obtain that we have that if $\mathbb{E}_\alpha[\mathbb{E}[u(\pi(s(\tilde{\alpha}),\tilde{\alpha}))]|N^*] \geq 0$, then:

$$1 - \exp(-\gamma K) \leq \mathbb{E}_\alpha \left[(1 - \exp(-\gamma \bar{\pi}(\alpha))) \cdot (1 - F(\alpha))^{N^*-1} \right]. \quad \text{(A.9)}$$

That is, the utility of having K dollars is lower than a bidder's expected utility of entering the auction at zero cost when there are N^* total entrants. Solving inequality (A.9) for K, we obtain that the certainty equivalent of entering the auction at zero cost given N bidders provides an upper bound on the cost of entry. □

A.3 Technical Details

A.3.1 Econometric Details

Let $b^d_{t,i,n}$ denote the unit bid observed by the econometrician for item t, by bidder i in auction n. Let $\theta = (\theta_1, \theta_2)$ be the vector of variables that parameterize the model prediction for each bid $b^*_{t,i,n}(\theta)$, as defined by equation 1.12. The subvector θ_1 refers to parameters estimated in the first stage, as detailed in section A.3.1. The subvector θ_2 refers to parameters estimated in the second stage, as detailed in section A.3.1. By Assumption 1, the residual of the optimal bid for each item-bidder-auction tuple with respect to its noisily observed bid: $\nu_{t,i,n} = b^d_{t,i,n} - b^*_{t,i,n}(\theta)$, is distributed identically and independently with a mean of zero across items, bidders and auctions. Furthermore, $\nu_{t,i,n}$ is orthogonal to the identity and characteristics of each item, bidder and auction.[2]

Our estimation procedure treats each auction n as a random sample from some unknown distribution. As such, auctions are exchangeable. Each auction n has an associated set of bidders who participate in the auction, $\mathcal{I}(n)$, as well as an associated set of items that receive bids in the auction, $\mathcal{T}(n)$. $\mathcal{I}(n)$ and $\mathcal{T}(n)$ are characteristics of auction n and so are drawn according to the underlying distribution over auctions themselves. For each bidder $i \in \mathcal{I}(n)$ and item $t \in \mathcal{T}(n)$, our model assigns a unique true bid $b^*_{t,i,n}(\theta)$ at the true parameter vector θ.

Items $t \in \mathcal{T}(n)$ are characterized by a $P \times 1$ vector, $X_{t,n}$, of features. Bidders $i \in \mathcal{I}(n)$ are characterized by a $J \times 1$ vector, $X_{i,n}$, of features. The construction of $X_{t,n}$ and $X_{i,n}$ is discussed in detail in section A.3.2. Estimation proceeds in two stages. In the first stage, we estimate θ_1, the subvector of parameters that governs bidders' beliefs over ex-post item quantities, using a best-predictor model estimated with Hamiltonian Monte Carlo. In the second stage, we estimate θ_2, which characterizes bidders' risk aversion and cost types, using a GMM estimator.

First Stage

In the first stage, we use the full dataset of auctions available to us in order to estimate a best-predictor model of expected item quantities conditional on DOT estimates and project-item characteristics, as well as the level of uncertainty that characterizes each projection.

[2]It is not strictly necessary to assume IIDness across bidders and items. However, allowing for further heterogeneity complicates estimation substantially and so we defer this to a robustness check using Bayesian methods in a future revision.

Each observation is an instance of a type of item t, being used in an auctioned project n. Each observation (t, n) is associated with a vector of item-auction characteristic features $X_{t,n}$, the construction of which is discussed in section A.3.2 below. For simplicity, we employ a linear model for the expected quantity of item t in auction n, $\widehat{q^b_{t,n}}$ as a function of the the DOT quantity estimate $q^e_{t,n}$ and $X_{t,n}$.[3] In order to model the level of uncertainty in the projection $\widehat{q^b_{t,n}}$, we model the distribution of the quantity model fit residuals ($\eta_{t,n} = q^a_{t,n} - \widehat{q^b_{t,n}}$) with a lognormal regression function of $q^e_{t,n}$ and $X_{t,n}$ as well. The full model specification is below. While we could fit this in two stages (first, fit the expected quantity and then fit the distribution of the residuals), we do this jointly using Hamiltonian Monte Carlo (HMC) with the MC Stan probabilistic programming language. We then take the posterior modes of the estimated distributions and use them as point estimates for the second stage.

$$q^a_{t,n} = \widehat{q^b_{t,n}} + \eta_{t,n} \text{ where } \eta_{t,n} \sim \mathcal{N}(0, \hat{\sigma}^2_{t,n}) \tag{A.10}$$

such that
$$\widehat{q^b_{t,n}} = \beta_{0,q} q^e_{t,n} + \vec{\beta}_q X_{t,n} \text{ and } \hat{\sigma}_{t,n} = \exp(\beta_{0,\sigma} q^e_{t,n} + \vec{\beta}_\sigma X_{t,n}). \tag{A.11}$$

Denote $\theta_1 = (\beta_{0,q}, \vec{\beta}_q, \beta_{0,\sigma}, \vec{\beta}_\sigma, \vec{\beta}_s, \vec{\sigma}_s)$ for the vector of first stage parameters, and let $\hat{\theta}_1$ be the posterior modes of θ_1, produced by the first stage HMC estimation. Thus, $\hat{\theta}_1$ specifies, for each item $t \in \mathcal{T}(n)$ in each auction n, the model estimate of bidders' predictions for the item's quantity: $\widehat{q^b_{t,n}}$ as well as the variance of that prediction, $\widehat{\sigma}^2_{t,n}$.

Second Stage

Denote $\theta_2 = (\gamma, \alpha^1, \ldots, \alpha^I, \beta^1_\alpha, \ldots, \beta^J_\alpha)$ for the vector of second stage parameters, where I is the number of unique firm IDs and J is the number of auction-bidder features. Note that θ_2 is $(1 + I + J)$-dimensional.

We estimate θ_2 in the second stage, using a GMM framework, evaluated at the first stage

[3] In principle, any statistical model (not necessarily a linear one) would be sound, and we intend to discuss robustness tests to the final results using different machine learning algorithms for the first stage in a future version of this paper.

estimates $\hat{\theta}_1$:

$$\theta_2 = \arg\min \mathbb{E}_n \left[g(\theta_2, \hat{\theta}_1)' W g(\theta_2, \hat{\theta}_1) \right]$$

where $g(\theta_2|\hat{\theta}_1)$ is a vector of moments, as a function of θ_2, evaluated at the estimates of θ_1 obtained in the first stage, and W is a weighting matrix. We make use of the following 3 types of moments, asymptotic in the number of auctions N. The first type of moment states that the average residual of a unit bid submitted by each (unique) bidder i is zero across auctions. There are I such moments, where I is the number of unique bidders.[4] The second type of moment states that the average residual of a unit bid submitted for a an item labeled as a"top skew item" by the DOT chief engineer's office is zero across auctions. There is one such moment. The third type of moment states that the average residual on a unit bid submitted in each auction is zero, independently of the auction-bidder characteristics of the bidder submitting the bid. There are J such moments—one for each of the auction-bidder characteristics. In total, there are $(1 + I + J)$ moments, so that the GMM estimator is just identified. As such, the choice of W does not affect efficiency, and we weight each moment equally as a default.

$$m_i^1(\theta_2|\hat{\theta}_1) = \mathbb{E}_n \left[\frac{1}{|\mathcal{T}(n)|} \sum_{t \in \mathcal{T}(n)} \tilde{v}_{t,i,n}(\theta_2|\hat{\theta}_1) \cdot \mathbf{1}_{i \in \mathcal{I}(n)} \right]$$

$$m_s^2(\theta_2|\hat{\theta}_1) = \mathbb{E}_n \left[\frac{1}{|\mathcal{I}(n)| \cdot |\mathcal{T}_s|} \sum_{i \in \mathcal{I}(n)} \sum_{t \in \mathcal{T}(n)} \tilde{v}_{t,i,n}(\theta_2|\hat{\theta}_1) \cdot \mathbf{1}_{i \in \mathcal{I}(n)} \cdot \mathbf{1}_{t \in \mathcal{T}_s} \right]$$

$$m_j^3(\theta_2|\hat{\theta}_1) = \mathbb{E}_n \left[\frac{1}{|\mathcal{I}(n)| \cdot |\mathcal{T}(n)|} \sum_{i \in \mathcal{I}(n)} \sum_{t \in \mathcal{T}(n)} \tilde{v}_{t,i,n}(\theta_2|\hat{\theta}_1) \cdot \mathbf{1}_{i \in \mathcal{I}(n)} \cdot X_{i,n}^j \right]$$

For each auction n, we denote $\mathcal{I}(n)$ as the set of bidders involved in n, $\mathcal{T}(n)$ as the set of items used in n, and $\mathcal{T}_s$ as the subset of items that were labeled as "top skew items" by the DOT chief engineer's office. All moments are formed with respect to the *de-meaned* bid

[4] To simplify notation, we do not distinguish between 'unique' bidders—e.g. bidders who appear in 30+ auctions—and rare bidders, whom we group into a single unique bidder ID for the purposes of this econometrics section. For the latter group, we treat all observations of rare bidders as observations of the same single bidder, who may enter a given auction more than once, with a different draw of auction-bidder characteristics, but the same bidder fixed effect determining his efficiency type.

residual:

$$\tilde{v}_{t,i,n}(\theta_2|\hat{\theta}_1) = b^d_{t,i,n} - \boldsymbol{\alpha^i_n(\theta_2)} \left(c_{t,n} - \frac{q^e_{t,n}}{\widehat{\sigma}^2_{t,n} \sum\limits_{p\in\mathcal{T}(n)} \left[\frac{(q^e_{p,n})^2}{\widehat{\sigma}^2_{p,n}}\right]} \left[\sum_{p\in\mathcal{T}(n)} c_{p,n} q^e_{p,n} \right] \right)$$
$$- \frac{1}{\boldsymbol{\gamma(\theta_2)}} \left(\frac{\widehat{q^b_{t,n}}}{\widehat{\sigma}^2_{t,n}} - \frac{q^e_{t,n}}{\widehat{\sigma}^2_{t,n} \sum\limits_{p\in\mathcal{T}(n)} \left[\frac{(q^e_{p,n})^2}{\widehat{\sigma}^2_{p,n}}\right]} \left[\sum_{p\in\mathcal{T}(n)} \frac{\widehat{q^b_{p,n}} q^e_{p,n}}{\widehat{\sigma}^2_{p,n}} \right] \right)$$
$$- \frac{q^e_{t,n}}{\widehat{\sigma}^2_{t,n} \sum\limits_{p\in\mathcal{T}(n)} \left[\frac{(q^e_{p,n})^2}{\widehat{\sigma}^2_{p,n}}\right]} \left[s^d_{i,n} \right].$$

The residual terms in the moments are *de-meaned* in the sense that they use the *observed* score $s^d_{i,n}$ in the formulation of the optimal bid for (t, i, n), rather than the *true* optimal score, $s^*_{i,n}$. That is, since $s^d_{i,n}$ is composed of noisily observed unit bids, the de-meaned residual $\tilde{v}_{t,i,n}(\theta_2|\hat{\theta}_1)$ omits an unobserved score error term:

$$\tilde{v}_{t,i,n} = \nu_{i,t,n} - \frac{q^e_{t,n}}{\sigma^2_{t,n} \sum\limits_{p\in\mathcal{T}(n)} \left[\frac{(q^e_{p,n})^2}{\sigma^2_{p,n}}\right]} \bar{v}_{i,n}, \tag{A.12}$$

where

$$\bar{v}_{i,n} = -\sum_{t=1}^{T_n} \nu_{t,i,n} q^e_{t,n}. \tag{A.13}$$

However, as bid residuals $\nu_{t,i,n}$ are assumed to be mean zero and independent of auction and item characteristics, $\mathbb{E}_n[\bar{v}_{i,n}]$, and the unobserved score error term is mean zero as well. Thus, the use of demeaned bid residuals does not pose a bias for our GMM estimation procedure.

In this draft, we compute standard errors for θ_2 at the point estimates of θ_1, without accounting for the uncertainty in the point estimates themselves. In this case, the asymptotic variance of θ_2 follows the standard just-identified GMM form:

$$\sqrt{n}(\hat{\theta}_2 - \theta^0_2) \xrightarrow{\text{d}} \mathcal{N}(0, V)$$

where $V = (\Gamma\Delta\Gamma)'$, for

$$\Gamma = \mathbb{E}\left[\frac{\partial g}{\partial \theta_2}(\theta_2^0, \theta_1^0)\right] \text{ and } \Delta = \mathbb{E}\left[g(\theta_2^0, \theta_1^0)g(\theta_2^0, \theta_1^0)'\right].$$

In an upcoming draft, we will revise the standard error computations to account for the uncertainty in the estimation of θ_1. In this case, the asymptotic variance will be given by the standard two-step GMM sandwich formula (see Chamberlain (1987) for reference). However, as we detail below, we compute the standard errors presented in the text by bootstrap, rather than in-sample asymptotic approximation.

Estimation Procedure

To summarize, we estimate our parameters in a two-stage procedure. In the first stage, we estimate the informational parameters that model bidders' expectations over item quantities and competing scores. In the second stage, we use a two-step optimal GMM estimator to estimate the economic parameters:

1. Estimate $\hat{\theta}_1 = (\hat{\beta}_{0,q}, \hat{\vec{\beta}}_q, \hat{\beta}_{0,\sigma}, \hat{\vec{\beta}}_\sigma, \hat{\vec{\beta}}_s, \hat{\vec{\sigma}}_s)$ and initialize θ_2

2. Solve:

$$\hat{\theta}_2 = \min_{\theta_2}\left\{\frac{1}{I}\sum_i m_i^1(\theta_2|\hat{\theta}_1)^2 + m_s^2(\theta_2|\hat{\theta}_1)^2 + \frac{1}{J}\sum_{j=1}^{J} m_j^3(\theta_2|\hat{\theta}_1)^2\right\}$$

   where I is the set of unique firm IDs, and J is the number of columns in $X_{i,n}$. This optimization problem is solved subject to the constraint that $\alpha_n^i(\theta_2)$ be non-negative for every i and n.[5]

We calculate standard errors by a bootstrap procedure. In the current version, we only bootstrap over step 2 (however, in an upcoming version, we will draw samples from the posterior distribution in step 1 so as to account for the uncertainty in the first stage

[5]This is a computationally efficient approach to impose the theoretical restriction that bidder costs are positive (so that bidders do not gain money from using materials). One could alternatively impose this through an additional moment condition. However, this would add a substantial computational burden as indicators for non-negativity are non-differentiable functions. We provide estimates without the non-negativity constraint as a robustness check. The results do not differ to an economically significant degree.

estimation). In particular, we draw auctions at random with replacement from the total set of auctions in our sample, and repeat the step 2 optimization procedure. We repeat this 1000 times. The confidence interval presented in the results section corresponds to the 5%th and 95%th percentile of the parameter estimates across the bootstrap draws.

A.3.2 Projecting Items and Bidder-Auction Pairs onto Characteristic Space

Our dataset consists of 440 bridge projects with a total of 218,110 unit bid observations. Of these, there are 2,883 unique bidder-project pairs and 29,834 unique item-project pairs. Each auction has an average of 6.55 bidders and 67.8 items. Of these, there are 116 unique bidders and 2,985 unique items (as per the DOT's internal taxonomy). In order to keep the computational burden of our estimator within manageable range, while still capturing heterogeneity across bidders and items within and across projects, we project item-project and bidder-project pairs onto characteristic space.

We first build a characteristic-space model of items as follows. The DOT codes each item observation in two ways: a 6-digit item id, and a text description of what the item is. Each item id comprises a hierarchical taxonomy of item classification. That is, the more digits two items have in common (from left to right), the closer the two items are. For example, item 866100 – also known as "100 Mm Reflect. White Line (Thermoplastic)" – is much closer to item 867100 – "100 Mm Reflect. Yellow Line (Thermoplastic)", than it is to item 853100 – "Portable Breakaway Barricade Type Iii", and even farther from item 701000 – "Concrete Sidewalk". To leverage the information in both the item ids and the description, we break the ids into digits, and tokenize the item description.[6] We then add summary statistics for each item: the relative commonness with which the item is used in projects, the average DOT cost estimate for that item, and dummies that indicate whether or not the item is frequently used in a single unit quantity, and whether the item is often ultimately not used at all.

[6]That is, we split each description up by words, clean them up and remove common “stop" words. Then we create a large dummy matrix in which entry i, j is 1 if the unique item indexed at i contains the word indexed by j in its description. We owe a big thanks to Jim Savage for suggesting this approach.

We create an item-project level characteristic matrix by combining the item characteristic matrix with project-level characteristics: the project category, the identities of the project manager, designer and engineer, the district in which the project is located, the project duration, the number of items in the project spec that the engineer has flagged for us as "commonly skewed", and the share of projects administered by the manager and engineer that over/under-ran.[7] The resulting matrix is very high dimensional, and so we project the matrix onto its principle components, and use the first 15.[8] In addition, we added 3 stand-alone project features: a dummy variable indicating whether the item is often given a single unit quantity (indicating that its quantity is particularly discrete), the historical share of observations of that item in which it was not used at all, and an indicator for whether or not the item itself is a "commonly skewed" item. The result is the matrix $X_{t,n}$, used in the estimation in equation (1.11).

To estimate the efficiency type $\alpha_{i,n}$ for each bidder-auction pair, we combine each bidder's unique firm ID with the matrix of project characteristics described above, and a matrix of project-bidder specific features. As a number of bidders only participate in a few auctions, we combine all bidders who appear in less than 10 auctions in our data set into a single firm ID. This results in 52 unique bidder IDs: 51 unique firms and one aggregate group. For project-bidder characteristics, we compute the bidder's *specialization* in each project type – the share of projects of the same type as the current project that the bidder has bid on – the bidder's *capacity* – the maximum number of DOT projects that the DOT has ever had open while bidding on another project – and the bidder's *utilization* – the share of the bidder's capacity that is filled when she is bidding on the current project. We also include dummies for whether or not the bidder is a *fringe* bidder, and whether or not the bidder's

[7]There are 11 items that have been flagged at our request by the cheif engineer: 120100: Unclassified Excavation;129600: Bridge Pavement Excavation; 220000: Drainage Structure Adjusted; 450900: Contractor Quality Control; 464000: Bitumen For Tack Coat; 472000: Hot Mix Asphalt For Miscellaneous Work; 624100: Steel Thrie Beam Highway Guard (Double Faced); 851000: Safety Controls For Construction Operations (Traffic Cones For Traffic Management); 853200: Temporary Concrete Barrier; 853403: Movable Impact Attenuator; 853800: Temporary Illumination For Work Zone (Temporary Illumination For Night Work)

[8]We have tried replicating this using more/less principle components and the results are very stable.

headquarters is located in the same district as the project at hand.[9] Our $X_{i,n}$ matrix has a total of 14 columns, and so we have a total of 66 efficiency-type parameters to identify. We use $X_{i,n}$ and the unique bidder ideas to model α_n^i in equation 1.15.

Finally, we make use of a project-level characteristic matrix X_n in our counterfactuals, in order to parametrize the distribution of efficiency types in each auction. In principle, we could use the bidder-auction matrix $X_{i,n}$ here. However, this would require each bidder to know the identities of her competitors. For the purpose of our main counterfactuals, we focus on the simpler case in which the distribution of scores is homogenous across the bidders participating in a given auction. Therefore, we construct X_n by taking an average of X_n with respect to the bidders in auction n.

[9] We define "fringe" similarly to BHT, as a firm that receives less than 1% of the total funds spent by the DOT on projects within the same project type as the auction being considered, within the scope of our dataset.

A.4 Estimation Results Tables

First Stage Model Fit

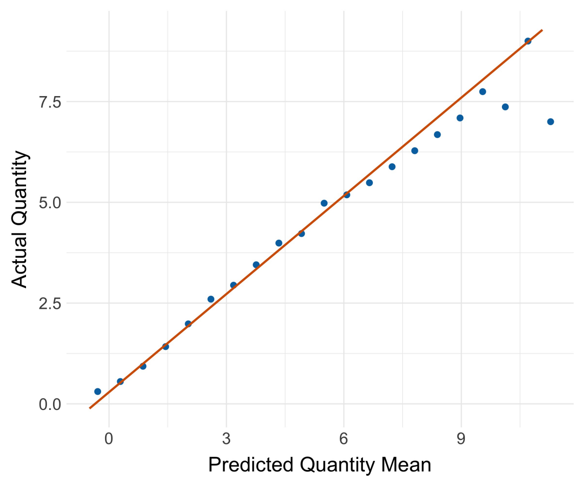


Figure A.2: *A bin scatter of actual quantities vs model predictions*

	Dependent variable:
	Actual Quantity
Predicted Quantity	0.812***
	(0.005)
Constant	0.291***
	(0.015)
Observations	29,834
R^2	0.476

Table A.1: *Regression report for figure A.1*

First Stage Parameter Estimates

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
$\beta_{0,\sigma}$	1.00	4000	-0.67	0.00	-0.67	-0.67	-0.66
$\beta_\sigma[1]$	1.00	1655	-0.05	0.01	-0.06	-0.05	-0.04
$\beta_\sigma[2]$	1.00	2120	0.02	0.00	0.01	0.02	0.03
$\beta_\sigma[3]$	1.00	3275	-0.02	0.00	-0.03	-0.02	-0.01
$\beta_\sigma[4]$	1.00	3516	0.00	0.00	-0.01	0.00	0.01
$\beta_\sigma[5]$	1.00	4000	0.02	0.00	0.01	0.02	0.03
$\beta_\sigma[6]$	1.00	3131	0.08	0.01	0.07	0.08	0.09
$\beta_\sigma[7]$	1.00	2275	0.03	0.01	0.02	0.03	0.04
$\beta_\sigma[8]$	1.00	1766	0.00	0.01	-0.01	0.00	0.01
$\beta_\sigma[9]$	1.00	1917	-0.01	0.01	-0.02	-0.01	0.00
$\beta_\sigma[10]$	1.00	1466	0.03	0.01	0.02	0.03	0.05
$\beta_\sigma[11]$	1.00	1952	-0.03	0.01	-0.04	-0.03	-0.02
$\beta_\sigma[12]$	1.00	2153	0.02	0.01	0.01	0.02	0.03
$\beta_\sigma[13]$	1.00	2590	0.04	0.01	0.03	0.04	0.05
$\beta_\sigma[14]$	1.00	2156	0.02	0.01	0.01	0.02	0.03
$\beta_\sigma[15]$	1.00	2992	0.00	0.00	-0.01	0.00	0.01
$\beta_\sigma[16]$	1.00	1856	-0.16	0.01	-0.18	-0.16	-0.15
$\beta_\sigma[17]$	1.00	4000	0.07	0.00	0.06	0.07	0.08
$\beta_\sigma[18]$	1.00	4000	0.02	0.00	0.02	0.02	0.03
$\beta_{0,q}$	1.00	4000	0.82	0.00	0.82	0.82	0.83
$\beta_q[1]$	1.00	3260	-0.02	0.00	-0.03	-0.02	-0.01
$\beta_q[2]$	1.00	4000	-0.01	0.00	-0.02	-0.01	-0.01
$\beta_q[3]$	1.00	4000	-0.03	0.00	-0.04	-0.03	-0.02
$\beta_q[4]$	1.00	4000	0.02	0.00	0.01	0.01	0.02
$\beta_q[5]$	1.00	4000	-0.02	0.00	-0.03	-0.02	-0.01
$\beta_q[6]$	1.00	4000	0.01	0.00	0.00	0.01	0.01
$\beta_q[7]$	1.00	4000	0.01	0.00	0.00	0.01	0.02
$\beta_q[8]$	1.00	2744	-0.03	0.00	-0.04	-0.03	-0.02
$\beta_q[9]$	1.00	4000	-0.03	0.00	-0.03	-0.03	-0.02
$\beta_q[10]$	1.00	2374	-0.02	0.00	-0.03	-0.02	-0.01
$\beta_q[11]$	1.00	4000	0.01	0.00	-0.00	0.01	0.01
$\beta_q[12]$	1.00	4000	-0.00	0.00	-0.01	-0.00	0.00
$\beta_q[13]$	1.00	4000	0.01	0.00	-0.00	0.01	0.01
$\beta_q[14]$	1.00	3366	0.03	0.00	0.02	0.03	0.03
$\beta_q[15]$	1.00	4000	0.01	0.00	0.00	0.01	0.02
$\beta_q[16]$	1.00	2890	0.01	0.00	0.01	0.01	0.02
$\beta_q[17]$	1.00	4000	-0.18	0.00	-0.19	-0.18	-0.17
$\beta_q[18]$	1.00	4000	-0.01	0.00	-0.02	-0.01	-0.00

Table A.2: *First Stage Parameter Estimates*

Second Stage Parameter Estimates

	Parameter Estimate	95Pct CI
$\widehat{\gamma}$	0.046	(0.032,0.264)
$\widehat{\beta}[1]$	-0.011	(-0.167,0.137)
$\widehat{\beta}[2]$	-0.003	(-0.105,0.084)
$\widehat{\beta}[3]$	0.027	(-0.138,0.063)
$\widehat{\beta}[4]$	0.017	(-0.142,0.106)
$\widehat{\beta}[5]$	-0.055	(-0.014,0.214)
$\widehat{\beta}[6]$	0.021	(-0.014,0.175)
$\widehat{\beta}[7]$	0.017	(-0.153,0.259)
$\widehat{\beta}[8]$	0.051	(-0.025,0.079)
$\widehat{\beta}[9]$	-0.060	(-0.022,0.063)
$\widehat{\beta}[10]$	-0.006	(-0.151,0.037)
$\widehat{\beta}[11]$	-0.040	(-0.027,0.107)
$\widehat{\beta}[12]$	-0.023	(-0.161,0.152)
$\widehat{\beta}[13]$	0.097	(-0.09,0.233)
$\widehat{\beta}[14]$	0.085	(-0.242,0.176)

Table A.3: *Parameter estimates for the Second Stage GMM estimation*

Second Stage Model Fit

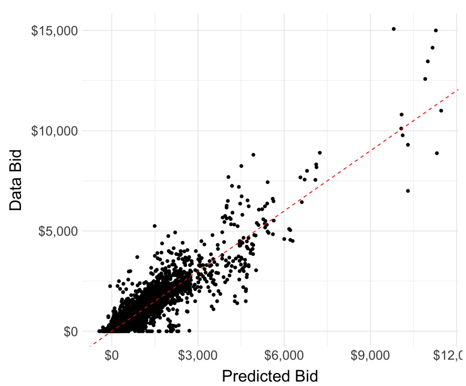
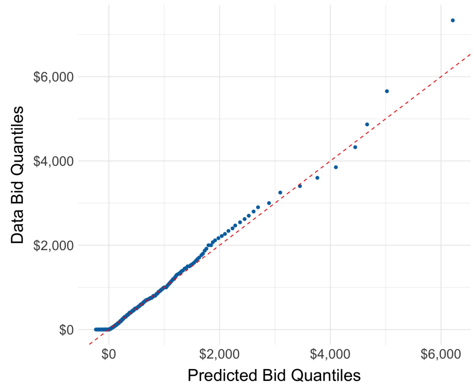


Figure A.3: *A scatter plot of actual quantities vs model predictions.*
Note: Unit bids are scaled so as to standardize quantities so exact dollar values are not representative.

Table A.4: *Regression report for figure A.3*

	Dependent variable:
	Data Bid
Predicted Bid	0.992***
	(0.001)
Constant	251.170
	(163.912)
Observations	215,332
R^2	0.879
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure A.4: *Quantile-Quantile plot of predicted bids against data bids*



Note: Quantiles are presented at the 0.0001 level and truncated at the top and bottom 0.01% for clarity. The 45-degree line is dashed in red for reference. Unit bids are scaled so as to standardize quantities so exact dollar values are not representative.

A.5 Bayesian Sampling

We make the following additional assumptions for a Bayesian approach.

First, we choose priors on the structural parameters. Note that the particular choice of priors has/is being experimented with and results do not appear to be very sensitive to it thus far.[10]

$$\gamma = \frac{1}{\gamma_{raw}} \text{ where } \gamma_{raw} \sim \mathcal{N}_+(10, 3)$$

$$\alpha_n^i \sim \mathcal{N}_+(1, 0.5)$$

The key additional assumption is the modeling of the measurement error on observed bids. For GMM, we assumed only that $b^d_{t,i,n} = b^*_{t,i,n} + \nu_{t,i,n}$ with $\mathbb{E}[\nu_{t,i,n}] = 0$. For the Bayesian approach, we model the distribution of IID draws:

$$\nu_{t,i,n} \sim \mathcal{N}(0, \sigma_b^2),$$

where σ_b is given a prior distribution and estimated.[11]

[10] We model γ as an exponential transformation to allow for higher flexibility in its level estimate while keeping the raw parameters on a similar scale for computational efficiency.

[11] We use the prior $\sigma_b \sim \mathcal{N}(0, 3)$ at the moment.

Note, however, that by the formula for $b^*_{i,t,n}$, the optimal bid (given the auction data and structural parameters), the optimal bid for each item is a function of the optimal total score $s^*_{i,n}$. We do not observe the optimal score, however - we observe only an "observed" score

$$s^d_{i,n} = \sum_{t\in\mathcal{T}(n)} b^d_{i,t,n} q^e_{t,n} \equiv s^*_{i,n} - \sum_{t\in\mathcal{T}(n)} \nu_{i,t,n} q^e_{t,n}.$$

Note that by construction, the distribution of the error on the observed score is known given the assumptions above:

$$\sum_{t\in\mathcal{T}(n)} \nu_{i,t,n} q^e_{t,n} \sim \mathcal{N}\left(0, \left(\sigma_b^2 \sum_{t\in\mathcal{T}(n)} \left[(q^e_{t,n})^2\right]\right)\right) \text{ given } q^e_{t,n} \text{ and } X_{t,n}$$

Putting these together, we model:

$$s^*_{i,n} \sim \mathcal{N}\left(s^d_{i,n}, \left(\sigma_b^2 \sum_{t\in\mathcal{T}(n)} \left[(q^e_{t,n})^2\right]\right)\right) \text{ and } b^d_{i,t,n} \sim \mathcal{N}\left(b^*_{i,t,n}(s^*_{i,n}), \sigma_b^2\right)$$

Posterior Mode Results for Bridge Auctions The following are summary statistics of the posterior mode of the HMC samples, analogous to those in section 1.6. Note that while the estimated CARA coefficient here is higher than the GMM estimate, this is in part due to the level of aggregation in the GMM estimate. While we aggregate bidders who appear in less than 10 auctions together for GMM – assigning them the same bidder-specific fixed effect – we treat each bidder-auction pair as an independent draw from the distribution of efficiency types in this estimation procedure. In an upcoming revision, we will present results for an extended Bayesian analysis in which relationships between bidder-auction draws are modeled in a hierarchical fashion, and correlations between bid errors are allowed.

$1/\widehat{\gamma}$	SE
2.097	0.165

Table A.5: *Estimates for the CARA coefficient. Note that the modal $\widehat{\gamma}$ here is 1/2.097 $\approx$ 0.48.*

Table A.6: *Summary statistics of α_n^i estimates by project type*

Project Type	Mean	Sd	Q1	Median	Q3
Bridge Reconstruction/Rehab	1.149	0.354	0.89	1.083	1.389
Bridge Replacement	1.137	0.319	0.89	1.091	1.326
Structures Maintenance	1.041	0.32	0.84	1.005	1.228

Note: Estimated $\hat{\alpha}_n^i$ are truncated at 1% before summarizing so that means do not reflect outliers.

Table A.7: *Summary statistics of estimated winning bidders' markups given alpha $\hat{\alpha}_n^i$*

Project Type	Mean	Sd	Q1	Median	Q3
Bridge Reconstruction/Rehab	9.85%	27.17%	-5.74%	3.23%	14.54%
Bridge Replacement	2.32%	19.26%	-10.83%	-0.55%	13.64%
Structures Maintenance	20.12%	47.42%	-6.56%	5.87%	30.85%

Note: Estimated $\hat{\alpha}_n^i$ are truncated at 1% before summarizing so that means do not reflect outliers.

A.6 Additional Tables and Figures

A.6.1 Distribution of Projects by Year in Our Data

A.6.2 Shares of Projects with "Unbalanced" Bids

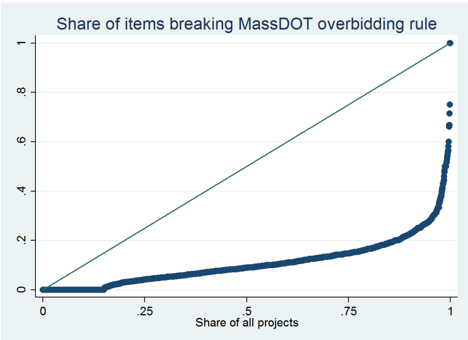
Most projects have a substantial portion of unit bids that should trigger a mathematical unbalancedness flag.

A.6.3 Discussion of Quantity Uncertainty vis-a-vis Designer and Project Manager Identities

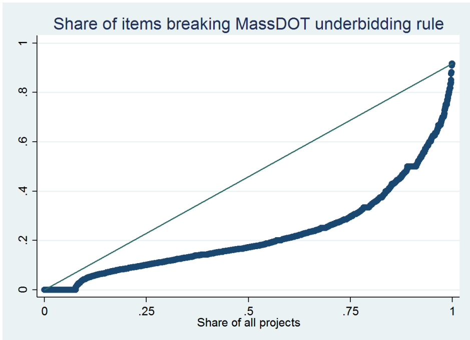
Although many factors could influence the percent overrun for each item, one factor of note is the identity of the designer, resident engineer and project manager in charge. A designer who is less experienced, for example, might be more prone to mis-estimates in the project specification. A project manager who is less experienced might be more prone to making mistakes that necessitate changes. Figures A.6a and A.6b show the average absolute value

	Year	Num Projects	Percent	Cumul Percent
1	1998	1	0.227	0.227
2	1999	5	1.136	1.364
3	2000	5	1.136	2.500
4	2001	20	4.545	7.045
5	2002	27	6.136	13.182
6	2003	26	5.909	19.091
7	2004	25	5.682	24.773
8	2005	37	8.409	33.182
9	2006	21	4.773	37.955
10	2007	32	7.273	45.227
11	2008	53	12.045	57.273
12	2009	46	10.455	67.727
13	2010	61	13.864	81.591
14	2011	32	7.273	88.864
15	2012	24	5.455	94.318
16	2013	19	4.318	98.636
17	2014	6	1.364	100

Table A.8: *Distribution of projects by year in our data*

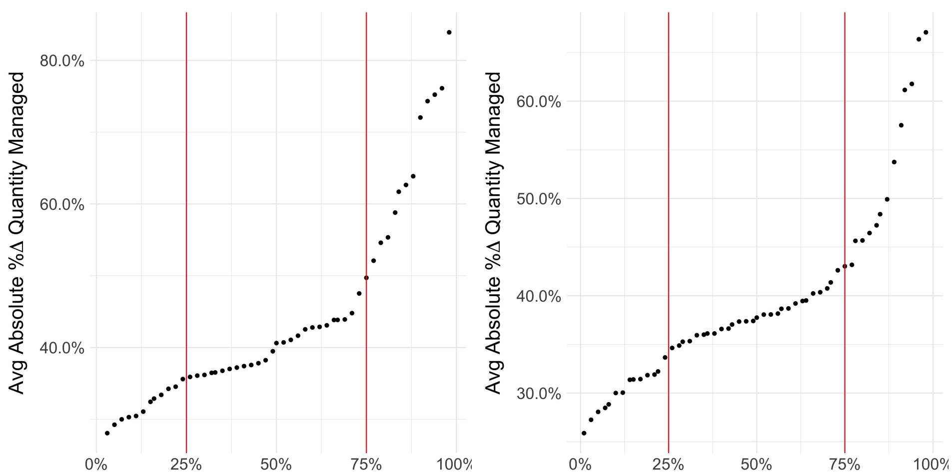


(a) *Share of projects (x-axis) that have a particular share of their items breaking the MassDOT overbidding rule (y-axis)*



(b) *Share of projects (x-axis) that have a particular share of their items breaking the MassDOT underbidding rule (y-axis)*

Figure A.5: *Plot of proportion of projects with mathematically unbalanced bids* of percent quantity overruns across items in projects managed by each project manager or designed by each designer, respectively. There are 53 unique project managers and 57 unique designers. The median project manager worked on 6 projects in our data set, with a mean of 8.4 and a maximum of 38. The median designer worked on 3 projects, with a mean of 7.8 and a maximum of 147 (this is the in-house MassDOT designer team, in contrast to the others, who are consultants). While there is not a clear relationship between absolute overruns and experience, and it is possible that the variation in overruns stems from differences in the projects that each project manager/designer is involved with, the heterogeneity in overruns across project managers and designers suggests that the choice or training of the staff employed by MassDOT could be an avenue for reducing levels of uncertainty.[12]



(a) *Project Managers* **(b)** *Designers*

Figure A.6: *Average absolute value of percent quantity overruns across items managed by each project manager (a) and each designer (b)*

A.6.4 Robustness Checks for Figure 1.10a

For robustness, we replicate figure 1.10a, without controlling for $\%\Delta q_t$:

[12]The full distributions of the number of auctions that each project manager and designer participated in, as well as a plot of average absolute overruns against the number of auctions are included in the appendix, for reference.

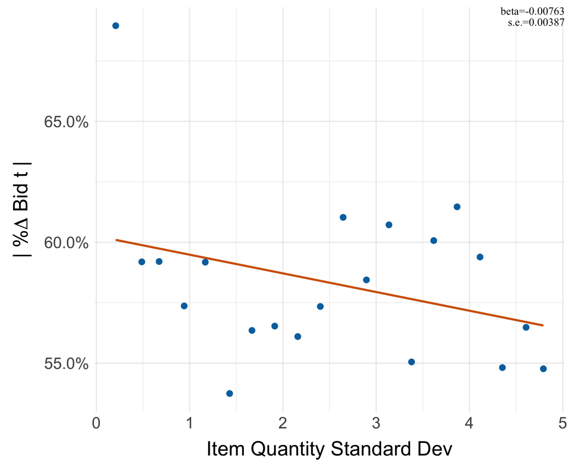


Figure A.7: *Residualized bin-scatter of item-level percent absolute overbid against the square root of estimated item quantity variance—without controlling for* $\%\Delta q_t$

A.7 Additional Discussion of the Toy Model

A.7.1 Savings from Eliminating Risk by Risk and Risk Aversion

In this section, we present additional simulation results for the toy model discussed in section 1.2. The parameters of the example are described in Table A.9 below. In Table A.10,

	DOT Estimates q^e	Bidders Expect $\mathbb{E}[q^a]$	Noise Var σ^2	Bidder Cost $\alpha \times c$
Concrete	10	12	2	12
Traffic Cones	20	16	1	18

Table A.9: *Auction parameters from the toy model*

we present the percent difference between the baseline and the counterfactual across CARA coefficients and the magnitude of the quantity noise variance. Each column corresponds to the percent savings to the DOT from the No Quantity Risk counterfactual when the baseline quantity variance term for each item is multiplied by the factor heading the column. For example, in the column labeled 0.5, the baseline equilibrium is computed with $\sigma_c^2 = 0.5 \times 2 = 1$ and $\sigma_r^2 = 0.5 \times 1 = 0.5$. Similarly, the bolded column corresponds to the last column of Table 1.2, and in the column labeled 2, the baseline equilibrium is computed with $\sigma_c^2 = 2 \times 2 = 4$ and $\sigma_r^2 = 2 \times 1 = 2$.[13]

	Magnitude of Prediction Noise			
CARA Coeff	0.1	0.5	**1**	2
0	0%	0%	**0**%	0%
0.001	0.01%	0.06%	**0.13**%	0.26%
0.005	0.06%	0.32%	**0.64**%	1.30%
0.01	0.13%	0.63%	**1.29**%	2.62%
0.05	0.60%	3.17%	**6.64**%	10.38%
0.10	1.19%	6.42%	**10.71**%	5.65%

Table A.10: *Percent DOT savings from eliminating quantity uncertainty under different levels of baseline uncertainty and bidder risk aversion*

A.7.2 Worked Out Example of Risk Neutral Bidding

Two risk-neutral bidders compete for a project that requires two types of inputs to complete: concrete and traffic cones. The DOT estimates that 10 tons of concrete and 20 traffic cones will be necessary to complete the project. However, the bidders (both) anticipate that the actual quantities that will be used – random variables that we will denote q_c^a and q_r^a for concrete and traffic cones, respectively – are distributed with means $\mathbb{E}[q_c^a] = 12$ and $\mathbb{E}[q_r^a] = 10$. We will assume that the actual quantities are exogenous to the bidding process, and do not depend on who wins the auction in any way.

The bidders differ in their private costs for the materials (including overhead, etc.): each bidder i incurs a privately known flat unit cost c_c^i for each unit of concrete and c_r^i for each traffic cone used. Thus, at the time of bidding, each bidder i expects to incur a total cost

$$\theta^i \equiv \mathbb{E}\left[q_c^a c_c^i + q_r^a c_r^i\right] = 12c_c^i + 10c_r^i,$$

if she were to win the auction. Each bidder i submits a unit bid for each of the items: b_c^i and

[13]Note that while the savings from eliminating risk are generally higher as prediction noise and risk aversion get higher, the relationship may not always be monotonic. This is because when risk and risk aversion in an auction is very high, bidders are incentivized to bid close to their costs across items so as to minimize their exposure. That is, the variance term overwhelms the prediction term. Note that this is, in part, a result of the CARA functional form.

b_r^i. The winner of the auction is then chosen on the basis of her *score*: the sum of her unit bids multiplied the DOT's quantity estimates:

$$s^i = 10b_c^i + 20b_r^i.$$

Once a winner is selected, she will implement the project and earn net profits of her unit bids, less the unit costs of each item, multiplied by the *realized* quantities of each item that are ultimately used. At the time of bidding, these quantities are unrealized samples of random variables. However, as the bidders are risk-neutral, they consider the expected value of profits to make their bidding decisions:

$$\begin{aligned} E[\pi(b_c^i, b_r^i)|c_c^i, c_r^i] &= \underbrace{\mathbb{E}\left[(q_c^a b_c^i + q_r^a b_r^i) - (q_c^a c_c^i + q_r^a c_r^i)\right]}_{\text{Expected profits conditional on winning}} \times \underbrace{\text{Prob}(s^i < s^j)}_{\text{Probability of winning}} \\ &= \left((12b_c^i + 10b_r^i) - \theta^i\right) \times \text{Prob}\left((10b_c^i + 20b_r^i) < (10b_c^j + 20b_r^j)\right). \end{aligned}$$

The key intuition for bid skewing is as follows. Suppose that the bidders' expectations of the actual quantities to be used are accurate. Then for any score s that bidder i deems competitive, she can construct unit bids that maximize her ex-post profits if she wins the auction. For example, suppose that bidder i has unit costs $c_c^i = \$70$ and $c_r^i = \$3$, and she has decided to submit a score of \$1000. She could bid her costs with a \$5 markup on concrete and a \$9.50 markup on traffic cones: $b_c^i = \$75$ and $b_r^i = \$12.50$, yielding a net profit of \$155. However, if instead, she bids $b_c^i = \$99.98$ and $b_r^i = \$0.01$, bidder i could submit the same score, but earn a profit of nearly \$330 if she wins.

This logic suggests that the DOT's inaccurate estimates of item quantities enable bidders to extract surplus profits without ceding a competitive edge. If the DOT were able to predict the actual quantities correctly, it would eliminate the possibility of bid skewing. In order for bidder i to submit a score of \$1000 in this case, she would need to choose unit bids such that $12b_c^i + 20b_r^i = \$1000$—the exact revenue that she would earn upon winning the auction. She could still bid $b_r^i = \$0.01$, for example, but then she would need to bid $b_c^i = \$83.33$, resulting in a revenue of \$1000 and a profit of \$130 if she wins the auction. A quick inspection shows that no choice of b_c^i and b_r^i could improve her expected revenue at the same score.

It would follow that when bidders have more accurate assessments of what the actual item quantities will be – as is generally considered to be the case – bids with apparent skewing *are materially* more costly to the DOT. If the bidders were to share their expectations truthfully with the DOT, it appears that a lower total cost might be incurred without affecting the level of competition.

However, this intuition does not take into account the equilibrium effect that a change in DOT quantity estimates would have on the competitive choice of score. It is not true that if a score of \$1000 is optimal for bidder i under inaccurate DOT quantity estimates, then it will remain optimal under accurate DOT estimates as well. As we demonstrate below, when equilibrium score selection is taken into consideration, the apparent possibility of extracting higher revenues by skewing unit bids is shut down entirely.

To illustrate this point, we derive the equilibrium bidding strategy for each bidder in our example. In order to close the model, we need to make an assumption about the bidders' beliefs over their opponents' costs. Note that bidder i's expected total cost for the project θ^i is fixed at the time of bidding, and does not depend on her unit bids. For simplicity, we will assume that these expected total costs are distributed according to some commonly known distribution: $\theta \sim F[\underline{\theta}, \overline{\theta}]$.

By application of Asker and Cantillon (2010), there is a unique (up to payoff equivalence) monotonic equilibrium in which each bidder of type θ submits a unique equilibrium score $s(\theta)$, using unit bids that maximize her expected profits conditional on winning, and add up to $s(\theta)$. That is, in equilibrium, each bidder i submits a vector of bids $\{b_c(\theta^i), b_r(\theta^i)\}$ such that:

$$\{b_c(\theta^i), b_r(\theta^i)\} = \arg\max_{\{b_c, b_r\}} \left\{12b_c + 40b_r - \theta^i\right\} \text{ s.t. } 10b_c + 50b_r = s(\theta^i).$$

Solving this, we quickly see that at the optimum, $b_r(\theta^i) = 0$ and $b_c(\theta^i) = s(\theta^i)/10$ (to see this, note that if $b_r = 0$, then the bidder earns a revenue of $\frac{12}{10} \cdot s(\theta^i)$ whereas if $b_c = 0$, then the bidder earns a revenue of $\frac{40}{50} \cdot s(\theta^i)$.)

The equilibrium can therefore be charcterized by the optimality of $s(\theta)$ with respect to the expected profits of a bidder with expected total cost θ:

$$E[\pi(s(\theta^i))|\theta^i] = \left(\frac{12}{10} \cdot s(\theta^i) - \theta^i\right) \cdot \text{Prob}\left(s(\theta^i) < s(\theta^j)\right) \tag{A.14}$$

$$= \left(\frac{12}{10} \cdot s(\theta^i) - \theta^i\right) \cdot \left(1 - F(\theta^i)\right), \tag{A.15}$$

where the second equality follows from the strict monotonicity of the equilibrium.[14]

As in a standard first price auction, the optimality of the score mapping is characterized by the first order condition of expected profits with respect to $s(\theta)$:

$$\frac{\partial \mathbb{E}[\pi(\tilde{s},\theta)]}{\partial \tilde{s}}|_{\tilde{s}=s(\theta)} = 0.$$

Solving the resulting differential equation, we obtain:

$$s(\theta) = \frac{10}{12}\left[\theta + \frac{\int_{\theta}^{\bar{\theta}} \left[1 - F(\tilde{\theta})\right] d\tilde{\theta}}{1 - F(\theta)}\right].$$

Thus, each bidder i will bid $b_c(\theta^i) = \frac{s(\theta^i)}{10}$ and $b_r(\theta) = 0$. If bidder i wins the auction, she expects to earn a markup of:

$$E[\pi(\theta^i)] = 12 \cdot \frac{s(\theta^i)}{10} - \theta^i \tag{A.16}$$

$$= \frac{\int_{\theta^i}^{\bar{\theta}} \left[1 - F(\tilde{\theta})\right] d\tilde{\theta}}{1 - F(\theta^i)}. \tag{A.17}$$

More generally, no matter *what* the quantities projected by the DOT are – entirely correct or wildly inaccurate – the winner of the auction and the markup that she will earn in equilibrium will be the same.

In particular, writing q_c^e and q_r^e for the DOT's quantity projections (so that a bidder's score is given by $s = b_c q_c^e + b_r q_r^e$) and q_c^b and q_r^b for the bidders' expectations for the actual

[14]More concretely, a monotonic equilibrium requires that for any $\theta' > \theta$, $s(\theta') > s(\theta)$. Therefore, the probability that $s(\theta^i)$ is lower than $s(\theta^j)$ is equal to the probability that θ^i is lower than θ^j.

quantities, the equilibrium score function can be written:

$$s(\theta) = \min\left\{\frac{q_c^e}{q_c^b}, \frac{q_r^e}{q_r^b}\right\} \cdot \left[\theta + \frac{\int_\theta^{\overline{\theta}} \left[1 - F(\tilde{\theta})\right] d\tilde{\theta}}{1 - F(\theta)}\right]. \tag{A.18}$$

Suppose that $\frac{q_r^e}{q_r^b} \leq \frac{q_c^e}{q_c^b}$. Then bidder i will bid $b_r^*(\theta^i) = \frac{s(\theta^i)}{q_r^e}$ and $b_c^*(\theta^i) = 0$. Consequently, if bidder i wins, she will be paid $q_r^b \cdot b_r^*(\theta^i) = \left[\theta^i + \frac{\int_{\theta^i}^{\overline{\theta}} \left[1-F(\tilde{\theta})\right] d\tilde{\theta}}{1-F(\theta^i)}\right]$ as in our example.

Note that the probability of winning is determined by the probability of having the lowest cost type, in equilibrium, and so this too is unaffected by the DOT's quantity estimates. That is, the level of competition and the degree of markups extracted by the bidders is determined entirely by the density of the distribution of expected total costs among the competitors. The more likely it is that bidders have similar costs, the lower the markups that the bidders can extract. However, regardless of whether the DOT posts accurate quantity estimates – in which case, bidders cannot benefit from skewing their unit bids at any score – or not, the expected cost of the project to the DOT will be the same in equilibrium. Therefore, a mathematically unbalanced bid, while indicative of a discrepancy in the quantity estimates made by the bidders and the DOT, is not indicative of a material loss to the government.

Appendix B

Appendix to Chapter 2

B.1 Appendix

B.1.1 Market Size Approximation

Method

We use Huang and Rojas (2013); Huang, Dongling and Rojas, Christian (2014) to calibrate the potential market size using a simpler logit demand model. With a logit specification, we have:

$$\ln q_{jt} - \ln q_{0mt} = \alpha_{m(j)} \ln p_{jt} + \beta_{m(j)} g_j + \lambda_{m(j)} x_j + \phi_j + \mu_{m(j)t} + \xi_{jt}$$

with $M_{mt} = q_{0t} + \sum_{j=1}^{J_m} q_{jt}$.

As q_{0mt} or M_{mt} are not observed, we can use the difference across inside goods to identify some of the parameters of the model:

$$\ln q_{jt} - \ln q_{j't} = \alpha_{m(j)} \left(\ln p_{jt} - \ln p_{j't} \right) + \beta_{m(j)} \left(g_j - g_{j'} \right) + \left(\phi_j - \phi_{j'} \right) + \left(\xi_{jt} - \xi_{j't} \right)$$

which does not depend on unobserved q_{0mt} or M_{mt} in order to identify α_m and β_m that are denoted $\hat{\alpha}_m$, $\hat{\beta}_m$ from these last specifications. For a given M_{mt} we have

$$\ln q_{jt} - \ln \left(M_{mt} - \sum_{j=1}^{J_m} q_{jt} \right) = \alpha_m \ln p_{jt} + \beta_m g_j + \lambda_m x_j + \phi_j + \mu_{mt} + \xi_{jt}$$

whose estimation with two stage least squares using the same instruments as with our BLP demand model leads to the estimates $\hat{\alpha}_m\left(M_{mt}\right)$, $\hat{\beta}_m\left(M_{mt}\right)$, $\hat{\lambda}_m\left(M_{mt}\right)$.

Then, we look for M_{mt} that solves the following minimization problem:

$$\min_{M_{mt}\geq\sum_{j=1}^{J_m} q_{jt}} \sum_{t=1}^{T}\left(\hat{\alpha}_m\left(M_{mt}\right)-\hat{\alpha}_m\right)^2+\left(\hat{\beta}_m\left(M_t\right)-\hat{\beta}_m\right)^2+\left(\hat{\lambda}_m\left(M_t\right)-\hat{\lambda}_m\right)^2$$

B.1.2 Descriptive Statistics

Table B.1: *Average Prices in the US and Canada*

		All		Patented		Branded Off		Generic	
ATC4		CA	US	CA	US	CA	US	CA	US
A10C1	H INSUL+ANG FAST ACT	12.05	37.13	12.05	37.13	0.00	0.00	0.00	0.00
A2B2	ACID PUMP INHIBITORS	0.76	2.44	0.83	2.75	0.67	3.43	0.59	0.63
B1B2	FRACTIONATED HEPARINS	15.65	37.61	15.65	38.01	0.00	30.42	0.00	28.47
C10A1	STATINS (HMG-COA RED	1.32	2.23	1.77	3.43	1.99	2.32	0.49	0.50
C2A2	ANTIHYPER.PL MAINLY PERI	0.60	1.35	46.08	13.16	2.79	2.09	0.16	1.06
C7A0	B-BLOCKING AGENTS,PLAIN	0.22	0.87	0.30	3.37	1.11	2.14	0.18	0.60
C8A0	CALCIUM ANTAGONIST PLAIN	0.93	3.40	1.30	2.32	0.83	25.38	0.49	1.17
C9A0	ACE INHIBITORS PLAIN	0.52	0.51	0.68	1.68	0.51	1.70	0.26	0.30
C9C0	ANGIOTEN-II ANTAG, PLAIN	0.97	2.31	1.10	2.72	1.19	2.64	0.25	0.47
L1A0	ALKYLATING AGENTS	17.69	135.17	24.69	229.55	1.53	109.79	14.51	48.55
L1B0	ANTIMETABOLITES	16.27	124.41	18.00	382.37	17.90	209.39	11.15	17.12
L1C0	VINCA ALKALOIDS	270.87	443.02	468.30	999.85	110.03	350.44	86.50	73.89
L1D0	ANTINEOPLAS. ANTIBIOTICS	164.08	322.95	250.68	1350.26	360.70	998.83	77.61	108.92
L1X4	A-NEO PROTEIN KINASE INH	66.23	112.85	66.35	112.77	65.16	0.00	25.55	146.05
L1X9	ALL OTH. ANTINEOPLASTICS	20.64	138.60	642.79	741.55	0.94	0.00	1.64	15.12
L2B2	CYTO ANTI-ANDROGENS	2.19	10.08	10.43	30.28	1.45	9.63	0.69	1.31
L2B3	CYTOSTAT AROMATASE INHIB	4.81	11.29	4.88	11.75	3.80	17.52	2.23	0.52
L4X0	OTHER IMMUNOSUPPRESSANTS	3.07	23.46	3.01	18.79	0.75	5.92	5.29	59.29
M1A1	ANTIRHEUMATICS NON-S PLN	0.19	0.27	0.67	3.68	0.50	0.95	0.13	0.23
M5B3	BISPHOSPH OSTEOPOROSIS	2.11	18.17	2.43	27.81	3.21	19.98	1.40	2.49
N1A1	INHAL GEN ANAESTHETICS	0.81	0.82	0.79	0.92	0.92	0.76	0.26	0.54
N1A2	INJECT GEN ANAESTHETICS	5.00	9.03	11.73	76.79	5.43	18.96	4.53	5.96
N1B1	ANAESTH LOCAL MEDIC INJ	4.48	4.35	10.97	15.64	4.63	6.76	3.19	2.84
N1B3	ANAESTH LOCAL TOPICAL	0.92	1.11	6.68	23.42	1.04	3.90	0.39	0.85
N2A0	NARCOTIC ANALGESICS	0.59	1.36	0.77	3.20	1.18	3.86	0.48	1.14
N2B0	NON-NARCOTIC ANALGESICS	0.31	0.53	0.56	14.59	0.34	1.34	0.30	0.39
N3A0	ANTI-EPILEPTICS	0.26	1.49	1.36	4.30	0.20	5.65	0.20	0.79
N5A1	ATYPICAL ANTIPSYCHOTICS	1.77	9.09	1.82	10.36	3.09	4.92	0.43	3.43
N5A9	CONVNTL ANTIPSYCHOTICS	0.46	1.27	2.90	2.76	1.17	16.30	0.25	1.11
N6A4	SSRI ANTIDEPRESSANTS	0.48	1.61	1.35	3.61	1.44	4.22	0.28	0.48
N6A9	ANTIDEPRESSANTS ALL OTH	0.19	0.63	0.40	2.95	0.58	3.47	0.13	0.32

Note: Average price by ATC-4, country, in US$ per std. unit.

Estimates

B.1.3 Supply sides estimates

Table B.2: *Outside Good Market Share Estimates by country and ATC-4*

		s_{0mt}	
ATC4		US	Canada
A10C1	H INSUL+ANG FAST ACT	0.11	0.34
A2B2	ACID PUMP INHIBITORS	0.44	0.19
B1B2	FRACTIONATED HEPARINS	0.09	0.10
C10A1	STATINS (HMG-COA RED	0.27	0.10
C2A2	ANTIHYPER.PL MAINLY PERI	0.31	0.21
C7A0	B-BLOCKING AGENTS,PLAIN	0.10	0.25
C8A0	CALCIUM ANTAGONIST PLAIN	0.18	0.09
C9A0	ACE INHIBITORS PLAIN	0.65	0.60
C9C0	ANGIOTEN-II ANTAG, PLAIN	0.44	0.10
L1A0	ALKYLATING AGENTS	0.11	0.17
L1B0	ANTIMETABOLITES	0.13	0.27
L1C0	VINCA ALKALOIDS	0.59	0.39
L1D0	ANTINEOPLAS. ANTIBIOTICS	0.09	0.21
L1X4	A-NEO PROTEIN KINASE INH	0.17	0.54
L1X9	ALL OTH. ANTINEOPLASTICS	0.14	0.14
L2B2	CYTO ANTI-ANDROGENS	0.46	0.12
L2B3	CYTOSTAT AROMATASE INHIB	0.35	0.17
L4X0	OTHER IMMUNOSUPPRESSANTS	0.47	0.20
M1A1	ANTIRHEUMATICS NON-S PLN	0.12	0.14
M5B3	BISPHOSPH OSTEOPOROSIS	0.17	0.44
N1A1	INHAL GEN ANAESTHETICS	0.10	0.27
N1A2	INJECT GEN ANAESTHETICS	0.11	0.66
N1B1	ANAESTH LOCAL MEDIC INJ	0.10	0.26
N1B3	ANAESTH LOCAL TOPICAL	0.84	0.51
N2A0	NARCOTIC ANALGESICS	0.22	0.18
N2B0	NON-NARCOTIC ANALGESICS	0.09	0.09
N3A0	ANTI-EPILEPTICS	0.48	0.12
N5A1	ATYPICAL ANTIPSYCHOTICS	0.18	0.14
N5A9	CONVNTL ANTIPSYCHOTICS	0.19	0.63
N6A4	SSRI ANTIDEPRESSANTS	0.11	0.92
N6A9	ANTIDEPRESSANTS ALL OTH	0.21	0.39

Note: Estimated outside good market shares obtained form the market size estimates by ATC-4 , country and quarter. Table presents average across quarters.

Table B.3: *Margins Estimates by ATC-4*

Margins		Canada				US		
ATC4	Label	All	On Patent	Branded Off Patent	Generics	All	On Patent	Branded Off Patent
A10C1	H INSUL+ANG FAST ACT	18.46	18.46			85.71	85.71	
A2B2	ACID PUMP INHIBITORS	37.81	38.58	28.47	83.87	54.04	53.87	63.67
B1B2	FRACTIONATED HEPARINS	38.61	40.25	0.00	0.00	96.56	96.95	100.00
C10A1	STATINS (HMG-COA RED	60.64	59.13	43.99	82.24	54.01	58.66	62.88
C2A2	ANTIHYPER.PL MAINLY PERI	49.10	96.35	96.62	16.52	11.30	37.65	11.83
C7A0	B-BLOCKING AGENTS,PLAIN	19.68	4.64	16.35	25.92	19.62	48.86	47.56
C8A0	CALCIUM ANTAGONIST PLAIN	25.12	73.56	10.69	21.73	39.46	66.31	45.52
C9A0	ACE INHIBITORS PLAIN	62.75	46.83	93.93	36.20	20.56	23.31	29.53
C9C0	ANGIOTEN-II ANTAG, PLAIN	47.54	44.63	96.05	43.82	51.47	53.49	10.64
L1A0	ALKYLATING AGENTS	11.69	13.03	1.41	10.95	39.69	47.46	46.62
L1B0	ANTIMETABOLITES	8.10	6.93	5.70	19.50	42.26	46.80	45.23
L1C0	VINCA ALKALOIDS	50.78	45.88	50.09	88.92	42.84	47.19	37.88
L1D0	ANTINEOPLAS. ANTIBIOTICS	31.08	45.37	20.93	25.08	35.24	46.35	49.96
L1X4	A-NEO PROTEIN KINASE INH	32.23	31.83	55.66	3.73	52.71	52.90	0.00
L1X9	ALL OTH. ANTINEOPLASTICS	13.40	14.06	76.78	3.61	42.34	46.55	0.00
L2B2	CYTO ANTI-ANDROGENS	35.18	29.57	50.79	77.99	49.43	54.39	48.54
L2B3	CYTOSTAT AROMATASE INHIB	50.67	50.58	29.19	73.32	63.92	64.26	47.59
L4X0	OTHER IMMUNOSUPPRESSANTS	19.66	33.59	0.65	2.73	28.20	47.87	48.10
M1A1	ANTIRHEUMATICS NON-S PLN	55.06	42.59	88.46	47.57	9.52	25.50	45.53
M5B3	BISPHOSPH OSTEOPOROSIS	6.78	4.27	17.31	27.39	57.62	62.06	48.38
N1A1	INHAL GEN ANAESTHETICS	62.35	41.64	94.54	17.15	64.71	73.89	45.85
N1A2	INJECT GEN ANAESTHETICS	20.27	13.40	16.56	24.06	25.12	47.01	58.07
N1B1	ANAESTH LOCAL MEDIC INJ	79.76	67.59	99.11	63.82	23.45	51.97	24.55
N1B3	ANAESTH LOCAL TOPICAL	68.36	47.28	71.38	28.14	5.34	1.82	6.54
N2A0	NARCOTIC ANALGESICS	40.15	49.76	47.88	37.55	10.30	11.35	46.03
N2B0	NON-NARCOTIC ANALGESICS	56.38	6.47	93.59	70.32	13.78	46.96	43.05
N3A0	ANTI-EPILEPTICS	29.69	20.77	23.32	40.51	25.82	45.45	44.93
N5A1	ATYPICAL ANTIPSYCHOTICS	18.67	9.06	92.70	20.99	50.42	53.94	4.12
N5A9	CONVNTL ANTIPSYCHOTICS	12.59	60.06	15.83	8.27	6.64	18.12	45.48
N6A4	SSRI ANTIDEPRESSANTS	2.35	1.37	2.15	6.29	46.86	58.64	47.95
N6A9	ANTIDEPRESSANTS ALL OTH	18.77	12.00	9.91	28.85	27.27	48.15	50.76

Note: Average margins in percentage of US average price by ATC-4 across all quarters. Average across drugs within category is weighted by market share. For generics in the US we impose price equal to marginal costs and do not estimate margins but they are taken into account in the average margin for all drugs in the US.

Table B.4: *Estimates of ρ_{jm} by ATC-4*

ATC4		On Patent	Branded Off	Generic
A10C1	H INSUL+ANG FAST ACT	0.62		
A2B2	ACID PUMP INHIBITORS	0.55	0.90	0.87
B1B2	FRACTIONATED HEPARINS	0.70		
C10A1	STATINS (HMG-COA RED	0.54	1.00	0.77
C2A2	ANTIHYPER.PL MAINLY PERI	1.00	1.00	0.94
C7A0	B-BLOCKING AGENTS,PLAIN	0.72	1.00	1.00
C8A0	CALCIUM ANTAGONIST PLAIN	0.56	0.89	0.86
C9A0	ACE INHIBITORS PLAIN	0.47	0.95	1.00
C9C0	ANGIOTEN-II ANTAG, PLAIN	0.60	0.94	0.50
L1A0	ALKYLATING AGENTS	0.91	0.50	1.00
L1B0	ANTIMETABOLITES	0.64	0.50	1.00
L1C0	VINCA ALKALOIDS	0.50	0.50	0.98
L1D0	ANTINEOPLAS. ANTIBIOTICS	0.99	0.50	0.50
L1X4	A-NEO PROTEIN KINASE INH	1.00	0.50	0.50
L1X9	ALL OTH. ANTINEOPLASTICS	0.92	0.50	0.57
L2B2	CYTO ANTI-ANDROGENS	0.83	0.94	0.61
L2B3	CYTOSTAT AROMATASE INHIB	0.70	0.79	0.58
L4X0	OTHER IMMUNOSUPPRESSANTS	0.95	0.91	1.00
M1A1	ANTIRHEUMATICS NON-S PLN	0.44	0.91	1.00
M5B3	BISPHOSPH OSTEOPOROSIS	0.93	0.95	0.54
N1A1	INHAL GEN ANAESTHETICS	0.45	0.57	1.00
N1A2	INJECT GEN ANAESTHETICS	1.00	1.00	0.92
N1B1	ANAESTH LOCAL MEDIC INJ	0.96	1.00	0.75
N1B3	ANAESTH LOCAL TOPICAL	0.50	0.50	0.58
N2A0	NARCOTIC ANALGESICS	0.51	0.78	0.89
N2B0	NON-NARCOTIC ANALGESICS	0.50	0.96	0.88
N3A0	ANTI-EPILEPTICS	0.87	0.93	1.00
N5A1	ATYPICAL ANTIPSYCHOTICS	0.86	0.86	0.94
N5A9	CONVNTL ANTIPSYCHOTICS	0.64	0.97	0.94
N6A4	SSRI ANTIDEPRESSANTS	0.80	0.99	0.91
N6A9	ANTIDEPRESSANTS ALL OTH	0.27	0.89	0.99

B.1.4 Theoretical Result

B.1.5 Theoretical Result

This section is meant to show that an international reference pricing policy can only increase price in the referenced country and decrease it in the referencing country. We show it under "regularity" conditions of the profit function and conditions where the same drugs are present in the referencing and referenced country. We start by showing it when we have a monopoly drug in each country, then when we have a duopoly. It is straightforward from the proof in duopoly that it extends to markets with N firms.

Monopoly case

Let's start with monopoly firms in each country A and B.

Consider one firm producing a product, at marginal costs c. Denote $D_A(p_A)$ and $D_B(p_B)$ the demands in countries A and B, respectively, when their prices are p_A and p_B. We assume that each profit function $\Pi_A(p_A) \equiv (p_A - c) D_A(p_A)$ and $\Pi_B(p_B) \equiv (p_B - c) D_B(p_B)$ is strictly concave in price and have a finite maximum above marginal cost.

Under regulation, we suppose that a governmental agency negotiates price by engaging in Nash bargaining with the firm. The governmental's objective function takes the general form $W(p_B)$ in country B, where $W(.)$ is decreasing over $[c, +\infty)$. For instance, $W(p_B)$ could be consumer surplus, social welfare or coverage.

Thus, the unregulated price in country A solves

$$p_A^* = \arg\max_{c \leq p_A} \Pi_A(p_A)$$

and the price in country B under regulation solves the following maximization program:

$$p_B^* = \arg\max_{p_B \geq c} \Pi_B(p_B)^{1-\alpha} \Delta W(p_B)^{\alpha}$$

where $\Delta W(p_B) \equiv W(p_B) - W(\infty)$ is decreasing in p_B and $\alpha \in (0, 1]$ captures the bargaining power of the governmental agency.

Now with international reference pricing imposing that the firm can sell in country A only if $p_A \leq p_B$, the new price equilibrium (p_A^{**}, p_B^{**}) simultaneously solves:

$$\begin{cases} p_A^{**} = \tilde{p}_A(p_B^{**}) \equiv \underset{c \leq p_A \leq p_B^{**}}{\arg\max} \Pi_A\left(p_A\right) \\ p_B^{**} = \underset{p_B \geq c}{\arg\max} \left[\Pi_A\left(\tilde{p}_A(p_B)\right) + \Pi_B\left(p_B\right) - \Pi_A\left(p_A^*\right)\right]^{1-\alpha} \Delta W\left(p_B\right)^{\alpha} \end{cases}$$

where $\Pi_A\left(\tilde{p}_A(p_B)\right) + \Pi_B\left(p_B\right)$ is the firm profit in A and B if selling if both countries and $\Pi_A\left(p_A^*\right)$ is the firm profit in A only if disagreeing with B.

Proposition The international reference pricing policy implies that the price in country A decreases and the price in country B increases:

$$p_A^{**} \leq p_A^* \quad \text{and} \quad p_B^{**} \geq p_B^*$$

Proof *Let's start with proving that* $p_A^{**} \leq p_A^*$*:*

From its definition, $p_A^{**} \equiv \tilde{p}_A(p_B^{**}) = p_A^*$ if $p_A^* \leq p_B^{**}$. If $p_A^* > p_B^{**}$, then $p_A^{**} \equiv \tilde{p}_A(p_B^{**}) \leq p_B^{**}$ because $\tilde{p}_A(p) \leq p$ for all p and thus $p_A^{**} < p_A^*$. This proves that in all cases $p_A^{**} \leq p_A^*$.

Let's prove now that $p_B^{**} \geq p_B^*$*:*

Let's define

$$\Delta\Pi_A\left(p_A^*, p_B)\right) \equiv \Pi_A\left(\tilde{p}_A(p_B)\right) - \Pi_A\left(p_A^*\right)$$

$\Delta\Pi_A\left(p_A^*, p_B)\right)$ is negative increasing in p_B and equal to zero when $p_B \geq p_A^*$:
It is negative because $p_A^* = \underset{p_A \geq c}{\arg\max} \Pi_A\left(p_A\right)$ and thus $\Pi_A\left(\tilde{p}_A(p_B)\right) \leq \Pi_A\left(p_A^*\right)$. By concavity of $\Pi_A\left(.\right)$, it is increasing on $[0, p_A^*]$, $\tilde{p}_A(p_B)$ is also weakly increasing in p_B, thus $\Pi_A\left(\tilde{p}_A(p_B)\right)$ is increasing in p_B because $\tilde{p}_A(p_B) \leq \tilde{p}_A(p_A^*) \leq p_A^*$.

Then, using

$$p_B^{**} = \underset{p_B \geq c}{\arg\max} \left[\Pi_B\left(p_B\right) + \Delta\Pi_A\left(p_A^*, p_B\right)\right] \Delta W\left(p_B\right)^{\frac{\alpha}{1-\alpha}},$$

$$p_B^* = \underset{p_B \geq c}{\arg\max} \Pi_B\left(p_B\right) \Delta W\left(p_B\right)^{\frac{\alpha}{1-\alpha}},$$

we have

$$\begin{aligned}
& \Pi_B\left(p_B^{**}\right) \Delta W\left(p_B^{**}\right)^{\frac{\alpha}{1-\alpha}}+\Delta \Pi_A\left(p_A^*, p_B^{**}\right) \Delta W\left(p_B^{**}\right)^{\frac{\alpha}{1-\alpha}} \\
= & \left[\Pi_B\left(p_B^{**}\right)+\Delta \Pi_A\left(p_A^*, p_B^{**}\right)\right] \Delta W\left(p_B^{**}\right)^{\frac{\alpha}{1-\alpha}} \\
\geq & \left[\Pi_B\left(p_B^*\right)+\Delta \Pi_A\left(p_A^*, p_B^*\right)\right] \Delta W\left(p_B^*\right)^{\frac{\alpha}{1-\alpha}} \text { because of the definition of } p_B^{**} \\
= & \Pi_B\left(p_B^*\right) \Delta W\left(p_B^*\right)^{\frac{\alpha}{1-\alpha}}+\Delta \Pi_A\left(p_A^*, p_B^*\right) \Delta W\left(p_B^*\right)^{\frac{\alpha}{1-\alpha}} \\
\geq & \Pi_B\left(p_B^{**}\right) \Delta W\left(p_B^{**}\right)^{\frac{\alpha}{1-\alpha}}+\Delta \Pi_A\left(p_A^*, p_B^*\right) \Delta W\left(p_B^*\right)^{\frac{\alpha}{1-\alpha}} \text { because of the definition of } p_B^*
\end{aligned}$$

Thus

$$\Delta \Pi_A\left(p_A^*, p_B^{**}\right) \Delta W\left(p_B^{**}\right)^{\frac{\alpha}{1-\alpha}} \geq \Delta \Pi_A\left(p_A^*, p_B^*\right) \Delta W\left(p_B^*\right)^{\frac{\alpha}{1-\alpha}} .$$

If $p_B^* \geq p_B^{**}$ then

$$\Delta \Pi_A\left(p_A^*, p_B^*\right)) \Delta W\left(p_B^*\right)^{\frac{\alpha}{1-\alpha}} \geq \Delta \Pi_A\left(p_A^*, p_B^*\right)) \Delta W\left(p_B^{**}\right)^{\frac{\alpha}{1-\alpha}}$$

because $\Delta \Pi_A\left(p_A^*, p_B^*\right)) \leq 0$ and $\Delta W(.)$ is positive decreasing. Using the above inequality, it implies

$$\Delta \Pi_A\left(p_A^*, p_B^{**}\right) \geq \Delta \Pi_A\left(p_A^*, p_B^*\right)$$

and thus $p_B^{**} \geq p_B^*$ because $\Delta \Pi_A\left(p_A^*, p_B\right)$ is increasing in p_B, which contradicts $p_B^* \geq p_B^{**}$ implying that it must be that $p_B^{**} \geq p_B^*$.

Duopoly case

Consider two firms competing against each other and producing two differentiated products, 1 and 2, at marginal costs c, respectively. Denote $D_{1c}\left(p_{1c}, p_{2c}\right)$ and $D_{2c}\left(p_{1c}, p_{2c}\right)$ as demands for products 1 and 2 in country c, respectively, when their prices are given by p_{1c} and p_{2c}. We assume that each firm i's profit function $\Pi_{ic} \equiv \left(p_{ic}-c\right) D_{ic}\left(p_{ic}, p_{-ic}\right)$ is strictly concave in its own price, weakly increasing in the rival's price, and that its best-response price is increasing in its rival's price (i.e., prices are strategic complements). We suppose further that a Nash equilibrium $\left(p_{1c}^*, p_{2c}^*\right)$ to the Bertrand game exists and is unique.

Under regulation, we suppose that a governmental agency negotiates prices by engaging in simultaneous Nash bargaining with both firms. We assume that the governmental agency's objective function of country B takes the general form $W(p_{1B}, p_{2B})$, where $W(.,.)$ is decreasing over $[c, +\infty) \times [c, +\infty)$. For instance, $W(p_{1B}, p_{2B})$ could be consumer surplus, social welfare or coverage.

The prices that arise in country A solve the Bertrand-Nash equilibrium

$$\begin{cases} p_{1A}^* = \arg\max_{p_{1A} \geq c} \Pi_{1A}\left(p_{1A}, p_{2A}^*\right) \\ p_{2A}^* = \arg\max_{p_{2A} \geq c} \Pi_{2A}\left(p_{1A}^*, p_{2A}\right) \end{cases}$$

and in country B, the regulation solves the following system of maximization programs:

$$\begin{cases} p_{1B}^* = \arg\max_{p_{1B} \geq c} \Pi_{1B}\left(p_{1B}, p_{2B}^*\right)^{1-\alpha_1} \Delta W_1\left(p_{1B}, p_{2B}^*\right)^{\alpha_1} \\ p_{2B}^* = \arg\max_{p_{2B} \geq c} \Pi_{2B}\left(p_{1B}^*, p_{2B}\right)^{1-\alpha_2} \Delta W_2\left(p_{1B}^*, p_{2B}\right)^{\alpha_2} \end{cases} \tag{B.1}$$

where

$$\Delta W_1\left(p_{1B}, p_{2B}^*\right) \equiv W\left(p_{1B}, p_{2B}^*\right) - W\left(\infty, p_{2B}^*\right),$$

$$\Delta W_2\left(p_{1B}^*, p_{2B}\right) \equiv W\left(p_{1B}^*, p_{2B}\right) - W\left(p_{1B}^*, \infty\right)$$

$$\alpha_1, \alpha_2 \in (0, 1]$$

capture the bargaining power of the governmental agency in its negotiation with firms 1 and 2, respectively. We assume that the pair (p_{1B}^*, p_{2B}^*) solving the system exists and is unique.

We now consider the international reference pricing equilibrium that satisfies

$$\begin{cases} p_{1A}^{**} = \tilde{p}_{1A}\left(p_{1B}^{**}, p_{2A}^{**}\right) \equiv \arg\max_{p_{1A} \leq p_{1B}^{**}} \Pi_{1A}\left(p_{1A}, p_{2A}^{**}\right) \\ p_{2A}^{**} = \tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}^{**}\right) \equiv \arg\max_{p_{2A} \leq p_{2B}^{**}} \Pi_{2A}\left(p_{1A}^{**}, p_{2A}\right) \\ p_{1B}^{**} = \arg\max_{p_{1B} \geq c} \left[\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}, p_{2A}^{**}\right), p_{2A}^{**}\right) + \Pi_{1B}\left(p_{1B}, p_{2B}^{**}\right)\right. \\ \qquad\qquad \left. -\Pi_{1A}\left(p_{1A}^*, p_{2A}^{**}\right)\right]^{1-\alpha_1} \Delta W_1\left(p_{1B}, p_{2B}^{**}\right)^{\alpha_1} \\ p_{2B}^{**} = \arg\max_{p_{2B} \geq c} \left[\Pi_{2A}\left(p_{1A}^{**}, \tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}\right)\right) + \Pi_{2B}\left(p_{1B}^{**}, p_{2B}\right)\right. \\ \qquad\qquad \left. -\Pi_{2A}\left(p_{1A}^{**}, p_{2A}^*\right)\right]^{1-\alpha_2} \Delta W_2\left(p_{1B}^{**}, p_{2B}\right)^{\alpha_2} \end{cases}$$

Remark that imposing the reference pricing constraint on one product only would generate the same proposition, but for simplicity of exposition we consider the symmetric case.

Proposition The international reference pricing policy implies that the prices in country A decrease and the prices in country B increase:

$$p_{iA}^{**} \leq p_{iA}^{*} \quad \text{and} \quad p_{iB}^{**} \geq p_{iB}^{*} \quad \text{for } i = 1,2$$

Proof *Let's start with proving that $p_{iA}^{**} \leq p_{iA}^{*}$ for $i = 1,2$:*

By definition of the solution of

$$\begin{cases} p_{1A}^{*} = \tilde{p}_{1A}\left(\infty, p_{2A}^{*}\right) = \arg\max_{p_{1A}} \Pi_{1A}\left(p_{1A}, p_{2A}^{*}\right) \\ p_{2A}^{*} = \tilde{p}_{2A}\left(p_{1A}^{*}, \infty\right) = \arg\max_{p_{2A}} \Pi_{2A}\left(p_{1A}^{*}, p_{2A}\right) \end{cases}$$

and

$$\begin{cases} p_{1A}^{**} = \tilde{p}_{1A}\left(p_{1B}^{**}, p_{2A}^{**}\right) \equiv \arg\max_{p_{1A} \leq p_{1B}^{**}} \Pi_{1A}\left(p_{1A}, p_{2A}^{**}\right) \\ p_{2A}^{**} = \tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}^{**}\right) \equiv \arg\max_{p_{2A} \leq p_{2B}^{**}} \Pi_{2A}\left(p_{1A}^{**}, p_{2A}\right) \end{cases}$$

Then

$$\begin{aligned} p_{1A}^{**} &= \tilde{p}_{1A}\left(p_{1B}^{**}, p_{2A}^{**}\right) \leq \tilde{p}_{1A}\left(\infty, p_{2A}^{**}\right) \leq \tilde{p}_{1A}\left(\infty, p_{2A}^{*}\right) = p_{1A}^{*} \text{ if } p_{2A}^{**} \leq p_{2A}^{*} \\ p_{2A}^{**} &= \tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}^{**}\right) \leq \tilde{p}_{2A}\left(p_{1A}^{**}, \infty\right) \leq \tilde{p}_{2A}\left(p_{1A}^{*}, \infty\right) = p_{2A}^{*} \text{ if } p_{1A}^{**} \leq p_{1A}^{*} \end{aligned}$$

If $p_{1A}^{**} > p_{1A}^{*}$ then $p_{2A}^{**} = \tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}^{**}\right) \geq \tilde{p}_{2A}\left(p_{1A}^{*}, p_{2B}^{**}\right) = p_{2A}^{*}$ if $p_{2B}^{**} \geq p_{2A}^{*}$. Thus $p_{1A}^{**} > p_{1A}^{*}$ implies $p_{2A}^{**} > p_{2A}^{*}$ if $p_{2B}^{**} \geq p_{2A}^{*}$, but both prices increasing is not possible by definition of the unconstrained Nash equilibrium. Thus, it must be that if $p_{1A}^{**} > p_{1A}^{*}$ then $p_{2B}^{**} < p_{2A}^{*}$, but then $p_{2A}^{**} \leq p_{2B}^{**} < p_{2A}^{*}$. But we have shown that if $p_{2A}^{**} \leq p_{2A}^{*}$ then $p_{1A}^{**} \leq p_{1A}^{*}$ which proves that we must have both $p_{iA}^{**} \leq p_{iA}^{*}$ for $i = 1,2$.

Let's prove now that $p_{iB}^{**} \geq p_{iB}^{*}$ for $i = 1,2$:

Remark that $\tilde{p}_{1A}(p_{1B}, p_{2A})$ is weakly increasing in the second argument p_{2A} because of strategic complementarity in profit, and symmetrically for $\tilde{p}_{2A}(.,.)$.

Moreover, $\tilde{p}_{1A}(p_{1B}, p_{2A})$ is weakly increasing in the first argument p_{1B} because of the concavity of the profit function in its own price.

Moreover, $\tilde{p}_{1A}(p_{1B}, p_{2A}^{**}) \leq \tilde{p}_{1A}(p_{1B}, p_{2A}^{*})$ and $\tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}\right) \leq \tilde{p}_{2A}\left(p_{1A}^{*}, p_{2B}\right)$ since $p_{iA}^{**} \leq p_{iA}^{*}$.

Then, $\tilde{p}_{1A}(p_{1B}, p_{2A}^{*}) \leq p_{1A}^{*}$ and thus $\tilde{p}_{1A}(p_{1B}, p_{2A}^{**}) \leq p_{1A}^{*}$ which implies that

$$\Delta\Pi_{1A}\left(p_{1B}, p_{1A}^{*}, p_{2A}^{**}\right) \equiv \Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}, p_{2A}^{**}\right), p_{2A}^{**}\right) - \Pi_{1A}\left(p_{1A}^{*}, p_{2A}^{**}\right) \leq 0$$

because the reaction function of firm 2 is increasing in the price of firm 1. Similarly $\Pi_{2A}\left(p_{1A}^{**}, \tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}\right)\right) - \Pi_{2A}\left(p_{1A}^{**}, p_{2A}^{*}\right) \leq 0$.

Moreover, $\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}, p_{2A}^{**}\right), p_{2A}^{**}\right) - \Pi_{1A}\left(p_{1A}^{*}, p_{2A}^{**}\right)$ is then weakly increasing in p_{1B} as well as $\Pi_{2A}\left(p_{1A}^{**}, \tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}\right)\right) - \Pi_{2A}\left(p_{1A}^{**}, p_{2A}^{*}\right)$ in p_{2B}.

$\Delta W_1\left(p_{1B}, p_{2B}^{*}\right) \equiv W\left(p_{1B}, p_{2B}^{*}\right) - W\left(\infty, p_{2B}^{*}\right) \geq 0$ is decreasing in p_{1B} and $\Delta W_2\left(p_{1B}^{*}, p_{2B}\right) \equiv W\left(p_{1B}^{*}, p_{2B}\right) - W\left(p_{1B}^{*}, \infty\right) \geq 0$ is decreasing in p_{2B}.

Define

$$\tilde{\Pi}_{1B}\left(p_{1B}, p_{1A}^{*}, p_{2A}^{**}, p_{2B}^{**}\right) = \Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}, p_{2A}^{**}\right), p_{2A}^{**}\right) + \Pi_{1B}\left(p_{1B}, p_{2B}^{**}\right) - \Pi_{1A}\left(p_{1A}^{*}, p_{2A}^{**}\right)$$

and

$$\tilde{\Pi}_{2B}\left(p_{2B}, p_{2A}^{*}, p_{1A}^{**}, p_{1B}^{**}\right) = \Pi_{2A}\left(p_{1A}^{**}, \tilde{p}_{2A}\left(p_{1A}^{**}, p_{2B}\right)\right) + \Pi_{2B}\left(p_{1B}^{**}, p_{2B}\right) - \Pi_{2A}\left(p_{1A}^{**}, p_{2A}^{*}\right)$$

As $\Pi_{1B}\left(p_{1B}, p_{2B}\right)$ is increasing in p_{1B} for $p_{1B} \leq \bar{p}_{1B}\left(p_{2B}\right)$ and increasing in p_{2B}, we have that $\tilde{\Pi}_{1B}\left(p_{1B}, p_{1A}^{*}, p_{2A}^{**}, p_{2B}^{**}\right)$ is increasing in p_{1B} for $p_{1B} \leq \bar{p}_{1B}\left(p_{2B}\right)$ and increasing in p_{2B}^{**}. Symmetrically, $\tilde{\Pi}_{2B}\left(p_{2B}, p_{2A}^{*}, p_{1A}^{**}, p_{1B}^{**}\right)$ is increasing in p_{2B} for $p_{2B} \leq \bar{p}_{2B}\left(p_{1B}\right)$ and increasing in p_{1B}^{**}.

Moreover, because of the previous inequalities, $\tilde{\Pi}_{1B}\left(p_{1B}, p_{1A}^{*}, p_{2A}^{**}, p_{2B}^{**}\right) \leq \Pi_{1B}\left(p_{1B}, p_{2B}^{**}\right)$ and $\tilde{\Pi}_{2B}\left(p_{2B}, p_{2A}^{*}, p_{1A}^{**}, p_{1B}^{**}\right) \leq \Pi_{2B}\left(p_{1B}^{**}, p_{2B}\right)$.

Then

$$
\begin{aligned}
&\left[\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}^{**},p_{2A}^{**}\right),p_{2A}^{**}\right)-\Pi_{1A}\left(p_{1A}^{*},p_{2A}^{**}\right)\right]\Delta W_1\left(p_{1B}^{**},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}} \\
&\quad+\Pi_{1B}\left(p_{1B}^{**},p_{2B}^{**}\right)\Delta W_1\left(p_{1B}^{**},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}} \\
&=\left[\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}^{**},p_{2A}^{**}\right),p_{2A}^{**}\right)+\Pi_{1B}\left(p_{1B}^{**},p_{2B}^{**}\right)-\Pi_{1A}\left(p_{1A}^{*},p_{2A}^{**}\right)\right]\Delta W_1\left(p_{1B}^{**},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}} \\
&\geq\left[\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}^{*},p_{2A}^{**}\right),p_{2A}^{**}\right)+\Pi_{1B}\left(p_{1B}^{*},p_{2B}^{**}\right)-\Pi_{1A}\left(p_{1A}^{*},p_{2A}^{**}\right)\right]\Delta W_1\left(p_{1B}^{*},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}} \\
&\qquad\text{because of the definition of } p_{1B}^{**} \\
&=\left[\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}^{*},p_{2A}^{**}\right),p_{2A}^{**}\right)-\Pi_{1A}\left(p_{1A}^{*},p_{2A}^{**}\right)\right]\Delta W_1\left(p_{1B}^{*},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}} \\
&\quad+\Pi_{1B}\left(p_{1B}^{*},p_{2B}^{**}\right)\Delta W_1\left(p_{1B}^{*},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}} \\
&\geq\left[\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}^{*},p_{2A}^{**}\right),p_{2A}^{**}\right)-\Pi_{1A}\left(p_{1A}^{*},p_{2A}^{**}\right)\right]\Delta W_1\left(p_{1B}^{*},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}} \\
&\quad+\Pi_{1B}\left(p_{1B}^{**},p_{2B}^{**}\right)\Delta W_1\left(p_{1B}^{**},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}} \\
&\qquad\text{because of the definition of } p_{1B}^{*}
\end{aligned}
$$

then, using the fact that $\Delta\Pi_{1A}\left(p_{1B}^{**},p_{1A}^{*},p_{2A}^{**}\right)=\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}^{**},p_{2A}^{**}\right),p_{2A}^{**}\right)-\Pi_{1A}\left(p_{1A}^{*},p_{2A}^{**}\right)$ and $\Delta\Pi_{1A}\left(p_{1B}^{*},p_{1A}^{*},p_{2A}^{**}\right)=\Pi_{1A}\left(\tilde{p}_{1A}\left(p_{1B}^{*},p_{2A}^{**}\right),p_{2A}^{**}\right)-\Pi_{1A}\left(p_{1A}^{*},p_{2A}^{**}\right)$ the previous inequality implies that

$$
\Delta\Pi_{1A}\left(p_{1B}^{**},p_{1A}^{*},p_{2A}^{**}\right)\Delta W_1\left(p_{1B}^{**},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}}\geq\Delta\Pi_{1A}\left(p_{1B}^{*},p_{1A}^{*},p_{2A}^{**}\right)\Delta W_1\left(p_{1B}^{*},p_{2B}^{**}\right)^{\frac{\alpha_1}{1-\alpha_1}}
$$

thus

$$
\left(\frac{\Delta W_1\left(p_{1B}^{**},p_{2B}^{**}\right)}{\Delta W_1\left(p_{1B}^{*},p_{2B}^{**}\right)}\right)^{\frac{\alpha_1}{1-\alpha_1}}\leq\frac{\Delta\Pi_{1A}\left(p_{1B}^{*},p_{1A}^{*},p_{2A}^{**}\right)}{\Delta\Pi_{1A}\left(p_{1B}^{**},p_{1A}^{*},p_{2A}^{**}\right)}
$$

because $\Delta\Pi_{1A}\left(p_{1B}^{**},p_{1A}^{*},p_{2A}^{**}\right)\leq 0$.

This inequality if not possible if $p_{1B}^{**}<p_{1B}^{*}$ because in such case $\frac{\Delta W_1\left(p_{1B}^{**},p_{2B}^{**}\right)}{\Delta W_1\left(p_{1B}^{*},p_{2B}^{**}\right)}>1$ because $\Delta W_1\left(p_{1B},p_{2B}\right)$ is decreasing in p_{1B}, and $\frac{\Delta\Pi_{1A}\left(p_{1B}^{*},p_{1A}^{*},p_{2A}^{**}\right)}{\Delta\Pi_{1A}\left(p_{1B}^{**},p_{1A}^{*},p_{2A}^{**}\right)}\leq 1$ because $\Delta\Pi_{1A}\left(p_{1B},p_{1A}^{*},p_{2A}^{**}\right)$ is increasing in p_{1B} but negative. This implies that necessarily $p_{1B}^{**}\geq p_{1B}^{*}$. Symmetrically $p_{2B}^{**}\geq p_{2B}^{*}$.

B.1.6 Additional Counterfactual Variations

Required Comparison

In this section, we consider a case where the United States regulator additionally requires that any on-patent drug sold in the United States must also be sold in Canada in order to provide a reference price.[1] This eliminates the potential for pharmaceutical companies to exit the Canadian market and set an unrestricted price in the United States. While it is never an equilibrium outcome for the firm to exit the Canadian market and only supply in the United States, the prohibition on doing so improves the bargaining position of the Canadian regulator.

As in Section 2.5.1, we consider the case in which prices are negotiated first in Canada and then chosen in the US subject to the price ceiling constraint. Negotiations between the firm and the Canadian regulator account for the impact that the Canadian price (p_j^{CA}) has on US profits. The sole difference from Section 2.5.1 is that the disagreement profit is now zero:

$$\Delta\Pi_j(p_j^{CA}, \mathbf{p}_{-j}^{US}, \mathbf{p}_{-j}^{CA}) = \underbrace{\Pi_j^{US}(p_j^{US}(p_j^{CA}, \mathbf{p}_{-j}^{US}), \mathbf{p}_{-j}^{US}) + \Pi_j^{CA}(p_j^{CA}, \mathbf{p}_{-j}^{CA})}_{\text{global profit under agreement}}. \tag{B.2}$$

After substituting (B.2) for (2.10) in Section 2.5.1, (2.12) yields the analogous equilibrium conditions when the regulator requires comparisons.

Index Pricing

In this section, we consider the case in which the United States implements a reference pricing rule requiring the price of on-patent drugs sold in the United States to be lower than its average price in all other countries in which the product is sold:

$$p_j^{US} \leq \overline{p_j^{\mathcal{C}}} \equiv \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}} p_j^c, \tag{B.3}$$

[1] In our simulations, we apply this rule only to on-patent drugs that are empirically sold in both markets.

where $\mathcal{C}$ denotes the set of other countries in which j is sold. Analogously to Section 2.5.1, we model firms as first simultaneously negotiating prices abroad and then setting prices in the United States.[2] Given its own negotiated prices abroad ($\{p_j^c\}_{c\in\mathcal{C}}$) and its competitors' prices in the United States ($\mathbf{p}_{-j}^{US}$), the firm chooses its price in the United States (p_j^{US}) to maximize its profitability in the United States subject to the reference index constraint (B.3):

$$p_j^{US}(\overline{p_j^{\mathcal{C}}}, \mathbf{p}_{-j}^{US}) \equiv \underset{p\in[0,\overline{p_j^{\mathcal{C}}}]\cup\{\infty\}}{\arg\max}\ \Pi_j^{US}\left(p, \mathbf{p}_{-j}^{US}\right)\mathbf{1}_{\{\overline{p_j^{\mathcal{C}}}\geq p\}}. \tag{B.4}$$

Negotiations between the pharmaceutical company and the regulator in country c account for the impact that the price p_j^c has on US profits through the index price constraint (B.3).

$$\begin{aligned}\Delta\Pi_j^c(p_j^c, \{p_j^{c'}\}_{c'\in\mathcal{C}\backslash c}, \mathbf{p}_{-j}^c, \mathbf{p}_{-j}^{US}) \equiv{}& \underbrace{\Pi_j^{US}(p_j^{US}(\overline{p_j^{\mathcal{C}}}(p_j^c), \mathbf{p}_{-j}^{US}), \mathbf{p}_{-j}^{US}) + \sum_{c'\in\mathcal{C}}\Pi_j^{c'}(p_j^c, \mathbf{p}_{-j}^{c'})}_{\text{global profit under agreement}} \\ -\max&\left\{\underbrace{\Pi_j^{US}(p_j^{US}(\infty, \mathbf{p}_{-j}^{US}), \mathbf{p}_{-j}^{US})}_{\text{profit if only in US}}, \underbrace{\Pi_j^{US}(p_j^{US}(\overline{p_j^{\mathcal{C}\backslash c}}, \mathbf{p}_{-j}^{US}), \mathbf{p}_{-j}^{US}) + \sum_{c'\in\mathcal{C}\backslash c}\Pi_j^{c'}(p_j^c, \mathbf{p}_{-j}^{c'})}_{\text{global profit if in US and } \mathcal{C}\backslash c}\right\},\end{aligned} \tag{B.5}$$

where the notation $\overline{p_j^{\mathcal{C}}}(p_j^c)$ emphasizes that $\overline{p_j^{\mathcal{C}}}$ is a function of p_j^c. Negotiations with each country c results in a price p_j^c that maximizes the Nash product:

$$p_j^c(\{p_j^{c'}\}_{c'\in\mathcal{C}\backslash c}, \mathbf{p}_{-j}^c, \mathbf{p}_{-j}^{US}) \equiv \arg\max_p \left(\underbrace{\Delta\Pi_j^c(p, \{p_j^{c'}\}_{c'\in\mathcal{C}\backslash c}, \mathbf{p}_{-j}^c, \mathbf{p}_{-j}^{US})}_{\text{profit gain from agreement}}\right)^{\rho_j}\left(\underbrace{\Delta_j W_c(p, \mathbf{p}_{-j}^c)}_{\text{welfare gain in } c \text{ if agrees}}\right)^{1-\rho_j}$$

[2] Also as in Section 2.5.1, it is equivalent to say that firms cannot commit not to decrease prices as the result of negotiations with other countries' regulators.

B.1.7 Additional Tables of counterfactuals

Counterfactuals with Canada as Price Ceiling for the US

Table B.5: *Counterfactual Quantity Changes on All Drugs when Canada as Price Ceiling for the US*

	ρ_{jm}			Canada			US		
ATC4	On Patent	Branded Off	Generic	Before	After	Δ (%)	Before	After	Δ (%)
A10C1	0.62			349	249	-28.6	3302	3317	0.5
A2B2	0.55	0.90	0.87	19362	16020	-17.3	113244	114653	1.2
B1B2	0.70			2598	2306	-11.2	35354	35556	0.6
C10A1	0.54	1.00	0.77	11349	10431	-8.1	79186	80203	1.3
C2A2	1.00	1.00	0.94	2384	2384	0.0	26882	26882	-0.0
C7A0	0.72	1.00	1.00	23492	23401	-0.4	167276	167278	0.0
C8A0	0.56	0.89	0.86	12760	12477	-2.2	73390	73697	0.4
C9A0	0.47	0.95	1.00	18050	14721	-18.4	101954	103864	1.9
C9C0	0.60	0.94	0.50	4801	4458	-7.1	27227	27982	2.8
L1A0	0.91	0.50	1.00	795	675	-15.1	1793	1795	0.1
L1B0	0.64	0.50	1.00	2320	1591	-31.4	3737	3791	1.4
L1C0	0.50	0.50	0.98	332	299	-10.2	1424	1460	2.5
L1D0	0.99	0.50	0.50	123	116	-5.2	522	522	0.0
L1X4	1.00	0.50	0.50	785	556	-29.3	722	723	0.1
L1X9	0.92	0.50	0.57	755	753	-0.3	994	994	0.0
L2B2	0.83	0.94	0.61	1791	1739	-2.9	677	690	2.0
L2B3	0.70	0.79	0.58	1928	1648	-14.6	917	1033	12.7
L4X0	0.95	0.91	1.00	12181	8232	-32.4	11599	11666	0.6
M1A1	0.44	0.91	1.00	8374	5944	-29.0	99443	99592	0.1
M5B3	0.93	0.95	0.54	1225	1079	-11.9	1823	1826	0.2
N1A1	0.45	0.57	1.00	19107	17776	-7.0	665328	665934	0.1
N1A2	1.00	1.00	0.92	2865	2828	-1.3	69548	69548	0.0
N1B1	0.96	1.00	0.75	1096	1080	-1.4	20315	20315	0.0
N1B3	0.50	0.50	0.58	16254	16254	0.0	145882	145882	-0.0
N2A0	0.51	0.78	0.89	36395	36395	-0.0	343829	343829	0.0
N2B0	0.50	0.96	0.88	10159	9870	-2.8	153266	153266	-0.0
N3A0	0.87	0.93	1.00	42619	39004	-8.5	274813	275012	0.1
N5A1	0.86	0.86	0.94	41625	32705	-21.4	115139	115470	0.3
N5A9	0.64	0.97	0.94	9269	9267	-0.0	36694	36694	0.0
N6A4	0.80	0.99	0.91	13805	12720	-7.9	89527	89578	0.1
N6A9	0.27	0.89	0.99	11944	11890	-0.4	72854	72876	0.0
Total				330892	298869	-9.6	2738661	2745930	.2

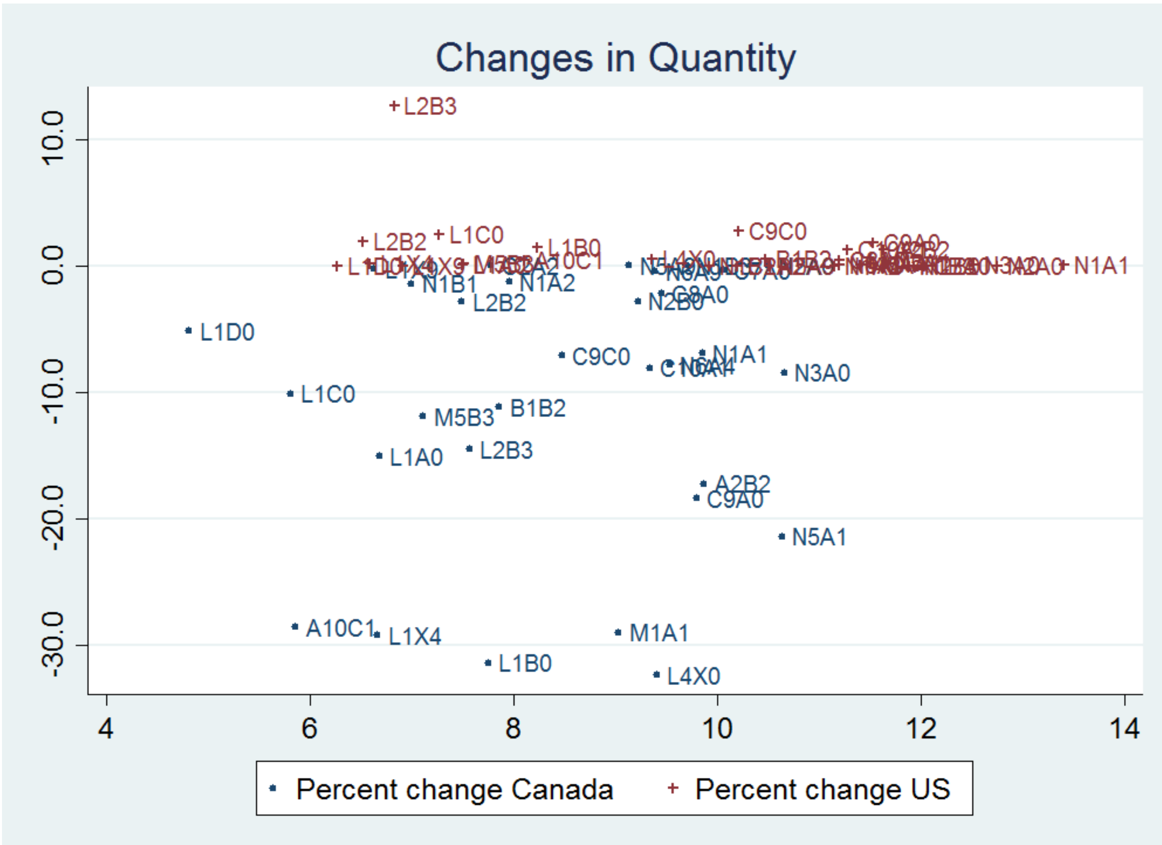
Note: Quantity are average yearly standard units (on period 2002-2013). Δ stands for the change of quantity between after and before in percentage of initial quantity. The parameter ρ_{jm} is the one estimated from the supply model in Canada and used for counterfactual simulations.

Table B.6: *Counterfactual Profits on All Drugs with Canada as Price Ceiling for the US*

ATC4	ρ_{jm} On Patent	ρ_{jm} Branded Off	ρ_{jm} Generic	Canada Before	Canada After	Canada Δ (%)	US Before	US After	US Δ (%)
A10C1	0.62			3147	5123	62.8	97767	95925	-1.9
A2B2	0.55	0.90	0.87	11212	23988	114.0	143282	142252	-0.7
B1B2	0.70			37401	75708	102.4	1279709	1243295	-2.8
C10A1	0.54	1.00	0.77	12037	21953	82.4	89953	88070	-2.1
C2A2	1.00	1.00	0.94	1439	1439	-0.0	4260	4260	0.0
C7A0	0.72	1.00	1.00	4853	4905	1.1	27738	27734	-0.0
C8A0	0.56	0.89	0.86	9790	12714	29.9	97479	96582	-0.9
C9A0	0.47	0.95	1.00	5921	7716	30.3	14233	13947	-2.0
C9C0	0.60	0.94	0.50	3139	6715	113.9	32509	31673	-2.6
L1A0	0.91	0.50	1.00	12000	21508	79.2	91496	90799	-0.8
L1B0	0.64	0.50	1.00	24193	51007	110.8	190112	170679	-10.2
L1C0	0.50	0.50	0.98	50263	72748	44.7	269879	265900	-1.5
L1D0	0.99	0.50	0.50	14564	17221	18.2	61282	61162	-0.2
L1X4	1.00	0.50	0.50	50602	64535	27.5	43552	43253	-0.7
L1X9	0.92	0.50	0.57	17276	17571	1.7	60304	60223	-0.1
L2B2	0.83	0.94	0.61	3468	4594	32.5	3503	3462	-1.2
L2B3	0.70	0.79	0.58	8559	15482	80.9	6585	5623	-14.6
L4X0	0.95	0.91	1.00	36855	102785	178.9	77871	77508	-0.5
M1A1	0.44	0.91	1.00	1319	4670	254.1	2592	2519	-2.8
M5B3	0.93	0.95	0.54	2151	2532	17.7	19601	19477	-0.6
N1A1	0.45	0.57	1.00	12530	16762	33.8	349549	347992	-0.4
N1A2	1.00	1.00	0.92	13656	13792	1.0	157280	157277	-0.0
N1B1	0.96	1.00	0.75	4678	4904	4.8	20696	20695	-0.0
N1B3	0.50	0.50	0.58	9221	9221	0.0	20075	20075	0.0
N2A0	0.51	0.78	0.89	19276	19278	0.0	45855	45855	0.0
N2B0	0.50	0.96	0.88	2984	3842	28.8	10947	10948	0.0
N3A0	0.87	0.93	1.00	10688	21015	96.6	116621	116554	-0.1
N5A1	0.86	0.86	0.94	69988	133776	91.1	526368	520215	-1.2
N5A9	0.64	0.97	0.94	3851	3852	0.0	3148	3147	-0.0
N6A4	0.80	0.99	0.91	5084	5371	5.6	70404	70171	-0.3
N6A9	0.27	0.89	0.99	1878	1913	1.9	13149	13123	-0.2
Total				464022	768639	65.59	3947799	3870395	-1.9

Note: Profits are average yearly expenses in 1000 US$ (from the period 2002-2013). Δ stands for the change in profits between after and before in percentage of initial profits. The parameter ρ_j is the one estimated from the supply model in Canada and used for counterfactual simulations.

Figure B.1: *Changes in quantity by ATC-4 in counterfactual against log quantity observed*



Note: Each point corresponds to one ATC-4 labeled with its code. Percentage changes on vertical axis (not drawing if change larger than 200%). Log quantity on horizontal axis. Both US and Canada on same graph.

Table B.7: *Counterfactual Expenses on Patented Drugs when Canada as Price Ceiling for the US*

i		ρ_{jm}		Canada			US		
ATC4	On Patent	Branded Off	Generic	Before	After	Δ (%)	Before	After	Δ (%)
A10C1	0.62			4161	5777	38.8	113984	112471	-1.3
A2B2	0.55	0.90	0.87	9018	17984	99.4	182955	184476	0.8
B1B2	0.70			40084	78711	96.4	1261933	1228934	-2.6
C10A1	0.54	1.00	0.77	11938	19425	62.7	139298	141079	1.3
C2A2	1.00	1.00	0.94	761	761	-0.0	8330	8330	0.0
C7A0	0.72	1.00	1.00	311	319	2.9	27376	27387	0.0
C8A0	0.56	0.89	0.86	7317	8807	20.4	47992	48490	1.0
C9A0	0.47	0.95	1.00	6881	7465	8.5	23967	26246	9.5
C9C0	0.60	0.94	0.50	4056	7452	83.7	61146	61340	0.3
L1A0	0.91	0.50	1.00	10281	18019	75.3	170467	169746	-0.4
L1B0	0.64	0.50	1.00	25159	45386	80.4	392044	364805	-6.9
L1C0	0.50	0.50	0.98	74104	84637	14.2	524380	545884	4.1
L1D0	0.99	0.50	0.50	8457	10106	19.5	65298	65527	0.4
L1X4	1.00	0.50	0.50	49887	63476	27.2	83505	83278	-0.3
L1X9	0.92	0.50	0.57	18322	18545	1.2	129433	129361	-0.1
L2B2	0.83	0.94	0.61	2638	3473	31.7	6229	6442	3.4
L2B3	0.70	0.79	0.58	9158	15905	73.7	10671	10439	-2.2
L4X0	0.95	0.91	1.00	35121	95430	171.7	152335	152260	-0.0
M1A1	0.44	0.91	1.00	438	1476	237.0	2867	3268	14.0
M5B3	0.93	0.95	0.54	1238	1413	14.1	27912	27806	-0.4
N1A1	0.45	0.57	1.00	6900	8619	24.9	350707	350088	-0.2
N1A2	1.00	1.00	0.92	1192	1171	-1.8	70722	70740	0.0
N1B1	0.96	1.00	0.75	1456	1508	3.5	31019	31035	0.0
N1B3	0.50	0.50	0.58	923	923	-0.0	1362	1362	-0.0
N2A0	0.51	0.78	0.89	346	346	-0.0	870	871	0.1
N2B0	0.50	0.96	0.88	180	273	51.3	18975	18976	0.0
N3A0	0.87	0.93	1.00	3541	6502	83.6	195096	195655	0.3
N5A1	0.86	0.86	0.94	32795	67883	107.0	966832	962088	-0.5
N5A9	0.64	0.97	0.94	1716	1716	0.0	1153	1154	0.1
N6A4	0.80	0.99	0.91	2473	2371	-4.1	113105	113115	0.0
N6A9	0.27	0.89	0.99	341	344	0.8	4674	4738	1.4
Total				371191	596221	60.62	5186638	5147393	-.75

Note: Expenses are average yearly expenses in 1000 US$ (on period 2002-2013). Patented drugs only.

Table B.8: *Counterfactual Price Changes by ATC-4 when Canada as Price Ceiling for the US*

	ρ_{jm}			Price Change All drugs		Price Change Patented		Price Change Branded Off		Price Change Generic	
ATC4	On Patent	Branded Off	Generic	CA (%)	US (%)	CA (%)	US (%)	CA (%)	US (%)	CA (%)	US (%)
A10C1	0.62			81.1	-1.7	81.1	-1.7				
A2B2	0.55	0.90	0.87	132.1	-0.7	197.6	-1.3	88.2	-0.1	23.9	-0.1
B1B2	0.70			122.1	-3.1	122.1	-3.2		-2.1		0.0
C10A1	0.54	1.00	0.77	78.5	-0.8	117.8	-2.6	166.8	-1.7	36.0	-0.2
C2A2	1.00	1.00	0.94	-0.0	0.0	-0.0	0.0	0.0	0.0	-0.0	0.0
C7A0	0.72	1.00	1.00	1.5	0.0	3.3	-0.1	0.5	0.0	1.1	0.0
C8A0	0.56	0.89	0.86	25.9	-1.1	43.0	-2.8	16.3	0.4	17.7	-0.2
C9A0	0.47	0.95	1.00	29.1	1.9	66.1	-6.5	1.8	0.0	5.4	-0.1
C9C0	0.60	0.94	0.50	92.6	-2.3	104.9	-2.9	19.5	0.4	1.7	-0.8
L1A0	0.91	0.50	1.00	111.6	-0.6	212.7	-1.3	16.8	-0.6	133.9	0.0
L1B0	0.64	0.50	1.00	156.4	-8.5	199.3	-21.7	-0.5	-3.0	118.4	-0.4
L1C0	0.50	0.50	0.98	26.8	0.6	66.9	-4.6	-1.5	0.2	6.6	-0.1
L1D0	0.99	0.50	0.50	21.0	0.1	173.6	-0.5	-3.0	-0.0	2.3	-0.0
L1X4	1.00	0.50	0.50	81.6	-0.4	85.4	-0.4	3.8		0.0	0.0
L1X9	0.92	0.50	0.57	1.6	-0.1	52.8	-0.5	1.5		2.2	0.0
L2B2	0.83	0.94	0.61	31.3	1.6	245.1	-3.9	15.5	0.3	10.1	0.1
L2B3	0.70	0.79	0.58	101.7	-13.4	104.6	-14.0	19.0	-0.4	11.2	-0.0
L4X0	0.95	0.91	1.00	312.0	-0.9	329.7	-1.2	5.2	-0.1	126.9	0.0
M1A1	0.44	0.91	1.00	358.0	1.1	104.4	-17.1	63.6	-0.1	494.4	-0.0
M5B3	0.93	0.95	0.54	24.8	-0.5	56.6	-0.8	21.1	0.3	2.4	-0.1
N1A1	0.45	0.57	1.00	35.4	-0.4	48.1	-0.7	23.0	-0.1	25.0	0.0
N1A2	1.00	1.00	0.92	2.3	0.0	41.4	-0.1	1.8	-0.0	1.4	0.0
N1B1	0.96	1.00	0.75	6.7	0.0	16.2	-0.0	10.1	-0.0	4.3	0.0
N1B3	0.50	0.50	0.58	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
N2A0	0.51	0.78	0.89	0.0	0.0	0.1	-0.1	-0.0	-0.0	0.0	0.0
N2B0	0.50	0.96	0.88	29.0	0.0	4.5	0.0	22.8	0.0	30.3	0.0
N3A0	0.87	0.93	1.00	116.5	0.0	128.8	-0.5	79.2	-0.0	132.3	0.0
N5A1	0.86	0.86	0.94	136.5	-0.7	374.2	-1.0	52.3	0.2	49.1	0.4
N5A9	0.64	0.97	0.94	0.0	-0.0	0.5	-0.3	-0.0	0.0	0.0	0.0
N6A4	0.80	0.99	0.91	9.5	-0.1	146.8	-0.4	0.4	-0.1	0.4	0.0
N6A9	0.27	0.89	0.99	1.1	0.0	6.4	-3.1	-0.1	-0.0	0.8	0.0

Note: Changes in % of initial price using market shares weighted average prices.

Table B.9: *Counterfactual Prices of Drugs on Patent present in both US and Canada when Canada as Price Ceiling for the US*

	Before		After			
	Canada	US	Canada		US	
ATC4	Price	Price	Price	Δ (%)	Price	Δ (%)
A10C1	12.8661	36.95631	28.68712	122.97	36.33642	-1.68
A2B2	.8079825	2.675094	2.697367	233.84	2.638698	-1.36
B1B2	15.55126	38.01187	37.48726	141.06	36.81441	-3.15
C10A1	1.770697	3.82358	3.857452	117.85	3.676275	-3.85
C2A2	54.75095	13.18811	54.74297	-0.01	13.18826	0.00
C7A0	1.244643	7.688396	2.174808	74.73	7.643762	-0.58
C8A0	1.268361	2.297664	2.244284	76.94	2.229781	-2.95
C9A0	.5956708	1.685162	1.66685	179.83	1.553767	-7.80
C9C0	1.094559	2.528681	2.301818	110.30	2.441019	-3.47
L1A0	24.69844	198.238	77.59502	214.17	195.6298	-1.32
L1B0	17.69059	248.5182	67.17113	279.70	191.2764	-23.03
L1C0	471.2581	1033.659	941.1678	99.71	978.6742	-5.32
L1D0	292.0294	780.7533	1259.364	331.25	778.822	-0.25
L1X4	66.59953	99.22975	125.9008	89.04	98.87067	-0.36
L1X9	647.0916	745.1591	993.0546	53.46	737.4075	-1.04
L2B2	10.43049	32.88236	35.99053	245.05	30.63242	-6.84
L2B3	4.868706	11.67673	10.1707	108.90	9.953807	-14.76
L4X0	3.17295	9.958714	14.10345	344.49	9.872184	-0.87
M1A1	.6490896	2.914528	2.132971	228.61	2.231462	-23.44
M5B3	6.017537	14.48667	17.12974	184.66	14.33339	-1.06
N1A1	.662663	.919805	1.02603	54.83	.9137184	-0.66
N1A2	12.79765	75.2755	31.65229	147.33	75.17899	-0.13
N1B1	13.03696	16.28716	16.25504	24.68	16.27825	-0.05
N1B3						
N2A0	1.003563	3.317234	3.377676	236.57	3.313498	-0.11
N2B0						
N3A0	1.384019	3.735585	3.399423	145.62	3.694843	-1.09
N5A1	1.760912	8.886744	8.998435	411.01	8.755308	-1.48
N5A9	.8207119	1.356838	1.303776	58.86	1.313062	-3.23
N6A4	1.334195	3.653124	3.66085	174.39	3.634008	-0.52
N6A9	.6356424	3.21982	1.843886	190.08	3.059738	-4.97

Note: Market shares weighted average price of patented drugs by ATC-4, country for drugs present in both only. Percentage changes are changes with respect to the initial situation.

Table B.10: *Counterfactual Consumer Welfare Changes on All Drugs when Canada as Price Ceiling for the US*

	ρ_{jm}			Canada			US		
ATC4	On Patent	Branded Off	Generic	Before	After	Δ (%)	Before	After	Δ (%)
A10C1	0.62			885	984	11.3	11159	11317	1.4
A2B2	0.55	0.90	0.87	56646	42825	-24.4	337119	339633	0.7
B1B2	0.70			8804	6711	-23.8	120918	123494	2.1
C10A1	0.54	1.00	0.77	37661	30663	-18.6	242072	245518	1.4
C2A2	1.00	1.00	0.94	7111	7111	0.0	88937	88936	-0.0
C7A0	0.72	1.00	1.00	62951	62610	-0.5	675640	675647	0.0
C8A0	0.56	0.89	0.86	44044	40952	-7.0	230677	232046	0.6
C9A0	0.47	0.95	1.00	51194	46325	-9.5	335249	337842	0.8
C9C0	0.60	0.94	0.50	16875	14186	-15.9	72193	73773	2.2
L1A0	0.91	0.50	1.00	2395	1835	-23.4	6204	6220	0.3
L1B0	0.64	0.50	1.00	6643	4688	-29.4	12862	13171	2.4
L1C0	0.50	0.50	0.98	855	773	-9.6	4061	4119	1.4
L1D0	0.99	0.50	0.50	363	336	-7.5	2161	2162	0.0
L1X4	1.00	0.50	0.50	2126	1725	-18.9	2109	2116	0.3
L1X9	0.92	0.50	0.57	2431	2413	-0.8	3561	3562	0.0
L2B2	0.83	0.94	0.61	6150	5700	-7.3	2158	2181	1.1
L2B3	0.70	0.79	0.58	5894	4575	-22.4	2478	2806	13.3
L4X0	0.95	0.91	1.00	35194	20803	-40.9	29593	29734	0.5
M1A1	0.44	0.91	1.00	27343	16265	-40.5	389901	390104	0.1
M5B3	0.93	0.95	0.54	3244	2908	-10.4	5887	5913	0.4
N1A1	0.45	0.57	1.00	55089	49074	-10.9	2307142	2313437	0.3
N1A2	1.00	1.00	0.92	9216	9144	-0.8	260259	260260	0.0
N1B1	0.96	1.00	0.75	2969	2898	-2.4	76524	76527	0.0
N1B3	0.50	0.50	0.58	48316	48316	0.0	736051	736050	-0.0
N2A0	0.51	0.78	0.89	108236	108233	-0.0	1145022	1145022	0.0
N2B0	0.50	0.96	0.88	34712	31972	-7.9	618698	618692	-0.0
N3A0	0.87	0.93	1.00	138685	109393	-21.1	830380	830730	0.0
N5A1	0.86	0.86	0.94	128719	97672	-24.1	335035	337240	0.7
N5A9	0.64	0.97	0.94	31051	31047	-0.0	128193	128195	0.0
N6A4	0.80	0.99	0.91	124721	123535	-1.0	327380	327671	0.1
N6A9	0.27	0.89	0.99	31311	31172	-0.4	240522	240599	0.0
Total				1091834	956844	-12.3	9580145	9604719	.2

Note: Welfare values are average yearly on period 2002-2013 scaled by market size. Δ stands for the change of welfare between after and before in percentage of initial welfare. The parameter ρ_{jm} is the one estimated from the supply model in Canada and used for counterfactual simulations.

Table B.11: *Counterfactual Expenses, profits and Consumer Welfare Global Changes on All Drugs when Canada as Price Ceiling for the US*

	Expenses			Profits		
ATC4	Before	After	Δ (%)	Before	After	Δ (%)
A10C1	118145	118248	0.1	100914	101048	0.1
A2B2	284788	298546	4.8	154494	166240	7.6
B1B2	1366755	1371139	0.3	1317110	1319003	0.1
C10A1	186216	196756	5.7	101991	110023	7.9
C2A2	38062	38063	0.0	5699	5699	0.0
C7A0	148647	148697	0.0	32590	32639	0.1
C8A0	259058	259921	0.3	107268	109296	1.9
C9A0	66957	69790	4.2	20155	21663	7.5
C9C0	68047	71655	5.3	35648	38387	7.7
L1A0	244904	253235	3.4	103496	112307	8.5
L1B0	482675	479282	-0.7	214305	221685	3.4
L1C0	720750	753126	4.5	320143	338649	5.8
L1D0	192101	195651	1.8	75846	78384	3.3
L1X4	135841	149569	10.1	94154	107787	14.5
L1X9	160825	160989	0.1	77580	77794	0.3
L2B2	10886	12162	11.7	6970	8056	15.6
L2B3	20259	26822	32.4	15144	21106	39.4
L4X0	313040	378839	21.0	114726	180293	57.2
M1A1	28109	32044	14.0	3911	7189	83.8
M5B3	33507	33770	0.8	21752	22009	1.2
N1A1	558890	561233	0.4	362079	364754	0.7
N1A2	642266	642396	0.0	170936	171068	0.1
N1B1	93051	93285	0.3	25374	25599	0.9
N1B3	167484	167484	-0.0	29296	29296	0.0
N2A0	486180	486182	0.0	65131	65133	0.0
N2B0	84433	85358	1.1	13931	14791	6.2
N3A0	450060	460845	2.4	127309	137569	8.1
N5A1	1113479	1172218	5.3	596356	653992	9.7
N5A9	51146	51144	-0.0	6999	6999	0.0
N6A4	155712	155563	-0.1	75488	75542	0.1
N6A9	48988	49026	0.1	15027	15036	0.1
Total	8731260	8973041	2.7	4411821	4639034	5.1

Note: All values are average yearly on period 2002-2013, summing US and Canada. Δ stands for the change between after and before in percentage of initial value.

Table B.12: *Counterfactual Expenses by ATC-4 with varying MFN rule (0, +33%, +50%) and Larger Reference Market (without ex ante commitment)*

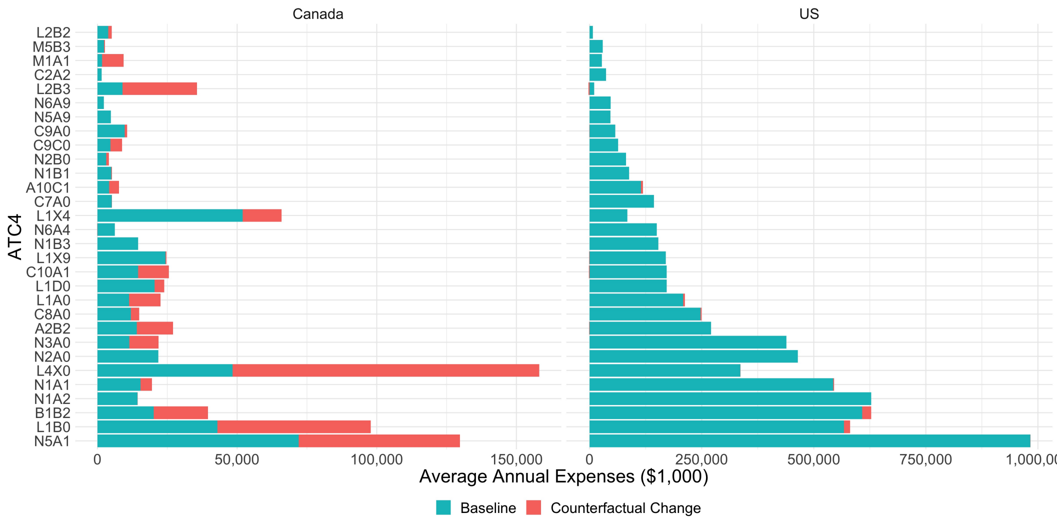
			ρ_{jm}			Canada			US		
ATC4	MFN	US market Share	On Patent	Branded Off	Generic	Before	After	Δ (%)	Before	After	Δ (%)
A2B2	0	0	0.55	0.90	0.87	14057	26529	88.7	270730	272016	0.5
A2B2	0	50	0.55	0.90	0.87	87651	156990	79.1	263927	271068	2.7
A2B2	0	100	0.55	0.90	0.87	185358	311647	68.1	268488	278951	3.9
A2B2	33	0	0.55	0.90	0.87	14057	23928	70.2	270730	271609	0.3
A2B2	50	0	0.55	0.90	0.87	14057	22349	59.0	270730	271328	0.2
B1B2	0	0	0.70			40084	78711	96.4	1326672	1292428	-2.6
B1B2	0	50	0.70			273470	462709	69.2	1324179	1152193	-13.0
B1B2	0	100	0.70			485702	767729	58.1	1176227	930594	-20.9
B1B2	33	0	0.70			40084	62512	56.0	1326889	1310894	-1.2
B1B2	50	0	0.70			38563	50214	30.2	1284627	1277971	-0.5
L1B0	0	0	0.64	0.50	1.00	33075	60786	83.8	449600	418495	-6.9
L1B0	0	50	0.64	0.50	1.00	27299	52851	93.6	481761	459276	-4.7
L1B0	0	100	0.64	0.50	1.00	42581	74518	75.0	365435	337062	-7.8
L1B0	33	0	0.64	0.50	1.00	31638	54254	71.5	426957	404204	-5.3
L1B0	50	0	0.64	0.50	1.00	29628	45343	53.0	396918	381692	-3.8
N1A1	0	0	0.45	0.57	1.00	15417	19358	25.6	543474	541875	-0.3
N1A1	0	50	0.45	0.57	1.00	220806	267795	21.3	543474	520339	-4.3
N1A1	0	100	0.45	0.57	1.00	441612	523748	18.6	543474	503817	-7.3
N1A1	33	0	0.45	0.57	1.00	15417	16602	7.7	543474	543058	-0.1
N1A1	50	0	0.45	0.57	1.00	15417	15634	1.4	543474	543459	-0.0
N1A2	0	0	1.00	1.00	0.92	14275	14395	0.8	627990	628001	0.0
N1A2	0	50	1.00	1.00	0.92	65445	66059	0.9	627990	628042	0.0
N1A2	0	100	1.00	1.00	0.92	130890	132107	0.9	627990	628091	0.0
N1A2	33	0	1.00	1.00	0.92	14275	14354	0.6	627990	627998	0.0
N1A2	50	0	1.00	1.00	0.92	14275	14308	0.2	627990	627994	0.0
N2A0	0	0	0.51	0.78	0.89	21736	21737	0.0	464444	464445	0.0
N2A0	0	50	0.51	0.78	0.89	108392	108398	0.0	464444	464446	0.0
N2A0	0	100	0.51	0.78	0.89	216785	216799	0.0	464444	464448	0.0
N2A0	33	0	0.51	0.78	0.89	21736	21737	0.0	464444	464445	0.0
N2A0	50	0	0.51	0.78	0.89	21736	21737	0.0	464444	464445	0.0
N3A0	0	0	0.87	0.93	1.00	11366	21739	91.3	438695	439107	0.1
N3A0	0	50	0.87	0.93	1.00	70808	154260	117.9	438695	441048	0.5
N3A0	0	100	0.87	0.93	1.00	141617	305398	115.7	438695	442809	0.9
N3A0	33	0	0.87	0.93	1.00	11366	19879	74.9	438695	438915	0.1
N3A0	50	0	0.87	0.93	1.00	11366	17628	55.1	438695	438772	0.0
N5A1	0	0	0.86	0.86	0.94	74422	138244	85.8	1039056	1033974	-0.5
N5A1	0	50	0.86	0.86	0.94	104691	197505	88.7	1039056	1031817	-0.7
N5A1	0	100	0.86	0.86	0.94	212643	399326	87.8	1064367	1049939	-1.4
N5A1	33	0	0.86	0.86	0.94	74422	125483	68.6	1039056	1035314	-0.4
N5A1	50	0	0.86	0.86	0.94	74422	114947	54.5	1039056	1036307	-0.3

Table B.13: *Counterfactual Prices when US is using Canada as Maximum Reference Price*

	All				Patented				Branded Off				Generic			
	Before		After		Before		After		Before		After		Before		After	
ATC4	CA	US	CA	US	CA	US	CA	US	CA	US	CA	US	CA	US	CA	US
A10C1	12.05	37.13	21.83	36.51	12.05	37.13	21.83	36.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A2B2	0.76	2.44	1.76	2.42	0.83	2.75	2.46	2.71	0.67	3.43	1.27	3.43	0.59	0.63	0.73	0.63
B1B2	15.65	37.61	34.77	36.45	15.65	38.01	34.77	36.81	0.00	30.42	0.00	29.78	0.00	28.47	0.00	28.47
C10A1	1.32	2.23	2.36	2.21	1.77	3.43	3.86	3.34	1.99	2.32	5.32	2.28	0.49	0.50	0.66	0.50
C2A2	0.60	1.35	0.60	1.35	46.08	13.16	46.07	13.16	2.79	2.09	2.79	2.09	0.16	1.06	0.16	1.06
C7A0	0.22	0.87	0.22	0.87	0.30	3.37	0.31	3.37	1.11	2.14	1.12	2.14	0.18	0.60	0.18	0.60
C8A0	0.93	3.40	1.17	3.36	1.30	2.32	1.86	2.25	0.83	25.38	0.97	25.50	0.49	1.17	0.58	1.16
C9A0	0.52	0.51	0.67	0.52	0.68	1.68	1.13	1.57	0.51	1.70	0.52	1.70	0.26	0.30	0.28	0.30
C9C0	0.97	2.31	1.87	2.26	1.10	2.72	2.26	2.64	1.19	2.64	1.42	2.65	0.25	0.47	0.26	0.46
L1A0	17.69	135.17	37.44	134.39	24.69	229.55	77.20	226.60	1.53	109.79	1.79	109.11	14.51	48.55	33.95	48.56
L1B0	16.27	124.41	41.71	113.83	18.00	382.37	53.88	299.26	17.90	209.39	17.81	203.15	11.15	17.12	24.34	17.04
L1C0	270.87	443.02	343.38	445.82	468.30	999.85	781.79	953.96	110.03	350.44	108.43	350.99	86.50	73.89	92.17	73.82
L1D0	164.08	322.95	198.49	323.12	250.68	1350.26	685.96	1342.95	360.70	998.83	350.01	998.49	77.61	108.92	79.41	108.91
L1X4	66.23	112.85	120.27	112.42	66.35	112.77	123.04	112.33	65.16	0.00	67.65	0.00	25.55	146.05	25.56	146.05
L1X9	20.64	138.60	20.98	138.49	642.79	741.55	982.29	738.00	0.94	0.00	0.95	0.00	1.64	15.12	1.67	15.12
L2B2	2.19	10.08	2.87	10.24	10.43	30.28	35.99	29.09	1.45	9.63	1.67	9.66	0.69	1.31	0.75	1.31
L2B3	4.81	11.29	9.70	9.78	4.88	11.75	9.99	10.11	3.80	17.52	4.52	17.45	2.23	0.52	2.49	0.52
L4X0	3.07	23.46	12.67	23.25	3.01	18.79	12.93	18.57	0.75	5.92	0.79	5.91	5.29	59.29	12.01	59.32
M1A1	0.19	0.27	0.88	0.28	0.67	3.68	1.38	3.05	0.50	0.95	0.82	0.95	0.13	0.23	0.75	0.23
M5B3	2.11	18.17	2.64	18.08	2.43	27.81	3.81	27.59	3.21	19.98	3.89	20.04	1.40	2.49	1.43	2.49
N1A1	0.81	0.82	1.09	0.82	0.79	0.92	1.18	0.91	0.92	0.76	1.13	0.76	0.26	0.54	0.33	0.54
N1A2	5.00	9.03	5.11	9.03	11.73	76.79	16.59	76.72	5.43	18.96	5.53	18.96	4.53	5.96	4.59	5.96
N1B1	4.48	4.35	4.78	4.35	10.97	15.64	12.75	15.63	4.63	6.76	5.10	6.76	3.19	2.84	3.32	2.84
N1B3	0.92	1.11	0.92	1.11	6.68	23.42	6.68	23.42	1.04	3.90	1.04	3.90	0.39	0.85	0.39	0.85
N2A0	0.59	1.36	0.59	1.36	0.77	3.20	0.77	3.20	1.18	3.86	1.18	3.86	0.48	1.14	0.48	1.14
N2B0	0.31	0.53	0.40	0.53	0.56	14.59	0.58	14.59	0.34	1.34	0.42	1.34	0.30	0.39	0.39	0.39
N3A0	0.26	1.49	0.57	1.50	1.36	4.30	3.12	4.28	0.20	5.65	0.35	5.65	0.20	0.79	0.45	0.79
N5A1	1.77	9.09	4.20	9.03	1.82	10.36	8.63	10.25	3.09	4.92	4.71	4.93	0.43	3.43	0.65	3.44
N5A9	0.46	1.27	0.46	1.27	2.90	2.76	2.91	2.75	1.17	16.30	1.17	16.30	0.25	1.11	0.25	1.11
N6A4	0.48	1.61	0.53	1.61	1.35	3.61	3.32	3.60	1.44	4.22	1.45	4.22	0.28	0.48	0.29	0.48
N6A9	0.19	0.63	0.19	0.63	0.40	2.95	0.43	2.86	0.58	3.47	0.58	3.47	0.13	0.32	0.13	0.32

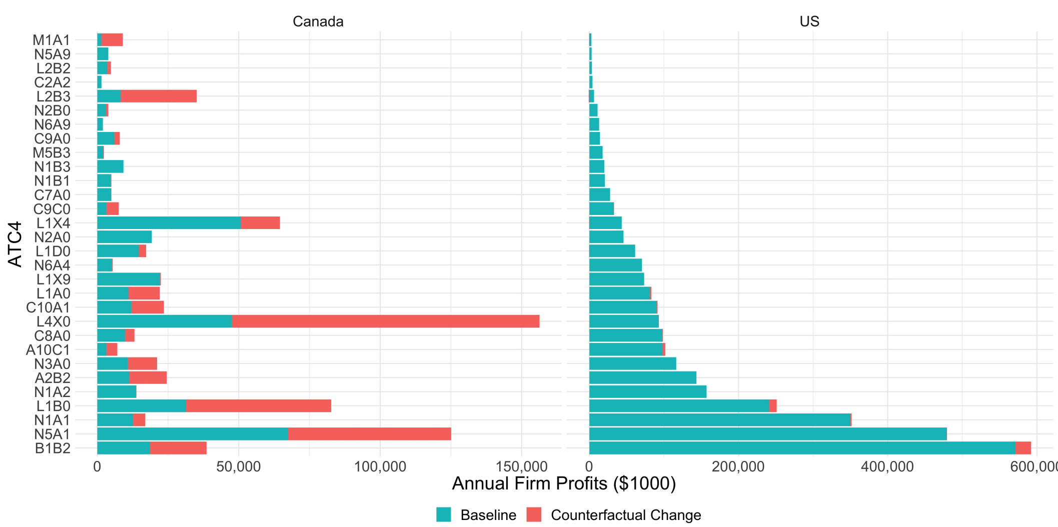
Note: Market shares weighted average price by ATC-4, country.

Figure B.2: *Counterfactual Expenditure under Price Ceiling*



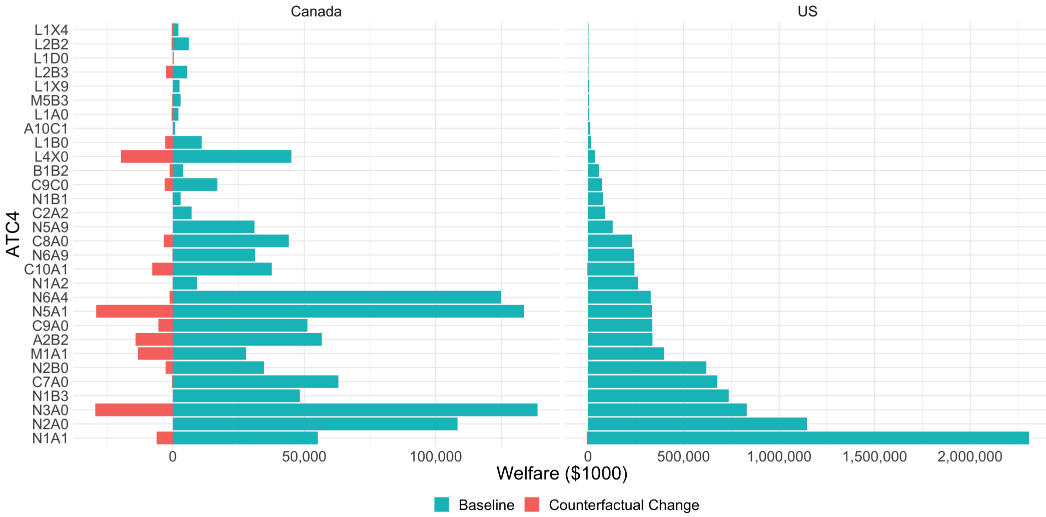
Note: Each blue bar indicates the average annual expenditure in each ATC-4 class in the baseline without reference pricing. The red bar indicates the change in average annual expenditure resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that expenditure increased by the length of the red bar. A red bar to the left of the blue bar indicates that expenditure decreased by the length of the red bar.

Figure B.3: *Counterfactual Profit Changes on All Drugs with US as Price Floor for Canada*



Note: Each blue bar indicates the average annual profits in it ATC-4 class in the baseline without reference pricing. The red bar indicates the change in average annual profits resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that profits increased by the length of the red bar. A red bar to the left of the blue bar indicates that profits decreased by the length of the red bar.

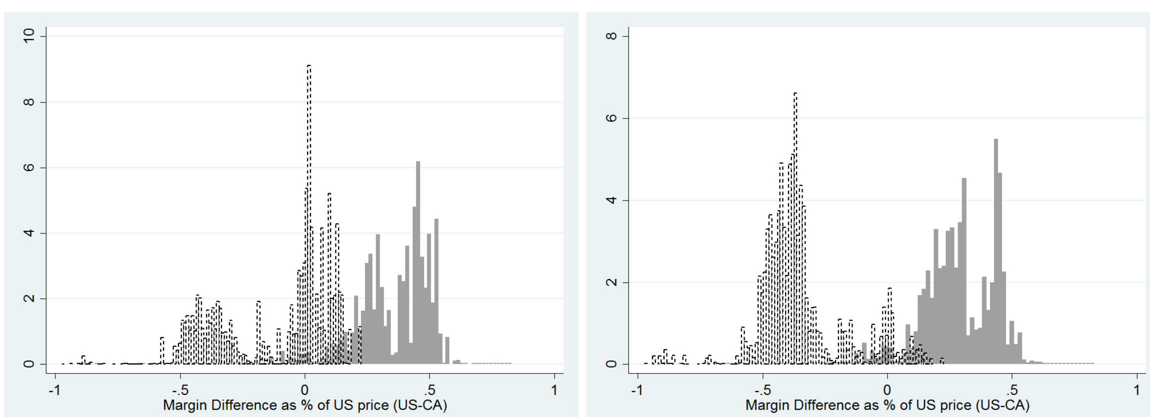
Figure B.4: *Counterfactual Welfare Changes on All Drugs with US as Price Floor for Canada*



Note: Each blue bar indicates the average annual welfare in it ATC-4 class in the baseline without reference pricing. The red bar indicates the change in average annual welfare resulting from imposing reference pricing. A red bar to the right of the blue bar indicates that welfare increased by the length of the red bar. A red bar to the left of the blue bar indicates that welfare decreased by the length of the red bar.

Counterfactuals with US as Price Floor for the Canada

Figure B.5: *Counterfactual Margins Differences between US and Canada for on Patent Drugs when US is Price Floor for Canada*



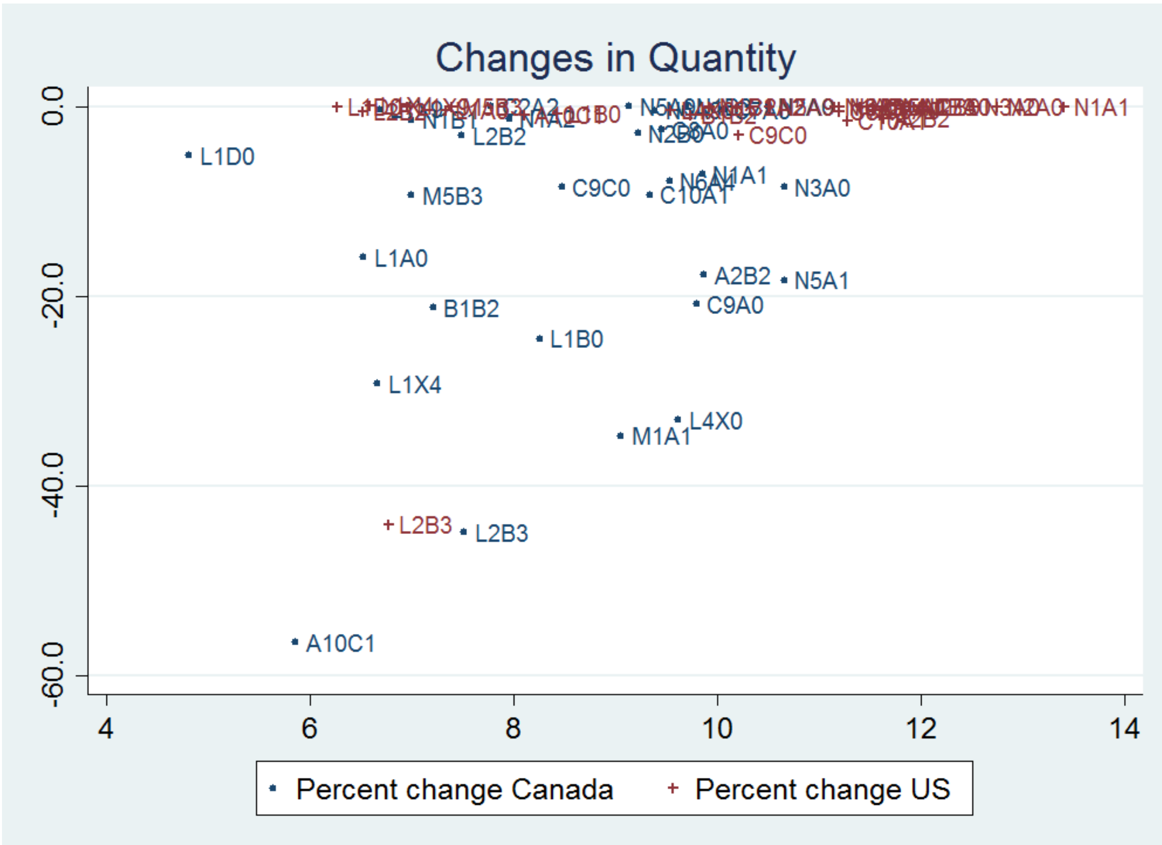
Note: This distribution of margins differences by drug is weighted by the US quantities of the drug. This distribution is for the sample of on patent drugs present in both the US and Canada.

Table B.14: *Counterfactual Quantity on All Drugs when US as Price Floor for Canada*

	ρ_{jm}			Canada			US		
ATC4	On Patent	Branded Off	Generic	Before	After	Δ (%)	Before	After	Δ (%)
A10C1	0.62			349	152	-56.5	3304	3275	-0.9
A2B2	0.55	0.90	0.87	19362	15912	-17.8	113244	111835	-1.2
B1B2	0.70			1356	1068	-21.3	16922	16763	-0.9
C10A1	0.54	1.00	0.77	11349	10290	-9.3	79186	78021	-1.5
C2A2	1.00	1.00	0.94	2384	2384	0.0	26882	26882	-0.0
C7A0	0.72	1.00	1.00	23492	23400	-0.4	167276	167274	-0.0
C8A0	0.56	0.89	0.86	12760	12441	-2.5	73390	73001	-0.5
C9A0	0.47	0.95	1.00	18050	14289	-20.8	101961	101544	-0.4
C9C0	0.60	0.94	0.50	4801	4391	-8.5	27227	26413	-3.0
L1A0	0.91	0.50	1.00	680	572	-15.9	1521	1515	-0.4
L1B0	0.64	0.50	1.00	3855	2910	-24.5	4578	4543	-0.8
L1D0	0.99	0.50	0.50	123	116	-5.2	522	522	0.0
L1X4	1.00	0.50	0.50	785	556	-29.3	722	723	0.1
L1X9	0.92	0.50	0.57	799	796	-0.3	1039	1039	-0.0
L2B2	0.83	0.94	0.61	1791	1735	-3.1	677	673	-0.6
L2B3	0.70	0.79	0.58	1838	1012	-44.9	866	484	-44.1
L4X0	0.95	0.91	1.00	15038	10061	-33.1	13925	13868	-0.4
M1A1	0.44	0.91	1.00	8560	5585	-34.8	101620	101592	-0.0
M5B3	0.93	0.95	0.54	1096	993	-9.4	1614	1614	0.0
N1A1	0.45	0.57	1.00	19107	17729	-7.2	665328	664659	-0.1
N1A2	1.00	1.00	0.92	2865	2828	-1.3	69548	69548	0.0
N1B1	0.96	1.00	0.75	1096	1080	-1.5	20315	20314	-0.0
N1B3	0.50	0.50	0.58	16254	16254	0.0	145882	145882	-0.0
N2A0	0.51	0.78	0.89	36395	36395	-0.0	343829	343829	-0.0
N2B0	0.50	0.96	0.88	10159	9870	-2.8	153266	153266	-0.0
N3A0	0.87	0.93	1.00	42619	38975	-8.5	274813	274734	-0.0
N5A1	0.86	0.86	0.94	42666	34843	-18.3	112308	112259	-0.0
N5A9	0.64	0.97	0.94	9269	9267	-0.0	36694	36693	-0.0
N6A4	0.80	0.99	0.91	13805	12712	-7.9	89527	89520	-0.0
N6A9	0.27	0.89	0.99	11944	11888	-0.5	72854	72851	-0.0
Total				334645	300503	-10.2	2720842	2715136	-.2

Note: Quantity are average yearly standard units (on period 2002-2013). Δ stands for the change of quantity between after and before in percentage of initial quantity. The parameter ρ_{jm} is the one estimated from the supply model in Canada and used for counterfactual simulations.

Figure B.6: *Changes in quantity by ATC-4 in counterfactual against log quantity observed*



Note: Each point corresponds to one ATC-4 labeled with its code. Percentage changes on vertical axis (not drawing if change larger than 200%). Log quantity on horizontal axis. Both US and Canada on same graph.

Table B.15: *Counterfactual Expenses on Patented Drugs when US as Price Floor for Canada*

	ρ_{jm}			Canada			US		
ATC4	On Patent	Branded Off	Generic	Before	After	Δ (%)	Before	After	Δ (%)
A10C1	0.62			4161	7698	85.0	114540	118714	3.6
A2B2	0.55	0.90	0.87	9018	18401	104.0	182955	181462	-0.8
B1B2	0.70			20119	39542	96.5	478597	495483	3.5
C10A1	0.54	1.00	0.77	11938	20641	72.9	139298	136735	-1.8
C2A2	1.00	1.00	0.94	761	761	-0.0	8330	8330	0.0
C7A0	0.72	1.00	1.00	311	320	2.9	27376	27375	-0.0
C8A0	0.56	0.89	0.86	7317	9002	23.0	47992	47297	-1.4
C9A0	0.47	0.95	1.00	6881	7511	9.1	23987	23454	-2.2
C9C0	0.60	0.94	0.50	4056	8191	102.0	61146	60823	-0.5
L1A0	0.91	0.50	1.00	9394	18724	99.3	152325	155363	2.0
L1B0	0.64	0.50	1.00	31685	70195	121.5	496763	508524	2.4
L1D0	0.99	0.50	0.50	8457	10107	19.5	65298	65493	0.3
L1X4	1.00	0.50	0.50	49887	63475	27.2	83505	83278	-0.3
L1X9	0.92	0.50	0.57	23506	23767	1.1	156957	155799	-0.7
L2B2	0.83	0.94	0.61	2638	3668	39.1	6229	6064	-2.6
L2B3	0.70	0.79	0.58	8706	35275	305.2	10127	7450	-26.4
L4X0	0.95	0.91	1.00	45041	143863	219.4	182953	183078	0.1
M1A1	0.44	0.91	1.00	450	2518	459.9	3195	2979	-6.8
M5B3	0.93	0.95	0.54	1077	1212	12.6	25993	25994	0.0
N1A1	0.45	0.57	1.00	6900	8686	25.9	350707	351757	0.3
N1A2	1.00	1.00	0.92	1192	1171	-1.8	70722	70740	0.0
N1B1	0.96	1.00	0.75	1456	1508	3.6	31019	30997	-0.1
N1B3	0.50	0.50	0.58	923	923	-0.0	1362	1362	-0.0
N2A0	0.51	0.78	0.89	346	346	-0.0	870	870	-0.0
N2B0	0.50	0.96	0.88	180	273	51.3	18975	18976	0.0
N3A0	0.87	0.93	1.00	3541	6527	84.3	195096	194883	-0.1
N5A1	0.86	0.86	0.94	29323	59479	102.8	865930	866139	0.0
N5A9	0.64	0.97	0.94	1716	1716	0.0	1153	1152	-0.1
N6A4	0.80	0.99	0.91	2473	2370	-4.2	113105	113095	-0.0
N6A9	0.27	0.89	0.99	341	344	0.8	4674	4668	-0.1
Total				293793	568212	93.40	3921179	3948336	.69

Note: Expenses are average yearly expenses in 1000 US$ (on period 2002-2013). Patented drugs only.

Table B.16: *Counterfactual Prices by ATC-4 when US as Price Floor for Canada*

	ρ_{jm}			Price Change All drugs		Price Change Patented		Price Change Branded Off		Price Change Generic	
ATC4	On Patent	Branded Off	Generic	CA (%)	US (%)	CA (%)	US (%)	CA (%)	US (%)	CA (%)	US (%)
A10C1	0.62			328.3	5.2	328.3	5.2				
A2B2	0.55	0.90	0.87	138.8	0.6	209.4	1.2	92.1	0.0	24.2	0.1
B1B2	0.70			149.5	4.2	149.5	4.7		2.7		-0.0
C10A1	0.54	1.00	0.77	91.0	0.7	141.9	2.6	184.1	1.7	40.0	0.4
C2A2	1.00	1.00	0.94	-0.0	0.0	-0.0	0.0	0.0	0.0	-0.0	0.0
C7A0	0.72	1.00	1.00	1.5	0.0	3.3	0.1	0.5	-0.0	1.1	0.0
C8A0	0.56	0.89	0.86	29.3	1.5	49.3	3.6	18.4	-0.2	19.9	0.2
C9A0	0.47	0.95	1.00	34.3	-0.5	80.1	1.5	2.1	-0.0	8.3	0.0
C9C0	0.60	0.94	0.50	112.7	2.3	128.2	3.0	23.0	-0.2	1.8	0.5
L1A0	0.91	0.50	1.00	147.1	2.5	262.0	5.7	15.5	0.1	160.3	-0.0
L1B0	0.64	0.50	1.00	214.8	3.2	289.8	12.4	-1.4	0.0	164.9	-0.1
L1D0	0.99	0.50	0.50	21.0	0.0	174.1	-0.5	-3.0	-0.0	2.3	-0.0
L1X4	1.00	0.50	0.50	81.6	-0.4	85.4	-0.4	3.8		0.0	0.0
L1X9	0.92	0.50	0.57	1.5	-0.6	48.9	0.1	1.4		2.1	-0.0
L2B2	0.83	0.94	0.61	37.9	-2.7	320.5	0.5	16.9	0.2	11.1	-0.2
L2B3	0.70	0.79	0.58	612.1	29.2	648.0	33.1	24.4	0.1	15.5	0.0
L4X0	0.95	0.91	1.00	392.3	0.6	414.7	0.8	5.6	0.0	158.4	-0.0
M1A1	0.44	0.91	1.00	769.3	-0.7	607.0	6.0	186.8	0.0	936.3	-0.0
M5B3	0.93	0.95	0.54	22.4	-0.0	53.8	-0.0	21.3	0.0	2.4	0.0
N1A1	0.45	0.57	1.00	36.7	0.4	50.1	0.8	23.5	-0.0	26.2	-0.1
N1A2	1.00	1.00	0.92	2.3	0.0	41.4	-0.1	1.8	-0.0	1.4	0.0
N1B1	0.96	1.00	0.75	6.7	-0.0	16.4	0.1	10.1	-0.0	4.3	-0.0
N1B3	0.50	0.50	0.58	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
N2A0	0.51	0.78	0.89	0.0	0.0	0.1	0.0	-0.0	0.0	0.0	0.0
N2B0	0.50	0.96	0.88	29.0	0.0	4.5	0.0	22.8	0.0	30.3	0.0
N3A0	0.87	0.93	1.00	117.4	-0.0	132.8	0.2	79.6	0.0	133.3	-0.0
N5A1	0.86	0.86	0.94	121.3	0.1	372.1	0.1	47.8	-0.1	51.4	-0.0
N5A9	0.64	0.97	0.94	0.0	0.0	0.5	0.3	-0.0	-0.0	0.0	0.0
N6A4	0.80	0.99	0.91	9.5	0.0	149.6	0.1	0.5	0.0	0.4	-0.0
N6A9	0.27	0.89	0.99	1.2	0.0	6.7	0.5	-0.1	0.0	0.8	0.0

Note: Changes in % of initial price using market shares weighted average prices.

Table B.17: *Counterfactual Prices of Drugs on Patent present in both US and Canada when US as Price Floor for Canada*

	Before		After			
	Canada	US	Canada		US	
ATC4	Price	Price	Price	Δ (%)	Price	Δ (%)
A10C1	12.14421	37.09733	61.24129	404.28	39.06341	5.30
A2B2	.8079825	2.675094	2.821949	249.26	2.70827	1.24
B1B2	15.36494	37.85803	41.56478	170.52	39.64621	4.72
C10A1	1.770697	3.82358	4.284205	141.95	3.982836	4.17
C2A2	54.75095	13.18811	54.72902	-0.04	13.18826	0.00
C7A0	1.244643	7.688396	2.185894	75.62	7.714972	0.35
C8A0	1.268361	2.297664	2.40667	89.75	2.385341	3.82
C9A0	.5887031	1.685948	2.012465	241.85	1.717792	1.89
C9C0	1.094559	2.528681	2.574439	135.20	2.621561	3.67
L1A0	25.97724	205.1478	95.36224	267.10	216.579	5.57
L1B0	17.3866	268.2582	97.04	458.13	306.8802	14.40
L1D0	292.0294	780.7533	1263.382	332.62	779.0521	-0.22
L1X4	66.59953	99.22975	125.893	89.03	98.87096	-0.36
L1X9	753.4308	886.57	1125.542	49.39	887.5515	0.11
L2B2	10.43049	32.88236	43.86275	320.52	33.14288	0.79
L2B3	4.865076	11.75005	39.00875	701.81	16.18691	37.76
L4X0	3.320533	10.59573	17.55219	428.60	10.66112	0.62
M1A1	.6708143	2.973108	6.520396	872.01	3.234343	8.79
M5B3	6.807281	16.13616	18.41846	170.57	16.13214	-0.02
N1A1	.662663	.919805	1.044167	57.57	.9271347	0.80
N1A2	12.79765	75.2755	31.65223	147.33	75.17893	-0.13
N1B1	13.03696	16.28716	16.27862	24.87	16.30099	0.08
N1B3						
N2A0	1.003563	3.317234	3.382701	237.07	3.317749	0.02
N2B0						
N3A0	1.384019	3.735585	3.464821	150.34	3.752379	0.45
N5A1	1.869504	9.782676	9.571216	411.97	9.7985	0.16
N5A9	.8207119	1.356838	1.388645	69.20	1.396926	2.95
N6A4	1.334195	3.653124	3.710609	178.12	3.656124	0.08
N6A9	.6356424	3.21982	2.009834	216.19	3.244288	0.76

Note: Market shares weighted average price of patented drugs by ATC-4, country for drugs present in both only. Percentage changes are changes with respect to the initial situation.

Table B.18: *Counterfactual Consumer Welfare Changes on All Drugs when US as Price Floor for Canada*

ATC4	ρ_{jm} On Patent	ρ_{jm} Branded Off	ρ_{jm} Generic	Canada Before	Canada After	Canada Δ (%)	US Before	US After	US Δ (%)
A10C1	0.62			885	820	-7.3	11159	10912	-2.2
A2B2	0.55	0.90	0.87	56646	42488	-25.0	337119	335034	-0.6
B1B2	0.70			3914	2712	-30.7	55450	53879	-2.8
C10A1	0.54	1.00	0.77	37661	29862	-20.7	242072	238483	-1.5
C2A2	1.00	1.00	0.94	7111	7112	0.0	88937	88936	-0.0
C7A0	0.72	1.00	1.00	62951	62608	-0.5	675640	675626	-0.0
C8A0	0.56	0.89	0.86	44044	40615	-7.8	230677	229045	-0.7
C9A0	0.47	0.95	1.00	51194	45660	-10.8	335249	334683	-0.2
C9C0	0.60	0.94	0.50	16875	13793	-18.3	72193	70600	-2.2
L1A0	0.91	0.50	1.00	2092	1570	-24.9	5245	5204	-0.8
L1B0	0.64	0.50	1.00	10952	8058	-26.4	15958	15786	-1.1
L1D0	0.99	0.50	0.50	363	336	-7.5	2161	2162	0.0
L1X4	1.00	0.50	0.50	2126	1725	-18.9	2109	2116	0.3
L1X9	0.92	0.50	0.57	2486	2467	-0.8	3771	3769	-0.1
L2B2	0.83	0.94	0.61	6150	5665	-7.9	2158	2149	-0.4
L2B3	0.70	0.79	0.58	5427	2814	-48.1	2336	1665	-28.7
L4X0	0.95	0.91	1.00	45075	25378	-43.7	35499	35382	-0.3
M1A1	0.44	0.91	1.00	27881	14616	-47.6	397633	397520	-0.0
M5B3	0.93	0.95	0.54	2917	2656	-9.0	5131	5131	0.0
N1A1	0.45	0.57	1.00	55089	48904	-11.2	2307142	2299762	-0.3
N1A2	1.00	1.00	0.92	9216	9144	-0.8	260259	260260	0.0
N1B1	0.96	1.00	0.75	2969	2897	-2.4	76524	76521	-0.0
N1B3	0.50	0.50	0.58	48316	48316	0.0	736051	736050	-0.0
N2A0	0.51	0.78	0.89	108236	108233	-0.0	1145022	1145022	-0.0
N2B0	0.50	0.96	0.88	34712	31972	-7.9	618698	618692	-0.0
N3A0	0.87	0.93	1.00	138685	109231	-21.2	830380	830237	-0.0
N5A1	0.86	0.86	0.94	133543	104407	-21.8	333726	333418	-0.1
N5A9	0.64	0.97	0.94	31051	31046	-0.0	128193	128192	-0.0
N6A4	0.80	0.99	0.91	124721	123527	-1.0	327380	327334	-0.0
N6A9	0.27	0.89	0.99	31311	31166	-0.5	240522	240506	-0.0
Total				1104596	959797	-13.1	9524394	9504077	-.2

Note: Welfare values are average yearly on period 2002-2013 scaled by market size. Δ stands for the change of welfare between after and before in percentage of initial welfare. The parameter ρ_{jm} is the one estimated from the supply model in Canada and used for counterfactual simulations.

Table B.19: *Counterfactual Expenses, profits and Consumer Welfare Global Changes on All Drugs when US as Price Floor for Canada*

	Expenses			Profits		
ATC4	Before	After	Δ (%)	Before	After	Δ (%)
A10C1	118701	126412	6.5	101159	108738	7.5
A2B2	284788	296368	4.1	154494	168190	8.9
B1B2	628192	667698	6.3	589713	630485	6.9
C10A1	186216	195399	4.9	101991	114855	12.6
C2A2	38062	38063	0.0	5699	5699	0.0
C7A0	148647	148695	0.0	32590	32648	0.2
C8A0	259058	264450	2.1	107268	111676	4.1
C9A0	66978	67364	0.6	20164	22164	9.9
C9C0	68047	71946	5.7	35648	40708	14.2
L1A0	219964	235239	6.9	91762	105201	14.6
L1B0	610474	679005	11.2	272135	333723	22.6
L1D0	192101	195639	1.8	75846	78404	3.4
L1X4	135841	149569	10.1	94154	107787	14.5
L1X9	194257	193392	-0.4	95351	95906	0.6
L2B2	10886	12016	10.4	6970	8310	19.2
L2B3	19263	43231	124.4	14346	40579	182.9
L4X0	384207	494727	28.8	140623	249700	77.6
M1A1	28816	36290	25.9	4139	11752	183.9
M5B3	31274	31617	1.1	19780	20121	1.7
N1A1	558890	564717	1.0	362079	368412	1.7
N1A2	642266	642396	0.0	170936	171068	0.1
N1B1	93051	93261	0.2	25374	25602	0.9
N1B3	167484	167484	-0.0	29296	29296	0.0
N2A0	486180	486182	0.0	65131	65133	0.0
N2B0	84433	85358	1.1	13931	14791	6.2
N3A0	450060	460358	2.3	127309	137743	8.2
N5A1	1055035	1112993	5.5	546149	604642	10.7
N5A9	51146	51148	0.0	6999	7001	0.0
N6A4	155712	155639	-0.0	75488	75805	0.4
N6A9	48988	49002	0.0	15027	15067	0.3
Total	7419016	7815658	5.3	3401549	3801205	11.7

Note: All values are average yearly on period 2002-2013, summing US and Canada. Δ stands for the change between after and before in percentage of initial value.

Table B.20: *Counterfactual Prices when US is using Canada as Maximum Reference Price and ex ante commitment*

	All				Patented				Branded Off				Generic			
	Before		After		Before		After		Before		After		Before		After	
ATC4	CA	US	CA	US	CA	US	CA	US	CA	US	CA	US	CA	US	CA	US
A10C1	12.05	37.27	51.61	39.22	12.05	37.27	51.61	39.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A2B2	0.76	2.44	1.81	2.45	0.83	2.75	2.56	2.78	0.67	3.43	1.30	3.43	0.59	0.63	0.74	0.64
B1B2	15.14	36.08	37.79	37.60	15.14	37.86	37.79	39.65	0.00	30.42	0.00	31.25	0.00	28.47	0.00	28.47
C10A1	1.32	2.23	2.53	2.24	1.77	3.43	4.28	3.52	1.99	2.32	5.67	2.36	0.49	0.50	0.68	0.50
C2A2	0.60	1.35	0.60	1.35	46.08	13.16	46.06	13.16	2.79	2.09	2.79	2.09	0.16	1.06	0.16	1.06
C7A0	0.22	0.87	0.22	0.87	0.30	3.37	0.31	3.37	1.11	2.14	1.12	2.14	0.18	0.60	0.18	0.60
C8A0	0.93	3.40	1.20	3.45	1.30	2.32	1.94	2.40	0.83	25.38	0.98	25.33	0.49	1.17	0.59	1.17
C9A0	0.52	0.51	0.70	0.51	0.68	1.68	1.22	1.71	0.51	1.70	0.52	1.70	0.26	0.30	0.29	0.30
C9C0	0.97	2.31	2.06	2.36	1.10	2.72	2.52	2.80	1.19	2.64	1.47	2.63	0.25	0.47	0.26	0.47
L1A0	18.64	145.03	46.07	148.60	25.26	248.29	91.44	262.43	1.52	105.28	1.76	105.42	15.51	57.18	40.36	57.18
L1B0	16.21	127.49	51.05	131.58	17.50	401.51	68.22	451.30	28.88	223.27	28.49	223.35	11.06	16.72	29.30	16.70
L1D0	164.08	322.95	198.53	323.10	250.68	1350.26	687.10	1344.01	360.70	998.83	350.01	998.54	77.61	108.92	79.41	108.91
L1X4	66.23	112.85	120.26	112.42	66.35	112.77	123.04	112.33	65.16	0.00	67.65	0.00	25.55	146.05	25.56	146.05
L1X9	25.11	159.47	25.49	158.48	749.23	866.80	1115.65	867.61	0.95	0.00	0.96	0.00	1.56	15.22	1.59	15.22
L2B2	2.19	10.08	3.01	9.81	10.43	30.28	43.86	30.42	1.45	9.63	1.69	9.65	0.69	1.31	0.76	1.31
L2B3	4.80	11.34	34.18	14.66	4.88	11.83	36.49	15.75	3.80	17.52	4.72	17.54	2.23	0.52	2.58	0.52
L4X0	3.22	23.88	15.87	24.03	3.13	18.62	16.10	18.78	0.76	5.84	0.80	5.85	6.01	61.97	15.53	61.95
M1A1	0.19	0.28	1.67	0.27	0.68	3.61	4.77	3.83	0.50	0.94	1.45	0.94	0.13	0.22	1.31	0.22
M5B3	2.17	18.79	2.65	18.79	2.63	30.08	4.05	30.08	3.21	21.10	3.89	21.10	1.40	2.46	1.43	2.46
N1A1	0.81	0.82	1.10	0.82	0.79	0.92	1.19	0.93	0.92	0.76	1.14	0.76	0.26	0.54	0.33	0.54
N1A2	5.00	9.03	5.11	9.03	11.73	76.79	16.59	76.72	5.43	18.96	5.53	18.96	4.53	5.96	4.59	5.96
N1B1	4.48	4.35	4.78	4.34	10.97	15.64	12.76	15.65	4.63	6.76	5.10	6.76	3.19	2.84	3.32	2.84
N1B3	0.92	1.11	0.92	1.11	6.68	23.42	6.68	23.42	1.04	3.90	1.04	3.90	0.39	0.85	0.39	0.85
N2A0	0.59	1.36	0.59	1.36	0.77	3.20	0.77	3.20	1.18	3.86	1.18	3.86	0.48	1.14	0.48	1.14
N2B0	0.31	0.53	0.40	0.53	0.56	14.59	0.58	14.59	0.34	1.34	0.42	1.34	0.30	0.39	0.39	0.39
N3A0	0.26	1.49	0.58	1.49	1.36	4.30	3.17	4.31	0.20	5.65	0.35	5.65	0.20	0.79	0.46	0.79
N5A1	1.68	8.76	3.72	8.76	1.92	10.88	9.06	10.89	3.13	9.60	4.62	9.59	0.40	3.16	0.61	3.16
N5A9	0.46	1.27	0.46	1.27	2.90	2.76	2.91	2.76	1.17	16.30	1.17	16.30	0.25	1.11	0.25	1.11
N6A4	0.48	1.61	0.53	1.61	1.35	3.61	3.36	3.62	1.44	4.22	1.45	4.22	0.28	0.48	0.29	0.48
N6A9	0.19	0.63	0.19	0.63	0.40	2.95	0.43	2.97	0.58	3.47	0.58	3.47	0.13	0.32	0.13	0.32

Note: Market shares weighted average price by ATC-4, country.

Table B.21: *Counterfactual Expenses by ATC-4 with varying MFN rule (0, +33%, +50%) and Larger Reference Market (with ex ante commitment)*

			ρ_{jm}			Canada			US		
ATC4	MFN	Share US market	On Patent	Branded Off	Generic	Before	After	Δ (%)	Before	After	Δ (%)
A2B2	0	0	0.55	0.90	0.87	14057	27056	92.5	270730	269312	-0.5
A2B2	0	50	0.55	0.90	0.87	92679	415627	348.5	270730	229192	-15.3
A2B2	0	100	0.55	0.90	0.87	185358	860600	364.3	270730	218992	-19.1
A2B2	33	0	0.55	0.90	0.87	14057	24039	71.0	270730	270848	0.0
A2B2	50	0	0.55	0.90	0.87	14057	22464	59.8	270730	271066	0.1
B1B2	0	0	0.70			20119	39542	96.5	608073	628156	3.3
B1B2	0	50	0.70			135563	279338	106.1	678039	770136	13.6
B1B2	0	100	0.70			238686	620087	159.8	608073	741229	21.9
B1B2	33	0	0.70			20119	29369	46.0	608073	599771	-1.4
B1B2	50	0	0.70			20119	26030	29.4	608073	620713	2.1
L4X0	0	0	0.95	0.91	1.00	48328	158193	227.3	335879	336534	0.2
L4X0	0	50	0.95	0.91	1.00	36593	120167	228.4	346961	347399	0.1
L4X0	0	100	0.95	0.91	1.00	73186	241375	229.8	346961	347999	0.3
L4X0	33	0	0.95	0.91	1.00	53661	128286	139.1	366193	353273	-3.5
L4X0	50	0	0.95	0.91	1.00	53661	98516	83.6	366193	359094	-1.9
N1A1	0	0	0.45	0.57	1.00	15417	19477	26.3	543474	545239	0.3
N1A1	0	50	0.45	0.57	1.00	220806	290984	31.8	543474	569227	4.7
N1A1	0	100	0.45	0.57	1.00	441612	616921	39.7	543474	578469	6.4
N1A1	33	0	0.45	0.57	1.00	15417	16624	7.8	543474	543661	0.0
N1A1	50	0	0.45	0.57	1.00	15417	15631	1.4	543474	543431	-0.0
N1A2	0	0	1.00	1.00	0.92	14275	14395	0.8	627990	628001	0.0
N1A2	0	50	1.00	1.00	0.92	65445	66059	0.9	627990	628042	0.0
N1A2	0	100	1.00	1.00	0.92	130890	132107	0.9	627990	628091	0.0
N1A2	33	0	1.00	1.00	0.92	14275	14345	0.5	627990	628235	0.0
N1A2	50	0	1.00	1.00	0.92	14275	14306	0.2	627990	628182	0.0
N2A0	0	0	0.51	0.78	0.89	21736	21737	0.0	464444	464445	0.0
N2A0	0	50	0.51	0.78	0.89	108392	108398	0.0	464444	464444	-0.0
N2A0	0	100	0.51	0.78	0.89	216785	216796	0.0	464444	464444	-0.0
N2A0	33	0	0.51	0.78	0.89	21736	21737	0.0	464444	464445	0.0
N2A0	50	0	0.51	0.78	0.89	21736	21737	0.0	464444	464445	0.0
N3A0	0	0	0.87	0.93	1.00	11366	21817	91.9	438695	438541	-0.0
N3A0	0	50	0.87	0.93	1.00	70808	158094	123.3	438695	437694	-0.2
N3A0	0	100	0.87	0.93	1.00	141617	318854	125.2	438695	436773	-0.4
N3A0	33	0	0.87	0.93	1.00	11366	19574	72.2	438695	440600	0.4
N3A0	50	0	0.87	0.93	1.00	11056	17313	56.6	432920	432944	0.0
N5A1	0	0	0.86	0.86	0.94	72065	129771	80.1	982970	983222	0.0
N5A1	0	50	0.86	0.86	0.94	98298	181237	84.4	982970	983347	0.0
N5A1	0	100	0.86	0.86	0.94	196596	362947	84.6	982970	983687	0.1
N5A1	33	0	0.86	0.86	0.94	72065	116842	62.1	982970	968099	-1.5
N5A1	50	0	0.86	0.86	0.94	72065	106353	47.6	982970	945020	-3.9

Appendix C

Appendix to Chapter 3

C.0.1 Additional Tables and Figures

Figure C.1: *Examples of Telematics Devices in Auto Insurance*

	Basic Specification	**Primary Specification**	**Hold-Out Data**
Monitoring share (if eligible)	21.2%	17.9%	17.6%
Expected score	5.23	3.97	4.17
Selection effect (risk)	5.2%	23.7%	-
Coverage share			
30K	-	-	-
40K	9.4%	7.6%	7.2%
50K	66.3%	60.5%	58.1%
100K	13.4%	17.5%	19.6%
300K	9.7%	10.9%	12.8%
500K	1.3%	3.6%	2.4%
First renewal attrition	132.2%	104.2%	100.0%

Note: This table reports our cross-validation result. All measures are calculated analogously as table 3.5. For the state that changed mandatory minimum, the hold-out data include all post-period data. For the other two states, the hold-out data include all observations that are not in our demand estimation data.

Table C.1: *Cross Validation*

Table C.2: *Model Estimate - Heterogeneous Latent Parameters*

	Log Claim Risk (μ_λ)	Monitoring Disutility (ξ, \$)	Firm-switching Inertia (η, \$)
	(1)	(2)	(3)
Intercept	-3.294^{***}	96.773^{***}	228.559^{***}
	(0.080)	(2.813)	(6.213)
Private Risk		25.238^{***}	
		(1.657)	
Monitoring Ind.	0.404^{***}		
	(0.063)		
Monitoring Duration	-0.796^{***}		
	(0.081)		
Driver			
Driver Age	-0.240^{***}	-1.049^{**}	4.526^{***}
	(0.053)	(0.437)	(1.641)
– Square	0.156^{***}	-1.047^{***}	3.816^{**}
	(0.055)	(0.309)	(0.742)
Age < 25	0.081^{**}	0.326	-0.500
	(0.032)	(0.339)	(0.922)
Age > 21	-0.064	-0.059	3.195^{***}
	(0.053)	(0.403)	(0.449)
Age > 60	-0.046	-0.139	-0.275
	(0.068)	(1.689)	(0.340)
Year of Education	0.001	-2.452^{***}	-7.526^{***}
	(0.025)	(0.331)	(0.915)

College Ind.	−0.00001	−0.952***	0.234
	(0.038)	(0.339)	(0.237)
Post Grad Ind.	0.005	−0.728	−1.547
	(0.039)	(1.644)	(1.686)
Female Ind.	0.099***	−0.261	1.007
	(0.021)	(1.643)	(1.686)
Driver License Year	−0.018	−0.016	16.776***
	(0.019)	(0.905)	(0.338)
Home Ownership	−0.020	−0.039	0.058
	(0.038)	(0.447)	(1.653)
Out-of-State License	−0.104***	−0.380	−0.406
	(0.030)	(0.339)	(0.922)
Location			
Garage Verified Ind.	−0.069*	0.008	1.847**
	(0.036)	(0.521)	(0.922)
Population Density	0.076***	0.359	−4.902***
	(0.015)	(0.419)	(0.445)
Zipcode Income	−0.058***	0.610	−2.936*
	(0.017)	(1.615)	(1.677)
Log Zipcode Income	0.031***	0.284	−0.808
	(0.008)	(2.949)	(1.850)
Vehicle			
Length of Ownership	0.017	−0.918	−0.084
	(0.012)	(0.887)	(0.338)
Vehicle on Lease Ind.	0.092***	−1.058	4.789***
	(0.024)	(1.677)	(0.343)
Model Year	−0.026*	−1.621***	3.211***
	(0.014)	(0.421)	(0.445)
ABS Ind.	−0.058*	0.034	−1.626***
	(0.035)	(0.741)	(0.422)
Airbag Ind.	0.014	0.199	1.225
	(0.021)	(1.644)	(1.686)
Class C Ind.	0.023	0.079	3.843**
	(0.053)	(0.448)	(1.655)
Tier			
Credit Report Ind.	0.044	0.414	1.832***
	(0.035)	(0.429)	(0.448)
Delinq. Score	−0.016	2.114***	10.959***
	(0.014)	(0.331)	(0.917)
Prior Ins. Length	−0.038**	−2.293	−3.993***
	(0.017)	(1.648)	(0.338)
Has Prior Ins.	−0.067*	−1.183***	−0.759*
	(0.035)	(0.427)	(0.448)
– w/ Lapse	−0.050	0.204	0.001
	(0.043)	(1.686)	(0.620)
Violation Points	−0.032	1.084***	4.333***
	(0.030)	(0.337)	(0.429)
Clean Record Ind.	−0.097***	−0.909	−1.392***
	(0.035)	(0.916)	(0.342)
Total Accident Count	0.115***	0.470	−0.139

	(0.029)	(1.638)	(1.690)
Total DUI Count	-0.233^{***}	0.031	0.326
	(0.065)	(0.922)	(0.536)
Log Risk Class	0.275^{***}		
	(0.046)		
Risk Class	0.042		
	(0.074)		
– Square	-0.124^{*}		
	(0.073)		
– Cube	0.0002		
	(0.046)		
Seasonality	0.026^{**}	-0.764^{**}	-1.585^{***}
	(0.011)	(0.331)	(0.427)
– Square	0.063	-0.364	-0.519
	(0.046)	(0.340)	(0.430)
Trend Year	0.083^{*}	-1.570	7.417^{***}
	(0.043)	(1.660)	(0.338)
– Square	-0.102^{***}	-1.413	6.199^{***}
	(0.039)	(1.830)	(1.674)

*p<0.1; **p<0.05; ***p<0.01

Note: This table reports intercept and slope estimates for heterogeneous latent parameters of our structural model. The latent parameters are risk occurrence (λ), firm-switching inertia (η), and monitoring disutility (ξ). Continuous covariate variables are normalized (except λ and monitoring duration). Discrete variables with more than two values are normalized so that the minimum is zero. Deliq. Score is a measure of financial responsibility from the credit bureaus. The higher the score and worse the record.

C.0.2 Analysis of Actual Firm Pricing

Cream skimming effect Advantageous selection into monitoring may cream skim from the firm's unmonitored pool. As a result, firms may choose to raise prices in the unmonitored pool. In addition, they may also want to surcharge the unmonitored pool to indirectly encourage monitoring participation. To test the effect of monitoring introduction on the unmonitored pool more formally, we take advantage of the staggered introduction of monitoring across states. This gives rise to a regression discontinuity strategy that evaluates how prices and average cost changed in the *unmonitored* pool. We focus on a year before and after monitoring introduction; our observable characteristics also include state fixed effects and flexible controls for trends and seasonality. We only focus on the first semester

	$\mathbb{E}[R_{0,m=0,t=0}]$ (1)	μ_s (2)	$\mathbb{E}[R_{0,m=0,t=1}]$ (3)
Intercept	-0.362^{***}	11.367^{***}	-1.131^{***}
	(0.001)	(0.506)	(0.132)
Log Risk Class	-0.413^{***}	-0.384^{**}	-0.080^{***}
	(0.018)	(0.155)	(0.018)
Risk Class	0.367^{***}	-0.077	0.063
	(0.051)	(0.304)	(0.034)
– Square	-0.290^{***}	0.245	-0.155^{***}
	(0.054)	(0.308)	(0.036)
– Cube	-0.229^{***}	-0.039	0.031
	(0.022)	(0.140)	(0.019)
$\ln \lambda$		1.859^{***}	
		(0.094)	
log(Monitoring Score)			0.150^{***}
			(0.005)

*p<0.1; **p<0.05; ***p<0.01

Note: This table reports estimates for the renewal pricing and monitoring score model. Instead of modeling the Gamma shape parameters (α), we use a change-of-variables technique to directly estimate the expected renewal rate. It is modeled with a Sigmoid function between 0.5 (50% cheaper) and 2 (twice as expensive). That is, $\mathbb{E}[R_0] = \sigma(\mathbf{x}'\theta_R) \times 1.5 + 0.5$. We include the appropriate Jacobian adjustments in estimation, and winsorize away extremely large or small renewal price change.

Table C.3: *Estimation Results - Renewal Pricing and Monitoring Score*

($t = 0$) to avoid contamination from attrition[1]. We therefore drop the t subscript, and run the following regression

$$dep.\ var._i = \alpha + \gamma Qtr_i + \kappa \mathbf{1}_{post,i} + \theta \cdot Qtr_i \times \mathbf{1}_{post,i} + \mathbf{x}_i'\beta + \xi_{y,i} + \epsilon_i \tag{C.1}$$

We use price p_i and claim count C_i as our dependent variable. Qtr is the running variable, which denotes the calendar quarter when driver i arrived at the firm[2]. $\mathbf{1}_{post}$ is an indicator for whether i arrived at the firm after the introduction of monitoring. $\mathbf{x}$ and a coverage fixed effect ξ_y soak up compositional changes in observable risk class and coverage plans. The coefficient θ reveals treatment effect of monitoring introduction on prices and claims in the unmonitored pool.

Estimates for $\hat{\theta}$ across various specifications are reported in C.3. The firm did not raise prices around monitoring introduction. We also find no evidence that the average cost of the unmonitored pool deteriorated by more than 2%.

In reality, monitoring is only a small fraction of the market. As our demand estimates will reveal in the next section, even when monitored drivers are significantly better, its influence on the unmonitored pool is significantly limited by its small size. Further, the firm does not make follow-up offers to customers who initially opted out monitoring, which is necessary for unraveling to occur empirically. Lastly, monitoring programs are subject to approval by state commissioners. And a new program that affects baseline pricing may be subject to more regulatory scrutiny. On the flip side, this suggests that the current monitoring regime is largely welfare-neutral for unmonitored drivers.

Dynamic and non-uniform pricing The firm is not required to offer monitoring, it therefore must benefit from it to justify administrative and R&D costs. Indeed, monitored drivers have 35% higher profitability overall, controlling for observables. On top of reduced moral hazard (during monitoring) and better risk rating (going forward), this can also be a

[1] This regression does not include monitored drivers, so there is no contamination from moral hazard.

[2] It is normalized so that the quarter immediately after monitoring introduction is indexed as 0.

result of higher profit margin and retention rate when information is revealed. We provide descriptive evidence on pricing and dynamic retention in this section.

First, the firm faces a dynamic pricing problem as information is revealed at the end of the first period. It offers a opt-in discount to encourage all drivers to participate in monitoring. This averages to around 5% across states and time.

When monitoring information is revealed, the firm can use it to set non-uniform prices. Here, the firm's pricing schedule is based on a monitoring tier that measures how "surprising" a given driver's monitoring score is to the firm. In C.13, we plot the empirical distribution of monitoring tier, which is realized monitoring score divided by firm's expected score given observables[3]. Consistent with our findings above, the average monitored driver performed much better than expected[4].

C.4 presents the discount schedule the firm uses given the percentile of monitoring tier as defined above. Surprisingly good drivers are on the left, who are offered the highest renewal discount, while around 25% of drivers that performed poorly (compared to firm's expectation) received a surcharge.

C.5 plots the corresponding retention rate. It is clear that as discounts approach zero or negative, retention rate drops significantly. In fact, we can regress renewal choice (binary) on prices with monitoring discount, controlling for observables and price level without the discount. θ then measures the slope of the residual (retention) demand.

$$\mathbf{1}_{renew,i} = \alpha + \delta p_i + \theta disc_i + \mathbf{x}_i'\mathbf{fi} + \epsilon_i \tag{C.2}$$

The estimates for $\hat{\theta}$ are reported in C.6. Without monitoring discount, a $1 increase in price (decrease in discount given) causes the retention rate to drop by 0.07 percentage points (7 basis points). When firms give discounts, however, the slope of the demand decreases, and

[3]For monitored driver i, the expected score is derived based on the average driver in i's observable (x_i) group. It also does not take into account the fact that i has selected into monitoring. The graph has a long right tail and is truncated at 200%.

[4]It is important to note that a driver with a monitoring tier of 30% is not necessarily 70% safer than the average person in her pool, especially in renewal period. This is because monitoring score does not capture risk perfectly, and it is also stochastic. Our structural model quantifies these effects more formally.

by 56% when the discount given is larger than 10%.

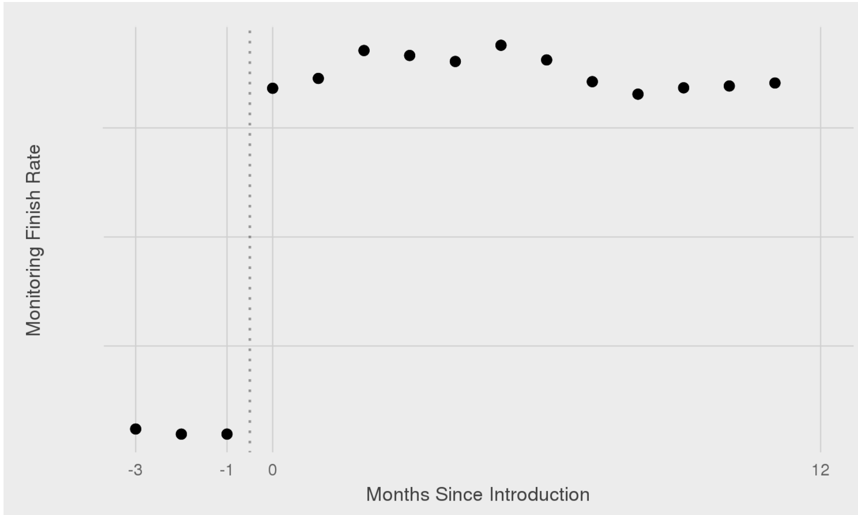


Figure C.2: *Monthly monitoring finish rate around monitoring introduction*

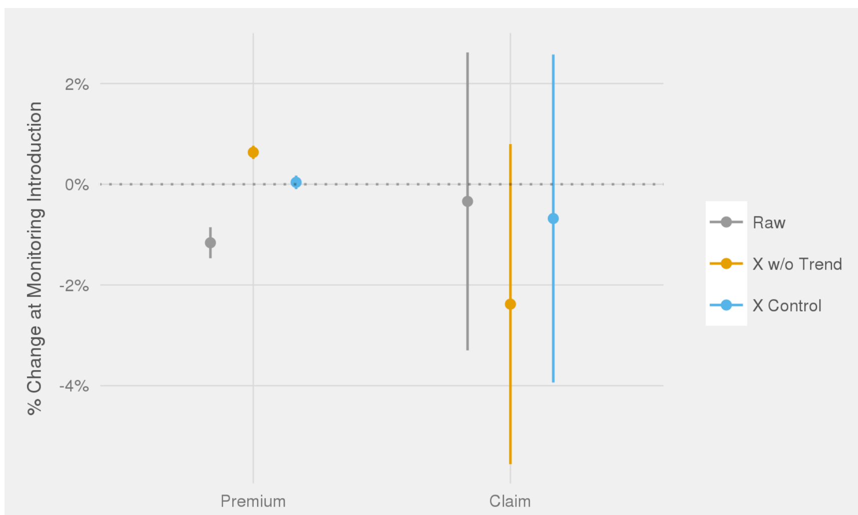


Figure C.3: *Event Study: treatment effect of monitoring introduction on the unmonitored pool*

Note: C.2 the progression of monthly monitoring finish rate around the introduction of monitoring. The monthly finish rate are below 0.1% in all months before monitoring introduction. The reason why it is not exactly zero before monitoring introduction is due to small-scale trial and experimentation. We throw out states that introduced monitoring in the first three months or the last 12 months of our research window. This ensures that the trend we see do not pick up changes in state composition. C.3 reports regression-discontinuity estimate θ of (**??**), where the horizontal axis distinguishes dependent variable used. These effects are translated in percentage terms by dividing the average of the dependent variable in the period immediately before monitoring introduction. We look at only first period outcomes, and include all *unmonitored* drivers arriving at the firm a year before or after the firm. States that introduced monitoring within a year after the beginning or a year before the end of our research window are excluded. The running variable is quarter since monitoring introduction. Different colors and positions represent different specifications of control variables (x_{it}). The grey (left-most) series represents estimates from regressions with the full set of x_{it}; the orange (middle) one includes a full set of observables, including flexible controls for trend and seasonality.

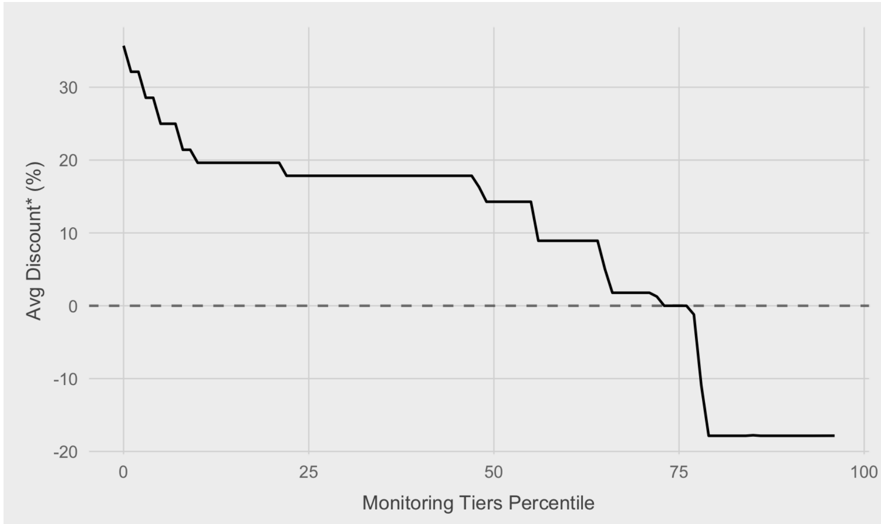


Figure C.4: *Monitoring Discount Schedule*

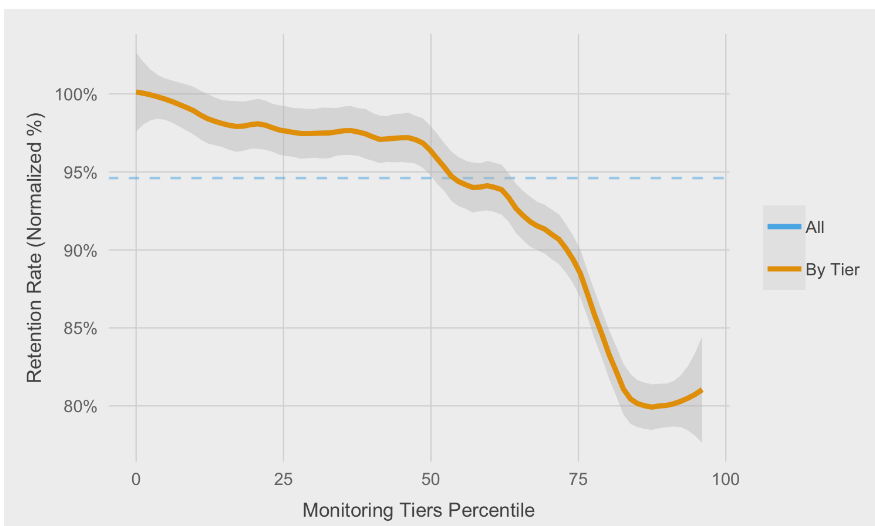
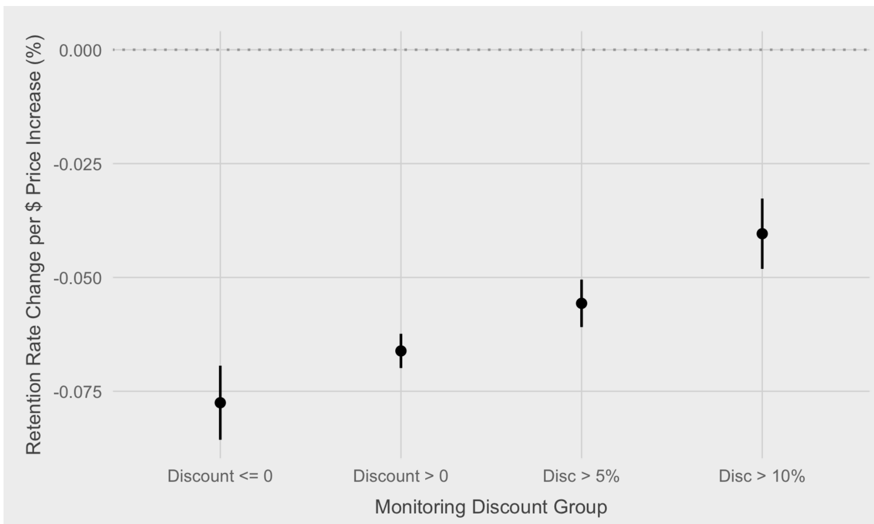


Figure C.5: *Indexed Retention Rate*

Note: C.4 plots the firm's pricing schedule for giving monitoring discount. On the horizontal axis, we plot the percentile of monitoring tier, which is monitoring score divided by that expected by the firm given observables. 74% of people received a discount. The vertical axis is scaled by a factor between 0.5 and 1.5. This is to protect the firm's identity while demonstrating the scale and shape of the pricing algorithm. The firm went through two pricing schedules. This graph plots the second pricing schedule. The first one is similar, except that no surcharge was given.
C.5 uses the same horizontal axis, and non-parametrically plots the retention rate for the semester immediately after drivers finish monitoring (and thus when they first got monitoring discounts). Bandwidth is set as 5, and all numbers are benchmarked/normalized against the mean retention rate of the lowest 5 monitoring tiers. For 93% of monitored drivers, this is the first renewal period.



Note: This figure plots the estimate of θ from (**??**) in various subsamples. These subsamples are represented on the horizontal axis. Notice that although we segment the data using discount percentage, we use the actual discount amount in the regression to measure demand elasticity. The results are scaled to percentage point terms. Therefore, -0.05 means that the slope of retention demand is such that a one dollar increase in price would lead to a 0.05 percentage point drop in retention rate.

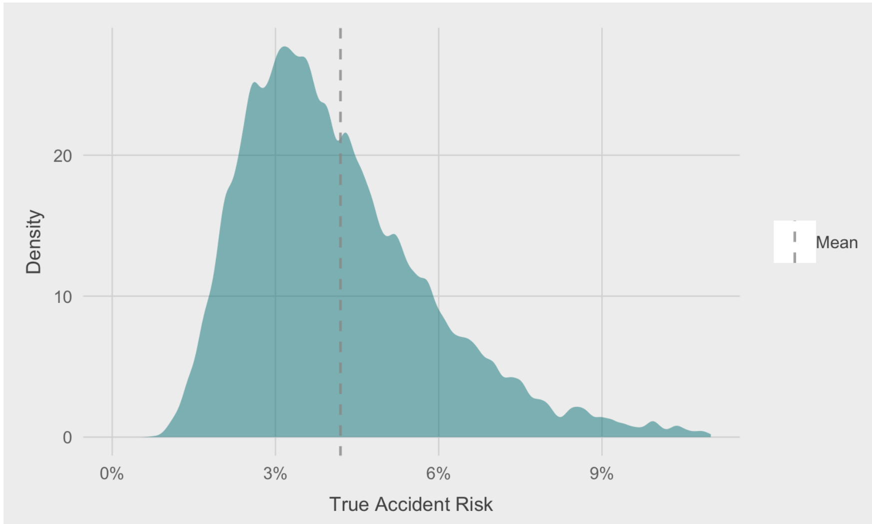
Figure C.6: *Comparison of subsequent claim cost across monitoring groups*

C.0.3 Simulation Analysis of the Informativeness of Monitoring Signal

We can conduct a simple simulation exercise to quantify the spread of private risk and monitoring's effectiveness. To do so, we first simulate a large risk pool by taking the mean of all observable characteristics and simulating each driver's private risk. Figure C.7 plots the density of simulated true risk.[5] Next, Figure C.8 plots the firm's prior mean for all drivers in the risk pool. The firm has a flat prior for all drivers in the first period, which is far from the perfect belief (represented by the dotted and zoomed in 45-degree line). In Figure C.9, we calculate the evolution of firm belief (posterior mean) in subsequent periods as the firm observes potential claim realization. The firm's belief evolves towards the truth as claim is a direct measure of risk. However, the sparsity of claims, especially among safe drivers, dramatically slows down the firm's belief updating.

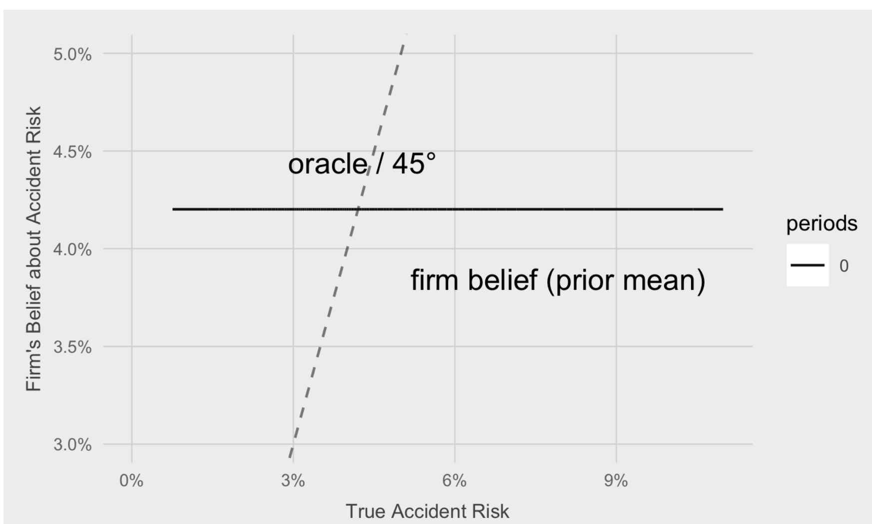
Monitoring score provides an immediate signal for driver risk after the first period. In Figure C.10, we plot, in orange, how the firm's belief updates after observing a one-time monitoring score. It is clear that monitoring is far more informative than observing a period of potential claim realization (dark grey line). Monitoring is especially useful in distinguishing the large mass of safe drivers, in which claims are even rarer. To quantify this measure, we can calculate the absolute deviation of firm belief from the true risk in our simulated risk pool. Overall, observing the monitoring score gets the firm 12.3% closer to the perfect belief (45-degree line).

[5] Our figures use private risk spread among new drivers for illustrative clarity.



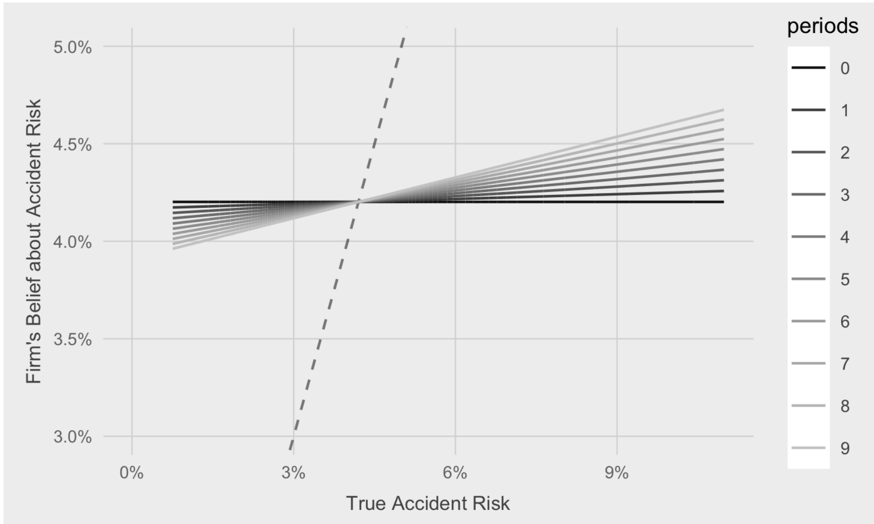
Note: This figure plots the distribution of a simulated mean risk pool given our cost estimates.

Figure C.7: *A simulated mean risk pool given our cost estimate*



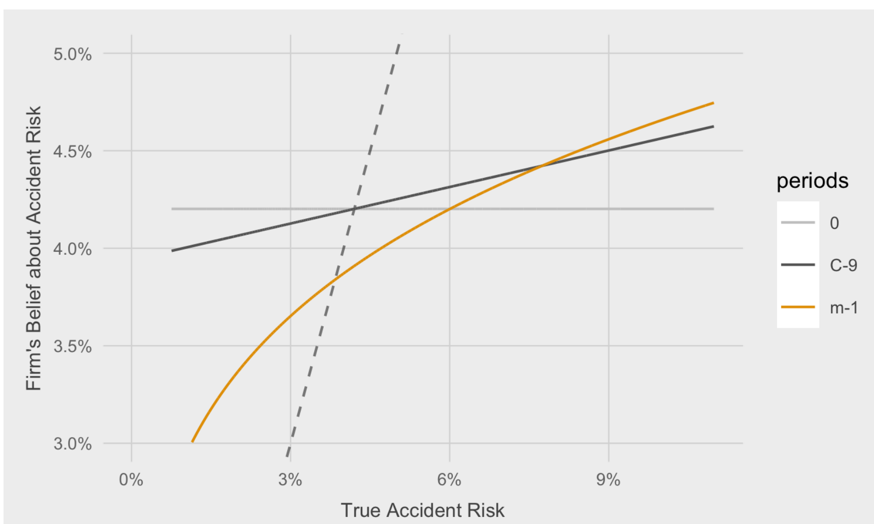
Note: This figure plots firm's belief (prior mean / risk rating) for drivers in our simulated pool. In the first period, they are by definition pooled together. Therefore, firm has a flat prior for all drivers in the pool. The dotted line is the 45 degree line, which represents perfect belief.

Figure C.8: *Firm's prior on simulated risk pool*



Note: This figure plots the evolution of firm belief (posterior mean) for drivers in our simulated pool based on liability claims alone. To make the updating analytically feasible, we first fit a gamma distribution on our risk pool by matching the mean and variance. Since gamma distribution is a conjugate prior for poisson updating, we are able to analytically derive the posterior mean.

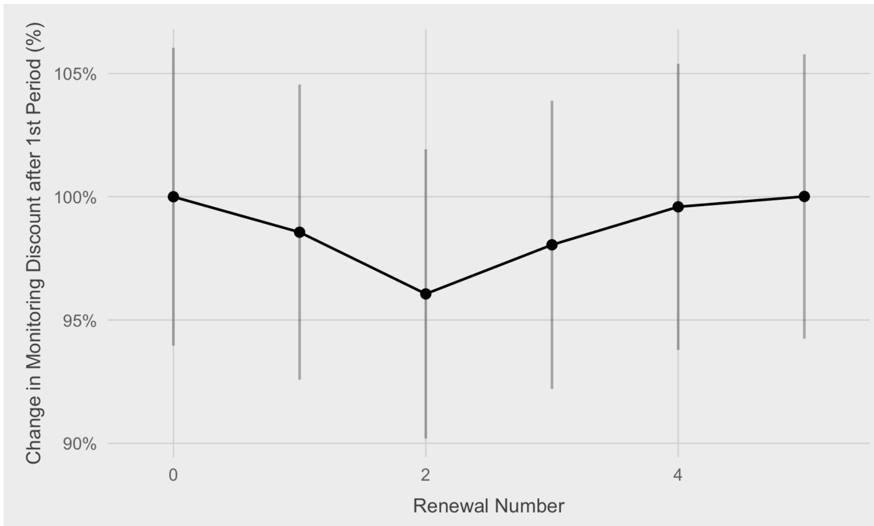
Figure C.9: *Firm's posterior updating based on claims*



Note: This figure plots the evolution of firm belief (posterior mean) for drivers in our simulated pool based on claims versus monitoring. Since lognormal distribution is a conjugate prior for lognormal updating, we are able to analytically derive the posterior mean.

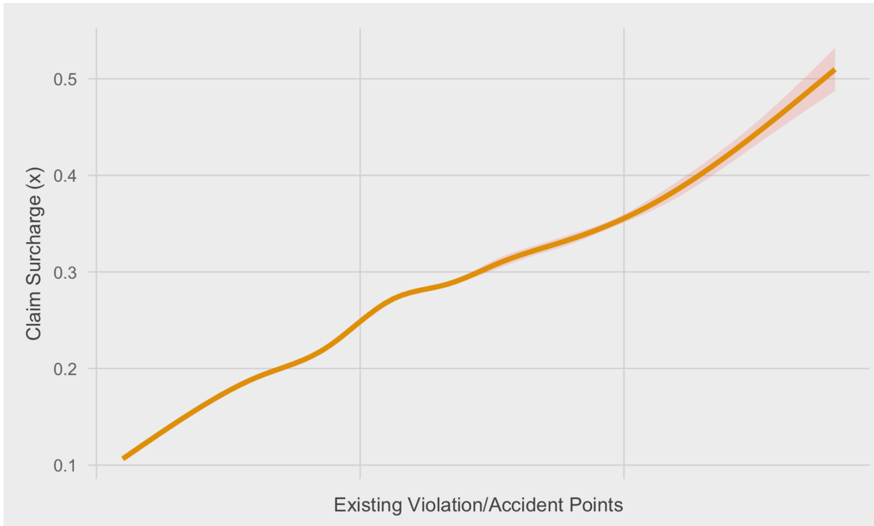
Figure C.10: *Firm's posterior updating based on monitoring vs. claims*

C.0.4 Additional Data Descriptive Graphs



Note: This graph plots the empirical progression of monitoring discount for all monitoring finishers in one state that stayed with the firm till at least the end of the 5th periods (so we observe monitoring discount in the renewal quote for the 6th period). The benchmark is monitoring discount in the first renewal quote ($t = 0$). Fluctuations and noises are due to ex-post adjustments. Firm may change their discount schedule slightly. Monitored drivers can also report mistakes in their records and have their discount adjusted.

Figure C.11: *Persistence of Monitoring Discount*



Note: This graph plots the empirical claim surcharge function for at-fault accidents. Claim surcharge varies with existing violation points and calendar time. 0.1 means 10% surcharge. This differs from the filed factors because the latter is applied on the base rate only, while this function represents the surcharge percentage on top of overall premium. This is done by regressing renewal price change on violation point last period and current period at-fault claim, controlling for all other observables.

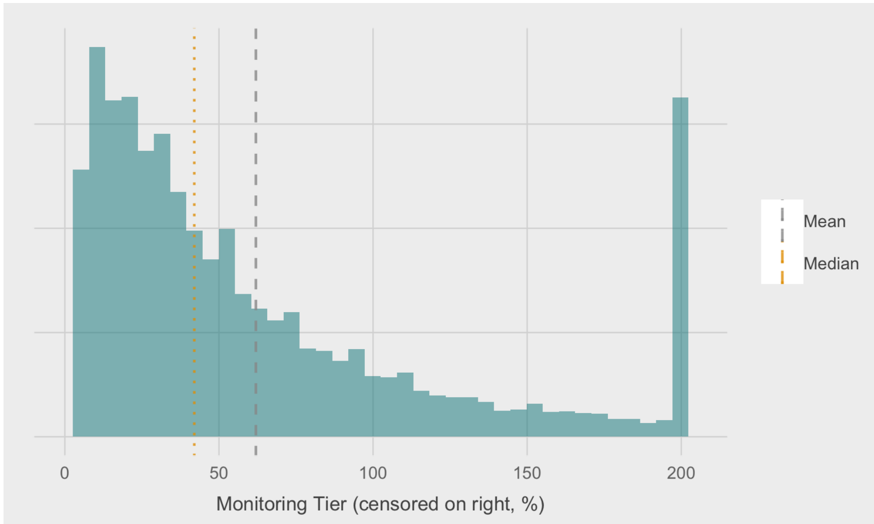
Figure C.12: *Renewal Price Claim Surcharge*

C.0.5 Additional Robustness Checks

	dependent variable: claim count (C)					
explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
constant	0.046***	0.003	0.004	0.046***	0.003	0.004
	(0.000)	(0.005)	(0.005)	(0.000)	(0.005)	(0.005)
post monitoring indicator	−0.001**	−0.003***	−0.003***	−0.001**	−0.003***	−0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monitoring start indicator (m_{start})	−0.009***	−0.008***	−0.008***	0.025***	0.020***	0.020***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
monitoring intensity (M)				−0.050***	−0.042***	−0.042***
				(0.004)	(0.004)	(0.004)
interaction ($\mathbf{1}_{post} \times m$)	0.006***	0.006***	0.006***	−0.014***	−0.014***	−0.014***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
interaction ($\mathbf{1}_{post} \times M$)				0.028***	0.029***	0.029***
				(0.005)	(0.005)	(0.005)
observables controls (x)	N	Y	Y	N	Y	Y
coverage fixed effects	N	N	Y	N	N	Y
implied risk reduction of a full period of monitoring (%)	28.0	29.4	29.5	27.5	29.4	29.6
pre- / post-periods for "first difference"	0 / 1-2					
treatment / control groups "second difference"	all starters / unmonitored					
number of drivers in balanced panel	812,924					

Note: This table reports results of (**??**), but instead look at all monitored drivers regardless of whether they finish or not. Again, the estimate on the interaction term ($\mathbf{1}_{post} \times m_{start}$ or z) measures the treatment effect of monitoring ending on claim count. We first balance our panel data to include all drivers who stay till the end of the third semester ($t = 3$). This gives us two renewal semesters ($t \in \{1, 2\}$) after the monitoring semester ($t = 0$). We control for a full set of observables, including driver and vehicle characteristics and tiers (past records of violations or claims). It also includes third-order polynomials of calendar year and month. Continuous observable characteristics are normalized. We report estimates with and without these controls.

Table C.4: *Estimates from Moral Hazard Regression*



Note: This figure plots the empirical density of monitoring tier for all monitored drivers who finished monitoring. It is calculated as the quotient of realized monitoring score over ex-ante expected monitoring score. For monitored driver i, the expected score is derived based on the average driver in i's observable (x_i) group. It does not take into account the fact that i has selected into monitoring. The graph has a long right tail and is truncated at 200%.

Figure C.13: *Distribution of monitoring tier*

C.0.6 Estimation and Simulation Details

Our model includes unobserved state variables (random coefficients) that enter utility non-linearly. Therefore, we use a random coefficient simulated maximum likelihood approach Train (2009); Handel (2013) to estimate the model.

For each parameter proposal θ, we simulate the model 50 times using Halton draws and compute the likelihood for all observations in the data. We then average over these to get the "simulated log likelihood", denoted as $\hat{\mathcal{L}}_{sim}(\theta)$. The estimator $\theta^{\star}$ maximizes the log likelihood. Simulated maximum likelihood suffer from simulation bias

Likelihood Function The log likelihood are sample analogs of four types of data likelihoods (denoted as $\mathcal{L}$) - claims, monitoring score, choices (of firm, coverage and monitoring participation), as well as renewal price. Utilities are history-dependent in our model. Therefore, we need to simulate choice sequence for each driver i. For notational simplicity, we suppress firm-dummy random effect ζ as in our baseline specification. The log likelihood function can then be expressed as follows.

$$\mathcal{L}_i \equiv \sum_{t \leq T_i} \int_\lambda \mathcal{L}(\underbrace{R_{it}, s_i, C_{it}, d_{it} | \lambda, \psi, x_{it}, \mathbf{p}_{it}, D_{it}, d_{i,t-1}; \mathbf{\Theta}}_{(A)\text{:obs. stoc outcome}}) \cdot \underbrace{g_\lambda(\lambda | x_{it}; \theta_\lambda, \sigma_\lambda)}_{(B)\text{:latent var.}} \, d\lambda$$

The simulation procedure allows us to numerically integrate over λ given parameter proposals θ_λ and σ_λ. We follow the timing of the model to decompose the likelihood component A as follows.

$$\begin{aligned}(A) = &\ln \Pr(d_{it} | \lambda, \mathbf{x}_{it}, \mathbf{p}_{it}, D_{it}, d_{i,t-1}; a, \psi_0, \psi_1, \grave{}_{\mathbf{J}}, \grave{}_{¸}, \alpha, \theta_\beta) + \\ &+ \ln \Pr(C_{it} | \lambda, \mathbf{x}_{it}) + \ln g(\ell_{it} | d_{it}, \mathbf{x}_{it}; \alpha, \grave{}_{\mathbf{fi}}) \\ &+ \ln g_s(s_i | \lambda, \mathbf{x}_{it}; \theta_s, \sigma_s) + \ln g_R(R_{idt} | C_{it}, s_i, \lambda, \mathbf{x}_{it}, \mathbf{p}_{it}; \grave{}_{\mathbf{R}}, \grave{}_{\mathbf{R,m}}, \sigma_R)\end{aligned}$$

Each component of (A) is modeled in the main text and given distributional assumptions.

Choice probability Our choice probability requires integration over all possible C, ℓ, R_0 and s. In our model, we assume away uncertainty in s, and our Poisson-Gamma model gives analytical solutions for expectation over C and ℓ.

For simplicity, in people's expectation, we only consider the possibility of one claim occurrence per term (Cohen and Einav 2007; Barseghyan et al. 2013). We can then capitalize on the attractive analytical property of gamma distributions and avoid numerical integration over C, ℓ, R_0 and s.

C.0.7 Filings

OHIO
VOLUNTARY PRIVATE PASSENGER AUTO
PREMIUM CALCULATION

ROUND AFTER EACH CALCULATION TO THE NEAREST PENNY

STEP #		AA	BB	CC	DD	HH	DNC*^	HNC*^
1	TERRITORIAL BASE RATE (RP-1BR)							
2	RATE ADJUSTMENT FACTOR (PENNY ROUND)	x 1.598	x 1.594	x 1.410	x 1.121	x 1.111	x 1.121	x 1.111
3	INCREASED LIMIT FACTOR/ADDEND (RP-3A)	x	+	x				
4	POLICY GROUP FACTOR (RP-4A-1 through RP-4A-2)	x	x	x	x	x	x	x
5	RATING TIER FACTOR (RP-5A)	x	x	x	x	x	x	x
6	ALLSTATE® YOUR CHOICE AUTO INSURANCE OPTION PACKAGE FACTOR (RP-15A)	x	x	x	x	x	x	x
7	POLICY CLASS FACTOR (RP-7A-1 through RP-7A-4)	x	x	x	x	x	x	x
8	HOUSEHOLD COMPOSITION FACTOR (RP-8A-1 and RP-8A-2)	x	x	x	x	x	x	x
9	SMART STUDENT DISCOUNT FACTOR (RP 10A and RP-11A)	x	x	x	x	x	x	x
10	DEFENSIVE DRIVER DISCOUNT FACTOR (RP-10A and RP-12A)	x	x	x	x	x	x	x
11	MULTIPLE POLICY DISCOUNT FACTOR (RP-15A)	x	x	x	x	x	x	x
12	HOMEOWNER DISCOUNT FACTOR (RP-15A)	x	x	x	x	x	x	x
13	THE GOOD HANDS PEOPLE® DISCOUNT FACTOR (RP-15A)	x	x	x	x	x	x	x
14	RESPONSIBLE PAYER DISCOUNT FACTOR (RP-15A)	x	x	x	x	x	x	x
15	FULLPAY DISCOUNT (RP-15A)	x	x	x	x	x	x	x
16	ALLSTATE EASY PAY PLAN DISCOUNT (RP-15A)	x	x	x	x	x	x	x
17	EARLY SIGNING DISCOUNT (RP-15A)	x	x	x	x	x	x	x
18	ALLSTATE AUTO/LIFE DISCOUNTSM (RP-15A)	x	x	x	x	x	x	x
19	ALLSTATE eSMARTSM DISCOUNT (RP-15A)	x	x	x	x	x	x	x
20	SAFE DRIVING CLUB (RP-10A and RP-13A through RP-14A)	x	x	x	x	x	x	x
21	PRIOR NON-STANDARD CARRIER SURCHARGE (RP-16A)	x	x	x	x	x	x	x
22	ACCIDENT SURCHARGE FACTOR (RP-17A)	x	x	x	x		x	
23	MAJOR VIOLATION SURCHARGE FACTOR (RP-18A)	x	x		x		x	
24	MINOR VIOLATION SURCHARGE FACTOR (RP-19A)	x	x		x		x	
25	MODEL YEAR FACTOR (RP-20A)				x	x	x	x
26	DEDUCTIBLE BY PGS FACTOR (RP-20A)				x	x	x	x
27	EXPERIENCE GROUP RATING FACTOR (EGR PAGES and RP-21A-24A)	x	x	x	x	x	x	x
28	ALLSTATE DRIVE WISE® ENROLLMENT DISCOUNT (RP-26A)	x	x	x	x	x	x	x
29	ALLSTATE DRIVE WISE® PERFORMANCE RATING (RP-26A)	x	x	x	x	x	x	x
30	ANNUAL VEHICLE MILEAGE FACTOR (RP-16A)	x	x	x	x	x	x	x
31	VEHICLE USAGE FACTOR (RP-16A)	x	x	x	x	x	x	x
32	FARM DISCOUNT FACTOR (RP-16A)	x	x		x		x	
33	ELECTRONIC STABILITY CONTROL DISCOUNT FACTOR (RP-16A)	x	x	x	x		x	
34	PASSIVE RESTRAINT DISCOUNT (RP-16A)			x				
35	ANTILOCK BRAKE DISCOUNT (RP-16A)	x	x		x		x	
36	NEW CAR DISCOUNT FACTOR (RP-16A)	x	x	x	x	x	x	x
37	CERTIFIED RISK SURCHARGE FACTOR (RP-16A)	x	x					
38	CAMPER UNIT ADDITIONAL PREMIUM (RP-25A)				+	+		
39	NEW CAR EXPANDED PROTECTION FACTOR (RP-25A)						x	x
40	RATE TRANSITION FACTOR (Rule 72)	x	x	x	x	x		
41	COMPLEMENTARY GROUP RATING (CGR) FACTOR (RP-9A-1 through RP-9A-13)	x	x	x	x	x	x	x
42	FIXED EXPENSE PREMIUM ^^ (RP-16A)	+						
43	SUB-TOTAL VEHICLE PREMIUM	=	=	=	=	=	=	=

	RENTAL REIMBURSEMENT (UU)	
	RENTAL REIMBURSEMENT BASE RATE (RP-52BR)	
	RENTAL REIMBURSEMENT INCREASED LIMIT FACTOR (RP-3A)	x
44	TOTAL RENTAL REIMBURSEMENT COVERAGE PREMIUM	=

	TOWING & LABOR COSTS (JJ) (RP-25A)	
	SOUND SYSTEMS (ZA) (RP-25A)	+
	TAPE (ZZ) (RP-25A)	+
45	TOTAL MISCELLANEOUS COVERAGES	=

	PER AUTO UM/UIM - PROPERTY DAMAGE COVERAGE (SSP)	
46	UM - PROPERTY DAMAGE PREMIUM RATE (RP-3A)	

	POLICY UM/UIM - BODILY INJURY COVERAGE (SS)	
	TERRITORIAL BASE RATE (RP-1BR)	
	RATE ADJUSTMENT FACTOR (PENNY ROUND)	x 0.872
	INCREASED LIMIT FACTOR/ADDEND (RP-3A)	x
	POLICY GROUP FACTOR (RP-4A-1 through RP-4A-2)	x
	RATING TIER FACTOR (RP-5A)	x
	POLICY CLASS FACTOR (RP-7A-1 through RP-7A-4)	x
	HOUSEHOLD COMPOSITION FACTOR (RP-8A-1 through RP-8A-2)	x
	SMART STUDENT DISCOUNT FACTOR (RP 10A and RP-11A)	x
	DEFENSIVE DRIVER DISCOUNT FACTOR (RP-10A and RP-12A)	x
	HOMEOWNER DISCOUNT FACTOR (RP-15A)	x
	RESPONSIBLE PAYER DISCOUNT FACTOR (RP-15A)	x
	FULLPAY DISCOUNT (RP-15A)	x
	SAFE DRIVING CLUB (RP-10A and RP-13A through RP-14A)	x
	ACCIDENT SURCHARGE FACTOR (RP-17A)	x
	MAJOR VIOLATION SURCHARGE FACTOR (RP-18A)	x
	MINOR VIOLATION SURCHARGE FACTOR (RP-19A)	x
	RATE TRANSITION FACTOR (Rule 72)	x
	COMPLEMENTARY GROUP RATING (CGR) FACTOR (RP-9A-1 through RP-9A-13)	x
47	TOTAL UM/UIM - BODILY INJURY COVERAGE	=

48	**TOTAL SEMI-ANNUAL VEHICLE 1 PREMIUM = 43 + 44 + 45 + 46 +47**	+
49	**TOTAL SEMI-ANNUAL VEHICLE 2 PREMIUM = 43 + 44 + 45 + 46 +47**	+
50	**TOTAL SEMI-ANNUAL VEHICLE 3 PREMIUM = 43 + 44 + 45 + 46 +47**	+
51	**TOTAL SEMI-ANNUAL VEHICLE 4 PREMIUM = 43 + 44 + 45 + 46 +47**	+
52	**TOTAL SEMI-ANNUAL POLICY PREMIUM = 48 + 49 + 50 +51**	=

A ALLSTATE FIRE AND CASUALTY INSURANCE COMPANY 04-

Note: This page is taken from an insurer's Ohio rate filing, which demonstrates their pricing algorithm.

Figure C.14: *Pricing Algorithm - Insurer 1 OH*

Progressive Direct Insurance Company (DI)
Progressive Specialty Insurance Company (AG)
Ohio Private Passenger Automobile Program
NB Effective Date: 9/05/2014

Exhibit: 11

RATE ORDER OF CALCULATION

The first step of the rate calculation formula is to determine the Household Risk Factor. The Household Risk Factor is the average of the Developed Driver Risk Factors for all eligible to be rated drivers up to the number of vehicles (or at least one in the case of a named operator policy). For policies where there are more drivers than vehicles, the Household Risk Factor is the average of the highest ranked drivers, up to the number of vehicles. The rank is determined by the Developed Driver Risk Factor for BI (higher factor = higher rank). The Developed Driver Risk Factor is determined as follows:

Driver Risk Factor Items	BI	PD	COMP	COLL	LOAN	MED	RENT	ROADSIDE	UMPD
Driver Classification Factor									
Years Licensed Factor	x	x	x	x	x	x	x	x	x
Driving Record Points Factor	+	+	+	+	+	+	+	+	+
Violation Leniency Factor [1]	-	-	-	-	-	-	-	-	-
Subtraction of One	-1	-1	-1	-1	-1	-1	-1	-1	-1
(1 - Distant Student Discount)	x	x	x	x	x	x	x	x	x
(1 - Minor Child Discount)	x	x	x	x	x	x	x	x	x
(1 - Good Student Discount)	x	x	x	x	x	x	x	x	x
(1 - Senior Citizen Discount)	x	x	x	x		x			
Household Member Factor	x	x	x	x	x	x	x	x	x
Driver Age Point Factor	x	x	x	x	x	x	x	x	x
Financial Responsibility by Clean Factor	x	x	x	x	x	x	x	x	x
Developed Driver Risk Factor									

The second step of the rate calculation formula uses the Household Risk Factor and follows

	BI	PD	COMP	COLL	LOAN	MED	RENT	ROADSIDE	UMPD
Household Risk Factor									
Base Rate	x	x	x	x	x	x	x	x	x
Financial Responsibility Factor	x	x	x	x	x	x	x	x	x
Financial Responsibility by Number of Drivers Factor	x	x	x	x	x	x			x
Deductible Savings Bank Factor	x	x	x	x		x			
Occupation/Education Rating Factor	x	x	x	x	x	x			x
Full Coverage Factor	x	x				x			
Household Structure Factor	x	x	x	x	x	x			x
Residency Rewards Factor	x	x	x	x	x	x	x		x
Luxury Vehicle Factor	x	x	x	x	x	x			x
Tier Factor	x	x	x	x	x	x	x	x	x
Policy Term Factor	x	x	x	x	x	x	x	x	x
Vehicle Age Factor [3]	x	x	x	x	x	x	x	x	x
Excess Vehicle Factor	x	x	x	x	x	x			x
Limit Factor	x	x			x	x	x	x	x
Deductible Factor			x	x					
Vehicle Age by Deductible Factor			x	x					
Vehicle Symbol Factor	x	x	x	x	x	x			x
Value Class Factor (for Vehicle symbols 67 & 68)			x	x	x				x
Vehicle Garaging Location Factor	x	x	x	x	x	x	x	x	x
(1 - Homeowner/Mobile Home/Multi-car Discount)	x	x	x	x		x	x	x	x
(1 - Advance Quote [1]/Three-year Safe Driving/Five-year Accident Free Discount)	x	x	x	x		x			x
(1 - Three-year Safe Driving Bonus) [1]	x	x	x	x		x			x
(1 - Agent Discount) [1]	x	x	x	x		x			x
(1 - Electronic Funds Transfer Discount)	x	x	x	x		x			x
(1 - Paid In Full Discount)	x	x	x	x		x			x
(1 - Online Quote Discount) [2]	x	x	x	x		x			x
(1 - Loyal Customer Discount) [2]	x	x	x	x		x			x
(1 - Paperless Discount)	x	x	x	x	x	x	x	x	x
(1 - Continuous Insurance Discount)	x	x	x	x	x	x	x	x	x
(1 - Multi-policy Discount)	x	x	x	x		x	x	x	x
(1 + Business Use Surcharge)	x	x	x	x		x			x
(1 + Financial Responsibility Filing Surcharge)	x	x	x	x		x			x
Bad Debt Factor	x	x	x	x		x			x
Apply Rate Capping Rule P23 [4]	x	x	x	x	x	x	x	x	x
Usage-based Insurance Factor	x	x	x	x	x	x	x	x	x
(1 - E-signature discount) [2]	x	x	x	x	x	x	x	x	x
Round to the Whole Dollar									
Operations Expense [5]	+		+						
Acquisition Expense [2,6]	+		+						
Developed Premium [8]									

[1] Applies to Progressive Specialty Insurance Company (AG) Only
[2] Applies to Progressive Direct Insurance Company Only (DI)
[3] If coverage is BI, PD, UM/UIM, MED, RENT, or ROADSIDE and Vehicle Symbol = 66, then Vehicle Age Factor = 1.0.
If coverage is COMP, COLL, LOAN, or UMPD and Vehicle Symbol = 66, 67, 68, or 69, then Vehicle Age Factor = 1.0.
[4] Policy level rate changes are capped at +/- 10% as described in Rule P23. The Snapshop Usage Based Insurance Program (UBI) is not taken into consideration when applying the Rate Capping Rule
[5] Operations expense is added to BI if BI is selected; if BI is not selected, then Operations Expense is added to COMP.
[6] Acquisition expense is added to BI if BI is selected; if BI is not selected, then Acquisition Expense is added to COMP.
[7] Average factors are determined by taking the average of Location, Symbol, Vehicle Age factors, and Business Use Surchage for each vehicle, respectively
[8] There is a minimum premium of $5 for each coverage selected for each vehicle.
[9] The trailer coverages will receive the factors associated with COMP and COLL, unless otherwise noted.

NOTES
x means factor is to be used multiplicatively
/ means factor is to be used as a divisor
+ means factor is to be added
- means factor or amount is to be subtracted

Note: These pages are taken from a an insurer's rate filing in Ohio, which demonstrate their pricing algorithm.

Figure C.15: *Pricing Algorithm - Insurer 2 OH 1/2*

Progressive Direct Insurance Company (DI)
Progressive Specialty Insurance Company (AG)
Ohio Private Passenger Automobile Program
NB Effective Date: 9/05/2014

Exhibit: 11

	UM/UIM
Base Rate	
Financial Responsibility Factor	x
Financial Responsibility by Number of Drivers Factor	x
Deductible Savings Bank Factor	x
Occupation/Education Rating Factor	x
Full Coverage Factor	x
Household Structure Factor	x
Residency Rewards Factor	x
Driver Count Factor	x
Luxury Vehicle Factor	x
Tier Factor	x
Policy Term Factor	x
Avg. Vehicle Age Factor [3, 7]	x
Excess Vehicle Factor	x
Limit Factor	x
Avg. Vehicle Symbol Factor [7]	x
Avg. Vehicle Garaging Location Factor [7]	x
(1 - Homeowner/Mobile Home/Multi-car Discount)	x
(1 - Advance Quote [1]/Three-year Safe Driving/Five-year Accident Free Discount)	x
(1 - Three-year Safe Driving Bonus) [1]	x
(1 - Agent Discount) [1]	x
(1 - Electronic Funds Transfer Discount)	x
(1 - Paid In Full Discount)	x
(1 - Online Quote Discount) [2]	x
(1 - Loyal Customer Discount) [2]	x
(1 - Paperless Discount)	x
(1 - Continuous Insurance Discount)	x
(1 - Multi-policy Discount)	x
(1 + Avg. Business Use Surcharge [7])	x
(1 + Financial Responsibility Filing Surcharge)	x
Bad Debt Factor	x
Apply Rate Capping Rule P23 [4]	x
(1 - E-signature discount) [2]	x
Round to the Whole Dollar	
Developed Premium [8]	

	ACPE	COMP-TRLR [1,9]	COLL-TRLR [1,9]	CONTENTS [1]	OPERATIONS EXPENSE [5]	ACQUISITION EXPENSE [2,6]
Base Rate				0.015 * Value		
Financial Responsibility Factor		x	x	x		
Deductible Savings Bank Factor		x	x			
Residency Rewards Factor		x	x			
Tier Factor		x	x	x		
Policy Term Factor	x	x	x	x	x	x
Limit Factor	x					
Deductible Factor		x	x			
Vehicle Symbol Factor		x	x			
Value Class Trailer Factor [1]		x	x			
Vehicle Garaging Location Factor	x	x	x			
(1 - Paperless Discount)	x	x	x	x	x	
(1 - Continuous Insurance Discount)		x	x			
(1 - Multi-policy Discount)						x
Operations Expense Factor 1					x	
Operations Expense Factor 2					x	
Operations Expense Factor 3					x	
Acquisition Expense Full Coverage Factor [2]						x
Acquisition Expense Homeowner Factor [2]						x
Acquisition Expense Online Quote Factor [2]						x
Acquisition Expense Prior Insurance Factor [2]						x
Acquisition Expense Vehicle Count Factor [2]						x
Number of Vehicles						/
Apply Rate Capping Rule P23 [4]	x	x	x	x	x	x
Bad Debt Factor		x	x			
Usage-based Insurance Factor	x				x	x
(1 - E-signature discount) [2]	x				x	x
Round to the Whole Dollar						
Developed Premium [8]						

Total Policy Premium = Sum of Developed Premiums

[1] Applies to Progressive Specialty Insurance Company (AG) Only
[2] Applies to Progressive Direct Insurance Company Only (DI)
[3] If coverage is BI, PD, UM/UIM, MED, RENT, or ROADSIDE and Vehicle Symbol = 66, then Vehicle Age Factor = 1.0.
If coverage is COMP, COLL, LOAN, or UMPD and Vehicle Symbol = 66, 67, 68, or 69, then Vehicle Age Factor = 1.0.
[4] Policy level rate changes are capped at +/- 10% as described in Rule P23. The Snapshop Usage Based Insurance Program (UBI) is not taken into consideration when applying the Rate Capping Rule
[5] Operations expense is added to BI if BI is selected; if BI is not selected, then Operations Expense is added to COMP.
[6] Acquisition expense is added to BI if BI is selected; if BI is not selected, then Acquisition Expense is added to COMP.
[7] Average factors are determined by taking the average of Location, Symbol, Vehicle Age factors, and Business Use Surchage for each vehicle, respectively
[8] There is a minimum premium of $5 for each coverage selected for each vehicle.
[9] The trailer coverages will receive the factors associated with COMP and COLL, unless otherwise noted.

NOTES
x means factor is to be used multiplicatively
/ means factor is to be used as a divisor
+ means factor is to be added
- means factor or amount is to be subtracted

Note: These pages are taken from an insurer's rate filing in Ohio, which demonstrate their pricing algorithm.

Figure C.16: *Pricing Algorithm - Insurer 2 OH 2/2*

GEICO Casualty Company - Voluntary Private Passenger Automobile Insurance
Ohio Rate Pages Effective: New Business 10/2/2009 Renewals 10/2/2009 Rate Gen 01
Rate Order of Calculation: Private Passenger

Machine Rated, Exception: Licensed/Registered Dune Buggies rated as PPV are Manually Rated

Oper	Step	BI	PD	MED	UM-UND	UMBI	UMPD	COMP	COLL	ERS	RR	MBI
Base Rate												
	Base Rate	X	X	X	X	X	X	X	X	X	X	X
*	Limit Factor	X	X	X	X	X	X			X	X	
*	Deductible Factor							X	X			
*	Term Factor	X	X	X	X	X	X	X	X	X	X	X
*	Upgraded Accident Forgiveness Factor	X	X	X	X	X	X	X	X			
Driver Level Rating Steps- Composite Relativities												
*	Driver Class Factor (Composite Relativity)	X	X	X	X	X	X		X			
	Accident Factor	X	X	X	X	X	X		X			
	* Minor Violation Factor	X	X	X	X	X	X		X			
	* Major Violation Factor	X	X	X	X	X	X		X			
	* Speeding Violation Factor	X	X	X	X	X	X		X			
	* DUI Violation Factor	X	X	X	X	X	X		X			
	* Unverifiable Driving Record Factor	X	X	X	X	X	X		X			
	= Merit Factor	X	X	X	X	X	X		X			
*	Merit Factor (Composite Relativity)	X	X	X	X	X	X		X			
Driver Level Discounts: Composite Relativities												
*	*Good Driver Discount (Composite Relativity)*	X	X	X	X	X	X		X			
*	*Student Away at School Discount (Composite Relativity)*	X	X	X	X	X	X		X			
*	*Driving Experience Discount (Composite Relativity)*	X	X	X	X	X	X		X			
*	*Good Student Discount (Composite Relativity)*	X	X	X	X	X	X		X			
*	*Defensive Driver Discount (Composite Relativity)*	X	X	X	X	X	X		X			
*	*Deployed Driver Discount (Composite Relativity)*	X	X	X	X	X	X		X			
Vehicle Level Rating Steps												
*	Vehicle Type Factor											
*	Annual Mileage/ Vehicle Use Factor	X	X	X	X	X	X	X	X			
*	Vehicle Classification Factor	X	X	X	X	X	X	X	X			
*	Vehicle Cost Factor	X	X	X	X	X	X	X	X			
*	Model Year Factor	X	X	X	X	X	X	X	X			
*	Vehicle Age Factor	X	X	X	X	X	X	X	X	X		
*	MBI Model Year Factor											X
*	MBI Coverage Age											X
Vehicle Level Discounts												
*	*Anti-Theft Discount*							X				
*	*New Vehicle Discount*	X	X	X	X	X	X	X	X			
*	*Extra Vehicle Discount*	X	X	X	X	X	X	X	X			
*	*Anti-Lock Brake Discount*	X	X	X	X	X	X	X	X			
*	*Restraint Discount*			X	X	X						
Policy Level Rating Steps												
*	Household Composite Factor	X	X	X	X	X	X	X	X			
*	Maximum Named Insured Age Factor	X	X	X	X	X	X	X	X			
*	Policy Occurrence Factor	X	X	X	X	X	X	X	X			
*	Risk Tier Factor	X	X	X	X	X	X	X	X	X	X	X
Policy Level Discounts												
*	*Financial Responsibility Discount*	X	X	X	X	X	X	X	X	X	X	X
*	*Seat Belt Discount*			X	X	X						
*	*Multi-Vehicle Discount*	X	X	X	X	X	X	X	X	X	X	X
*	*Continuous Insurance Discount*	X	X	X	X	X	X	X	X			
*	*Military Discount*	X	X	X	X	X	X	X	X			
*	*Multi-Line Discount*	X	X	X	X	X	X	X	X	X	X	X
*	*CDL Discount*											
Policy Level Discounts 2												
*	*Sponsored Marketing Discount*	X	X	X	X	X	X	X	X	X	X	X
*	*Associate Discount*	X	X	X	X	X	X	X	X	X	X	X
*	*E-Banking Discount*	X	X	X	X	X	X	X	X	X	X	X
Expense Constants												
+	Vehicle Expense Load	X	X									
+	Policy Expense Load	X	X									

Note: These pages are taken from an insurer's rate filing in Ohio, which demonstrate their pricing algorithm.

Figure C.17: *Pricing Algorithm - Insurer 3 OH*

OHIO
VOLUNTARY PRIVATE PASSENGER AUTO
POLICY CLASS FACTOR

POLICY CLASS FACTOR CALCULATION

Complete the steps below for all applicable coverages. Round to 4 decimals after each step.

Step 1: Based on number of male and female operators on the policy, obtain a value from Value Table 1a and find the corresponding factor from Factor Table 1a.

Step 2: Based on number of single operators, married operators, operators aged <25, and operators aged 25+ on the policy, obtain a value from Value Table 1b and find the corresponding factor from Factor Table 1b.

Step 3: Based on age and gender, obtain a value from Value Table 1c for **each** operator and calculate the geometric average of the corresponding factors from Factor Table 1c to obtain a factor for step 3.

Step 4: Multiply steps 1, 2, and 3 together to obtain the policy class factor to be applied to all vehicles on the policy.

See **RP-6A** for instructions on how to calculate the geometric average.

VALUE TABLE 1a - GENDER

Continuous Prior Insurance	# of Males	# of Females 0	1	2	3+
Yes	**0**	-	1	2	3
Yes	**1**	10	11	12	13
Yes	**2**	20	21	22	23
Yes	**3+**	30	31	32	33
No	**0**	-	4	5	6
No	**1**	14	15	16	17
No	**2**	24	25	26	27
No	**3+**	34	35	36	37

VALUE TABLE 1b - MARITAL STATUS AND AGE

Continuous Prior Insurance (Y/N)			Marital Status														
		# Married	0	0	0	1	1	1	1	2	2	2	2	3+	3+	3+	3+
		# Single	1	2	3+	0	1	2	3+	0	1	2	3+	0	1	2	3+
	# age >= 25	# age < 25															
Yes	0	1	1			16											
Yes	0	2		3			18			33							
Yes	0	3+			6			21	25		36	40	45	52	62	70	77
Yes	1	0	2			17											
Yes	1	1		4			19			34							
Yes	1	2			7			22			37			53			
Yes	1	3+			8				26			41	46	54	63	71	78
Yes	2	0		5			20			35							
Yes	2	1			9			23			38			55			
Yes	2	2			10				27			42		56	64		
Yes	2	3+			11				28				47	57	65	72	79
Yes	3+	0			12			24	29		39	43	48	58	66	73	80
Yes	3+	1			13				30			44	49	59	67	74	81
Yes	3+	2			14				31				50	60	68	75	82
Yes	3+	3+			15				32				51	61	69	76	83
No	0	1	84			99											
No	0	2		86			101			116							
No	0	3+			89			104	108		119	123	128	135	145	153	160
No	1	0	85			100											
No	1	1		87			102			117							
No	1	2			90			105			120			136			
No	1	3+			91				109			124	129	137	146	154	161
No	2	0		88			103			118							
No	2	1			92			106			121			138			
No	2	2			93				110			125		139	147		
No	2	3+			94				111				130	140	148	155	162
No	3+	0			95			107	112		122	126	131	141	149	156	163
No	3+	1			96				113			127	132	142	150	157	164
No	3+	2			97				114				133	143	151	158	165
No	3+	3+			98				115				134	144	152	159	166

VALUE TABLE 1c - AGE AND GENDER

Age	Male	Female
16 and Under	1	20
17	2	21
18	3	22
19	4	23
20	5	24
21	6	25
22	7	26
23	8	27
24	9	28
25-29	10	29
30-49	11	30
50-54	12	31
55-59	13	32
60-64	14	33
65-70	15	34
71-75	16	35
76-80	17	36
81-84	18	37
85+	19	38

RP-7A-1 ALLSTATE FIRE AND CASUALTY INSURANCE COMPANY 04-01-2014

Note: This is an excerpt from an insurer's rate filing on how observable information is used and interacted.

Figure C.18: *Variable Definition and Interactions*

GEICO Casualty Company - Voluntary Private Passenger Automobile Insurance
Ohio Rate Pages Effective: New Business 06/07/2013 Renewals 07/22/2013 Rate Gen 12
Driver Class Factors

** Risk Group: B = B10, B20, and B30; C = C10, C20, C30; D = D10, D20, D30
** Z in Risk Tier represents all Risk Tiers
** Driver Age 999 = 80 and older
** RV Factor = 1.0

Risk Group	Risk Tier	Rated Vehicle Type	Coverage	Named Insured Indicator	Gender	Marital Status	Driver Age	Factor
B	Z	PP	BI	N	F	S	24	1.1660
B	Z	PP	BI	Y	M	M	24	0.9460
B	Z	PP	BI	N	M	M	24	1.1976
B	Z	PP	BI	Y	M	S	24	0.9361
B	Z	PP	BI	N	M	S	24	1.1387
B	Z	PP	BI	Y	F	M	25	0.7939
B	Z	PP	BI	N	F	M	25	0.8392
B	Z	PP	BI	Y	F	S	25	0.9649
B	Z	PP	BI	N	F	S	25	1.1458
B	Z	PP	BI	Y	M	M	25	0.9460
B	Z	PP	BI	N	M	M	25	1.1633
B	Z	PP	BI	Y	M	S	25	0.9361
B	Z	PP	BI	N	M	S	25	1.1178
B	Z	PP	BI	Y	F	M	26	0.8060
B	Z	PP	BI	N	F	M	26	0.8520
B	Z	PP	BI	Y	F	S	26	0.9649
B	Z	PP	BI	N	F	S	26	1.0819
B	Z	PP	BI	Y	M	M	26	0.9460
B	Z	PP	BI	N	M	M	26	1.1360
B	Z	PP	BI	Y	M	S	26	0.9361
B	Z	PP	BI	N	M	S	26	1.0359
B	Z	PP	BI	Y	F	M	27	0.8060
B	Z	PP	BI	N	F	M	27	0.8520
B	Z	PP	BI	Y	F	S	27	0.9649
B	Z	PP	BI	N	F	S	27	1.0525
B	Z	PP	BI	Y	M	M	27	0.9460
B	Z	PP	BI	N	M	M	27	1.0460
B	Z	PP	BI	Y	M	S	27	0.9361
B	Z	PP	BI	N	M	S	27	1.0251
B	Z	PP	BI	Y	F	M	28	0.8060
B	Z	PP	BI	N	F	M	28	0.8520
B	Z	PP	BI	Y	F	S	28	0.9649
B	Z	PP	BI	N	F	S	28	1.0398
B	Z	PP	BI	Y	M	M	28	0.9460
B	Z	PP	BI	N	M	M	28	1.0260
B	Z	PP	BI	Y	M	S	28	0.9361
B	Z	PP	BI	N	M	S	28	1.0172
B	Z	PP	BI	Y	F	M	29	0.8060
B	Z	PP	BI	N	F	M	29	0.8530
B	Z	PP	BI	Y	F	S	29	0.9649
B	Z	PP	BI	N	F	S	29	1.0118
B	Z	PP	BI	Y	M	M	29	0.9460
B	Z	PP	BI	N	M	M	29	1.0110
B	Z	PP	BI	Y	M	S	29	0.9361
B	Z	PP	BI	N	M	S	29	0.9821
B	Z	PP	BI	Y	F	M	30	0.8060
B	Z	PP	BI	N	F	M	30	0.8440
B	Z	PP	BI	Y	F	S	30	0.9649
B	Z	PP	BI	N	F	S	30	1.0100
B	Z	PP	BI	Y	M	M	30	0.9460
B	Z	PP	BI	N	M	M	30	0.9900
B	Z	PP	BI	Y	M	S	30	0.9361
B	Z	PP	BI	N	M	S	30	0.9800
B	Z	PP	BI	Y	F	M	31	0.8060
B	Z	PP	BI	N	F	M	31	0.8360
B	Z	PP	BI	Y	F	S	31	0.9648
B	Z	PP	BI	N	F	S	31	1.0010
B	Z	PP	BI	Y	M	M	31	0.9415
B	Z	PP	BI	N	M	M	31	0.9760
B	Z	PP	BI	Y	M	S	31	0.9360
B	Z	PP	BI	N	M	S	31	0.9710
B	Z	PP	BI	Y	F	M	32	0.8060
B	Z	PP	BI	N	F	M	32	0.8270
B	Z	PP	BI	Y	F	S	32	0.9648
B	Z	PP	BI	N	F	S	32	0.9900
B	Z	PP	BI	Y	M	M	32	0.9421
B	Z	PP	BI	N	M	M	32	0.9670

Note: This is an excerpt from an insurer's rate filing on how observable information is translated into pricing factors.

Figure C.19: *Rating Factors based on Observables*

Progressive Direct Insurance Company
State of Ohio
New Business Effective: January 23, 2015
Renewals Effective: February 20, 2015

D06-Driving Violation Descriptions

The following chart lists the violation codes and their associated descriptions:

Violation Code	Violation Description
AAF	At Fault Accident
AFM	Accident found on MVR only at renewal - Not Chargeable
ANC	Waived Claim – Closed
ANO	Waived Claim – Open
ASW	Accident Surcharge Waived
CML	Commercial Vehicle Violation
CMP	Comprehensive Claim
CMU	Comprehensive Claim Less Than $1000
CRD	Careless or Improper Operation
DEV	Traffic Device/Sign
DR	Drag Racing
DWI	Drive Under Influence
FDL	Foreign Drivers Lic
FEL	Auto Theft/Felony Motor Vehicle
FFR	Failure to File Required Report
FLE	Fleeing from Police
FTC	Following Too Close
FTY	Failure to Yield
HOM	Vehicular Homicide
IP	Improper Passing
IT	Improper Turn
LDL	Operating Without Owner's Consent
LIC	License/Credentials Violation
LTS	Leaving the Scene
MAJ	Other Serious Violation
MMV	Minor Moving Violation
NAF	Not At Fault Accident
NFX	Waived Not At Fault Accident
PUA	Permissive Use At Fault Accident
PUN	Permissive Use Not At Fault Accident
RKD	Reckless Driving
SLV	Serious License Violations
SPD	Speeding
SUS	Driving Under Suspension
TMP	Dispute - At Fault Accident
UDR	Unverifiable Record
WSR	Wrong Way on a One Way Street

Note: This is an excerpt from an insurer's rate filing on the kinds of violations recorded in tier rating in Ohio.

Figure C.20: *Violation Captured in OH*

GEICO Casualty Company - Voluntary Private Passenger Automobile Insurance
Ohio Rate Pages Effective: New Business 06/07/2013 Renewals 07/22/2013 Rate Gen 12
Accident Factors

** Risk Group: B = B10, B20, and B30; C = C10, C20, C30; D = D10, D20, D30
** Z in Risk Tier represents all Risk Tiers
** For Coverages BI,PD, COLL, COLL PP, and COLL TL Driver Age 18 = 18 and younger; 999 = 80 and older. All other Coverages Driver Age 18 = 18 and you

Risk Group	Risk Tier	Rated Vehicle Type	Coverage	Driver Age	Number of Chargeable Occurrences	Months Since First Occurrence	Months Since Second Occurrence	Factor
B	Z	PP	BI	31	4	23	35	3.3112
B	Z	PP	BI	31	4	35	35	3.0748
B	Z	PP	BI	31	99	11	11	4.9426
B	Z	PP	BI	31	99	11	23	4.5307
B	Z	PP	BI	31	99	11	35	4.3248
B	Z	PP	BI	31	99	23	23	3.9644
B	Z	PP	BI	31	99	23	35	3.7842
B	Z	PP	BI	31	99	35	35	3.5140
B	Z	PP	BI	32	0	0	0	1.0000
B	Z	PP	BI	32	1	11	0	1.6375
B	Z	PP	BI	32	1	23	0	1.3267
B	Z	PP	BI	32	1	35	0	1.2320
B	Z	PP	BI	32	2	11	11	2.2925
B	Z	PP	BI	32	2	11	23	2.1014
B	Z	PP	BI	32	2	11	35	2.0059
B	Z	PP	BI	32	2	23	23	1.6550
B	Z	PP	BI	32	2	23	35	1.5797
B	Z	PP	BI	32	2	35	35	1.4669
B	Z	PP	BI	32	3	11	11	3.5525
B	Z	PP	BI	32	3	11	23	3.2565
B	Z	PP	BI	32	3	11	35	3.1083
B	Z	PP	BI	32	3	23	23	2.8493
B	Z	PP	BI	32	3	23	35	2.7199
B	Z	PP	BI	32	3	35	35	2.5256
B	Z	PP	BI	32	4	11	11	4.3248
B	Z	PP	BI	32	4	11	23	3.9644
B	Z	PP	BI	32	4	11	35	3.7842
B	Z	PP	BI	32	4	23	23	3.4689
B	Z	PP	BI	32	4	23	35	3.3112
B	Z	PP	BI	32	4	35	35	3.0748
B	Z	PP	BI	32	99	11	11	4.9426
B	Z	PP	BI	32	99	11	23	4.5307
B	Z	PP	BI	32	99	11	35	4.3248
B	Z	PP	BI	32	99	23	23	3.9644
B	Z	PP	BI	32	99	23	35	3.7842
B	Z	PP	BI	32	99	35	35	3.5140
B	Z	PP	BI	33	0	0	0	1.0000
B	Z	PP	BI	33	1	11	0	1.6375
B	Z	PP	BI	33	1	23	0	1.3267
B	Z	PP	BI	33	1	35	0	1.2320
B	Z	PP	BI	33	2	11	11	2.2925
B	Z	PP	BI	33	2	11	23	2.1014
B	Z	PP	BI	33	2	11	35	2.0059
B	Z	PP	BI	33	2	23	23	1.6550
B	Z	PP	BI	33	2	23	35	1.5797
B	Z	PP	BI	33	2	35	35	1.4669
B	Z	PP	BI	33	3	11	11	3.5525
B	Z	PP	BI	33	3	11	23	3.2565
B	Z	PP	BI	33	3	11	35	3.1083
B	Z	PP	BI	33	3	23	23	2.8493
B	Z	PP	BI	33	3	23	35	2.7199
B	Z	PP	BI	33	3	35	35	2.5256
B	Z	PP	BI	33	4	11	11	4.3248
B	Z	PP	BI	33	4	11	23	3.9644
B	Z	PP	BI	33	4	11	35	3.7842

Note: This is an excerpt from an insurer's rate filing on how tier information is rated.

Figure C.21: *Tier Factors*

Progressive Direct Insurance Company (DI)
Progressive Specialty Insurance Company (AG)
Ohio Private Passenger Automobile Program
Effective Date: January 23, 2015

Usage-based Insurance Factor Table - Initial Discount (DI Experience) **Exhibit: 9C**

UBI SCORE	BI/PD	COLL	COMP	LOAN	MED	RENT	ROADSIDE	UMPD	ACPE	OPERATIONS EXPENSE	ACQUISITION EXPENSE
0	0.56	0.56	0.96	0.96	0.56	0.56	0.96	0.56	0.96	1.00	1.00
1	0.61	0.61	0.96	0.96	0.61	0.61	0.96	0.61	0.96	1.00	1.00
2	0.65	0.65	0.97	0.97	0.65	0.65	0.97	0.65	0.97	1.00	1.00
3	0.75	0.74	0.97	0.97	0.75	0.74	0.97	0.75	0.97	1.00	1.00
4	0.79	0.79	0.97	0.97	0.79	0.79	0.97	0.79	0.97	1.00	1.00
5	0.83	0.83	0.97	0.97	0.83	0.83	0.97	0.83	0.97	1.00	1.00
6	0.86	0.87	0.97	0.97	0.86	0.87	0.97	0.86	0.97	1.00	1.00
7	0.89	0.89	0.97	0.97	0.89	0.89	0.97	0.89	0.97	1.00	1.00
8	0.89	0.90	0.97	0.97	0.89	0.90	0.97	0.89	0.97	1.00	1.00
9	0.89	0.91	0.97	0.97	0.89	0.91	0.97	0.89	0.97	1.00	1.00
10	0.90	0.90	0.97	0.97	0.90	0.90	0.97	0.90	0.97	1.00	1.00
11	0.90	0.90	0.97	0.97	0.90	0.90	0.97	0.90	0.97	1.00	1.00
12	0.90	0.90	0.98	0.98	0.90	0.90	0.98	0.90	0.98	1.00	1.00
13	0.91	0.89	0.98	0.98	0.91	0.89	0.98	0.91	0.98	1.00	1.00
14	0.91	0.88	0.98	0.98	0.91	0.88	0.98	0.91	0.98	1.00	1.00
15	0.91	0.90	0.98	0.98	0.91	0.90	0.98	0.91	0.98	1.00	1.00
16	0.92	0.90	0.98	0.98	0.92	0.90	0.98	0.92	0.98	1.00	1.00
17	0.92	0.91	0.98	0.98	0.92	0.91	0.98	0.92	0.98	1.00	1.00
18	0.92	0.91	0.98	0.98	0.92	0.91	0.98	0.92	0.98	1.00	1.00
19	0.92	0.92	0.98	0.98	0.92	0.92	0.98	0.92	0.98	1.00	1.00
20	0.92	0.92	0.98	0.98	0.92	0.92	0.98	0.92	0.98	1.00	1.00
21	0.92	0.92	0.98	0.98	0.92	0.92	0.98	0.92	0.98	1.00	1.00
22	0.92	0.92	0.98	0.98	0.92	0.92	0.98	0.92	0.98	1.00	1.00
23	0.92	0.92	0.98	0.98	0.92	0.92	0.98	0.92	0.98	1.00	1.00
24	0.93	0.93	0.98	0.98	0.93	0.93	0.98	0.93	0.98	1.00	1.00
25	0.93	0.93	0.98	0.98	0.93	0.93	0.98	0.93	0.98	1.00	1.00
26	0.93	0.93	0.98	0.98	0.93	0.93	0.98	0.93	0.98	1.00	1.00
27	0.93	0.93	0.99	0.99	0.93	0.93	0.99	0.93	0.99	1.00	1.00
28	0.93	0.94	0.99	0.99	0.93	0.94	0.99	0.93	0.99	1.00	1.00
29	0.93	0.94	0.99	0.99	0.93	0.94	0.99	0.93	0.99	1.00	1.00
30	0.94	0.94	0.99	0.99	0.94	0.94	0.99	0.94	0.99	1.00	1.00
31	0.94	0.94	0.99	0.99	0.94	0.94	0.99	0.94	0.99	1.00	1.00
32	0.94	0.94	0.99	0.99	0.94	0.94	0.99	0.94	0.99	1.00	1.00
33	0.94	0.94	0.99	0.99	0.94	0.94	0.99	0.94	0.99	1.00	1.00
34	0.95	0.95	0.99	0.99	0.95	0.95	0.99	0.95	0.99	1.00	1.00
35	0.95	0.95	0.99	0.99	0.95	0.95	0.99	0.95	0.99	1.00	1.00
36	0.95	0.95	0.99	0.99	0.95	0.95	0.99	0.95	0.99	1.00	1.00
37	0.95	0.95	0.99	0.99	0.95	0.95	0.99	0.95	0.99	1.00	1.00
38	0.95	0.95	0.99	0.99	0.95	0.95	0.99	0.95	0.99	1.00	1.00
39	0.95	0.96	0.99	0.99	0.95	0.96	0.99	0.95	0.99	1.00	1.00

Note:
-The premium-weighted average factor for the vehicle is calculated and applied to all coverages for the vehicle as indicated in the Rate Order of Calculation. This factor cannot be lower than 0.70 or greater than 1.0.

-If a vehicle does not participate in the Usage-based Insurance program it is assigned a 1.0 factor.

Note: This is an excerpt from an insurer's rate filing on how monitoring pricing is filed.

Figure C.22: *Violation Captured in OH*

Progressive Direct Insurance Company (DI) & Progressive Specialty Insurance Company (AG)
Private Passenger Automobile Program
Supporting Exhibits for the State of Ohio
Effective Date: September 5, 2014
Coverage: BI **Exhibit 10Y**

Limit Factor

Experience	Has Prior Insurance	Limit	Incurred Loss Capped	Indicated Factor	Proposed Factor	Current Factor	Percent Change
AG	N	$25,000/$50,000	243,943,611	1.00	1.00	1.00	0.0%
AG	N	$50,000/$100,000	102,950,757	1.16	1.08	1.08	0.0%
AG	N	$100,000 CSL	1,444,950	1.24	1.11	1.11	0.0%
AG	N	$100,000/$300,000	70,326,408	1.54	1.29	1.29	0.0%
AG	N	$300,000 CSL	3,758,408	2.04	1.50	1.50	0.0%
AG	N	$250,000/$500,000	9,874,286	2.15	1.68	1.68	0.0%
AG	N	$500,000 CSL	5,350,267	2.25	1.80	1.80	0.0%
AG	Y	$25,000/$50,000	302,253,249	1.00	1.00	1.00	0.0%
AG	Y	$50,000/$100,000	256,452,902	1.21	1.13	1.12	0.9%
AG	Y	$100,000 CSL	7,102,129	1.26	1.19	1.16	2.6%
AG	Y	$100,000/$300,000	388,729,047	1.53	1.37	1.33	3.0%
AG	Y	$300,000 CSL	25,394,374	1.85	1.45	1.46	-0.7%
AG	Y	$250,000/$500,000	85,216,412	2.10	1.69	1.80	-6.1%
AG	Y	$500,000 CSL	45,591,859	2.15	1.93	1.95	-1.0%
DI	N	$25,000/$50,000	94,310,074	1.00	0.95	0.95	0.0%
DI	N	$50,000/$100,000	71,807,198	1.16	1.00	1.00	0.0%
DI	N	$100,000 CSL	81,354	1.27	1.11	1.11	0.0%
DI	N	$100,000/$300,000	45,810,439	1.54	1.28	1.28	0.0%
DI	N	$300,000 CSL	254,864	1.56	1.41	1.41	0.0%
DI	N	$250,000/$500,000	10,296,001	2.00	1.49	1.49	0.0%
DI	N	$500,000 CSL	440,458	2.16	1.59	1.59	0.0%
DI	Y	$25,000/$50,000	182,880,315	1.00	1.00	1.00	0.0%
DI	Y	$50,000/$100,000	199,882,577	1.15	1.05	1.05	0.0%
DI	Y	$100,000 CSL	1,287,766	1.22	1.17	1.17	0.0%
DI	Y	$100,000/$300,000	286,763,971	1.40	1.33	1.33	0.0%
DI	Y	$300,000 CSL	4,867,338	1.74	1.39	1.39	0.0%
DI	Y	$250,000/$500,000	53,447,656	1.82	1.47	1.47	0.0%
DI	Y	$500,000 CSL	5,998,809	2.13	1.60	1.60	0.0%

Note: This is an excerpt from an insurer's rate filing on how limit choices influence pricing.

Figure C.23: *Tier Factors*