Regulation and Management of Innovative Technologies

By
Siddharth Prakash Singh

Dissertation

Submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in
Operations Management
at Tepper School of Business, Carnegie Mellon University

Pittsburgh, Pennsylvania

Doctoral Committee:

Jay Apt
Soo-Haeng Cho
Mohammad Delasay
Isa Hafalir
Alan Scheller-Wolf (Chair)
Darius Walczak

April 2019

Regulation and Management of Innovative Technologies

Abstract

Advances in technology have implications for the way businesses are regulated and managed. This dissertation examines three issues that have arisen in response to recent technological innovations. The first chapter studies the regulation of electricity markets with rooftop solar, a technology that has recently seen a sharp increase in adoption. Innovative internet-based technologies have given firms and customers the ability to harness more information in their decision-making. The second and third chapters of this dissertation study how this increased information availability affects firm-level decisions, and how additional information can be used as a strategic lever to compensate for capacity shortage.

The first chapter is motivated by the dramatic increase in the adoption of rooftop solar systems, driven in large part by cost reductions caused by technological advances in solar panel design. This has implications for regulators, who seek to induce an optimal level of rooftop solar adoption, trading off between its environmental benefits and the financial burden that it imposes, while simultaneously safeguarding the interests of utility companies, solar system installers, and customers. We formulate and analyze a social welfare maximization problem for the regulator, focusing on how the choice of tariff structure (that governs how customers pay the utility for their usage) interacts with its competing objectives. We uncover the structural properties of a successful tariff, finding that the tariff structures used in most states in the US are inadequate: to achieve welfare-optimal outcomes, a tariff must be able to discriminate among customer usage tiers and between customers with and without rooftop solar. We present a tariff structure with these two characteristics and show how it can be implemented as a simple buy-all, sell-all tariff while retaining its favorable properties. We illustrate our findings numerically using household-level data from Nevada and New Mexico.

The second chapter is motivated by the recent practice of service providers broadcasting real-time delay information to their customers. We consider the question: is announcing real-time delay necessarily a good idea? In a market with two service providers who compete for market share, we investigate whether one of the service providers (the technology leader L) should make the first move to announce her real-time delay information, when her competitor (the follower F) can opt to respond. We model and analyze this leader-follower setting as a sequential game, using continuous Markov chains to analyze the associated queueing dynamics. We find that L's optimal action depends crucially on the relative service capacities of the service providers: initiating delay announcements improves, in equilibrium, the market share of L if she is the lower-capacity service provider and worsens her market share otherwise. Therefore, for a lower-capacity service provider, delay announcements can be considered a strategic remedy for capacity shortage.

The third chapter is motivated by the availability of competitor fare information to firms such as airlines, who dynamically set prices to sell a fixed amount of inventory over a finite horizon. We study the impact of explicitly using competitor pricing information on equilibrium pricing strategies in a setting with two firms. Under a general model of demand that unifies the models used in prior work, we establish the non-existence of pure strategy subgame-perfect Nash equilibria when some strictly positive proportion of the customers is loyal, i.e., when the firms' product offerings are not perfectly substitutable. However, when all customers are flexible, i.e., the firms' product offerings are perfectly substitutable, we show the existence of a pure strategy subgame-perfect Nash equilibrium. Using these results, we study the strategic question of whether firms should explicitly use competitor price information in their pricing exercises, or continue to employ monopolistic models as they currently do. We argue that in general, the unique equilibrium strategy is for both firms to use competitor fare information. We find that typically, similar to Chapter 2, the lower capacity firm is benefited by a migration to competitive models, while the higher capacity firm is left worse-off. However, when total capacity relative to demand is very low and the firms have roughly the same capacity, this migration could improve both firms' profits.

Acknowledgments

It is only fitting that I begin by expressing my profound sense of gratitude to my advisor, Prof. Alan Scheller-Wolf. He has been, and continues to be, a role model for me in the truest sense of the term; he is a shining example of a creative, hard-working, organized and motivated academic who upholds the highest standards of integrity and professionalism, while always being friendly and approachable. Working with him has been a very fulfilling academic experience.

I thank my other co-authors, Prof. Mohammad Delasay and Dr. Darius Walczak, for their able guidance, support and mentorship through my doctoral studies. I have learned a lot about clarity in research, writing and presentation through my fruitful interactions with them. I also thank the other members of my doctoral committee, Prof. Jay Apt, Prof. Soo-Haeng Cho and Prof. Isa Hafalir, for their invaluable comments and guidance through my PhD. I have also benefited tremendously from the support, encouragement and advice of the other Operations Management faculty at Tepper: I am immensely thankful to Prof. Mustafa Akan, Prof. Sunder Kekre, Prof. Nicola Secomandi and Prof. Sridhar Tayur. I had the pleasure of working closely with Prof. Rein Vesilo, from Macquaire University, in the Fall of 2018, on some interesting problems in queueing; I thank him for all his support and guidance. I am also immensely thankful to Prof. Mor Harchol-Balter and the SQUALL group for all the feedback, the great talks, and the tasty lunches! My heartfelt and sincere thanks goes to Laila and Lawrence whose friendly and ready assistance and guidance has helped me navigate my PhD smoothly.

Tepper has provided me a great environment that has fostered my sense of academic and personal well-being. I have had the benefit of the support of some wonderful Tepper alumni during the course of my PhD. Interactions with them have enriched my understanding of academia greatly. Special thanks to OM alumni Sherwin Doroudi, Ersin Korpeoglu, Leela Nageswaran, Vince Slaugh and Xin Wang who have always been very generous with their time and advice. My appreciation also extends to non-OM alumni Stelios Despotakis, Tarek Elgindy, Francisco Cisternas, Alex Kazachkov, Thiago Serra and Christian Tjandraatmadja and Wenting Yu for our OR discussions, puzzle hunts, and squash games. I am also extremely grateful for the various interactions I've had with current Tepper students, in particular, with Mehmet Aydemir, Gerdus Benade, Franco Berbeglia, Neda Mirzaeian and Bo Yang. These associations have enriched my experience greatly.

Carnegie Mellon University also provides a vibrant campus culture, and I have benefited from having great conversations with members of the community outside of Tepper. A special mention is due to the members of the Indian Graduate Student Association, and in particular, to the members of "The Peace Band," who have filled my life with music for the past six years. Out of this group was born a sub-group called "The Club," which has succeeded in providing valuable inter-disciplinary academic feedback that I have greatly benefited from. My heartfelt thanks in particular to Saurabh Kadekodi and Vaishnavh Nagarajan, whose profoundly clear style of thinking and affinity for arguments I have learned much from. It is also through these groups that I met my band-mate, best friend and partner Purvasha Chakravarti who has supported me, taught me, made me laugh, nurtured my sense of cheer and optimism, and generally enriched my academic and personal life in ways that I am strapped for words to express.

I have not, through the course of my PhD, been allowed to feel isolated. This is in large part to the credit of my non-Pittsburgh friends like Hridya Ravimohan, Murthy Arlekatti, Ranjitha Shivaram, Nivi Samudrala and for lack of a better term, my brother-from-another-mother, Siddharth Prabhu. With them, I've created some indelible memories over the past few years.

I am immensely thankful to my sisters Shruti and Sheetal for their love, support and helpful conversations and to my brother-in-law Sridhar for his readiness to discuss literally anything. From the bottom of my heart, I am abundantly thankful to my parents, Viji and Mike, for their unwavering support, guidance and encouragement throughout my life. They continue to be the greatest source of inspiration for me, and motivate me, by example, to always work hard and stay honest. I am what I am because of them.

To my parents, Viji and Mike

Contents

List of Co-Authors

CHAPTER 1: ALAN SCHELLER-WOLF[1]
CHAPTER 2: MOHAMMAD DELASAY[2], ALAN SCHELLER-WOLF[1]
CHAPTER 3: DARIUS WALCZAK[3]

[1]Tepper School of Business, Carnegie Mellon University
[2]College of Business, Stony Brook University
[3]PROS Inc.

List of Figures

List of Tables

0
Introduction

Rapid technological innovation has spurred the growth of many businesses across the globe. These innovative technologies often drive fundamental changes in firm-level decision making for companies that use these technologies, and more broadly, in policy-level decision making for regulatory bodies. This dissertation examines three problems that have arisen in response to recent advances in technology.

Rooftop solar adoption has increased sharply in recent years, with residential rooftop solar installations consistently growing at over 20% year-on-year [Mike Munsell, 2017]. Under a widely prevalent scheme (in the United States and other parts of the world) called *net metering*, owners of rooftop solar panels, by regulation, can sell any excess electricity their panels generate to their utility company (the grid) for *full retail credit*. Effectively, such a customer pays only for her "net" usage. While this scheme has served as a strong incentive for rooftop solar adoption, increased rooftop solar penetration levels are eroding utility profitability. In response, utility regulators are grappling to devise alternative compensation schemes for customers who sell rooftop solar generation back to the grid; they seek to support renewable energy adoption while safeguarding the interests of utilities, solar system installers, and both solar and non-solar customers. This is a difficult balance: Tariff changes in Nevada introduced to protect NV Energy (a Nevada utility) induced SolarCity, the market leader in solar systems, to suspend operations in Nevada. Apparently, regulators failed to predict how the new tariff would influence the market players and ultimately their social objectives. Regulators in other states have also been responding to this issue by altering utility tariffs in various ways, most commonly by changing the rate at which excess generation is compensated. But it is unclear whether simply changing the compensation rate is sufficient to protect the welfare of all parties.

In Chapter 1, we take the perspective of a socially interested utility regulator seeking to regulate an electricity market with rooftop solar. We formulate and analyze a model to study the effect of tariff *structure* on the rooftop solar market, with implications for consumers, regulators and industry. Using a sequential game, we analyze the regulator's social welfare maximization problem in a market with a regulated utility, an unregulated, monopolistic, profit-maximizing solar company, and customers who endogenously determine whether to adopt solar. We show that two tariff features are essential: the ability to discriminate among customer usage tiers and the ability to discriminate between customers with and without rooftop solar. Absent any one, we show that the regulator might not be able maximize her social welfare objectives while avoiding *cross-subsidization*, wherein rate changes leave some customers worse off than before. We then present a tariff with these two characteristics—featuring full retail price repurchasing from solar customers—that always guarantees feasibility of the regulator's optimization problem and also avoids cross-subsidization. We illustrate our findings numerically using publicly available data from Nevada and New Mexico, two states currently grappling with this issue. We find, in both cases, that our suggested tariff structure outperforms the tariff structures currently used in these states.

The second and third chapters of this dissertation take the perspective of profit-maximizing firms competing in duopolistic settings. In these chapters, we study the role of increased information availability on firm-level decisions. In two specific contexts, we uncover the relationship between capacity and additional information, finding that information augmentation is generally to the benefit of the firm with lower capacity.

Advances in internet-based technology have enabled service providers to disseminate real-time delay estimates to customers who are strategic and delay-sensitive. In the absence of such rich information, customers are generally guided by historical or moving average delay information, typically obtained through published reports or previous service encounters. While prior work documents the benefits of announcing real-time delay information for a service provider functioning in isolation (such as a call-center), the literature does not adequately establish whether announcing real-time delay is beneficial to a firm operating in a competitive setting (such as a restaurant or an emergency room).

In Chapter 2 we model a market with two service providers who compete for market share. We investigate whether one of the service providers (the technology leader L) should make the first move to announce her real-time delay information, when her competitor (the follower F) can opt to respond. We model and analyze this leader-follower setting as a sequential game. We use continuous time Markov chains to analyze the associated queueing dynamics. To analyze our sequential game, we define and study three information regimes: in Regime 0, neither of the two service providers announces real-time delay information, and customers are guided by historical average delay information. In Regime 1, one of the two service providers announces real-time delay information, while only historical average delay information is available for the other provider. This historical average delay information is updated from time-to-time; therefore, we model and analyze a *dynamic* model to study this Regime. In Regime 2, both service providers announce real-time delay information, and the resulting system is akin to a stationary Join-the-shortest queue system, for which analytical results in the literature are limited.

Using a mixture of analytical results and numerical experiments, we find that L's optimal action depends crucially on the relative service capacities of the service providers: initiating delay announcements improves, in equilibrium, the market share of L if she is the lower-capacity service provider and worsens her market share otherwise. Therefore, for a lower-capacity service provider, delay announcements can be considered a strategic remedy for capacity shortage.

Motivated by the availability of competitor fare information for airlines, in Chapter 3, we study the impact of explicitly using this fare information on equilibrium pricing strategies under a flexible model of customer demand. We build on prior studies that suggest that failing to explicitly account for competitors' decisions in a dynamic pricing exercise could result in revenue losses. We study equilibrium pricing strategies if these decisions are explicitly accounted for in a duopolistic setting. We consider a flexible demand model—that unifies the models used in prior literature—and captures three realistic behaviors: (i) Uncertain market size; (ii) Uncertain customer valuations; and (iii) The presence of loyal customers who have a preference between the firms' offerings *and* the presence of flexible customers who always buy down to the lowest price in the market.

The first natural question to ask is whether we can guarantee the existence of a subgame-perfect Nash equilibrium in pure strategies (SPNE-P). We establish the non-existence of an SPNE-P when some strictly positive proportion of the customers is loyal, i.e., when the firms' product offerings are not perfectly substitutable. However, when all customers are flexible, i.e., the firms' product offerings are perfectly substitutable, we show the existence of an SPNE-P. An example of such a situation is competition between two low-cost-airline carriers, such as *Spirit* and *Frontier* airlines. We then propose a refinement of the subgame-perfect pure strategy Nash equilibrium to produce a practically-relevant, (typically) unique equilibrium. Our equilibrium result is an extension of the result in Martínez-de Albéniz and Talluri [2011] for a more general model of demand.

Using these results, we study the strategic question of whether firms should explicitly use competitor price information in their pricing decisions, or continue to employ monopolistic models. We demonstrate that in general the unique equilibrium strategy is for both firms to use competitor fare information. We find that typically, the lower capacity firm is benefited by a migration to competitive models, while the higher capacity firm is left worse-off. However, when total capacity (relative to demand) is very low and the firms have roughly the same starting capacity, this migration could improve both firms' profits.

In Chapter 4, we summarize the contributions of this dissertation and suggest avenues for future research.

1

That's not Fair: Tariff Structures for Electricity Markets with Rooftop Solar

1 Introduction

Rooftop solar has seen a boom in recent years, with residential rooftop solar installations consistently growing at over 20% year-on-year [Mike Munsell, 2017]. One of the main catalysts for this growth has been the practice of utility companies offering so-called "retail net-metering" to customers with rooftop solar panels. Under this scheme, owners of rooftop solar panels can sell any excess electricity their panels generate to their utility company (the grid) for *full retail credit*. Effectively, such a customer pays only for her "net" usage.

While such an incentive is useful because a move to rooftop solar is environmentally desirable, retail net-metering threatens the profitability of utility companies, who are forced to buy excess energy from customers at retail rates which are significantly higher than their prevailing wholesale rates. Two common ways a utility might combat this erosion in profitability are by raising retail electricity rates for all users, or by reducing the rate at which utilities repurchase excess generation from solar households. However, both these solutions are problematic: If the utility company raises rates, (typically) poorer non-solar households would bear some part of the burden imposed by (typically) wealthier solar households [Krysti Shallenberger, 2017a]; this would result in *cross-subsidization*, a phenomenon under which one set of customers benefits (or is subsidized) at the cost of another set of customers. Alternately, if the repurchase rate is reduced, customers may no longer be incentivized to put panels on their roof, rooftop solar installers could be put out of business and the rooftop solar revolution could grind to a halt. This latter dynamic played out recently in Nevada [Buhayar, 2016].

In each of the thirty-three U.S. states with regulated electricity markets, a body called the Public Utilities Commission (PUC) has the charge of solving this complex problem: In each of these states, the PUC is tasked with balancing the welfare of the various stake-holders by regulating the rates and services of public utilities. The PUCs can therefore be thought of as social welfare maximizers; in the context of rooftop solar, this means protecting utility company profitability and ensuring fair electricity bills for customers, while providing a nourishing environment for rooftop solar in order to protect environmental interests (PUCs' stated objectives often explicitly include environmental stewardship; see California Public Utilities Commission [2017]). The PUC's task is further complicated by the fact that solar system installers (henceforth solar companies) are typically unregulated (because they are not providing a public good); therefore, regulatory interventions must account for such solar companies making *self-interested* decisions.

As might be expected, the effect of increased rooftop solar adoption on utility companies' profits has resulted in considerable regulatory flux: In the U.S., 42 of the 50 states took some action related to net-metering, rate design or solar ownership during the third quarter of 2015 alone [NC Clean Energy Technology Center and Meister Consultants Group, 2015]. The PUC's regulatory tight-rope walk of balancing customer, societal and utility welfare is a tricky affair that has, on occasion, gone awry: NV Energy, the utility company in Nevada, imposes a simple two-part tariff on their customers, who pay a monthly fixed charge and a variable "energy" charge per kWh of energy consumed. After being negatively impacted by increased rooftop solar adoption, NV Energy initiated what would become a prolonged dialogue with their PUC, in which they raised the spectre of cross-subsidization [Chediak and Buhayar, 2015]. The outcome of this dialogue was a ruling that solar customers would eventually pay *thrice* as high a fixed charge as non-solar customers and would be credited for excess generation at *wholesale* rates (significantly lower than the existing retail rate credit). This announcement prompted SolarCity, the market leader in solar systems, to suspend operations in the state of Nevada and cut over 500 jobs [Buhayar, 2016]. In December 2016, a year after this ruling was made, the PUC reversed its stand by voting to restore retail net-metering and the original rate schedule in the Sierra Pacific territory of Nevada [Pyper, 2016]. Meanwhile in February 2017 in Maine, the PUC passed a bill to phase down compensation paid to customers for their excess generation. In July 2017, a new bill that aimed to roll back this decision in order to boost solar growth was vetoed by the Governor, who cited cross-subsidization as the reason for his decision: he said that net-metering subsidizes the cost of solar panels "at the expense of the elderly and poor who can least afford it" [Krysti Shallenberger, 2017b].

There is no evident consensus on the structural properties a tariff should have in order to be effective: utilities in different states operate a variety of different tariff structures. For instance, utilities such as NV Energy in Nevada and Duke Energy in North Carolina have only a single tier in their tariff structures, whereas utilities such as PNM Energy in New Mexico and Idaho Power in Idaho have 3 tiers. Utility tariffs also vary in whether they discriminate between solar and non-solar customers. Nevada's policy changes permit NV Energy to have solar and non-solar customers on different rate schedules, but states such as New Mexico and Washington explicitly disallow this. Meanwhile, Arizona Public Service (APS), a utility in Arizona has a *tiered* tariff structure that pays solar customers less than the retail rate for excess energy sold back to the grid, effectively putting them on *different rate schedules from non-solar customers*.

Motivated by such developments, we explore this delicate problem faced by the PUCs, focusing on how effective different tariff structures that the PUC might impose on the utility are in enabling the PUC to induce socially optimal welfare outcomes in an electricity market with rooftop solar. We do so by explicitly modeling the regulator's social welfare optimization problem and showing that its feasibility is governed by the tariff structure in place, demonstrating in the process that some common tariff forms are potentially inadequate to the task. The model we analyze consists of a monopolistic, vertically integrated utility company (like NV Energy) whose tariffs are set by the PUC; a monopolistic, price-setting solar company (similar to SolarCity); and residential customers who are heterogeneous in their demands and generation capability (available roof space)[1]. After the PUC fixes a tariff (upon negotiation with the utility company), the solar company and customers play a sequential game: The solar company sets the price of solar systems to maximize its own profit, anticipating customers' decisions (made endogenously) to install solar or not. Customers make their self-interested installation decision based on their demand, rooftop solar generation potential, excess generation that they expect to sell back to the grid if they install solar (taken together, these determine a customer's "usage tier," or equivalently, "usage profile"), the tariff set by the PUC, and the solar company's declared price. Naturally, the regulator takes the behavior of the solar company and customers into account when deciding on a tariff.

A key element of our model is the endogenization of the (monopolistic) solar company's pricing decision, a departure from most existing literature. This is in keeping with the observation that over a third of the market share in the U.S. was claimed by one private company (SolarCity) in 2014 and 2015 [Roselund, 2015] that wields significant pricing power because of lower costs (Shahan [2016]). The endogenized solar pricing decision creates additional incentive compatibility constraints (that we specify and discuss formally in Section 3) that the regulator must account for, failing which the solar company may exit the market.

Using our model, we find two attributes that a tariff structure *must* have in order to guarantee effectiveness: the ability to discriminate between customers based on their usage tier and the ability to discriminate between customers with and without rooftop solar. While APS's tariff structure has both these features, at least one of them is absent in the tariff structures of many other utilities: for example, NV Energy in Nevada and PNM Energy in New Mexico. In the absence of either one of these attributes, we show that the regulator might not be able to induce a socially optimal outcome

[1]We ignore commercial customers for tractability.

and often will have to resort to an outcome with cross-subsidization. We then show that a simple two-part tariff *with* these attributes, featuring full retail price repurchasing from residential solar customers, always guarantees feasibility of the regulator's social welfare optimization problem. We also show that if solar adoption generates an overall customer surplus (possibly at the expense of the utility), this tariff structure can guarantee an outcome *with no cross-subsidization*.

Complementary to our analysis, using consumer survey data obtained at the household level (U.S. Energy Information Administration [2009]), we estimate customer usage profile parameters and numerically illustrate how our suggested tariff structure compares to the tariff structures currently in use in Nevada and New Mexico, two states wrestling with this issue. We find that both states' tariffs perform poorly compared to our suggested tariff: While our tariff is able to avoid cross-subsidization in all test cases, the current tariffs in both states are not. Notably, in Nevada, use of the existing tariff structure most adversely affects (possibly low income) customers living in the smallest houses.

Our base model assumes that all customer usage parameters are deterministic, that customers do not exhibit adverse behaviors such as exaggerating their demands or altering their generation capacity in order to generate bill savings, and that all customers within a tier behave identically with respect to their solar adoption decisions. We relax all of these assumptions in Section 6.

2 Literature Review

There is a substantial body of Operations Management literature exploring various aspects related to the management of renewable energy resources. Aflaki and Netessine [2017] and Hu et al. [2015] study capacity investment decisions for renewable resources such as wind and solar. The effect of tariff structures on such investments have been studied in Alizamir et al. [2016], Ritzenhofen et al. [2016], and Kök et al. [2016]. The operational aspects of managing renewable energy resources that are inherently variable are studied in Zhou et al. [2014], Wu and Kapuscinski [2013], and Al-Gwaiz et al. [2016].

The energy policy literature contains a stream of work investigating regulatory considerations arising from the increase in distributed generation. Some of these papers provide frameworks for regulation: Keyes and Rábago [2013] present a detailed framework with which a regulator may capture the costs and benefits of distributed generation, Lehr [2013] indicates that increased dependence on renewable energy necessitates regulatory overhauls that incentivize utility companies appropriately to move towards a renewable future and Linvill et al. [2013] qualitatively discuss the challenges a regulator might face when implementing net-metering or feed-in tariffs for compensating distributed generation.

Other papers in this literature explore policy issues related to net-metering. Blackburn et al. [2014] present a survey of utility companies to gauge their perspectives on net-metering, with the finding that most utility companies feel that they overcompensate customers for generation. Borlick and Wood [2014] show using a representative California customer that net-energy metering (in addition to the federal tax credit) does indeed provide too high a subsidy to customers and opines that these subsidies are flowing from less affluent customers to more affluent ones. Brown and Bunyan [2014] provide qualitative support to this finding and conclude that distributed solar energy is currently overvalued. Moore et al. [2016] find that the value of solar diminishes as more rooftop solar is added to the grid, and that a uniform net metering tariff is unlikely to be appropriate as solar photovoltaic (PV) penetrations increase. NC Clean Energy Technology Center and Meister Consultants Group [2015] provide a comprehensive view of recent solar-related regulatory changes that have been effected in the 50 states in the U.S.

Bird et al. [2013] describe the role of the regulator in a changing electricity landscape as: (1) Keeping the utility company viable, resulting in relatively stable cash flows and revenues from year to year; (2) Fairly apportioning the utility's cost of service among customers, without undue discrimination; and (3) Promotion of economic efficiency in the use of energy as well as competing products and services, without compromising on reliability. Our work draws on this description of the regulator's role: We formulate and solve an analytical model to explore which tariff structures enable the regulator to induce market outcomes in keeping with these three criteria.

Another stream within the energy policy literature studies the diffusion of solar among customers. Simulation approaches are common: Denholm et al. [2009] present SolarDS, a software tool that performs a simulation-based bottom-up analysis of solar adoption over time as a function of utility rates. This tool is used as a building block for many other pieces of research: For instance, Gagnon and Sigrin [2015] and Drury et al. [2013]. Villa et al. [2012] present a similar system-dynamics model that interfaces with data from the city of Albuquerque.

There is also a large body of empirical literature in this stream. Ong et al. [2010] investigate solar adoption in California under 55 different rate structures, finding that the financial benefit of owning a PV system decreases with increased demand charges. Lobel and Perakis [2011] estimate the dynamic diffusion process for solar adoption using a discrete-choice model for customers: Empirical analysis of data from Germany indicates that initial subsidies should be higher than they currently are (in contrast with the findings of Borlick and Wood [2014] who study subsidies in California), and that the phase-out of subsidies should be more rapid. Bauner and Crago [2015] empirically establish that because of uncertainty in bill savings, customers need to be offered larger financial benefits than their expected savings in order to induce them to adopt solar. In a similar vein, Darghouth et al. [2011] perform a computational study of the impact of flat, time-of-use, and real-time pricing on PV savings in California. They find that the interaction between increased adoption and retail rate changes can lead to substantial uncertainty on future bill savings, making customers' adoption choices less clear. Cai et al. [2013] study the feedback between retail rates and solar penetration, finding that the feedback effect is most strongly affected by the proportion of customers who adopt PV in any year, independent of future savings uncertainties—including changes in rate structures. Other diffusion models for customer adoption are presented in Bollinger and Gillingham [2012], Rai and Sigrin [2013], and Agarwal et al. [2015].

One possible response to solar adoption eroding their profits is for utility companies to raise electricity rates, making solar energy an even more attractive possibility and further undermining utility profitability. Satchwell et al. [2015] build a financial model for the utility company and solar adoption and model the feedback between adoption and retail rates. Their analysis suggests that such a "death-spiral" is unlikely but establishes the need for regulatory intervention. Costello and Hemphill [2014] also suggest that concerns of a death-spiral are premature, as distributed generation is yet to gain traction and establish itself. They do say that some leeway from the regulators will eventually be necessary to protect the interests of the utilities. Building on the SolarDS simulation model [Denholm et al., 2009], Darghouth et al. [2016] study two competing effects in this posited feedback loop between utility rates and adoption: They find that reduced demand drives rates up and shifting temporal consumption profiles drive rates down, in effect, cancelling each other out and precluding the possibility of a death-spiral.

Sunar and Swaminathan [2018] studies the impact of net metering policies on utility profits by quantifying the impact of net metered distributed generation on the utility's wholesale rate. In contrast to our approach, Sunar and Swaminathan treat the adoption level of distributed generation as exogenous.

Closer to our work, Babich et al. [2017] take the perspective of a government entity deciding between offering a feed-in-tariff and a tax-rebate policy for rooftop solar installation. They study how the policy in place affects the solar panel investment decisions of a representative household in the presence of exogenous shocks that affect generation efficiency, variability in electricity price and solar panel investment cost (i.e. they have a dynamic model, but with exogenously given solar prices). Similar to Babich et al. [2017], our work also deals with aspects of renewable energy that involve decisions by a principal (the PUC) and customers; however, this chapter presents a *static* (rather than dynamic) model of solar adoption among *heterogeneous* customers who make potentially heterogeneous investment decisions with a solar company that makes an endogenous pricing decision. Our model's static setting allows us to study the question of what the regulator's welfare-optimal choice of tariff should be, and our heterogeneous customer model allows us to study the customer equity implications of solar adoption. Goodarzi et al. [2018] take the perspective of a regulator who seeks to minimize utility costs by choosing an appropriate feed-in-tariff rate paid to customers who sell all rooftop generation back to the grid at that rate. In their model, customers are homogeneous in their demand characteristics but heterogeneous in their discount rates, and make a solar adoption decision based on the feed-in-tariff rate. Similar to our model, the solar system price in their model is also chosen endogenously by a profit maximizing solar company. However, in contrast to their paper, we model customers who are *heterogeneous* in their demands and generation capabilities, and who sell only excess generation back to the grid. As mentioned above, our heterogenous customer model allows us to study the customer equity implications of the policies in place. Furthermore, our regulator chooses both the tariff structure and parameters in order to maximize social welfare: This framework generalizes the feed-in-tariff rate that Goodarzi et al. [2018]'s analysis is restricted to.

Our work also relates to the extensive literature on uniform versus non-linear pricing: We study what features a tariff structure must have in order for a regulator to be able to induce a socially optimal outcome. Varian [1989] and the references therein provide an elaborate discussion on various issues related to price discrimination: Tiered tariffs are a tool for second-degree price discrimination. While other papers such as Sundararajan [2004] and Choudhary et al. [2005] discuss non-linear pricing for certain specific situations, their findings are not directly applicable to our setting because of our model's unique characteristics: The tariff chosen by the regulator interacts with customers' strategic behavior through the price of solar (which customers use to decide whether to adopt solar or not). This tariff must be chosen so as to induce the monopolistic solar company to set a price of solar that will induce a socially optimal outcome.

3 Model

Our model considers three decision making entities – residential customers who are heterogeneous in their usage profiles, a monopolistic solar company S, and a regulator R (the PUC). In addition, we model a vertically integrated utility company U that is subject to regulation by R (which makes decisions on behalf of U).

We define our base case scenario as one with no solar systems; i.e. where all customers depend on U to satisfy all their demand for electricity. Customers are subject to flat-rate (rather than time-of-use, for tractability reasons) pricing. After S makes its product available, customers have the option of continuing to depend solely on U for their energy requirements, or installing solar systems thereby reducing their dependence on U. We study how R's regulatory actions influence social welfare moving from the base case to the post-solar scenario. R's social welfare measurement takes into account financial and non-financial (i.e. environmental) considerations. However, customers and the solar company are modeled as being self-interested; i.e. they maximize their own financial objectives.

We first detail the parameters that characterize each entity and formally define their decision variables.

- **Customers:** Customers are heterogeneous and have different demands and potentials to generate solar electricity (e.g. because of heterogeneous availability of roof space). We consider I classes (equivalently, tiers) of customers indexed by $i \in \{1 \dots I\}$. Customer class i has annual demand d_i kWh and annual generation capability g_i kWh. We do not capture demand response in our model, and as such, treat d_i as fixed for a given tier i. A household can estimate g_i using tools such as Google's Project Sunroof [Google, 2018]. If a class i customer installs a solar system, her excess generation (the amount of her generation that she does not consume) is $e_i \leq g_i$; this excess generation is sold back to the grid. Modeling e_i as a separate parameter allows us to capture any potential temporal relationship between generation and demand; i.e., a customer's demand does not necessarily follow the same profile as her generation does, and their temporal relationship, whatever it is, determines e_i. Under our framework, a class i customer depends on the grid for an amount of energy $d'_i = d_i - (g_i - e_i)$ (her grid usage) and sells back an amount of energy e_i, pegging her net usage at $d'_i - e_i = d_i - g_i$. We arrange customer classes in order of increasing generation so that $g_i < g_{i+1}$. We do not impose a relationship between d_i and g_i. Therefore, a customer may be able to generate more than she demands. We let h_i denote the number of households belonging to class i.

A type i household has an adopt/do not adopt decision that we encode by binary variable s_i; s_i takes the value of 1 if the customer chooses to adopt solar and 0 otherwise. All customer parameters are assumed to be fixed, and known with certainty by all decision making entities. In Section 4 we discuss how our model may be extended to accommodate uncertainty in d, g, and e. Our base model also assumes that all households in a certain class make the same adoption decision. We also discuss in Section 4 how this assumption may be relaxed.

- **Solar Company:** We consider a self-interested, monopolistic solar company that sets prices for its solar systems. This assumption is driven by prevailing market conditions: In the U.S. SolarCity is the established market leader that has held a stable 34% market share in 2014 and 2015 [Roselund, 2015], about 3 times the market share of the nearest competitor) and lower costs, thereby giving it the power to set market prices. We assume that solar panels are infinitely divisible (i.e. we ignore the topography, or solar panel-roof compatibility).

  S's decision variable is p_s, the price that a customer who adopts solar must pay to the solar company per unit of electricity she generates using the installed solar system. SolarCity offers such a contract (this is called a Power Purchase Agreement), under which customers are only assessed a variable charge per kWh of generation, rather than having to pay a lump-sum amount for system purchase and installation [SolarCity, 2016d]. For customer-owned systems that involve upfront payments, this price p_s can be interpreted as a levelized price—the lifetime adjusted price per kWh that the system generates; this accounts for average sunlight received by the panel, efficiency considerations, down-payments, and maintenance costs. Since these are equivalent from a modeling standpoint, we consider a Power Purchase Agreement setup. Corresponding to the levelized price p_s that S chooses, we assume that the levelized cost to S is c_s per kWh of generation. Once set, p_s is assumed to remain fixed. Customers in the U.S. can avail themselves of an investment tax credit of 30% on solar system purchases [Energy.gov, 2017]. At the end of Section 3, we show how our model may be adjusted to accommodate this.

- **Regulator:** We consider a socially interested regulator R whose decision is a tariff function $T(d', e, s)$ that governs the annual rate that the utility company charges a customer who draws an amount of energy d' kWh/year from the grid, sells back e kWh/year to the grid and either adopts solar ($s = 1$), or does not adopt solar ($s = 0$). Note that both d' and e are measurable by U with an appropriate metering system. Also observe that if $s = 0$, d' (the amount drawn from the grid) is d, and $e = 0$.

- **Utility Company:** We consider a monopolistic utility company U that faces a fixed annual grid maintenance cost f_u and an average per unit cost of electricity c_u^x /kWh, where x is the amount of electricity it supplies. This framework allows us to capture the non-linear cost functions that utility companies typically face because changes in the amount of electricity they supply alter the mix of generation sources they use. The utility company uses its existing architecture to redistribute excess generation that it purchases from customers across the grid; we ignore any benefits that the decreased reliance on U's infrastructure will have on U. U does not take any decisions.

The game played by R, S and the customers proceeds in the following fashion: Let Period 0 be the base case scenario when no households have rooftop solar. Under the Period 0 tariff structure, these customers pay a per unit energy cost of p_{r0}, and an annual fixed cost that we normalize to 0, without loss of generality. We use this particular base case tariff structure for simplicity, but our approach readily extends to *any* general base case tariff structure. In Period 1, R imposes tariff structure $T(\cdot)$. In response to tariff $T(\cdot)$, S sets a per unit solar rate p_s in Period 2. In Period 3, individual customers, with knowledge of their demand d_i, generation capability g_i, and excess e_i, observe the tariff $T(\cdot)$ and the solar price p_s and then endogenously decide to adopt solar ($s_i^* = 1$) or not to adopt solar ($s_i^* = 0$). Since the agents take actions sequentially, each agent's decision is taken anticipating other agents' responses in future periods.

We now present the objective functions that govern the decisions of each of the decision-making entities in our model:

- **Customers:** Customers wish to minimize their spend on electricity. Therefore, a class i customer solves the following problem:

$$\max_{s_i \in \{0,1\}} \quad (1 - s_i)\, T(d_i, 0, 0) + s_i(\, T(d_i', e_i, 1) + p_s g_i) \tag{1.1}$$

- **Solar Company:** The solar company wishes to maximize profit by choosing an appropriate unit price of solar.

$$\max_{p_s > c_s} \quad (p_s - c_s) \sum_{i=1}^{I} s_i^* h_i g_i, \tag{1.2}$$

where s_i^* is the optimal adoption decision taken by a customer in tier i. We restrict p_s to being larger than c_s because S is unregulated and has the freedom to exit the market rather than make a non-positive profit.

- **Regulator:** The regulator R wishes to maximize social welfare improvement. Since R takes a systemic view, it is useful to think of the customers, U, and S as belonging to a "system," we will consider two components of this social welfare improvement: Financial and Environmental.

(1) Financial: Note that all cash flows except the purchase of electricity at the purchase costs c_u^x and c_s occur *within* the system and can therefore be ignored from the system's perspective. Define $E_0 = \sum_{i=1}^{I} h_i d_i$ as the total amount of energy that customers depend on the utility for in the base case, and $E_1 = E_0 - \sum_{i=1}^{I} s_i^* h_i g_i$ as the total amount of energy that customers depend on the utility for in the post-solar case (here, we assume all excess rooftop electricity is redistributed to other customers). It is useful to define $\Delta_E = \sum_{i=1}^{I} s_i^* h_i g_i$ as the amount of energy dependence migrated to rooftop solar. The net decrease in cash flows going out of the system (and hence the financial welfare improvement of the system) is therefore $c_u^{E_0} E_0 - c_u^{E_0-\Delta_E}(E_0 - \Delta_E) - c_s(\Delta_E) = (c_u^{E_0} - c_u^{E_0-\Delta_E})E_0 + (c_u^{E_0-\Delta_E} - c_s)\Delta_E$.

(2) Environmental: In addition to this financial welfare consideration, the regulator considers the environmental benefit accrued by sourcing Δ_E kWh of energy from rooftop solar rather than from the utility. Let m_u^x be the (monetized) average environmental cost of the utility generating one kWh of electricity when the total amount it generates is x, and m_s be the (monetized) environmental cost of a rooftop solar panel generating one kWh of electricity[2]. Again, using x to parameterize m_u^x allows us to capture the non-linear relationship between environmental cost imposed and amount of electricity supplied by the utility because migration to solar potentially changes the mix of generation sources for U. This environmental cost can, for instance, be estimated using the social cost of carbon. We can now synthesize this to write out R's objective function.

$$\max_{T(\cdot)} \quad \underbrace{c_u^{E_0} E_0 - c_u^{E_0-\Delta_E(T(\cdot))}(E_0 - \Delta_E(T(\cdot))) - c_s \Delta_E(T(\cdot))}_{\text{Financial}} + \underbrace{m_u^{E_0} E_0 - m_u^{E_0-\Delta_E(T(\cdot))}(E_0 - \Delta_E(T(\cdot))) - m_s \Delta_E(T(\cdot))}_{\text{Environmental}}, \tag{1.3}$$

where $\Delta_E(T(\cdot))$ is the extent of migration to rooftop solar induced by tariff choice $T(\cdot)$.

The financial benefit from solar adoption crucially depends on the values of $c_u^{E_0-\Delta_E(T(\cdot))}$ and $c_u^{E_0}$ relative to c_s. Based on publicly available information, we estimate that c_s is approximately \$0.064 (the details of this estimation procedure are presented in Section 5). The average wholesale rate of electricity (which, in our case is a good estimate of $c_u^{E_0}$) in the U.S. is roughly \$0.04 [U.S. Energy Information Administration, 2017]. Solar production typically peaks around mid-day, creating the so-called duck curve [Jeff, St. John, 2016], and thus does not generally shave off the peak load (which typically occurs in the late evening) or displace the base load generators during low load periods (early in the morning). Therefore, we do not expect $c_u^{E_0-\Delta_E(T(\cdot))}$ to be significantly different from \$0.04. If $c_u^{E_0-\Delta_E(T(\cdot))}$ and $c_u^{E_0}$ are equal, the financial benefit simplifies to $(c_u - c_s)\Delta_E(T(\cdot))$. Therefore, as long as $c_u^{E_0-\Delta_E(T(\cdot))}$ and $c_u^{E_0}$ are close enough, we expect the financial benefit from

[2]This does not depend on the total amount of electricity generated by solar.

solar adoption to be negative (because \$0.04 < \$0.065); this can be interpreted as a cost that society must bear in order to encourage solar adoption. As technological improvements cause c_s to drop, the sign of this benefit could flip. Our model is robust to either case.

It is worth pointing out that although we do not explicitly model the decisions of the utility company in response to R's actions, we implicitly capture any capacity changes that the utility company would need to make in response to R's actions through the parameters $c_u^{E_0-\Delta_E(T(\cdot))}$ and $m_u^{E_0-\Delta_E(T(\cdot))}$: Since these depend on $T(\cdot)$, the impact of these capacity decisions on utility and system welfare are captured. The utility company has another role in our model: While choosing an adoption level that maximizes social welfare improvement, regulator R must ensure a specified rate-of-return to the utility company. We codify this by denoting the permissible increase in utility profit going from the base case to the post-solar case as Δ_U (note that this could be negative)[3]. Further, she would like to ensure a profit of Δ_S for the solar company (for example, to encourage further technological innovation; we assume $\Delta_S > 0$). Since the regulator is also responsible for making sure that customers do not overpay for electricity, she would like to choose, from among the functions $T(\cdot)$ that maximize her objective and respect the other constraints she faces, the one(s) that minimize the maximum cash outflow seen by any class of customers. We call this the fairness constraint. This is a common fairness criterion used in game theory, ethics, and communication networks, generally credited to Rawls [Rawls, 1974]. If this minimized maximum cash outflow is negative, then it means that all classes of customers benefit, and there is no cross-subsidization because no class of customer is hurt by the introduction of solar to the market. Formally, we represent these as constraints in R's optimization problem as follows:

$$\begin{aligned}
&\sum_{i=1}^{I} h_i \left(s_i^* T(d_i', e_i, 1) + (1-s_i^*) T(d_i, 0, 0) - c_u^{E_0-\Delta_E(T(\cdot))} (s_i^*(d_i - g_i) + (1-s_i^*) d_i) \right) \\
&- \sum_{i=1}^{I} h_i (p_{r0} - c_u^{E_0}) d_i = \Delta_U
\end{aligned} \tag{1.4}$$

$$(p_s - c_s) \sum_{i=1}^{I} s_i^* h_i g_i = \Delta_S \tag{1.5}$$

$$\begin{aligned}
&\max_i \quad s_i^* \left(T(d_i', e_i, 1) + p_s g_i \right) + (1-s_i^*) T(d_i, 0, 0) - p_{r0} d_i \\
&= \min_{T(\cdot)\in\tau} \quad \max_i \quad s_i^* \left(T(d_i', e_i, 1) + p_s g_i \right) + (1-s_i^*) T(d_i, 0, 0) - p_{r0} d_i,
\end{aligned} \tag{1.6}$$

where τ is the set of tariff structures that maximize the regulator's objective and satisfy all other constraints that the regulator must account for. A precise characterization of τ is provided in Section 4.

While this set-up doesn't explicitly take into account the 30% investment tax credit from the federal government, this can easily be accommodated with minor adjustments to the model: We

[3]This value of Δ_U may be padded to account for transmission losses. For clarity of exposition, we do not model this.

can now treat p_s as the discounted rate that customers pay for solar power, and the solar company now obtains revenue at a rate $\frac{p_s}{70\%}$ per kWh of energy. All our results continue to hold under this modification, so we ignore the investment tax credit for the remainder.

4 Analysis

We begin our analysis by examining the optimization problem of a class i customer, as specified in (1.1). Such a customer favors adopting solar if and only if $p_s \leq \dfrac{T(d_i, 0, 0) - T(d'_i, e_i, 1)}{g_i} \triangleq t(i)$ (we break ties in favor of adoption). Note that these values of $t(i)$ depend on $T(\cdot)$: specifying $T(\cdot)$ induces an ordering among the $t(i)$ values. Note also that for a given solar price p_s, the set of classes that adopt is $\{i : t(i) \geq p_s\}$.

Now consider S's pricing decision. We can assert that S's optimal choice of p_s must be either $t(i)$ for some $i \in \{1, \ldots I\}$ or some price larger than $\max_i \; t(i)$: If S chose some price p_s between $t(i)$ and $t(j)$ for some i and j such that $t(j) = \min_k \; t(k) : t(k) > p_s$, she could do strictly better by choosing price $t(j)$, as doing so increases S's margin and does not alter her volume (see equation (1.2)). Therefore, S's optimization problem reduces to choosing an i^* such that the profit obtained by setting $p_s = t(i^*)$ is larger than the profit obtained from all other choices $j \neq i^*$ or choosing $p_s > \max_i \; t(i)$. We clarify that *all* adopting customers pay the chosen solar price p_s; $t(i)$ is simply a threshold value of p_s upto which a class i customer is induced to adopt solar.

These observations allow us to rewrite R's optimization problem, folding in the decisions of S and households to reflect the sequence in which they are taken. To do so, we define some more notation. Let set $D = \{x \in \mathbb{R} : x = \vec{s} \cdot (h_1 g_1, h_2 g_2, \ldots, h_i g_i, \ldots, h_I g_I)^T\}$, where $\vec{s}$ is any I dimensional vector whose entries are binary. Here, $\vec{s}$ is the adoption decision vector $(s_1, s_2, \ldots s_i, \ldots, s_I)$ whose entries indicate whether class i adopts or not. The set D therefore contains all possible values that Δ_E (which we refer to as the *migration quantity*) could take. We are interested in the case that R wishes for at least one class to adopt; that is, the case where $\exists i : s_i = 1$, otherwise the problem is trivial. This is the case we shall assume for the remainder. Apart from this, we do not restrict $\vec{s}$: R may choose any arbitrary subset of classes to be adopters in order to maximize her social welfare objective. Let z index into this set. Corresponding to each migration quantity $E^{(z)}$, there exists at least one adoption vector $\vec{s}^{(z)} = (s_1^{(z)}, s_2^{(z)}, \ldots, s_i^{(z)}, \ldots, s_I^{(z)})$. Define adoption set $A^{(z)} = \{i : s_i^{(z)} = 1\}$.

Now, notice that R's objective function (1.3) is affected by $T(\cdot)$ only through Δ_E. Therefore, R's decision can be equivalently modeled as choosing z optimally from the set $\{1, 2, \ldots, 2^I\}$. In order for this value of z to induce adoption outcome $\vec{s}^{(z)}$, it must be the case that p_s is chosen so that $t(i) \geq p_s, \forall i \in A^{(z)}$. Define a class i to be 'marginal' if $p_s = t(i)$. Since p_s is chosen from among the set of $t(i)$ values, which as functions of $T(\cdot)$ are themselves variables, R may allow any (indeed, more than one) of the adopting classes in $A^{(z)}$ to be marginal. Let M be the set of indices of the marginal adopting classes. Since we assumed that at least one class adopts, this set must be non-empty. We can now pose R's optimization problem, folding in the household and solar company decision as follows:

$$\max_{T(\cdot),z\in\{1,2,\ldots,2^I\},M} \quad \begin{aligned} & c_u^{E_0}E_0 - c_u^{E_0-E^{(z)}}(E_0 - E^{(z)}) - c_sE^{(z)} + \\ & m_u^{E_0}E_0 - m_u^{E_0-E^{(z)}}(E_0 - E^{(z)}) - m_sE^{(z)}, \end{aligned} \tag{1.7}$$

Subject to constraints:

$$t(i) = \frac{T(d_i,0,0) - T(d_i',e_i,1)}{g_i}, \quad \forall i \tag{1.8}$$

$$M \subseteq A^{(z)} \tag{1.9}$$

$$\begin{aligned} & \sum_{i=1}^{I} h_i\left(s_i^* T(d_i',e_i,1) + (1-s_i^*)T(d_i,0,0) - c_u^z(s_i^*(d_i - g_i) + (1-s_i^*)d_i)\right) \\ & - \sum_{i=1}^{I} h_i(p_{r0} - c_u^{E_0})d_i = \Delta_U \end{aligned} \tag{1.10}$$

$$\begin{aligned} & \max_i \quad s_i^{(z)}\left(T(d_i',e_i,1) + t(m)g_i\right) + (1-s_i^{(z)})T(d_i,0,0) - p_{r0}d_i \\ & = \min_{T(\cdot)\in\tau} \quad \max_i \quad s_i^{(z)}\left(T(d_i',e_i,1) + t(m)g_i\right) + (1-s_i^{(z)})T(d_i,0,0) - p_{r0}d_i \end{aligned}, \quad \forall m \in M \tag{1.11}$$

$$(t(m) - c_s)\sum_{i=1}^{I} s_i^{(z)}h_ig_i = \Delta_S, \quad \forall m \in M \tag{1.12}$$

$$(t(i) - c_s)\sum_{j=1}^{I} \mathbb{I}_{t(j)\geq t(i)}h_jg_j < \Delta_S, \quad \forall i \notin M \tag{1.13}$$

$$t(i) \geq t(m), \quad \forall i \in A^{(z)}, m \in M \tag{1.14}$$

$$t(i) < t(m), \quad \forall i \notin A^{(z)}, m \in M \tag{1.15}$$

Here, (1.8) defines $t(i)$ in terms of the regulator's decision variables, (1.9) ensures that the choice of marginal adopting classes is consistent with the choice of z, (1.10) ensures that U receives the specified rate of return implied by Δ_U, (1.11) imposes the fairness constraint on customer payments, (1.12) ensures that S achieves the specified profit Δ_S by choosing solar price $t(m)$, $m \in M$, (1.13) is a set of incentive compatibility constraints, which ensure that S can do no better than make profit Δ_S by setting $p_s = t(i)$ for $i \notin M$[4], and (1.14)-(1.15) impose the condition that customers in $A^{(z)}$ are induced to adopt solar.

This formulation is not expressed in a convenient form, and is therefore not amenable to analysis: τ is ill-defined in (1.11), and (1.9) and (1.13) are not expressed in canonical form because they deal with set containment and the indicator function, respectively. However, by imposing the following restriction on tariff function $T(\cdot)$, we can convert this problem into a form that is easier to analyze.

[4]We model these incentive compatibility constraints as being strict rather than weak, because if S deviates to a price $p_s \neq t(m)$, the outcome induced is different from the desired z. We will show in Section 4.3 that if the tariff structure is appropriately chosen, this strict inequality does not impair the feasibility of the problem.

Tariff restriction: Tariff function $T(\cdot)$ is chosen so that all $t(i)$ values are distinct.

Proposition 1 *The tariff restriction does not affect the regulator's ability to maximize her objective* (1.7)*. Further, this restriction does not compromise her ability to meet all constraints, including the fairness constraint* (1.11)*.*

Proof: Presented in Appendix A.1

Therefore, the regulator can, without loss of optimality, choose a tariff function that ensures that all $t(i)$ values are distinct. We can now decompose R's problem neatly by making the observation that the objective function expressed in (1.7) depends only on z and can therefore be solved independently of the constraints. Accordingly, we can carry out the following steps:

1. Find the value of z (call this z^*) that maximizes the objective function. We call this problem $\mathcal{P}_1$.

2. Corresponding to the value of z^* chosen, we can enumerate all underlying orderings over the $t(i)$ values that could have resulted in adoption outcome $\vec{s}^{(z^*)}$. These feasible underlying orderings can be obtained by permuting the ordering of adopters (which we can do in exactly $|A^{(z^*)}|!$ different ways), and for each of these orderings, permuting the non-adopters (which we can do in $(I-|A^{(z^*)}|)!$ different ways). Let $O^{(z^*)}$ be the set of these orderings and let o be an index into these orderings. Observe that $|O^{(z^*)}| = |A^{(z^*)}|!(I-|A^{(z^*)}|)!$. Note that fixing the ordering over the $t(i)$ values automatically fixes the marginal adopting class: Pick the index corresponding to the adopting class chosen to have the lowest $t(i)$ value according to o. Let $m(o)$ be the marginal adopting class.

3. Use the value of z^* so obtained to solve $|O^{(z^*)}|$ different optimization problems (which we can index by o), one for each possible ordering, setting (1.11) as the objective. Choose an ordering with the best objective value. We call this problem $\mathcal{P}_2$.

We can formally write these problems as follows:

Problem $\mathcal{P}_1$:

$$\max_{z\in\{1,2,\ldots,2^I\}} \quad \begin{aligned} & c_u^{E_0}E_0 - c_u^{E_0-E^{(z)}}(E_0-E^{(z)}) - c_sE^{(z)} + \\ & m_u^{E_0}E_0 - m_u^{E_0-E^{(z)}}(E_0-E^{(z)}) - m_sE^{(z)}. \end{aligned} \tag{1.16}$$

Once the optimal z^* is obtained, solve the following $|O^{(z^*)}|$ optimization problems (which are now in canonical form) and choose the solution with the best objective value.

Problem $\mathcal{P}_2$:

$$\min_{T(\cdot)} \quad \max_{i} \quad s_i^{(z^*)} \left(T(d_i', e_i, 1) + p_s g_i\right) + (1 - s_i^{(z^*)})\, T(d_i, 0, 0) - p_{r0}\, d_i \tag{1.17}$$

$$\text{Subject to constraints:} \tag{1.18}$$

$$t(i) = \frac{T(d_i, 0, 0) - T(d_i', e_i, 1)}{g_i}, \forall i \tag{1.19}$$

$$\begin{aligned} &\sum_{i=1}^{I} h_i \left(s_i^* T(d_i', e_i, 1) + (1 - s_i^*)\, T(d_i, 0, 0) - c_u^{z^*} (s_i^* (d_i - g_i) + (1 - s_i^*)\, d_i)\right) \\ &- \sum_{i=1}^{I} h_i (p_{r0} - c_u^{E_0})\, d_i = \Delta_U \end{aligned} \tag{1.20}$$

$$t(i) \text{ ordering consistent with } o \tag{1.21}$$

$$(t(m(o)) - c_s) \sum_{i=1}^{I} s_i^{(z^*)} h_i g_i = \Delta_S \tag{1.22}$$

$$(t(i) - c_s) \sum_{j=1}^{I} \mathbb{I}_{t(j) > t(i) \text{in ordering } o} \cdot h_j g_j < \Delta_S, \forall i \neq m(o) \tag{1.23}$$

Here, (1.21) is a set of $I - 1$ inequalities that impose an ordering over the $t(i)$ values consistent with o.

We summarize the notation used in our model in Table A.1 in Appendix A.8.

Observe that optimization problem $\mathcal{P}_1$ always has a solution because it is unconstrained. However, it is not clear that $\mathcal{P}_2$ is feasible. In particular, it is not immediately clear that the tariff parameters can ensure that the incentive compatibility constraints (1.23) and the ordering constraints (1.21) can hold together. In relation to this observation, we answer the following question: Let $\mathcal{T}$ be the set of allowable tariff functions from which $T(\cdot)$ must be chosen. How does the choice of $\mathcal{T}$ affect the feasibility of $\mathcal{P}_2$? We are also interested in the following additional question: If $\mathcal{P}_2$ is feasible, can it induce an outcome free from cross-subsidization?

Definition 1 ***Cross-subsidization (CS):*** *A market outcome is said to feature CS if the objective value of $\mathcal{P}_2$ is positive; i.e., at least one class of customer is financially worse off in the post-solar case. Similarly, an outcome is said to be free from CS if the objective value of $\mathcal{P}_2$ is non-positive.*

For an outcome to be free from CS, we naturally require that the *total* improvement in financial welfare for customers $\Delta_C = c_u^{E_0} E_0 - c_u^{E_0 - E^{(z)}} (E_0 - E^{(z)}) - c_s E^{(z)} - \Delta_S - \Delta_U \geq 0$. We examine how the choice of $\mathcal{T}$ affects the regulator's ability to induce outcomes free from CS when $\Delta_C \geq 0$; specifically, is $\Delta_C \geq 0$ sufficient to induce such a CS-free outcome?

4.1 Non-tiered tariff structure that discriminates between adopters and non-adopters

Many states in the U.S. (including Nevada, which we examine more closely in Section 5) have utility companies that administer non-tiered rate schedules for residential customers. Non-tiered tariff structures have the benefit of being simple to administer and therefore simple to modify in the rate case proceedings, the process by which utility companies petition for rate changes to the PUC. These non-tiered structures can, however, discriminate between solar adopters and non-adopters; i.e., these two types of customers may be subject to different rate schedules, as is the case with NV Energy in Nevada. We study such rate structures in this section.

Let $\mathcal{T}$ be the set of linear, non-tiered tariff structures that discriminate between adopters and non-adopters; i.e., they are on different rate schedules. Such a tariff function has the following general specification:

$$\begin{aligned} T(d', e, 0) &= r_d d + r_0, \\ T(d', e, 1) &= s_d d' + s_e e + s_0. \end{aligned} \tag{1.24}$$

We now present an analysis of the tariff structure (1.24). Specifically, we examine whether the feasibility of $\mathcal{P}_2$ is guaranteed under such a tariff structure. The following is the system of constraints in problem $\mathcal{P}_2$ for a given ordering o and a given adoption outcome z^*.

$$t(i) = \frac{r_d d_i + r_0 - (s_d(d_i - g_i + e_i) + s_e e_i + s_0)}{g_i}, \forall i \tag{1.25}$$

$$\begin{aligned} &\sum_{i=1}^{I} h_i \left(s_i^{(z^*)}(s_d(d_i - g_i + e_i) + s_e e_i - c_u^{z^*}(d_i - g_i) + s_0) + (1 - s_i^{(z^*)})((r_d - c_u^{z^*})d_i + r_0) \right) \\ &- \sum_{i=1}^{I} h_i(p_{r0} - c_u^{E_0})d_i = \Delta_U \end{aligned} \tag{1.26}$$

$$t(i) \text{ ordering consistent with } o \tag{1.27}$$

$$(t(m(o)) - c_s) \sum_{i=1}^{I} s_i^{(z^*)} h_i g_i = \Delta_S \tag{1.28}$$

$$(t(i) - c_s) \sum_{j=1}^{I} \mathbb{I}_{t(j) > t(i) \text{in ordering } o} \cdot h_j g_j < \Delta_S, \forall i \neq m(o) \tag{1.29}$$

For this tariff structure with non-tiered rates, we prove the following propositions:

Proposition 2 *Tariff structure* (1.24) *cannot guarantee the feasibility of* $\mathcal{P}_2$*: There exist parameters and outcomes* z^*, Δ_S, Δ_U *for which* $\mathcal{P}_2$ *is not feasible for* any *ordering* o*.*

Proof: Please see Appendix A.2

Proposition 2 implies that the system (1.25)-(1.29) does not always have a solution. While this is discouraging, the following proposition shows that under a restriction on z^*, there does exist a feasible ordering of $\mathcal{P}_2$ if we drop the IC constraints.

Proposition 3 *In the absence of the incentive compatibility constraints* (1.29)*, there exists an ordering o for which tariff structure* (1.24) *guarantees the feasibility of optimization problem* $\mathcal{P}_2$ *if* $z^* = \{i^*, i^* + 1, \ldots, I\}$ *for some* i^*.

Proof: Please see Appendix A.3

This has an important implication: If the solar price p_s were also controlled by the regulator R, and z^* prescribes that a contiguous block of high-generation customer tiers adopts, a linear tariff structure with non-tiered rates *would* suffice to satisfy system (1.25)-(1.29), as the regulator would not have to contend with IC constraints (1.29). Alternatively, if the utility company itself offered rooftop solar rather than an outside firm, the IC constraints could be ignored (as the solar price set by U would now be subject to regulation) and equation (1.26) suitably modified.

The intuition for this tariff's failure to achieve feasible outcomes is its limited ability to transfer welfare among customers in different tiers. In particular, its ability to selectively make solar unattractive to some tiers and not to others is limited by not having tier-dependent parameters. The conditions laid out in Proposition 3 remove some of these hurdles.

While dropping the IC constraints is a special case under which the linear tariff structure suffices, the current environment in the US is one with an unregulated solar company. Therefore, a richer class of tariff structures may be required to satisfy a system analogous to (1.25)-(1.29), modified based on the choice of $\mathcal{T}$.

4.2 Tiered tariff structure that does not discriminate between solar adopters and non-adopters

In states such as New Mexico (which we study in detail in Section 5) and Washington, the PUCs have mandated that solar customers may not be assessed any additional standby, capacity, interconnection, or other fee or charge by the utility [NC Clean Energy Technology Center, 2017a,b]. Such a rule serves as an incentive for solar adoption. These tariff structures may be tiered, but operate under a single rate schedule and feature retail net-metering, whereby customers who adopt solar sell back excess electricity at their retail rate: If the utility repurchased electricity at less than their retail rate this would be considered a fee to solar adopters and is thus prohibited. Let $\mathcal{T}$ be the set of such tariff functions. We will show that operating such a tariff structure, while guaranteeing feasibility, limits the ability of the regulator to induce a CS-free market outcome.

Under such a tariff structure, the appropriate *rate class* (not to be confused with usage class $i \in \{1, 2, \ldots I\}$) in the rate schedule is applied based on a household's *net demand*. Recall that this net demand is the household's demand less their generation if they adopt solar. For instance, if a tier 1 household does not adopt solar, demands an amount of electricity d_1, and is placed in rate class 1, a tier 2 household who adopts solar and has a net demand of $d_2 - g_2 = d_1$ also falls into rate class 1 and is billed as such.

Let $C = \{1, 2, \ldots, |C|\}$ be the set of indices corresponding to rate classes in U's rate schedule. Without loss of generality, let this set be arranged in order of increasing (net) demand; that is, rate class 1 corresponds to the lowest net demand and rate class $|C|$ corresponds to the highest net demand. In order to support customers making endogenous solar adoption decisions, this rate schedule must contain enough rate classes to support any possible adoption/non-adoption decision by customers. Therefore, $|C| \geq I + 1$.

Now, consider the tariff function $T(d', e, s) = T(c) = r_c n + f$, where c is the index of the rate class to which a customer with usage profile (d', e, s) belongs and n is the household's net demand; this is obtained by mapping the household's net energy usage ($n = d$ for a non-adopter and $n = d' - e = d - g$ for a non-adopter) to a class $c \in C$, and f is a fixed cost that all customers pay. The tariff function is fully defined by choosing $r_c, \forall c \in C$, and f. Note that having the fixed cost f also depend on class; i.e., having a different fixed cost f_c for every class c is equivalent to the system currently under consideration, as r_c can be suitably modified for each class c to compensate for the difference $f_c - f$.

Proposition 4 *Corresponding to every ordering o and outcome z^*, there exists a feasible rate schedule (r_c, f) that satisfies the constraints of $\mathcal{P}_2$.*

Proof: Presented in Appendix A.4

While this tariff structure can always feasibly induce an outcome characterized by z^*, Δ_S, Δ_U, we find that it cannot guarantee a CS-free outcomes when $\Delta_C \geq 0$.

Proposition 5 *The tiered tariff structure that does not discriminate between adopters and non-adopters cannot guarantee CS-free outcomes: There exist parameters and outcomes z^*, Δ_S, Δ_U for which no CS-free outcome can be generated for* any *ordering o even when $\Delta_C \geq 0$.*

Proof: Presented in Appendix A.5

Therefore, while this tariff structure is simple and guarantees feasibility, it does not have desirable properties with respect to customer equity. Intuitively, this tariff fares better than the non-tiered tariff studied in Section 4.1 because its tiered nature allows welfare transfer among tiers. However, its ability to shield customers from cross-subsidization is limited by the fact that adopters and non-adopters may be grouped into the same tier. Therefore, a tariff structure that can guarantee CS-free outcomes must (at least) be able to discriminate between solar adopters and non-adopters by placing them in different rate schedules.

4.3 Tiered tariff structure that differentiates between solar adopters and non-adopters

As an alternative to the class of tariff structures explored in Sections 4.1 and 4.2, consider a *tiered* tariff structure that presents *different rate schedules* to solar and non-solar customers. Both these attributes are present in the tariff structure operated by Arizona's APS. We propose and study a specific tariff structure with these attributes: one that is non-tiered for non-adopters[5] and tiered for adopters. Under our structure, solar and non-solar customers are charged the same variable energy charge based on their *net* energy consumption[6], but a solar customer is also assessed a fixed annual charge *that depends on her class i.* A general specification of such a rate structure follows; for all classes i,

$$\begin{aligned} T(d_i, 0, 0) &= p_r d_i, \\ T(d_i', e_i, 1) &= p_r(d_i') - p_r(e_i) + f_i \\ &= p_r(d_i - g_i) + f_i. \end{aligned} \tag{1.30}$$

Note that U can infer a solar customer's class by observing the value of d_i', the energy drawn from the grid and contract on a fixed cost that a customer would pay based on this amount. We now present an analysis of tariff structure (1.30): We show that the feasibility of $\mathcal{P}_2$ is guaranteed under such a tariff structure and that it can also guarantee no CS if $\Delta_C \geq 0$.

Consider problem $\mathcal{P}_2$ for a given ordering o and a given adoption outcome z^*. It will be useful to re-order the indices to be consistent with o; that is, the indices are chosen such that $t(i) < t(j), \forall i < j$. As a result of this re-ordering, classes $1, \dots m-1$ do not adopt solar, while classes $m, \dots, I$ adopt, where m is the index of the marginal adopter. Note that as a result of this re-ordering, we no longer have the property that $g_i < g_j, \forall i < j$.

$$t(i) = p_r - \frac{f_i}{g_i}, \quad \forall i \tag{1.31}$$

$$\begin{aligned} &\sum_{i=1}^{m-1} h_i(p_r - c_u^{(z^*)})d_i + \sum_{i=m}^{I} h_i(p_r(d_i - g_i + e_i) - c_u^{(z^*)}(d_i - g_i) - p_r e_i + f_i) \\ &- \sum_{i=1}^{I} h_i(p_{r0} - c_u^{E_0})d_i = \Delta_U \end{aligned} \tag{1.32}$$

$$t(i) < t(j), \quad \forall i < j \tag{1.33}$$

$$(t(m) - c_s) \sum_{i=m}^{I} h_i g_i = \Delta_S \tag{1.34}$$

$$(t(i) - c_s) \sum_{j=i}^{I} h_j g_j < \Delta_S, \forall i \neq m \tag{1.35}$$

Because this tariff structure has more parameters than the non-tiered tariff structure in (1.24), it might at first seem intuitive (and trivial) that this structure guarantees the feasibility of $\mathcal{P}_2$: One

[5]If the Period 0 tariff structure is tiered rather than having the same rate p_{r0} apply to all customers, our proposed tariff structure would require that non-adopters also face tiered rates.

[6]Note that since solar adopters are charged based on net energy use, we implicitly prescribe retail net-metering.

could correctly make the observation that the fixed costs for adopting customer classes i can be chosen arbitrarily to adjust Δ_U up and down. However, these fixed costs must *also* result in the solar price of $p_s = t(m)$ being an incentive compatible choice for S: Picking a solar price $p_s \neq t(m)$ must result in a smaller profit (the product of margin and volume, see equation (1.35)) than her profit from picking $p_s = t(i)$. This deviation can be made sufficiently unattractive by ensuring that values $t(i)$ for $i \neq m$ are low enough; R can achieve this by manipulating the values of f_i, $i \neq m$ (see equation (1.31)). But this, again, imposes restrictions on our choice of fixed costs f_i. Therefore, it is not immediately obvious that there exists a feasible tariff adhering to this tariff structure. However, it turns out that this tiered tariff structure *does* indeed guarantee feasibility of $\mathcal{P}_2$.

Proposition 6 *Corresponding to every ordering o and outcome* z^*, Δ_U, Δ_S*, there exists a feasible tariff function of the form* (1.30) *that satisfies the constraints of* $\mathcal{P}_2$*.*

Proof: First, find f_m as a function of p_r by solving equation (1.34) to obtain $t(m) = \left(c_s + \frac{\Delta_S}{\sum_{j=m}^{I} h_j g_j}\right)$ and $f_m = g_m\,(p_r - t(m))$. For all $i < m$, set $t(i) = \epsilon i$ for some arbitrarily small $\epsilon > 0$ by setting $f_i = g_i(p_r - \epsilon i)$. For all $i > m$, set $t(i) = \left(c_s - \epsilon + \frac{\Delta_S}{\sum_{j=i}^{I} h_j g_j}\right)$ and $f_i = g_i(p_r - t(i))$. Note that by definition, this set of $t(i)$ values satisfies (1.31) and (1.34). Further, IC constraints (1.35) are satisfied because the $t(i)$ values for $i > m$ are set ϵ smaller than what they would have to be to obtain a profit of Δ_S, and for $i < m$ are set so low that they generate a negative profit. Finally, the ordering constraints (1.33) are satisfied for $i < m$ by definition, and for $i \geq m$ (if the ϵ term is sufficiently small) because $h_j g_j > 0, \forall j$, leading to $\sum_{j=i}^{I} h_j g_j$ decreasing in i. Once these f_i values have been set, substitute them in (1.32) to obtain a value for p_r. ■

So, our proposed two-part tariff, like the tariff presented in Section 4.2, guarantees the feasibility of optimization problem $\mathcal{P}_2$.

This tariff (1.30) features the same energy (variable) cost across tiers, and varying fixed costs f_i by tier for solar customers. Similar to our discussion in Section 4.2, R could alternatively impose a tariff structure featuring the same fixed costs across tiers, with energy costs varying across adopting tiers. This is a commonly used tariff structure, where the rate schedule that a customer is subject to features varying energy rates per kWh of usage that depend on the usage level. Such a structure has the following specification.

$$\begin{aligned} T(d_i, 0, 0) &= p_{r2}\, d_i, \\ T(d_i', e_i, 1) &= p_r^{(i)}(d_i') - p_r^{(i)}(e_i) + f \\ &= p_r^{(i)}(d_i - g_i) + f. \end{aligned} \tag{1.36}$$

Such a tariff is, in every way, equivalent to the tariff (1.30). By setting $f = f_m$, $p_r^{(i)} = p_r + \frac{f_i - f_m}{d_i - g_i}, \forall i$, and $p_{r2} = p_r$, we can map from structure (1.30) to structure (1.36). We now study properties of this tariff structure.

Lemma 1 *Consider an adopting tier i. Then, we have that* $g_i p_r - g_i p_s - f_i \geq 0$.

Proof: If m is the index of the marginal class, $p_s = p_r - \frac{f_m}{g_m}$. For all tiers i that adopt, it must be the case that $t(m) \leq t(i) \Leftrightarrow \frac{f_m}{g_m} \geq \frac{f_i}{g_i}$ (from equations (1.31)). Therefore, we have $g_i p_r - g_i p_s - f_i = g_i p_r - g_i(p_r - \frac{f_m}{g_m}) - f_i = g_i \frac{f_m}{g_m} - f_i \geq 0$. ■

We now address properties that this tariff structure exhibits with respect to customer equity and CS. Lemma 1 will be useful in examining these properties. For tariff structure (1.30), we state and prove the following propositions:

Proposition 7 *When* $\Delta_C > 0$*, there exists an ordering o such that tariff structure* (1.30) *can induce an outcome that is CS-free. When* $\Delta_C = 0$*, tariff structure* (1.30) *can induce an outcome that is arbitrarily close to being CS-free.*

Proof: Presented in Appendix A.6.

Further, we can provide a closed form characterization of the solution to $\mathcal{P}_2$ when certain conditions hold:

Proposition 8 *The solution to* $\mathcal{P}_2$ *can be characterized in closed form when* $\Delta_C > 0$ *and* $l : d_l \leq d_i \forall i$ *is such that* $s_l^{(z^*)} = 0$*; that is, the lowest demand tier does not adopt solar.*

Proof: Presented in Appendix A.7.

Therefore, our proposed tariff structure (1.30) is guaranteed to be feasible, has desirable properties with respect to customer equity, and its solution can even be characterized in closed form when adopting solar creates a customer surplus.

4.3.1 Practical issues/extensions

Some practical hurdles stand in the way of implementing tariff structure (1.30). We now discuss these issues and how they may be remedied.

1. **Demand, generation and excess are not actually deterministic:** Our model assumes deterministic values for households h_i, demand d_i, generation capacity g_i, and excess e_i. However, this will not be true in practice: Households' usage and generation parameters are random variables. We can relax the deterministic assumption by allowing individual households to draw d, g, and e from a discrete distribution with I points in its support. We define a joint probability mass function $\phi(d, g, e)$ over the space of possible values of d, g, and e. Let us index the support of $\phi(\cdot)$ by i. Now, $h_i = H * \phi(d_i, g_i, e_i)$, where H is the total number of households in the market. In expectation h_i customers will have demand d_i, generation g_i and excess e_i. Thus, with this definition of h_i, we can reformulate programs $\mathcal{P}_1$ and $\mathcal{P}_2$ in terms of expectations, and h_i customers will make a decision consistent with the deterministic tier i decision.

2. **Not all customers belonging to a class that is induced to adopt solar will actually adopt:** For various reasons including access to liquidity and inertia to change, some customers belonging to a tier designated to adopt solar might not actually make the adoption decision. We can account for this behavior. Let π_i be an exogenously given expected fraction of households in tier i that would adopt solar if economically viable (endogenous determination of π_i is beyond the scope of this model). The remaining $1-\pi_i$ fraction of households in tiers $i : s_i^{(z^*)} = 1$ continue to fulfill their demand from the utility directly because they are unwilling to adopt solar. We can then reformulate our model by replacing h_i by $\pi_i h_i$ in the solar profit equations and adjusting the utility company's rate-of-return equation appropriately.

3. **The tariff structure may incentivize customers to exaggerate their demand to generate bill savings:** It is undesirable for a tariff structure to incentivize a type i customer to "spoof" a type j customer by exaggerating her demand in order to generate bill savings; that is, her total outflow (to U and S) is reduced by spoofing another class[7]. Therefore, we must consider the following four possible ways a customer type may spoof another customer type.

   (i) **Non-solar to non-solar:** Non-solar customer i can appear to be non-solar customer j by exaggerating her demand to $\tilde{d}_i = d_j > d_i$. However, doing so increases her bill from $p_r d_i$ to $p_r d_j > p_r d_i$, so she will not do so.

   (ii) **Non-solar to solar:** Non-solar customer i can appear to belong to a tier j that adopts solar by installing solar that generates g_i and appearing to have a grid usage of $d'_j = d_j - g_j + e_j$. To do so, she must alter her demand to $\tilde{d}_i = (d_j - g_j + e_j) + g_i - e_i$. This is undesirable if $\tilde{d}_i > d_i \Rightarrow (d_j - g_j + e_j) > (d_i - g_i + e_i)$. To prevent this from happening, we must ensure that $p_r d_i < p_r(\tilde{d}_i - g_i) + f_j + p_s g_i = p_r(d_j - g_j + e_j - e_i) + f_j + p_s g_i$. But we have that $p_r d_i < p_r(d_i - g_i) + f_i + p_s g_i$. Therefore, it is sufficient for us to specify parameters such that:

$$p_r(d_i - g_i) + f_i < p_r(d_j - g_j + e_j - e_i) + f_j. \tag{1.37}$$

   (iii) **Solar to non-solar:** Solar customer i can appear to be a non-solar customer j by altering her demand d_i to $\tilde{d}_i = d_j$. This is undesirable if $d_j > d_i$. To prevent this from happening, we must ensure that $p_r(d_i - g_i) + f_i + p_s g_i < p_r d_j$. But since tier i adopts, we have that $p_r(d_i - g_i) + f_i + p_s g_i \leq p_r d_i$. Since $d_j > d_i$, $p_r d_i < p_r d_j$, and therefore, such a customer i will not spoof a non-solar customer j.

   (iv) **Solar to solar:** Solar customer i can appear to be another solar customer j by appearing to have a grid usage of $d'_j = d_j - g_j + e_j$. To do so, she must alter her demand

[7] We do not discuss the case of a customer curtailing her demand in order to generate bill savings because we do not model demand response and the costs associated with curtailing demand.

to $\tilde{d}_i = (d_j - g_j + e_j) + g_i - e_i$. This is undesirable if $\tilde{d}_i > d_i \Rightarrow (d_j - g_j + e_j) > (d_i - g_i + e_i)$. To prevent this from happening, we must ensure that $p_r(d_i - g_i) + f_i + p_s g_i < p_r(\tilde{d}_i - g_i) + f_j + p_s g_i$, which is identical to the condition in inequality (1.37).

Such spoofing behaviors can be eliminated in various ways. For example, constraint (1.37) could be added to $\mathcal{P}_2$ for classes falling into (ii) and (iv) above; that is, for all classes i, j such that $d_j - g_j + e_j > d_i - g_i + e_i$ and either i is a non-adopter and j is an adopter, or both i and j are adopters. Whether $\mathcal{P}_2$ remains feasible under these constraints depends on the specific parameters under consideration. We show in Section 5 that $\mathcal{P}_2$'s feasibility and CS outcome is unaffected by the inclusion of these constraints for realistic parameter values. If the regulator chooses not to enforce this constraint in $\mathcal{P}_2$, she can use it to check if the solution obtained is exposed to such spoofing behavior. Alternatively, U could assign a class not just by measuring grid usage, but also by measuring net demand $d_i - g_i$ and assigning a class on the basis of both these measurements. Such a measurement would deter spoofing behavior, because except under pathological parameter values, simply exaggerating demand will not allow a type i customer to mimic a type j customer on both these dimensions.

4. **The tariff structure may incentivize customers to install solar capacity strictly less than g_i:** It is undesirable for this tariff structure to induce solar customers to install solar capacity smaller than g_i. Observe that if a tier i customer's generation is reduced to $\tilde{g}_i < g_i$, her excess also reduces to some $\tilde{e}_i \leq e_i$. This might or might not alter her tier (which, recall, is measured by measuring her grid usage $d - (g - e)$). Suppose it does not result in a tier alteration. In this case, we can impose the constraint $p_s < p_r$ so that installing less than capacity g_i will increase a customer's bill by an amount p_r per unit of generation foregone and this bill increase is not compensated for by having to purchase less from S. To ensure that $p_s < p_r$, we constrain f_m, the fixed cost of the marginal customer, to be larger than 0 so that $p_s = t(m) = p_r - f_m/g_m < p_r$.

   But what if the customer was able to change her tier, and therefore, the applied fixed cost by installing panels to less than capacity? Let us assume that for every possible $\tilde{g}_i < g_i$, R can compute a corresponding $\tilde{e}_i$ (the resulting excess) based on the customer's usage profile over the day. Such a customer can spoof a tier j customer by choosing a capacity $\tilde{g}_i$ such that $d_i - \tilde{g}_i + \tilde{e}_i = d_j - g_j + e_j$. For all such pairs of $\tilde{g}_i, \tilde{e}_i$ values, impose constraint (1.38), in addition to the constraint that $f_m > 0$.

$$p_r(d_i - g_i) + f_i + p_s g_i < p_r(d_i - \tilde{g}_i) + f_j + p_s \tilde{g}_i \tag{1.38}$$

   Note that we need not consider the possibility of a customer altering both d_i and g_i to exaggerate her grid usage d'_i: Decreasing g_i by one unit increases her cash outflow by $p_r - p_s$, while increasing d_i by one unit increases her cash outflow by p_r, which is larger than $p_r - p_s$.

We shall see in Section 5 that for realistic parameter values, $p_s < p_r$ because the prescribed fixed cost f_m for the marginal tier m is positive. We shall also see in Section 5 that for realistic parameter values, moving to a higher tier causes customers to incur higher fixed costs, and therefore, reducing g_i—even to effect a change in tier—is not beneficial. Therefore, there is no incentive for a customer to install less solar than her capacity g_i.

5 Numerical Analysis

Now, using data from the states of Nevada and New Mexico where regulatory changes threaten the rooftop solar industry, we study how the tariff structures in operation compare to the two-part tiered tariff structure presented in (1.30). We first discuss our approach to estimating the parameters $d, g, e,$ and h in Section 5.1. Using these parameters, in Section 5.2 we numerically investigate the performance of using the specific versions of the linear tariff structure proposed in the states of Nevada and New Mexico. We contrast these to results obtained if the states were to adopt our proposed tiered two-part tariff structure. As we will see, while both states' tariffs are able to feasibly generate the outcome z^* specified by $\mathcal{P}_1$, they induce poor customer equity outcomes in problem $\mathcal{P}_2$.

5.1 Estimating the parameters

Our starting point for estimating the parameters of the system is the household micro-data from the US Energy Information Administration's Residential Energy Consumption 2009 Survey [U.S. Energy Information Administration, 2009]. This survey contains responses from 112 single-family housing units in the states of Nevada and New Mexico (that appear as a single group in the data set). For each housing unit, we estimate the value g_i: We take the total area of each house in square feet, as reported in the data set, and divide by the number of stories to obtain an estimate of the total roof area. Then, we assume that each square foot of panel area can generate 9 Watts of electricity when the sun is shining [Solar-Estimate, 2012] to obtain the rated power output of solar panel installations. This estimate is an approximation of installed capacity because total household area reported in the survey also includes basements and attics, where they are present. Moreover, not all available roof space is typically usable for solar panels. Accordingly, we correct this estimate (as a first approximation) by a single multiplicative factor such that installation sizes so obtained are roughly in the 3 kWh - 10 kWh range [Fu et al., 2016]. We multiply this installation capacity by 2190, the estimated number of hours of sun in Nevada and by 2471 hours of sun for New Mexico to estimate the generation per year for households in each state [SolarDirect, 2016]. We group these households into 4 roughly equally sized buckets ($i = 1 \ldots 4$) based on their generation capacity. Within each bucket i, we compute the average energy demanded d_i and the average generation capacity g_i. Bucket size h_i is computed as the proportion of these households

Table 1.1: Estimated parameters for the state of Nevada

Generation Class (kWh)	Avg. Generation g_i (kWh)	Avg. Demand d_i (kWh)	Avg. Excess e_i (kWh)	% households h_i
0-6500	5272	8896	2636	24.11%
6500-8700	7653	10383	4577	25.00%
8700-11000	9907	13204	5995	24.11%
>11000	14483	14282	10252	26.78%

Table 1.2: Estimated parameters for the state of New Mexico

Generation Class (kWh)	Avg. Generation g_i (kWh)	Avg. Demand d_i (kWh)	Avg. Excess e_i (kWh)	% households h_i
0-7250	5778	8810	2765	21.42%
7250-9500	8326	9808	4971	24.12%
9500-12250	10947	13303	6397	26.78%
>12250	16212	14210	11352	27.68%

belonging to generation bucket i (we normalize the total number of households to 1). Since $\sum_i^I h_i$ is normalized to 1, Δ_S, Δ_C, and Δ_U are measured on a per-household basis.

Using the US Department of Energy data on hourly residential load in a typical meteorological year for cities in Nevada and New Mexico [US Department of Energy, 2013], we find that residences in Nevada typically consume about $\mu = 29\%$ of their demand between the hours of 11 a.m. and 6 p.m., typical hours for solar reliance. This figure is about $\mu = 34\%$ for New Mexico. Since we do not have this data broken up by household class, we assume that all houses consume this proportion of their demand when the sun is shining, and use this to estimate e_i, the excess generation they would be able to sell back to the grid as $\max(0, g_i - \mu d_i)$.

The results of this exercise are shown in Tables 1.1-1.2. We use these parameters in our analysis.

5.2 Comparison of tariff structures

We now wish to compare the tariff structures in operation in Nevada and New Mexico to our suggested two-part tiered tariff structure. We are particularly interested in understanding how these tariff structures perform with respect to their ability to induce CS-free outcomes. To do so, we set $\Delta_C = 0$ and compare tariff structures. (If $\Delta_C < 0$, then CS is unavoidable; as Δ_C increases above zero, CS-free outcomes become easier to induce.) We ignore the federal investment tax credit in our experiments because it will ramp down starting in 2019 [Energy.gov, 2017].

As noted in Section 4, not all customers who are financially incentivized to install solar do so; we attempt to capture this inertia in our experiments by letting π_i be the proportion of class i households that adopt solar if financially prudent. Equivalently, this may be interpreted as the probability that a class i household that is financially incentivized to adopt solar does so. We estimate a reference level of π_i for each class and state in our data set using data from the preliminary version of US Energy Information Administration's Residential Energy Consumption

Table 1.3: Estimated π values

Generation Class (i)	Nevada		New Mexico	
	Optimistic π_i	Pessimistic π_i	Optimistic π_i	Pessimistic π_i
1	4.05%	1.35%	3.51%	1.17%
2	4.15%	1.38%	3.81%	1.27%
3	3.83%	1.28%	3.33%	1.11%
4	3.35%	1.12%	3.01%	1.00%

2015 Survey [U.S. Energy Information Administration, 2015][8], which also has a flag indicating whether the surveyed households have installed rooftop solar. Specifically, we use data from the Pacific Census Region, where households have had significant federal and state incentives to adopt solar [Borlick and Wood, 2014].

We scale these π_i values to come up with values of π_i that reflect optimistic and pessimistic adoption scenarios. Under the optimistic scenario, if all classes adopt at the specified π_i levels, 3% of all residential energy, the current level in California, [California Distributed Generation Statistics, 2017, Ivan Penn, 2017] will be supplied by rooftop solar. Under the pessimistic scenario, if all classes adopt at the specified π_i levels, 1% of all residential energy will be supplied by rooftop solar. Table 1.3 shows the results of this exercise.

To study this situation with partial adoption within a class, we reformulate $\mathcal{P}_1$ so that $E^{(z)}$ values (this is the migration quantity corresponding to outcome z) accurately reflect the change in energy demanded from U given adoption levels π_i. Since $\pi_i < 1, \forall i$, we reformulate $\mathcal{P}_2$ as follows:

$$\min_{T(\cdot)} \quad \max_{i} \quad \left\{ T(d_i', e_i, 1) + p_s g_i - p_{r0}\, d_i,\, T(d_i, 0, 0) - p_{r0}\, d_i \right\}$$

Subject to constraints:

$$t(i) = \frac{T(d_i, 0, 0) - T(d_i', e_i, 1)}{g_i}, \forall i$$

$$\sum_{i=1}^{I} h_i \left(s_i^* \left(\pi_i T(d_i', e_i, 1) + (1 - \pi_i) \right) T(d_i, 0, 0) \right) + (1 - s_i^*)\, T(d_i, 0, 0)$$

$$-c_u^{z^*} \left(s_i^* \left(\pi_i (d_i - g_i) + (1 - \pi_i) d_i \right) + (1 - s_i^*) d_i \right) \Big) - \sum_{i=1}^{I} h_i (p_{r0} - c_u^{E_0})\, d_i = \Delta_U$$

$t(i)$ ordering consistent with o

$$(t(m(o)) - c_s) \sum_{i=1}^{I} s_i^{(z^*)} \pi_i h_i g_i = \Delta_S$$

$$(t(i) - c_s) \sum_{j=1}^{I} \mathbb{I}_{t(j) > t(i) \text{in ordering } o} \cdot \pi_j h_j g_j < \Delta_S, \forall i \neq m(o).$$

In our numerical study, we use a c_s value of \$0.059 for Nevada and \$0.052 for New Mexico. We

[8] The complete version of this data set is still not available.

compute this using a 30 year lifetime [SolarCity, 2016c] for solar systems that produce power at 200 Watts per panel [SolarCity, 2016b], with output degrading at a rate of 0.5% per year [Jordan et al., 2010] and a per Watt cost of $2.89 [SolarCity, 2016a].

Next, we turn to estimation of c_u^x. Using financial disclosures from NV Energy in Nevada [NV Energy, 2016a,b] and PNM Energy in New Mexico [PNM Energy, 2017], we estimate $c_u^{E_0}$ to be $0.035 in Nevada and $0.030 in New Mexico. Since natural gas and other fossil fuels are the major source of fuel in Nevada (81%) and New Mexico (66%), we estimate parameters for the post-solar case by assuming that rooftop solar generation displaces natural gas. We obtain the mix of energy sources used in Nevada from U.S. Energy Information Administration [2016a] and for New Mexico from U.S. Energy Information Administration [2016b]. In order to compute $c_u^{(z)}$), we use the levelized cost estimates of various sources of electricity from Lazard [2017], scaled by a multiplicative factor to make them consistent with $c_u^{E_0}$. In order to estimate the values of m_u^x, we use the lifecycle greenhouse gas emissions of different energy sources published in Intergovernmental Panel on Climate Change [2014] and a social cost of Carbon of $59.03 per metric ton of CO_2, which we obtain by correcting the 2030 social-cost-of-Carbon estimate of fifty 2007 dollars per metric ton of CO_2 (Environmental Protection Agency [2017]) for inflation. We use the 2030 value rather than the 2020 value in order to reflect the long-term nature of regulatory planning for rooftop solar; of the years for which estimates are available, 2030 falls closest to the middle of the typical 30 year useable lifetime of a solar panel.

We now individually consider the cases of Nevada and New Mexico. Since we are interested in studying customer equity outcomes, we set $\Delta_S = 0.0001$, and set Δ_U so that $\Delta_C = 0$ according to equation $\Delta_C = c_u^{E_0} E_0 - c_u^{E_0 - E^{(z)}}(E_0 - E^{(z)}) - c_s E^{(z)} - \Delta_S - \Delta_U$. We compare the performance of our tariff structure and the tariff structure in operation on both the pessimistic and the optimistic adoption scenarios. For each state, for each type of scenario, we carry out the following steps:

1. Enumerate the objective value of $\mathcal{P}_1$ for all 16 possible outcomes (each tier may either be incentivized to adopt or not adopt). Choose the outcome z^* with the best objective value.

2. Corresponding to this solution z^*, find the solution of $\mathcal{P}_2$ for all possible orderings o using the tariff structure in operation in the state. Choose the solution with the best objective value.

3. Corresponding to this solution z^*, find the solution of $\mathcal{P}_2$ for all possible orderings o using our suggested two-part tiered tariff structure, *including the constraints required to ensure that customers are not encouraged to spoof their demands or generation amounts to generate bill savings.* Choose the solution with the best objective value.

5.3 Nevada

For both the optimistic and pessimistic scenarios in Nevada, we find that the solution to $\mathcal{P}_1$ corresponds to all four classes of customers adopting solar. In the optimistic scenario, the objective

Table 1.4: Solutions - Nevada Optimistic Scenario

Two-part Tiered Tariff				
p_r	f_1	f_2	f_3	f_4
\$0.1018	\$225.355	\$327.132	\$423.481	\$619.084
Optimal ordering			**$\mathcal{P}_2$ Objective value**	
4, 1, 3, 2			$\$1.13291 \cdot 10^{-7}$	

Current Non-tiered Tariff				
r_d	s_d	s_e	r_0	s_0
\$0.0711	\$0.0711	-\$0.0545	\$360.3880	\$380.0950
Optimal ordering			**$\mathcal{P}_2$ Objective value**	
1, 3, 2, 4			\$87.3733	

Table 1.5: Bill Comparisons - Nevada Optimistic Scenario

		Two-part Tiered Tariff Outflows		Current Non-tiered Tariff Outflows	
Class	**Old Bill**	**Adopter**	**Non-adopter**	**Adopter**	**Non-adopter**
1	\$905.79	\$905.79	\$905.79	\$992.88	\$992.88
2	\$1,057.20	\$1,057.20	\$1,057.20	\$1,102.40	\$1,098.80
3	\$1,344.43	\$1,344.43	\$1,344.43	\$1,299.74	\$1,299.74
4	\$1,454.19	\$1,454.19	\$1,454.19	\$1,392.28	\$1,376.52

value of $\mathcal{P}_1$ is \$1.46, and for the pessimistic scenario the objective value is \$0.49. Recall that these are to be interpreted as normalized net welfare improvements on a per-household basis.

NV Energy in Nevada operates a non-tiered tariff structure. After the changes in 2015, NV Energy published a memo to its customers, providing an overview of the changes they would see in their bill [NV Energy, 2015]. According to these changes, solar customers would be compensated at less than retail rates for excess energy sold back to the grid and would pay different fixed costs from non-solar customers. However, there was to be no change to the energy charge (the per-unit) energy rate for solar and non-solar customers. Accordingly, we use tariff structure (1.24), restricting r_d to be equal to s_d. We use a value of $p_{r0} = \$0.10182$, the average value of the rates in Northern and Southern Nevada (NV Energy [2017]). The results of our experiments are summarized in Tables 1.4-1.7. We see that in both the optimistic and pessimistic scenarios, the objective value of $\mathcal{P}_2$ using the current non-tiered tariff is quite large—indicating the presence of CS—compared to that of our suggested tariff (where it is near 0). Therefore, this outcome constitutes a situation of CS. We see that tier 1 and 2 customers are adversely affected by a move to solar. Further, the most strongly impacted households are tier 1 customers, who are also potentially the poorest customers with smaller houses and lower demand.

5.4 New Mexico

As was the case with Nevada, for both the optimistic and pessimistic scenarios in New Mexico, we find that the solution to $\mathcal{P}_1$ corresponds to all four classes of customers adopting solar. In the optimistic scenario, the normalized objective value of $\mathcal{P}_1$ is \$1.97, and for the pessimistic scenario it is \$0.66.

Table 1.6: Solutions - Nevada Pessimistic Scenario

Two-part Tiered Tariff				
p_r	f_1	f_2	f_3	f_4
\$0.1018	\$225.354	\$327.132	\$423.480	\$619.0840
Optimal ordering			**$\mathcal{P}_2$ Objective value**	
3, 1, 4, 2			$\$1.16379 \cdot 10^{-7}$	

Current Non-tiered Tariff				
r_d	s_d	s_e	r_0	s_0
\$0.0606	\$0.0606	-\$0.0585	\$484.2210	\$486.7230
Optimal ordering			**$\mathcal{P}_2$ Objective value**	
1, 3, 2, 4			\$117.470	

Table 1.7: Bill Comparisons - Nevada Pessimistic Scenario

		Two-part Tiered Tariff Outflows		**Current Non-tiered Tariff Outflows**	
Class	**Old Bill**	**Adopter**	**Non-adopter**	**Adopter**	**Non-adopter**
1	\$905.79	\$905.79	\$905.79	\$1,023.36	\$1,023.36
2	\$1,057.20	\$1,057.20	\$1,057.20	\$1,113.93	\$1,113.48
3	\$1,344.43	\$1,344.43	\$1,344.43	\$1,284.45	\$1,284.45
4	\$1,454.19	\$1,454.19	\$1,454.19	\$1,351.76	\$1,349.78

New Mexico's net-metering regulations have also seen some recent opposition [Robert Walton, 2016]. As it stands, New Mexico operates a tiered retail net-metering tariff that does not allow discrimination between solar and non-solar households [NC Clean Energy Technology Center, 2017a]; the tariff structure treated in Section 4.2. In order to set up this tariff structure, recall that our estimates in Table 1.2 categorized customers in New Mexico into four tiers. We use values of $p_{r0} = \$0.088, \$0.090, \$0.097$, and $\$0.099$ corresponding to rates that would apply to tiers 1, 2, 3, and 4 respectively [PNM Energy, 2018]. For convenience, we choose rate class boundaries close to the average of the tier demand values d_i. Specifically, we let rate class 1 apply to customers who demand up to 9300 kWh, rate class 2 apply to customers who consume between 9300 and 11550 kWh, rate class 3 apply to customers who demand between 11550 and 13750 kWh, and rate class 4 apply to customers who demand more than 13750 kWh.

Now, we consider the modifications required to the tariff structure in the post-solar case. Recall that a class c customer's bill under this tariff structure is of the form $r_c n + f$, where n is the net demand of the customer being considered. From the estimates of d and g in Table 1.2, we compute the net demand $(d - g)$ values for all tiers, if they were to adopt. A tier 1 customer who adopts has a net demand of 3031 kWh, a tier 2 customer who adopts has a net demand of 1482 kWh, a tier 3 customer who adopts has a net demand of 2356 kWh, and a tier 4 customer who adopts has a net demand of -2001 kWh (tier 4 customers are net suppliers). These are very low usage levels. In order to give additional flexibility to the New Mexico tariff, we add a fifth rate class that applies to net demands of up to 5900 kWh. This fifth rate class will then apply to all households that adopt

Table 1.8: Rate Classes - New Mexico

	Class 1	Class 2	Class 3	Class 4	Class 5
Net demand (kWh)	0-5900	5900-9300	9300-11550	11550-13750	>13750
Rate	r_1	r_2	r_3	r_4	r_5

Table 1.9: Solutions - New Mexico Optimistic Scenario

Two-part Tiered tariff					
p_r	**Non-adopters**	f_1	f_2	f_3	f_4
		-\$99.9315	-\$91.6358	-\$31.1684	-\$4.8735
\$0.0993	**Adopters**	f_1	f_2	f_3	f_4
		\$171.5560	\$299.5730	\$483.1920	\$756.8710
Optimal ordering				**$\mathcal{P}_2$ Objective value**	
1, 2, 3, 4				$\$5.50725 \cdot 10^-7$	
Current Tiered Tariff					
r_1	r_2	r_3	r_4	r_5	f
\$0.0001	\$0.0340	\$0.0440	\$0.0430	\$0.06	555.56
Optimal ordering			**$\mathcal{P}_2$ Objective value**		
2,4, 3, 1			\$108.76		

solar[9]. Table 1.8 summarizes the rate classes under consideration.

We compare New Mexico's tiered tariff structure to a variant of our tariff structure discussed in Section 4.3, where we have fixed costs that vary by tier in order to accommodate the fact that p_{r0} varies by tier (please see footnote 5 in Section 4.3). We also impose constraints to make rates $r_i, \forall i$ positive. The results of our experiments are summarized in Tables 1.9-1.12. These show a similar situation to Nevada's where tiers 1 and 2 are affected. However unlike in Nevada, the most adversely affected customers are customers in tier 2.

In summary, although Proposition 2 showed that $\mathcal{P}_2$ might not be feasible under the non-tiered tariff structure (1.24), we did not encounter infeasibility in our experiments for Nevada. However, while feasible, the non-tiered linear tariff structures in Nevada and New Mexico performed poorly compared to our tiered tariff structure with respect to their ability to avoid CS: In Nevada, under both the pessimistic and optimistic scenarios, the customers most affected by the introduction of solar belonged to tier 1. These are the households with the smallest rooftops, possibly housing lower-income residents. In New Mexico, the most adversely affected customers were customers

[9] Adding a new rate class for each of the four net-demand usage levels will implicitly enable the New Mexico tariff to distinguish between solar and non-solar customers, making it equivalent to our tariff.

Table 1.10: Bill Comparisons - New Mexico Optimistic Scenario

		Two-part Tiered Tariff Outflows		**Current Tiered Tariff Outflows**	
Class	**Old Bill**	**Adopter**	**Non-adopter**	**Adopter**	**Non-adopter**
1	\$775.28	\$775.28	\$775.28	\$858.08	\$858.08
2	\$882.72	\$882.72	\$882.72	\$991.49	\$991.49
3	\$1,290.39	\$1,290.39	\$1,290.39	\$1,128.72	\$1,128.72
4	\$1,406.79	\$1,406.79	\$1,406.79	\$1,404.38	\$1,404.38

Table 1.11: Solutions - New Mexico Pessimistic Scenario

Two-part Tiered tariff					
p_r	**Non-adopters**	f_1	f_2	f_3	f_4
		-\$99.9315	-\$91.6358	-\$31.1684	-\$4.8735
\$0.0993	**Adopters**	f_1	f_2	f_3	f_4
		\$171.5560	\$299.5730	\$483.1920	\$756.8710
Optimal ordering				**$\mathcal{P}_2$ Objective value**	
1, 2, 3, 4				$\$5.50725 \cdot 10^{-7}$	
Current Tiered Tariff					
r_1	r_2	r_3	r_4	r_5	f
\$0.0001	\$0.0340	\$0.0440	\$0.0430	\$0.06	555.56
Optimal ordering			**$\mathcal{P}_2$ Objective value**		
2,4, 3, 1			\$108.76		

Table 1.12: Bill Comparisons - New Mexico Pessimistic Scenario

		Two-part Tiered Tariff Outflows		**Current Tiered Tariff Outflows**	
Class	**Old Bill**	**Adopter**	**Non-adopter**	**Adopter**	**Non-adopter**
1	\$775.28	\$775.28	\$775.28	\$858.08	\$858.08
2	\$882.72	\$882.72	\$882.72	\$991.49	\$991.49
3	\$1,290.39	\$1,290.39	\$1,290.39	\$1,128.72	\$1,128.72
4	\$1,406.79	\$1,406.79	\$1,406.79	\$1,404.38	\$1,404.38

in tier 2, which shows that no class of rate-payers can be assured of not being forced into cross-subsidization. In contrast, our tiered tariff structure was able to avoid CS even after we imposed constraints to prevent customers from exaggerating their demand to realize bill savings. These experiments serve to illustrate the crucial role that tariff structure plays in helping to ensure equity across customers.

6 Conclusions and Future Work

There has been considerable regulatory flux associated with rooftop solar energy in the past few years. The regulator's task of trading off the interests of the utility company, the solar company, different consumers and society at large is clearly a challenging one. In some cases regulators have made decisions with dire consequences: Regulatory changes introduced in the state of Nevada all but killed the solar industry in the state.

In this chapter, we study the regulator's problem of choosing a tariff function to induce a socially optimal outcome. The regulator takes into account financial and environmental considerations and operates in a setting with a monopolistic, price-setting solar company and customers who individually decide whether or not to install solar. We pose the regulator's decision as an optimization problem which we show can conveniently be hierarchically separated into two subproblems. We show analytically that the tariff structure chosen must have the ability to discriminate across customer usage tiers and the ability to discriminate between solar and non-solar customers in order to guarantee feasibility. We present a tariff structure with both these features and show that in addition to guaranteeing feasibility, it also can guarantee outcomes free from cross-subsidization when solar adoption generates a surplus for the customer base. The implication of our work is that utility regulators must migrate to tiered tariff structures and put solar and non-solar customers on different rate schedules in order to induce socially optimal outcomes. While some states such as Arizona already seem to be moving in this direction, most utility tariffs in the U.S. do not have *both* these features.

This work lays the foundation for several research extensions. For instance, our model considers a flat-rate pricing scheme, rather than a time-of-use (TOU) pricing scheme such as the one being rolled out in California. While TOU is yet to gain significant traction in the U.S., modeling this explicitly could eventually become critical to analyzing policy decisions related to solar, as having a TOU pricing scheme incentivizes customers to shift their demand profiles temporally (demand response). Capturing demand response would also modify our model's current assumption that the aggregate demand for electricity is unaffected by solar adoption. In addition, our work models the solar company as being monopolistic. Another possible avenue for future research is to explicitly model competition in the solar system marketplace, with the utility itself potentially being a competitor in the solar domain. Similarly, while our model implicitly captures the impact of utility capacity investments in response to solar adoption through parameters c_u^x and m_u^x, one could consider a model in which these investment decisions are explicit outputs of the model. And, on the customer side, we could modify the assumption that all customers (or an exogenously defined proportion) who are incentivized to adopt solar do so. An interesting extension would be to *endogenously* determine the proportion of customers in a tier that would adopt solar as a function of bill savings generated.

While this work explores a static setting, related questions can be explored in a dynamic setting, complementing existing literature such as Babich et al. [2017] and Lobel and Perakis [2011]. Such work would require a significantly different model that captures the diffusion of solar among customers, the interaction between solar adoption penetration and solar cost, and potentially utility capacity investment decisions. If such work also continued to capture heterogeneity in the customer base, this would necessitate more nuanced modeling of the regulator's optimization problem.

2

Evaluating the First-mover's Advantage in Announcing Real-time Delay Information

1 Introduction

Some services provide delay information to manage congestion by influencing their customers' patronage decisions; e.g., in call centers callers decide whether to continue to wait, call back later, or request a call-back based on the announced delays. Recent advances in internet-based technology have enabled customers to be informed about delay information at multiple competing service providers simultaneously—even before physical interaction with any of them—to decide which provider to patronize. For example, the Florida Hospital system publishes the real-time estimated delays at its 22 emergency rooms (ERs) on a smart-phone application [Florida Hospital, 2016], and the paid *ERtexting* service allows hospitals to text their expected delays to a central server that broadcasts the information to the community [Sadick, 2012]. Restaurants have also taken advantage of internet-based applications for delay information dissemination; for example, the application *NoWait*, available as a paid service to restaurants, disseminates the restaurants' expected time-to-seat [Perez, 2014], enabling diners to choose one of several restaurants based on their anticipated delays. *NoWait* has recently been acquired by and integrated into *Yelp* [Sawers, 2017].

When a service provider functions in *isolation*, the Operations Management literature has documented the advantages of delay announcements for the service provider and customers [e.g., Whitt, 1999]. Furthermore, settings with multiple service providers operating in a *network* setting, where *all* announce equally-rich delay information have received attention. For example, the implication of announcing real-time delay information on network synchronization has been studied in the context of ambulance diversion and emergency rooms [Deo and Gurvich, 2011, Dong et al., 2018]. However, service providers are often not identical in their propensity to adopt the required technology to disseminate real-time delay information. In the absence of real-time information for a service provider, customers often use historical information, typically reported in a moving average form [TechCrunch, 2015]. Therefore, it is not atypical that customers are required to make patronage decisions based on heterogeneously-rich delay information from competing providers.

In this chapter, we consider such a network setting where customers make patronage decisions based on (potentially) heterogeneously-rich delay information from two service providers with comparable services (but not necessarily equal service rates). We take the perspective of the service provider that is more tech-savvy; we call this provider the *technology leader* L. We study when L should invest in the infrastructure to make the first move in disseminating truthful[1] *real-time* delay in an attempt to attract more market share. The other service provider, the *follower* F, can opt to respond to L's announcement. We study the impact on *competition* rather than network *coordination* (or, *synchronization*). We cast our model as a combination of a sequential game and continuous-time Markov chains (CTMCs). We choose a sequential game structure, rather than a simultaneous one, because of two observations:

1. Not all service providers are equally willing to invest in new technology and infrastructure, as evidenced by the remarkably low penetration rate of such technology among service providers. Only 12% of restaurants consider themselves to be technology leaders. Our own exploration on the *Yelp* app shows that only 5 out of the 383 restaurants within a six-block radius of *Yelp*'s downtown office in San Francisco have *Yelp*'s *NoWait* infrastructure in place to disseminate real-time delay information (see Fig. 2.1). Similar adoption heterogeneity is also prevalent among Emergency Departments: In San Jose, only two of five urgent care centers announce *real-time* delay information, while the other three only provide *historical average* delay (updated monthly or yearly, for example) [HospitalStats, 2017].

2. This modeling choice allows us to also provide insights to firms that sell infrastructure for delay-information dissemination (like *NoWait*); is it possible for such firms to efficiently target one of the service providers and allow market competition to induce the other service provider to also adopt their technology without the extra marketing effort and cost?

We summarize our findings, from a combination of analytical derivations and numerical experiments, as follows:

[1]The backlash from systematic information misrepresentation could result in loss of customer goodwill and even legal actions; O'Donnell [2014] reports about the repercussions of delay information falsification at a Veterans' Administration Hospital.

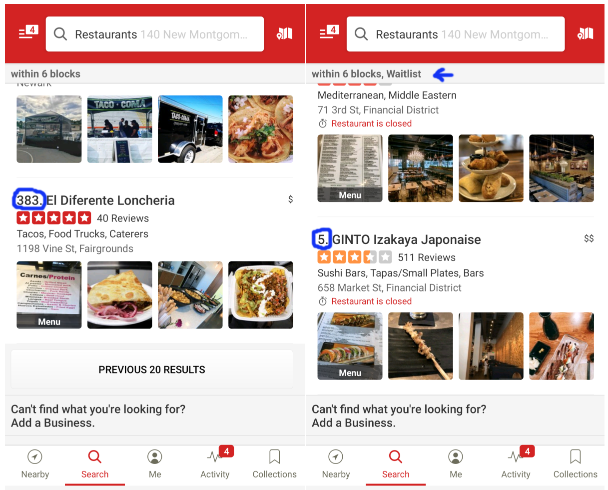


Figure 2.1: Penetration of real-time delay announcement infrastructure is not significantly high

- L should initiate delay announcements only if she is not the higher-capacity service provider as it improves her market share in equilibrium. If L is the higher-capacity service provider, initiating delay announcements would result in market share loss.

- When L initiates delay announcements in equilibrium, F's optimal response is mixed; for some range of system load, F's optimal response is to follow L in announcing real-time delay, but for some intermediate system loads, it is optimal for F not to respond.

- From the perspective of a delay-announcement infrastructure firm such as *NoWait*, it is prudent to target the lower-capacity service provider.

- Social welfare, measured by average customer delay, improves in equilibrium when L initiates delay announcement, and remains unchanged otherwise.

2 Literature Review

Ibrahim [2018] provides a comprehensive review of the delay announcement literature, including various delay estimators. The queue length (QL) estimator (*expected delay* $\approx$ *queue length* $\times$ *average service time*) is commonly used in Markovian first-come-first-served (FCFS) systems, because of its computational simplicity and accuracy [Ibrahim and Whitt, 2009a]. Extensions are available for systems with customer abandonment, priority service, and time-varying arrivals [Ibrahim and Whitt, 2009b, Jouini et al., 2009, Ibrahim and Whitt, 2011, Jouini et al., 2015]. Other estimators include delay of the last customer to enter service, head-of-the-line customer, most recently served customer, and data-driven estimators [Ibrahim and Whitt, 2009a,b, Ibrahim et al., 2015, Ang et al., 2015]. We employ the QL estimator in our analysis.

The impact of delay information provision has been studied extensively in *single-service provider* settings. Whitt [1999] and Armony and Maglaras [2004], among others, document the benefits of QL announcements to improve wait times and system utilization when customers use them to make balk/join decisions. Armony et al. [2009] model the equilibrium balk/join behavior of utility-maximizing customers in a multi-server queueing environment, and Akşin et al. [2016] show empirically that callers to a call center react to longer delay announcements by abandoning earlier. Hassin [1986] shows that it could be socially suboptimal for a service provider to suppress her QL delay. Jouini et al. [2011] show that firms can control congestion by announcing different percentiles of the waiting time distribution. Guo and Zipkin [2007] identify conditions under which broadcasting more precise information could degrade system performance and customer experience; therefore, a self-interested firm may intentionally provide vague information [Allon et al., 2011].

We consider a *multi-service provider* (or, a *network*) setting. A related example in the literature is the impact of ambulance diversion, in which ERs can request diversion of ambulances to other hospitals during overcrowding periods, on network coordination. Considering decentralized threshold diversion policies, Enders [2010] establishes the optimality condition for the "never divert" policy, and Deo and Gurvich [2011] establish the Pareto optimality of the defensive equilibrium in which the ERs are always on diversion. Do and Shunko [2015] propose a centralized threshold policy that is Pareto improving over the decentralized policy. He and Down [2009], Ramirez-Nafarrate et al. [2014], and Dong et al. [2018] show that delay announcements improve network synchronization and customers' wait times even when a small proportion of customers decide based on the delay information.

The paper most relevant to our work is Hassin [1996], which models two gas stations with equal service rates on a highway where drivers only observe the nearer station's queue, and infer the farther station's expected delay conditioned on the expected delay at the nearer station. He concludes that the station with the observable queue always attracts more demand, and thus, it helps to be the server with the observable queue when your competitor's queue is unobservable. Altman et al. [2004] extend the model to heterogeneous service rates; they hypothesize that the emerging equilibrium is not always of threshold type [unlike in Hassin, 1996]. They support this hypothesis using a mixture of numerical and analytical arguments.

In Hassin [1996] and Altman et al. [2004], the service providers do not have the choice of revealing/hiding their delay information; the congestion level is observable for one service provider and not for the other one. In contrast, we allow the service providers to choose their best delay announcement strategy as a lever for competitive advantage in a sequential game setting. We also allow for unequal service capacities. Notably, we find that it is the difference in service capacities that drives the players' choices. Further, customers in Hassin [1996] and Altman et al. [2004] are assumed to be sophisticated enough to compute the conditional expected delay at the service provider with the unobservable queue, given the state of the observable queue, in *real time*; this requires customers to know the exact operating parameters of the service providers and perform complex equilibrium calculations. We model customers as naturally less sophisticated; when only one service provider announces real-time delay information, customers are not able to exactly infer the expected congestion level of the unobservable (non-announcing) service provider. Instead, they construct and periodically update a belief about her historical delay (for example, through previous service encounters), or they can access her moving average delay (for example, through published reports).

Dong et al. [2018] is a recent work that explores the operational implications of delay announcement in a network. They focus on the impact on network coordination, and their numerical and simulation results indicate that the level of achievable coordination depends on several factors, including heterogeneity in service capacities. They also study the impact of different delay estimating methods (moving averages with different time windows), however, they do not consider heterogeneous granularity of delay information among the service providers; all service providers employ delay estimators with similar settings. In a similar vein, Pender et al. [2018] investigate the impact of announcing moving average delay to customers on the resulting queueing dynamics; as with Dong et al. [2018], all service providers employ the same delay announcement strategy. Using a fluid-model approximation, Pender et al. [2018] show analytically that providing moving average announcements can result in oscillations in the delay. This is similar to the oscillations empirically recorded in Dong et al. [2018].

Unlike Dong et al. and Pender et al. [2018], we investigate a network where the *competition* effect is dominant, and therefore, we investigate the impact of delay announcements on *market shares*. Furthermore, as mentioned before we allow for the service providers to have different delay announcement policies.

3 Model Setup

We consider a system with two single-server service providers, L and F, with service capacities $\mu^{(L)}$ and $\mu^{(F)}$. we do not impose any relationship between $\mu^{(F)}$ and $\mu^{(L)}$. Therefore, the leader L may be the faster or the slower of the two providers. L and F offer comparable services and compete for market share. Delay-sensitive customers arrive according to a Poisson distribution with rate Λ. For system stability, we assume:

Assumption 1 $\Lambda < \mu^{(L)} + \mu^{(F)}$.

Customers patronize the service provider with whom they expect a shorter queue delay[2], which in the absence of delay announcements, results in respective effective *status-quo* arrival rates $\lambda_0^{(L)}$ and $\lambda_0^{(F)}$ ($\Lambda = \lambda_0^{(L)} + \lambda_0^{(F)}$). We denote the status-quo average queue delays and market shares by $D_0^{(i)}$ and $M_0^{(i)} = \lambda_0^{(i)}/\Lambda$. Given customers' delay-minimizing behavior, we endogenously determine $\lambda_0^{(L)}$ and $\lambda_0^{(F)}$ such that $D_0^{(L)} = D_0^{(F)}$; i.e., routing based on customers' beliefs about the service providers' delays leads to equal delays; this corresponds to the *Wardrop equilibrium* [Hassin, 2016, p. 207], and matches the routing mechanism in Hassin [1996, p. 623] when both queues are unobservable.

[2]Alternatively, we could model customers as making patronage decisions based on sojourn time. These results would be similar.

Without loss of generality, we scale time such that $\mu^{(L)} = 1$. Considering exponential service times and no delay announcements from the two service providers, L and F are independent stable $M/M/1$ queues with expected delays,

$$D_0^{(L)} = \frac{\lambda_0^{(L)}}{1-\lambda_0^{(L)}} \quad \text{and} \quad D_0^{(F)} = \frac{\lambda_0^{(F)}}{\mu^{(F)}\left(\mu^{(F)} - \lambda_0^{(F)}\right)}. \tag{2.1}$$

By setting $D_0^{(L)} = D_0^{(F)}$, we derive $M_0^{(L)}$ as in (2.2), and the status-quo market share for F follows $M_0^{(F)} = 1 - M_0^{(L)}$. Replacing the two status-quo market shares in (2.1), we derive the status-quo delay as in (2.3).

$$M_0^{(L)} = \begin{cases} \dfrac{1+\mu^{(F)2} + \Lambda\left(1-\mu^{(F)}\right) - \sqrt{\left(1+\mu^{(F)2} + \Lambda\left(1-\mu^{(F)}\right)\right)^2 - 4\Lambda\left(1-\mu^{(F)}\right)}}{2\Lambda\left(1-\mu^{(F)}\right)} & \mu^{(F)} \neq 1, \\ \dfrac{1}{2} & \mu^{(F)} = 1, \end{cases} \tag{2.2}$$

$$D_0^{(L)} = D_0^{(F)} = \begin{cases} \dfrac{1+\mu^{(F)2} - \Lambda\left(1+\mu^{(F)}\right) - \sqrt{\left(1+\mu^{(F)2} - \Lambda\left(\mu^{(F)}-1\right)\right)^2 + 4\Lambda\left(\mu^{(F)}-1\right)}}{2\mu^{(F)}\left(\Lambda-\mu^{(F)}-1\right)} & \mu^{(F)} \neq 1, \\ \dfrac{\Lambda}{2-\Lambda} & \mu^{(F)} = 1. \end{cases} \tag{2.3}$$

We refer to the status-quo setting, in which neither of the service providers broadcasts real-time delay, as Regime 0 or Period 0.

3.1 First Mover's Game

Service provider L, the *L*eader, is considering announcing real-time delay information. Being the leader, L might expect to accrue immediate short-term benefits. However, the potential short-term benefits (if any) might not sustain in the long term as F, the *F*ollower, could respond after realizing (or anticipating) a potential market share loss. Therefore, L must account for F's possible response to evaluate the long-term effects of real-time delay announcements.

We model L's decision as a sequential game. We demarcate two information regimes that arise. In Regime 1, F opts not to respond to L, and therefore, only L announces real-time delay information. In Regime 2, F responds to L with a time lag after which both L and F announce real-time delay information. Fig. 2.2, in which $M_{R_1}^{(i)}$ and $M_{R_2}^{(i)}$, $i \in \{L, F\}$ represent the long-term market shares in Regimes 1 and 2, depicts the game structure: L anticipates F's best response if she opts to initiate real-time delay announcements and compares the resulting market share to her status-quo market share $M_0^{(L)}$. We consider market share as the main decision factor in the game as it influences profitability directly. We investigate the decision effects on customer delay as well, but we exclude it from the game to avoid the need to impose a specific cost function to account for its indirect effect on profitability. A similar approach is taken in Hassin [1996], for example.

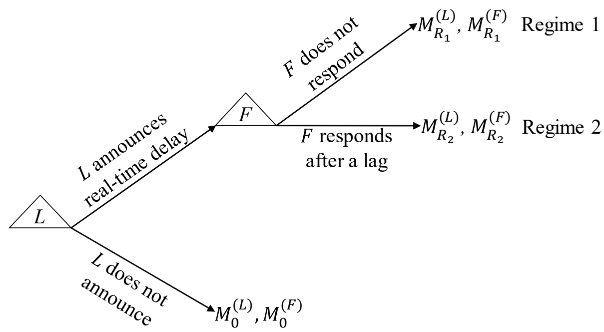


Figure 2.2: First mover's sequential game

3.2 Customer Patronage Decisions and CTMCs under Regimes 1 and 2

We detail the customer patronage decision models and the resulting CTMCs for service providers L and F to analyze their market shares under Regimes 1 and 2 in Sections 3.2.1-3.2.2.

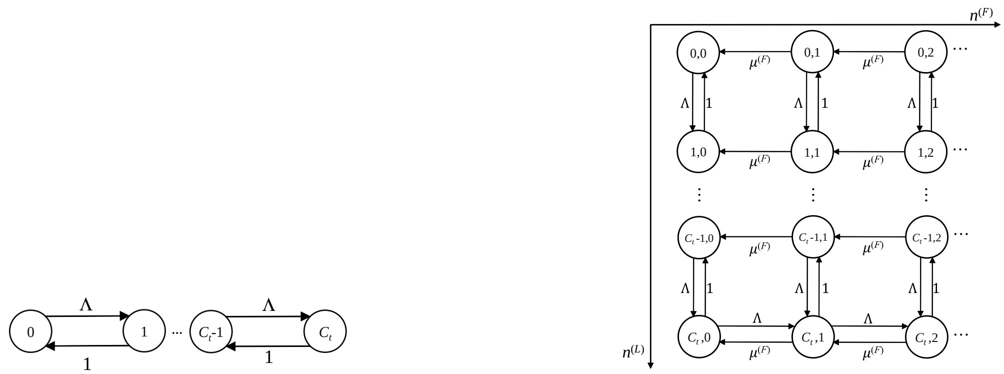
3.2.1 Regime 1 Model

Under this information regime, L announces real-time QL delay: The announced expected delay to a customer is $d_n = n^{(L)}/\mu^{(L)} = n^{(L)}$ (as $\mu^{(L)} = 1$) where $n^{(L)}$ is L's current number of customers. Since F does not announce real-time delay, customers make patronage decisions based on their beliefs of F's delay (or F's historical average delay, if available).

We consider the notion of *updating periods*: Customers' beliefs (or knowledge) about F's *average delay* are updated at the end of each period, assuming the periods are long enough for system stationarity (for example, ERs' published delays are based on annual averages [Groeger et al., 2014]). We index the updating periods by t, and we refer to the initial state of the system (the status quo) as Period 0. We denote the effective arrival rate, market share, and average delay of service provider i in Period t by $\lambda_t^{(i)}$, $M_t^{(i)}$, and $D_t^{(i)}$, respectively.

In any Period $t \geq 1$, customers make joining decisions based on the available information; they join L if $d_n \leq D_{t-1}^{(F)}$ and join F otherwise. We break ties in favor of L taking into account her more reliable and up-to-date information. This induces a threshold structure for the arrivals: When $n^{(L)} < C_t = \left\lfloor D_{t-1}^{(F)} + 1 \right\rfloor$, where $\lfloor\ \rfloor$ is the floor function, the arrival rate to L is Λ and to F is 0; otherwise, the arrival rate to L is 0 and to F is Λ. Figs. 2.3(a)-2.3(b) present the resulting CTMCs. In each period, we analyze Model L to compute L's effective arrival rate, which relies on $D_{t-1}^{(F)}$ (found in the previous period), and to compute Period t market shares. Then, we analyze Model

F to compute F's average delay in Period t, which is used by the customers to make Period $t+1$ decisions.



(a) Model L **(b)** Model F

Figure 2.3: CTMCs in Period $t \geq 1$

Model L is a birth-death process with the states representing $n^{(L)}$. Model F's state space is two-dimensional ($\{(n^{(L)}, n^{(F)}) : n^{(L)} = 0, \ldots, C_t; n^{(F)} = 0, 1, \ldots\}$) as we need to keep track of both $n^{(L)}$ and $n^{(F)}$ to determine F's effective arrival rate. The transitions from a general state $(n^{(L)}, n^{(F)})$ follow:

- Service completion at L when $n^{(L)} > 0$, with rate $\mu^{(L)} = 1$, resulting in a transition to $(n^{(L)} - 1, n^{(F)})$,
- Service completion at F when $n^{(F)} > 0$, with rate $\mu^{(F)}$, resulting in a transition to $(n^{(L)}, n^{(F)} - 1)$,
- Arrival to L when $n^{(L)} < C_t$, with rate Λ, resulting in a transition to $(n^{(L)} + 1, n^{(F)})$,
- Arrival to F when $n^{(L)} = C_t$, with rate Λ, resulting in a transition to $(C_t, n^{(F)} + 1)$.

3.2.2 Regime 2 Model

Under this information regime, both L and F announce real-time delay. The CTMC in this Regime is akin to a Join-the-shortest-queue (JSQ) system, wherein a customer compares the queue length-based delay estimates at L and F (i.e., $n^{(L)}$ vs. $n^{(F)}/\mu^{(F)}$), and chooses the service provider with the shorter expected delay. Fig. 2.4 presents the CTMC when $\mu^{(F)} = 2$. The transitions from a general state $(n^{(L)}, n^{(F)})$ follow:

- Service completion at L when $n^{(L)} > 0$, with rate $\mu^{(L)} = 1$, resulting in a transition to $(n^{(L)} - 1, n^{(F)})$,
- Service completion at F when $n^{(F)} > 0$, with rate $\mu^{(F)}$, resulting in a transition to $(n^{(L)}, n^{(F)} - 1)$,

- Arrival to L when $n^{(L)} < n^{(F)}/\mu^{(F)}$, with rate Λ, resulting in a transition to $(n^{(L)}+1, n^{(F)})$,
- Arrival to F when $n^{(L)} > n^{(F)}/\mu^{(F)}$, with rate Λ, resulting in a transition to $(n^{(L)}, n^{(F)}+1)$,
- When $n^{(L)}\mu^{(F)} = n^{(F)}$, an arriving customer chooses a service provider randomly, resulting in a transitions to $(n^{(L)}, n^{(F)}+1)$ and $(n^{(L)}+1, n^{(F)})$, each rate $\Lambda/2$. We choose this tie-breaking rule because customers in our model are sensitive to delay; accordingly, equal delay announcements should result in equal arrival rates at both providers.

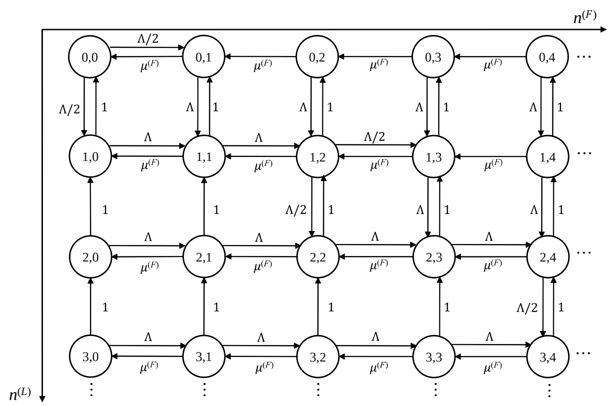


Figure 2.4: CTMC for Model LF in Period $t \geq 1$ when $\mu^{(F)} = 2$

4 Analysis

Before providing the full analysis of the models under Regimes 1 and 2 in Sections 4.1 and 4.2, we first provide a high-level summary of the two sections below:

For Regime 1, we first evaluate L's short-term market share, i.e., her market share immediately after she initiates real-time delay announcements. We then analytically characterize parameter settings under which the short-term behavior persists in the long term. We also present analytical results towards characterizing long-term market shares for more general parameter settings. Based on a conjecture supported by exhaustive numerical trials, we obtain closed-form bounds for long-term market shares under Regime 1.

For Regime 2, we carry out our analysis by constructing two CTMCs that, we conjecture and validate numerically, give us bounds on the market shares of the two service providers.

4.1 Analysis of Regime 1

We first analyze Period 1 (the period immediately following L's announcement initiation) under Regime 1; we refer to this as short-term analysis. We then build on the short-term results to analyze the long-term market shares.

Let π_i denote the limiting probability of state i in Model L (Fig. 2.3(a)). By solving the balance equations, we can derive the probabilities π_i, and then L's market share in Period 1 as follows:

$$M_1^{(L)} = \frac{\Lambda \sum_{i=0}^{C_1-1} \pi_i}{\Lambda} = \frac{\Lambda^{C_1} - 1}{\Lambda^{C_1+1} - 1}, \tag{2.4}$$

where $C_1 = \left\lfloor D_0^{(F)} + 1 \right\rfloor$, and $D_0^{(F)}$ follows (2.3). We have the following proposition for the short term:

Proposition 9

- *If L is the lower-capacity service provider* ($\mu^{(L)} < \mu^{(F)}$), $M_1^{(L)} > M_0^{(L)}$.
- *If service capacities are equal* ($\mu^{(F)} = \mu^{(L)}$), $M_1^{(L)} \geq M_0^{(L)}$.
- *If L is the higher-capacity service provider* ($\mu^{(L)} > \mu^{(F)}$), $M_1^{(L)}$ *may be larger or smaller than* $M_0^{(L)}$*, depending on the value of* Λ.

All proofs are presented in Appendix B.1.

The following characterization of market shares for extremely low and high values of load (proved in Appendix B.1.4) is also useful to develop intuition for the underlying mechanics that govern market shares in Regime 1.

$$M_1^{(L)} \rightarrow \begin{cases} 1 & \text{when } \Lambda \rightarrow 0, \\ M_0^{(L)} = \dfrac{1}{1+\mu^{(F)}} & \text{when } \Lambda \rightarrow 1+\mu^{(F)}. \end{cases} \tag{2.5}$$

When $\Lambda \rightarrow 0$, $C_1 = 1$ and since traffic is low, L is very rarely in state $C_1 = 1$, attracting almost all customers. At the other extreme, when $\Lambda \rightarrow \mu^{(L)} + \mu^{(F)}$, $M_1^{(L)} \rightarrow M_0^{(L)}$; in this case, L and F serve at capacity and their expected delays approach ∞ in both periods ($\lambda_0^{(L)} = \lambda_1^{(L)} \rightarrow \mu^{(L)}$)

The crux of our long-term analysis involves understanding how C_t evolves over time. First, we investigate whether we can obtain insight about L's long-term market share *analytically* via an analysis of Period 2. This Period 2 analysis requires analytical characterization of Model F in Period 1, which we are able to accomplish for selected parameter settings. This enables us to analytically characterize two of the patterns in the evolution of C_t, and hence Model F that could emerge in Period 2: instability and convergence. The onset of each of these two patterns in Period 2 allows us to predict the evolution of C_t in the following periods, and hence the long-term effects of L's announcements in Regime 1.

Onset of instability in Period 2. This occurs when L's announcements in Period 0 causes F to experience instability in Period 1. Note that instability can arise even when the whole system is stable.

Service provider F becomes unstable in Period 1 if and only if $\lambda_1^{(F)} \geq \mu^{(F)}$, where $\lambda_1^{(F)} = \Lambda - \lambda_1^{(L)}$ (see Appendix B.1.2). This directly implies the following proposition.

Proposition 10 *F becomes unstable in Period 1 if and only if* $\lambda_1^{(F)} = \Lambda - \frac{\Lambda^{C_1+1} - \Lambda}{\Lambda^{C_1+1} - 1} \geq \mu^{(F)}$*, where* $C_1 = \left\lfloor D_0^{(F)} + 1 \right\rfloor$*, and* $D_0^{(F)}$ *follows* (2.3)*.*

We noted in Proposition 9 that L could lose market share in the short term when $\mu^{(F)} < \mu^{(L)}$. Since F is stable in Period 0, F becomes unstable in Period 1 only if L loses too much market share. On the other hand, Proposition 9 also notes that L always gains market share when $\mu^{(F)} \geq \mu^{(L)}$. Therefore, the following sufficient condition guarantees F's stability.

Proposition 11 *When* $\mu^{(F)} \geq \mu^{(L)}$*, F is stable in Period 1.*

For the remainder of the chapter, we focus on cases where L and F remain stable in all periods.

Onset of convergence in Period 2. When $C_2 = C_1$, the systems in Periods 1 and 2 are identical, and therefore, $C_t = C_2 = C_1, \forall t > 2$, implying that the system converges. Analytical characterization of the parameters under which convergence occurs in Period 2 would inform us of the conditions under which the market share expression in (2.4) (and therefore the findings of Proposition 9) continues to hold in the long term. Unfortunately, Model F (Fig. 2.3(b)) is analytically intractable for an arbitrarily high number of phases (a general C_1). But we are able to derive closed-form analytical expressions for C_2 when $C_1 = 1$ or 2 by solving a simultaneous non-linear system of equations. We use these expressions to provide sufficient conditions for convergence in Period 2. We present the details in Appendix B.2. (This procedure remains valid for $C_1 > 2$, but it results in higher-order simultaneous non-linear systems of equations, which cannot be solved in closed form.) We obtain the following explicit condition for convergence in Period 2.

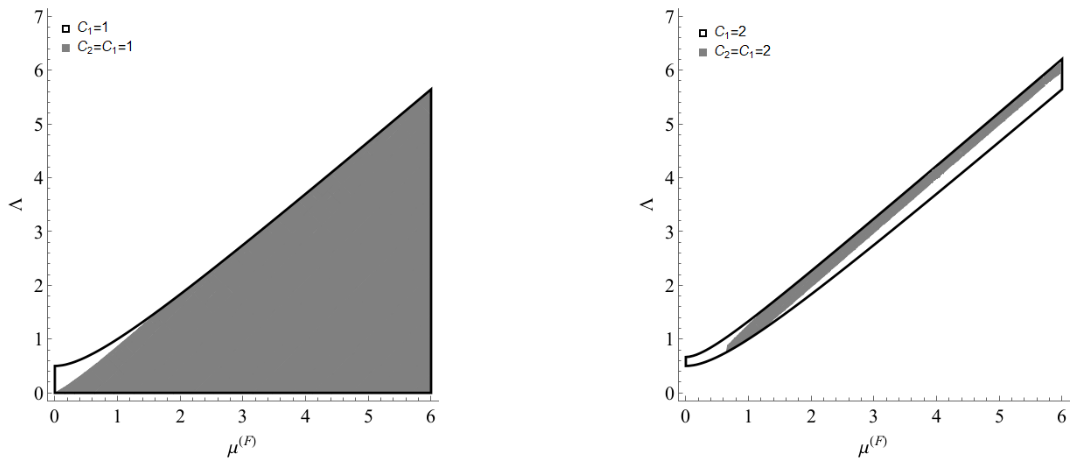
Proposition 12 *Convergence occurs at* $C_1 = C_2 = 1$ *in Period 2 if*

$$\Lambda < \frac{1}{2} \min \left\{ \mu^{(F)} \frac{{\mu^{(F)}}^2 - 1 + \sqrt{\left(\mu^{(F)} + 1\right)\left(\mu^{(F)} + 5\right)}}{\mu^{(F)} + 1}, \frac{2{\mu^{(F)}}^2 + \mu^{(F)} + 1}{\mu^{(F)} + 1} \right\}.$$

A similar convergence condition relating Λ *and* $\mu^{(F)}$ *exists for convergence at* $C_1 = C_2 = 2$*. This condition is extremely large and unwieldy to present.*

Figs. 2.5(a)-2.5(b) plot the convergence conditions in Period 2 for $C_1 = C_2 = 1$ and $C_1 = C_2 = 2$, respectively. The regions with thick black boundaries specify the parameters for which $C_1 = 1$ (Fig. 2.5(a)) and $C_1 = 2$ (Fig. 2.5(b)). The shaded regions specify the parameters for which the convergence conditions in Period 2 hold, and therefore, any findings we obtain for the short term will continue to be valid in the long term, i.e., for some parameter settings, we can characterize the long-term market shares in Regime 1 analytically.

This leaves open the question of what happens under more general parameters. To address this gap, we need to have an understanding of the long-term behavior of C_t. Towards that goal, we prove the following relationship between threshold C_t and the delay $D_t^{(F)}$ at F:



(a) Convergence to $C_t = 1$ **(b)** Convergence to $C_t = 2$

Figure 2.5: Parameter settings that result in convergence in Period 2

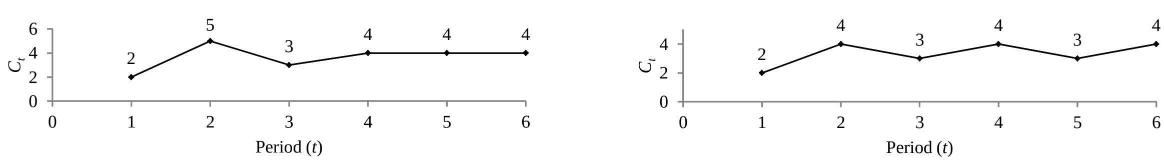
Lemma 2 $D_t^{(F)}$ *is non-increasing in* C_t.

Lemma 2 allows us to make a powerful assertion about the long-term behavior of the system for general parameters.

Proposition 13 *As long as L and F remain stable, only* ***two*** *possible long-term patterns are possible in the evolution of* C_t*:*

1. *Convergence, which arises when* $\exists t > 1 : C_t = C_{t-1}$*, and therefore,* $C_{t'} = C_t, \forall t' > t$.
2. *Stable oscillation, which arises when* $\exists t > 2 : C_t = C_{t-2}$ *and* $C_t \neq C_{t-1}$*, resulting in* $C_{t'} = C_{t'-2}, \forall t' \geq t$.

To illustrate Proposition 13, Fig. 2.6 shows representative examples of these long-term patterns. Fig. 2.6(a) shows the evolution of C_t for $\Lambda = 0.89$ and $\mu^{(F)} = 0.5$. In this example, C_t converges to 4 in Period 5. Fig. 2.6(b) shows the evolution of C_t for $\Lambda = 0.87$ and $\mu^{(F)} = 0.5$. In this example, C_t's stable oscillation between 3 and 4 is established in Period 4.



(a) Example of convergence **(b)** Example of stable oscillation

Figure 2.6: Representative examples of long-term patterns of C_t

In addition to Proposition 13, we need to answer the following questions to fully characterize the evolution of C_t: (i) Which of these two outcomes occurs for which values of Λ and $\mu^{(F)}$? (ii) What value(s) of C_t does the system stabilize at? The following conjecture, based on exhaustive numerical experiments with 10000 randomly chosen values of Λ and $\mu^{(F)}$, provides answers to both questions:

Conjecture 1 *Defining* $C_\infty = \lfloor \inf_{C_t} D_t^{(F)} + 1 \rfloor$*, one of the following occurs:*

1. *C_t converges to C_∞, or*
2. *C_t converges to $C_\infty + 1$, or*
3. *C_t oscillates between C_∞ and $C_\infty + 1$.*

From Lemma 2, we know that $D_t^{(F)}$ is non-increasing in C_t. Therefore, $\inf_{C_t} D_t^{(F)}$ is the tightest possible lower bound to the value of $D_t^{(F)}$ at the highest possible value of C_t; $D_t^{(F)}$ approaches this value asymptotically as C_t is increased. Accordingly, we can obtain C_∞ by computing the limiting value of $D_t^{(F)}$ as C_t approaches ∞.

Based on Conjecture 1 we can analytically obtain upper and lower bounds on the long-term market share of F in closed-form, which yield lower and upper bounds, respectively, for L's market share.

Proposition 14 *Under Conjecture 1, we have:*

$$M_{R_1}^{(F)} \leq \overline{M}_{R_1}^{(F)} = \begin{cases} \dfrac{\Lambda^2 - \Lambda}{\Lambda^2 - 1}, & \text{if } \Lambda < \dfrac{2{\mu^{(F)}}^2 + 3\mu^{(F)} + 2}{2\mu^{(F)} + 2}, \\ \dfrac{\Lambda^{\lfloor \omega+2 \rfloor} - \Lambda^{\lfloor \omega+1 \rfloor}}{\Lambda^{\lfloor \omega+2 \rfloor} - 1}, & \text{otherwise.} \end{cases} \tag{2.6}$$

$$M_{R_1}^{(F)} \geq \underline{M}_{R_1}^{(F)} = \begin{cases} \dfrac{\Lambda^{\lfloor \Omega_1+3 \rfloor} - \Lambda^{\lfloor \Omega_1+2 \rfloor}}{\Lambda^{\lfloor \Omega_1+3 \rfloor} - 1}, & \text{if } \Lambda < 1, \\ \dfrac{\Lambda^{\lfloor \Omega_2+3 \rfloor} - \Lambda^{\lfloor \Omega_2+2 \rfloor}}{\Lambda^{\lfloor \Omega_2+3 \rfloor} - 1}, & \text{if } \Lambda > 1, \end{cases} \tag{2.7}$$

where the details of the derivation and the expressions for ω, Ω_1 and Ω_2 are presented in Appendix B.1.7.

Propositions 12 and 14, along with the similar results that we derive for Regime 2 in Section 4.2, allow us to make analytical comparisons of the market shares across the different information regimes and determine the outcome for some instances of the game in Fig. 2.2. However, our ability to assert the game outcome based on these results is limited: Proposition 12 applies to a restricted region of the parameter space and Proposition 14 only provides *bounds* on the market shares. Therefore, we supplement our analysis using a set of comprehensive numerical experiments, described in Section 5.1.

4.2 Analysis of Regime 2

To analyze Regime 2, we need to analyze Model *LF* (Fig. 2.4). As noted in Section 3.2.2, Model *LF* is a JSQ-like system with asymmetric service rates. This system is analytically intractable and the exact characterization of its stationary probabilities, even for the symmetric case, is unknown. For the symmetric case, some work has been done on finding provable bounds for the stationary probabilities (e.g., Halfin [1985]), but these techniques do not extend to the asymmetric case. In fact, the asymmetric JSQ has received only limited attention: Existing literature focuses on numerical techniques that accurately approximate the chain's stationary probabilities and relevant metrics (e.g., Adan et al. [1990] and Selen et al. [2016]).

As our focus is finding the market shares in Regime 2, we do not require all stationary probabilities in Model LF; if we simply know the probability that one of the servers is idle, we can find the market shares. For ease of exposition, we set the service rate of the slower service provider (this could be L or F) to 1. Let Q denote the faster server with service rate, effective arrival rate, and idle probability of $\mu^{(Q)} > 1$, $\lambda^{(Q)}$, and $I^{(Q)}$, respectively. For Q as a work-conserving server, we have $\lambda^{(Q)} = (1 - I^{(Q)})\mu^{(Q)}$ from flow balance. Knowing $\lambda^{(Q)}$, we can find $M^{(Q)} = \lambda^{(Q)}/\Lambda$, and the market share of the slower service provider as $1 - M^{(Q)}$. However, even analytically characterizing $I^{(Q)}$ is not possible. Instead, we propose two models and conjecture that they provide lower and upper bounds on $I^{(Q)}$. We analytically solve these models to obtain conjectured analytical lower and upper bounds on $M^{(Q)}$ (that we verify numerically): obtaining a lower (upper) bound $M^{(Q)}$ is equivalent to obtaining an upper (lower) bound for $I^{(Q)}$.

Lower bound for $M^{(Q)}$: To get a lower bound for $M^{(Q)}$, we modify Model LF to allow centralized queueing: an arrival joins a centralized queue when both servers are busy, joins the empty server when only one server is idle, and joins randomly with equal probability to each server when both are idle. Fig. 2.7 shows the Markov chain for this modified model (Model $LF1$), which is similar to an $M/M/2$ queue except that the service rates of the two servers are different. Therefore, we need to distinguish whether the busy server is the faster server (state $(1, 0)$) or the slower server (state $(0, 1)$) when system size is one.

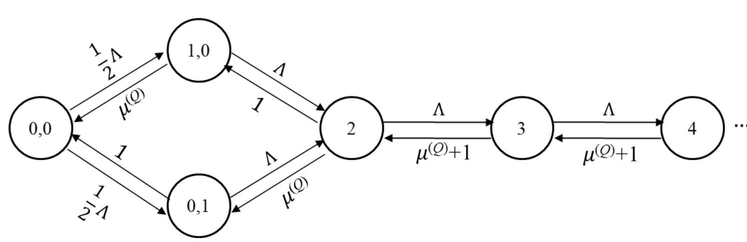


Figure 2.7: Model $LF1$ results in a lower bound for $M^{(Q)}$

We conjecture that owing to the pooling effect and more efficient use of servers in Model $LF1$, the faster server in Model $LF1$ is less busy compared to Model LF. We analytically derive $I^{(Q)}$ in Model $LF1$ to obtain a conjectured lower bound on $M^{(Q)}$.

Upper bound for $M^{(Q)}$: To get an upper bound for $M^{(Q)}$, we modify Model LF to allow for jockeying: when the faster server finishes a service and there is no other customer in the central queue, we move the customer being served by the slower server to the faster server. Fig. 2.8 shows the Markov chain for this modified model (Model $LF2$), with the state variables being defined identically to those in Model $LF1$. We intuitively conjecture that this jockeying procedure makes the faster server less idle. We analytically derive $I^{(Q)}$ in Model $LF2$ to obtain a conjectured upper bound on $M^{(Q)}$.

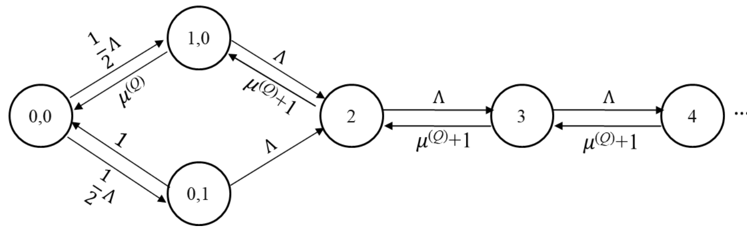


Figure 2.8: Model $LF2$ results in an upper bound for $M^{(Q)}$

Based on the derived $I^{(Q)}$ for Models $LF1$ and $LF2$, we make the following conjecture, which we validate based on an exhaustive set of 10000 experiments, with randomly generated values of $\mu^{(Q)}$ and Λ:

Conjecture 2 *The Market share for Q is bounded as:*

$$M_{R_2}^{(Q)} \geq \underline{M}_{R_2}^{(Q)} = \frac{\mu^{(Q)}(\Lambda\mu^{(Q)} + \mu^{(Q)} + 1)}{\Lambda + \mu^{(Q)}((\Lambda + 2)\mu^{(Q)} + 2)}. \quad (2.8)$$

$$M_{R_2}^{(Q)} \leq \overline{M}_{R_2}^{(Q)} = \frac{\mu^{(Q)}(\Lambda(3\mu^{(Q)} + 2) + \mu^{(Q)} + 1)}{2\Lambda^2 + \Lambda(\mu^{(Q)}(3\mu^{(Q)} + 2) + 1) + 2\mu^{(Q)}(\mu^{(Q)} + 1)}. \quad (2.9)$$

As was the case for Regime 1, our ability to predict the outcome of the game using Conjecture 2 is limited. We supplement our analysis with numerical experiments, described in detail in Section 5.1.

5 Game Outcome

In Section 4, we described how to analyze the long-term market shares under Regimes 1 and 2. For Regime 1, we provided an analytical characterization of the long-term market shares under some parameter settings (Proposition 12), and provided analytical support that—taken together with Conjecture 1—yields lower and upper bounds for the long-term market shares (Proposition 14). For Regime 2, we conjectured lower and upper bounds for the market shares (Conjecture 2).

As noted in Section 4, these analytical characterizations give us a limited ability to assert the outcome of the game. We therefore obtain a comprehensive understanding of the game through numerical analysis.

5.1 Numerical Experiment Setup

For Regime 1, we generate 2200 experiments by varying $\mu^{(F)}$ over the set $\{1/6, 1/5, \ldots, 1/2, 1, 2, \ldots, 6\}$ and 200 equally-spaced values of Λ in the range $\left(0, 1 + \mu^{(F)}\right)$. Model F has a repeating structure, and therefore, Matrix-analytic methods can be employed to compute stationary probabilities to an arbitrary degree of accuracy. We showed in Proposition 11 that F is stable in Period 1 when $\mu^{(F)} \geq \mu^{(L)}$; our experiments confirm that F will remain stable in the following periods as well. However, F could become unstable for some values of $\mu^{(F)} < \mu^{(L)}$ and Λ. This occurs in only 4 of the 1000 experiments, leaving us with 996 valid experiments.

For Regime 2, the experimental setting is similar to Regime 1 with a small modification to minimize the impact of the tie-breaking rule. To modify our experiments, we define a negligibly small irrational $\delta > 0$, and perform experiments with $\mu^{(F)}$ varying over sets $S_1 = \{1/6 + \delta, 1/5 + \delta, \ldots, 1/2 + \delta, 2 + \delta, \ldots, 6 + \delta\}$ and $S_2 = \{1/6 - \delta, 1/5 - \delta, \ldots, 1/2 - \delta, 2 - \delta, \ldots, 6 - \delta\}$. Adding or subtracting an irrational δ prevents $n^{(F)}$ from being an integer multiple of $\mu^{(F)}$. Therefore, the experiments in set S_1 favor F, while the experiments in S_2 favor L. When both service providers are empty, ties continue to be broken using a 50-50 rule.

In order to draw conclusions that are agnostic of the tie-breaking rule, we make our comparisons against experiments from S_1 and S_2. For instance, for the Regime 1 experiment with $\Lambda = 1$ and $\mu^{(F)} = 2$, we make a comparison with (i) the corresponding experiment in S_1, with $\Lambda = 1$ and $\mu^{(F)} = 2 + \delta$; and (ii) the corresponding experiment in S_2, with $\Lambda = 1$ and $\mu^{(F)} = 2 - \delta$.

Our modifications lead to a total of $2200 \times 2 - 400 = 4000$ experiments. (We exclude experiments for $\mu^{(F)} = 1$ as in this case the market will be equally divided between L and F.)

To compute stationary probabilities for Model LF, we truncate its state space along both dimensions. We denote the truncation bound for $n^{(L)}$ by κ, and we set the truncation bound for $n^{(F)}$ (the number of customers at F) to $\lceil \mu^{(F)} \rceil \times \kappa$ (recall that F is $\mu^{(F)}$ times as fast as L). To find the appropriate truncation bound, we initially set $\kappa = 10$, and compute the stationary probabilities of the corresponding truncated chain. For a given desired accuracy level $\epsilon > 0$, we increase κ and iterate the procedure if the sum of the probabilities of the boundary states (φ) is more than ϵ. We stop this procedure as soon as $\varphi < \epsilon$.

Table 2.1: Summary of numerical experiment settings for Regimes 1 and 2

	Regime 1 Experiments	**Regime 2 Experiments**	
		S_1	S_2
$\mu^{(F)} > 1$	1000	1000	1000
$\mu^{(F)} = 1$	200	Not Applicable	
$\mu^{(F)} < 1$	996	1000	1000

Table 2.1 summarizes our experimental settings; the values in the cells represent the number of valid experiments for the corresponding setting. We use these experiments to make pairwise comparisons between the information regimes: Regime 1 vs. 0 (Section 5.2), Regime 2 vs. 0 (Section 5.3), and Regime 1 vs. 2 (Section 5.4). We summarize the findings of Sections 5.2-5.4 in Table 2.2. In Section 5.5, we synthesize these results to find the outcome of the sequential game. In Section 5.6, we demonstrate that the analytical characterizations provided in Section 4 imply

the correct outcome of the sequential game for a large region of the parameter space.

Table 2.2: Summary of pairwise comparison results from Sections 5.2-5.4

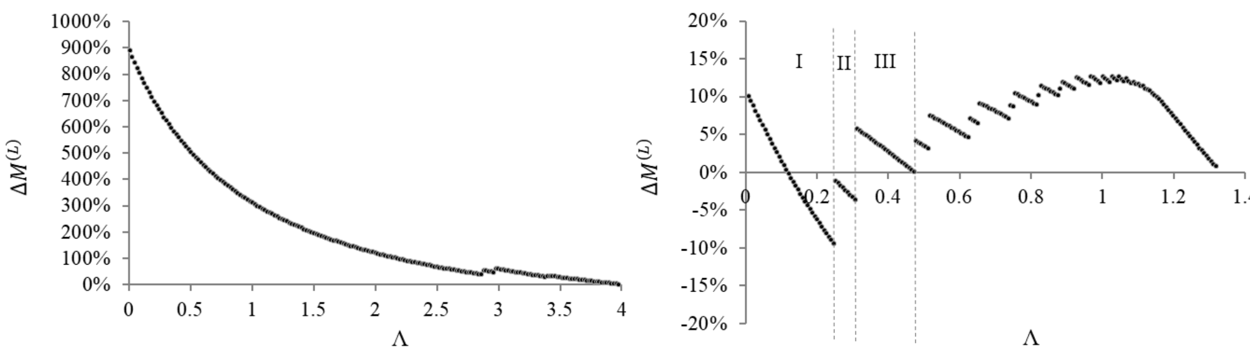
	Regime 1 vs. Regime 0	**Regime 2 vs. Regime 0**	**Regime 2 vs. Regime 1**
	L **prefers**		F **prefers**
$\mu^{(F)} > 1$	Regime 1	Regime 2	Mixed
$\mu^{(F)} = 1$	Regime 1	Indifferent	Regime 2
$\mu^{(F)} < 1$	Mixed	Regime 0	Regime 2

5.2 Regime 1 vs. Regime 0

Proposition 9 states that L's market share (weakly) increases in the short-term when she is not the higher-capacity service provider, but if L is the higher-capacity service provider, she may gain or lose market share in the short term. Proposition 12 characterizes a subset of the parameter space for which the short-term behavior, and hence the findings of Proposition 9, sustain in the long term. Our numerical experiments indicate that the findings of Proposition 9 hold more generally. For 100% (all 1200) experiments in which L is not the higher-capacity service provider, L's market share in the long-term is (weakly) higher than her status-quo market share ($M_{R_1}^{(L)} \geq M_0^{(L)}$). Among the experiments in which L is the higher-capacity service provider, we have that $M_{R_1}^{(L)} \geq M_0^{(L)}$ in 85% (846/996 experiments).

The intuition behind these results is as follows. In Regime 0, the higher-capacity service provider enjoys a disproportionate (relative to her service capacity) market share advantage in the absence of any delay announcements: L's relative service capacity is $1/(1+\mu^{(F)})$, and we can verify from Eq. (2.2) that $M_0^{(L)} \geq 1/(1+\mu^{(F)})$. This can be seen as a 'benefit-of-doubt' advantage, i.e., the faster service provider enjoys the advantage of attracting a significantly high market share when customers are only informed about *average* delay. To see this, note that if L serves twice as fast as F, and receives arrivals at twice the rate F receives them, her customers experience a lower average delay than F's do. This is immediate from the formula of delay in an $M/M/1$ queue with arrival rate λ and service rate μ, i.e., $Delay = \frac{\lambda/\mu}{\mu-\lambda}$; therefore, the way that customers split in Regime 0 tends to favor the higher-capacity service provider in the absence of richer delay information. Similarly, when L is the lower-capacity service provider, she has a significant disadvantage in Regime 0.

In Regime 1, there are two counteracting forces that govern L's market share position relative to Regime 0: the positive effect of signalling states in which she is relatively empty, and the negative effect of signalling states in which she is relatively full. The net effect of real-time announcements depends on load Λ in a non-monotone fashion. Figs. 2.9(a)-2.9(b) plot the percentage increase in L's long-term market share in Regime 1 (compared to Regime 0) for two values of $\mu^{(F)}$. As Fig. 2.9(a) shows, the net effect of real-time announcements is positive when L is the lower-capacity provider. However, as Fig. 2.9(b) shows, the net effect is mixed if L is the higher-capacity provider, depending on Λ. The discontinuities (which are the underlying reason for the non-monotonicity) in the plots corresponding to changes in the long-term equilibrium value(s) of C_t. These changes form "regions," each corresponding to a particular long-term pattern of C_t. In each region, C_t either converges to a specific value or oscillates between two values (Proposition 13). For example, in Fig. 2.9(b), C_t converges to 1 in region I, oscillates between 1 and 2 in region II, and converges to 2 in region III. The regions corresponding to oscillatory behavior are smaller: they represent the transition between two convergence regions. *Within* a region, we note that $\Delta M^{(L)}$ decreases with increasing Λ. This is intuitive: at the lowest value of Λ corresponding to the region, there is a unit jump in the long-run value of C_t, allowing L to attract a chunk of customers. As Λ increases within the same region, the value of C_t does not increase, tilting the balance in favor of F.



(a) $\mu^{(F)} = 3$: L serves one-third as fast as F **(b)** $\mu^{(F)} = 1/3$: L serves thrice as fast as F

Figure 2.9: L's market share gain in Regime 1 vs. Regime 0

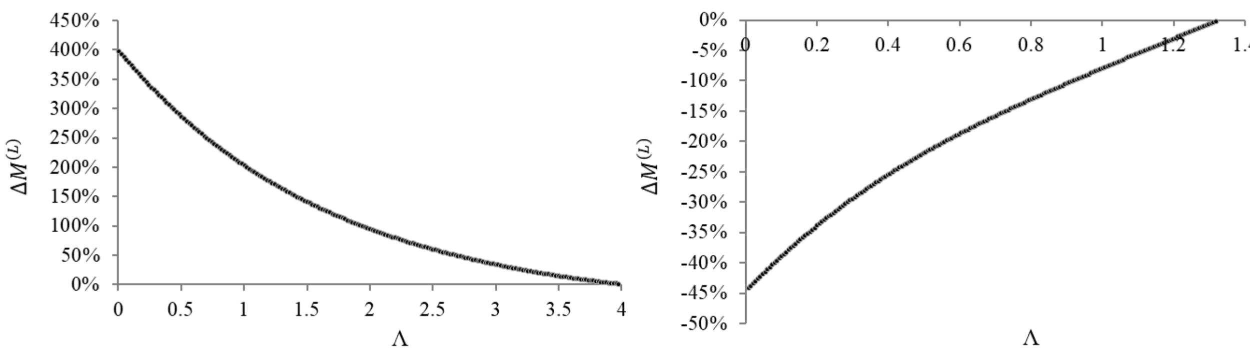
We observe in all our experiments that for very low values of load Λ, Regime 1 is significantly preferable to Regime 0 for *all* relative service rates. The system is almost always empty for very low values of Λ; therefore, an arriving customer who initiates a busy period is likely to observe an announcement of zero delay from L, which is necessarily smaller than F's (finite, non-zero) steady-state average delay. Since the threshold C_t up to which L receives arrivals is at least 1, this arriving customer chooses to route to L. Since the arrival rate Λ is very low, the customer is likely to finish service at L, and finish a busy period, before the next customer arrives. Therefore, L commands the majority of the market in this situation.

5.3 Regime 2 vs. Regime 0

In Regimes 2 and 0, providers L and F employ symmetric announcement strategies. Intuitively, given the Regime 0 advantage enjoyed by the higher-capacity service provider (as discussed in Section 5.2), we expect this advantage to be eroded in Regime 2, where the lower-capacity server is able to explicitly signal the states where she is relatively empty.

This intuition turns out to be true. For 100% (all 2000) experiments in which L is the higher-capacity provider ($\mu^{(F)} < 1$), we find that $M_{R_2}^{(L)} \leq M_0^{(L)}$. Similarly, for 100% (all 2000) experiments in which F is the lower-capacity provider ($\mu^{(F)} > 1$), we have that $M_{R_2}^{(L)} \geq M_0^{(L)}$. When service rates are equal, the providers split the market evenly in both regimes.

In Figs. 2.10(a)-2.10(b), we plot the percentage increase in L's market share in Regime 2 (compared to Regime 0) for two values of $\mu^{(F)}$. The market share benefit (loss) is monotone in Λ, with the strongest effect at the lowest load, and the weakest effect at the highest load. At very low loads, the higher capacity provider claims most of the market in Regime 0, while the two providers claim roughly equal market shares in Regime 2 because both systems are almost always empty and ties are broken arbitrarily. Therefore, at very low loads, the lower-capacity provider enjoys a strong advantage from moving to Regime 2. At very high loads, both providers serve almost at capacity in both regimes, and therefore, claim the same market shares in both regimes. As expected, the benefit (loss) plots in Fig. 2.10 are smoothly varying in load (in contrast to the discontinuous behavior in the Regime 1 vs. Regime 0 comparison discussed in Section 5.2), because there are no sources of discontinuity (such as the periodic updates in Regime 1) in the Regime 2 model. The scale difference in the vertical axes between the figures is attributed to a different denominator for computing percentage change.



(a) $\mu^{(F)} = 3 + \delta$: L serves ∼one-third as fast as F **(b)** $\mu^{(F)} = 1/3 - \delta$: L serves ∼thrice as fast as F

Figure 2.10: L's market share gain in Regime 2 vs. Regime 0

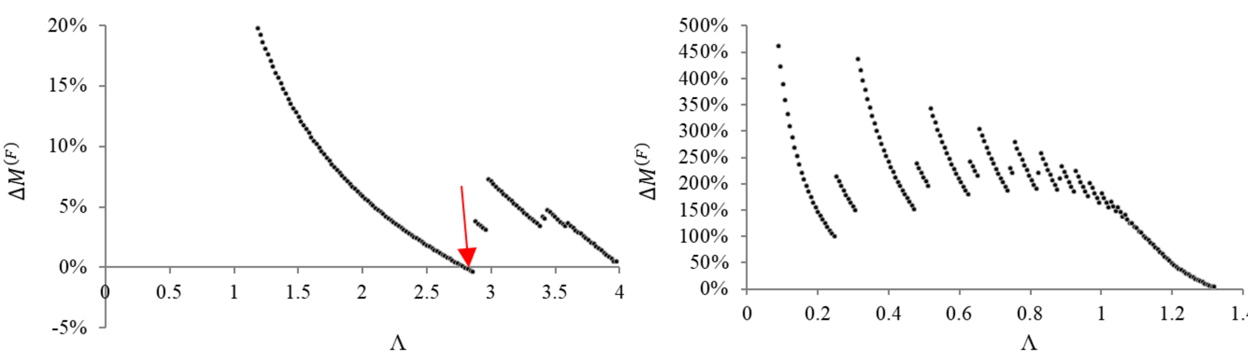
As an aside, we note in all experiments that regardless of relative service capacities, L's (and by symmetry, F's) average delay in Regime 2 is shorter than her Period 0 delay: She experiences the benefits of being able to signal to customers when she is relatively more congested than F. When the service rates are equal, this result has already been documented in the literature [see Hordijk and Koole, 1990].

5.4 Regime 1 vs. Regime 2

When L initiates announcements, does F prefer to respond or not? Again, our answer depends on the relative service capacities. In 100% (all 1992, including experiments in S_1 and S_2) of experiments in which F is the lower-capacity service provider, we have $M_{R_2}^{(F)} \geq M_{R_1}^{(F)}$. For 100% (all 200) of experiments in which F and L have equal capacities, $M^{(F)}_{R_2} \geq M^{(F)}_{R_1}$. However, when F is the higher-capacity service provider, we obtain mixed results that are (at a high level) consistent between experiments in S_1 (in which F is favored in Regime 2) and S_2 (in which L is favored in Regime 2). Among the experiments in set S_1, we find that $M^{(F)}_{R_2} \geq M^{(F)}_{R_1}$ in 96% (964/1000 experiments). Among the experiments in set S_2, we find that $M^{(F)}_{R_2} \geq M^{(F)}_{R_1}$ in 93% (940/1000 experiments). As can be expected based on the findings reported in 5.2, F prefers not to respond for some intermediate values of Λ, because of the unit jumps in the underlying threshold C_t that governs Regime 1 behavior.

It is intuitive that in most cases, F prefers to respond to L's announcements by initiating real-time announcements as well: in Regime 1, she only receives *overflow* customers, i.e., customers who arrive to the system when provider L's congestion is at threshold C_t. In contrast, in Regime 2, F is explicitly signals states in which she is empty. However, this comparison is made non-trivial by the fact that her market share in Regime 1 is determined by complex forces: the amount of time spent per visit to L's threshold C_t, which is easy to characterize, and the amount of time between visits to state C_t, which is difficult to characterize because it depends on the long-run behavior of the value(s) of C_t.

In Figs. 2.11(a)-2.11(b), we plot the percentage increase in F's market share in Regime 1 (compared to Regime 2) for two values of $\mu^{(F)}$. We truncate the vertical axis appropriately for ease of exposition. As is evident, there is a small region in Fig. 2.11(a) where F is actually worse off in Regime 2 than in Regime 1.



(a) $\mu^{(F)} = 3 + \delta : L$ serves ∼one-third as fast as F **(b)** $\mu^{(F)} = 1/3 - \delta : L$ serves ∼thrice as fast as F

Figure 2.11: F's market share gain in Regime 2 vs. Regime 1. The arrow in Fig. 2.11(a) indicates values of Λ for which $M^{(F)}_{R_2} < M^{(F)}_{R_1}$.

5.5 Synthesis

We now synthesize the findings in Sections 5.2-5.4 to find the outcome of the sequential game in Fig. 2.2.

Outcome for $\mu^{(F)} < \mu^{(L)}$: In Section 5.4, we find that F prefers Regime 2 to Regime 1 when she is the lower-capacity service provider, and will therefore respond to L initiating delay announcements. Therefore, the relevant comparison for L to make is between her market shares in Regime 2 and Regime 0. From the results in Section 5.3, a higher capacity service provider always prefers Regime 0 to Regime 2. Therefore, it is in L's best interests not to initiate delay announcements at all, as doing so would ultimately cause an erosion in her market-share if F responded, and she would forego her prevailing market share advantage. It is worthy of note, however, that by not announcing delay information, higher-capacity service providers are foregoing the benefits of improved delays that would be obtained if both L and F announced delay information.

Outcome for $\boldsymbol{\mu}^{(F)} \geq \boldsymbol{\mu}^{(L)}$: In Section 5.4, we find that F prefers Regime 2 to Regime 1 (she responds to L's delay announcements) for most (but not all) values of load when she is not the lower-capacity service provider. In spite of this mixed result for F's response, we find that L's action is consistent: L always has the incentive to initiate real-time delay announcements when she is not the higher-capacity service provider. This is because $M_{R_2}^{(L)} \geq M_0^{(L)}$ if F responds, and $M_{R_1}^{(L)} > M_0^{(L)}$ if F does not respond. As a bonus, this action also improves L's delay performance, as discussed in Section 5.3.

5.6 Analytical Support

Our numerical experiments suggest that L's optimal action follows:

- If she is the *higher-capacity* service provider, she *should not* initiate delay announcements
- If she is *not the higher-capacity* service provider, she *should* initiate delay announcements

Under what parameter settings is this optimal action of L implied by our analytical results from Section 4? To answer this question, we use:

- The closed-form expressions for Regime 0 market shares; $M_0^{(i)}, i \in \{L, F\}$, given in Eq. (2.2).
- The conjectured lower and upper bounds for Regime 2 market shares; $\underline{M}_{R_2}^{(i)}$ and $\overline{M}_{R_2}^{(i)}, i \in \{L, F\}$, given in Conjecture 2.
- The short-term market shares given in Eq. (2.4) for cases where convergence is established in Period 2 (indicated by $\mathbb{I}_C = 1$), based on Proposition 12: $M_{R_1}^{(i)}, i \in \{L, F\}$. For other cases, the lower and upper bounds for Regime 1 market shares: $\underline{M}_{R_1}^{(i)}$ and $\overline{M}_{R_1}^{(i)}, i \in \{L, F\}$ given in Proposition 14.

Outcome for $\boldsymbol{\mu}^{(F)} < \boldsymbol{\mu}^{(L)}$: When L is the higher-capacity service provider, our analytical results imply that L will not initiate delay announcements when either of the following conditions holds:

(i) L's market share in Regime 0 is higher than the highest market share she can obtain under either Regime 1 or Regime 2:

$$M_0^{(L)} > \max\left\{\overline{M}_{R_2}^{(L)}, \mathbb{I}_C M_{R_1}^{(L)} + (1 - \mathbb{I}_C)\overline{M}_{R_1}^{(L)}\right\}$$

(ii) L's market share in Regime 0 is higher than (a) the upper bound on her Regime 2 market share if F prefers Regime 2 to Regime 1; or (b) the upper bound on her Regime 1 market share if F prefers Regime 1 to Regime 2:

$$M_0^{(L)} > \begin{cases} \overline{M}_{R_2}^{(L)} & \text{if } \underline{M}_{R_2}^{(F)} > \mathbb{I}_C M_{R_1}^{(F)} + (1-\mathbb{I}_C)\overline{M}_{R_1}^{(F)} \\ \mathbb{I}_C M_{R_1}^{(L)} + (1-\mathbb{I}_C)\overline{M}_{R_1}^{(L)} & \text{if } \mathbb{I}_C M_{R_1}^{(F)} + (1-\mathbb{I}_C)\underline{M}_{R_1}^{(F)} > \overline{M}_{R_2}^{(F)} \end{cases}$$

Outcome for $\boldsymbol{\mu^{(F)} \geq \mu^{(L)}}$: When L is not the higher-capacity service provider, our analytical results imply that L will initiate delay announcements when either of the following conditions holds. These conditions can be interpreted analogously to the conditions presented above.

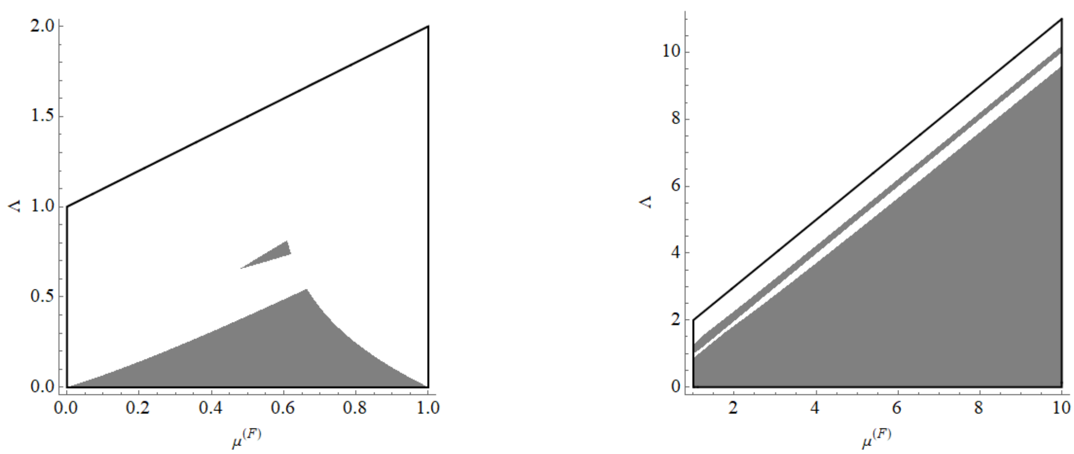
(i)

$$M_0^{(L)} < \min\left\{\underline{M}_{R_2}^{(L)}, \mathbb{I}_C M_{R_1}^{(L)} + (1-\mathbb{I}_C)\underline{M}_{R_1}^{(L)}\right\}$$

(ii)

$$M_0^{(L)} < \begin{cases} \underline{M}_{R_2}^{(L)} & \text{if } \underline{M}_{R_2}^{(F)} > \mathbb{I}_C M_{R_1}^{(F)} + (1-\mathbb{I}_C)\overline{M}_{R_1}^{(F)} \\ \mathbb{I}_C M_{R_1}^{(L)} + (1-\mathbb{I}_C)\underline{M}_{R_1}^{(L)} & \text{if } \mathbb{I}_C M_{R_1}^{(F)} + (1-\mathbb{I}_C)\underline{M}_{R_1}^{(F)} > \overline{M}_{R_2}^{(F)} \end{cases}$$

The shaded area of Fig. 2.12 represents the regions of the parameter space in which the analytical results can predict the game outcome. The feasible region of the parameter space (in which $\Lambda < (1+\mu^{(F)})$) is outlined by the dark trapezium. The discontiguous shaded regions are an artifact of the above compound rules that generate these plots. Evidently, there is a large region of the relevant parameter space for which the required results are implied by our analytical findings. This further supports the findings suggested by our exhaustive numerical experiments.



(a) $\mu^{(F)} < 1$ **(b)** $1 < \mu^{(F)} < 10$

Figure 2.12: Regions of the parameter space for which the game's outcome has analytical support

6 Conclusions and Future Work

Advances in technology have enabled service providers to disseminate their delay information to delay-sensitive customers. In this work, we study the implications of technology leader L initiating real-time delay announcements when her competitor, follower F, may initiate real-time delay announcements in response. Within this leader-follower framework, we consider three information regimes: In Regime 0, the status-quo, customers only have access to historical delay information for both providers. In Regime 1, F opts not to respond to L's delay announcements, making L the sole announcer in this regime. In Regime 2, F opts to respond, and therefore, both L and F announce delay information.

For each of the regimes, we specify how customers make decisions and for each regime, we present CTMCs whose analysis determines market shares in the regime. We analyze Regime 0 analytically. For Regime 1, we present some analytical results, conjectured bounds, and an exhaustive set of numerical experiments. For Regime 2, we present conjectured bounds and an exhaustive set of numerical experiments. Our analysis suggests the following outcome of the sequential game: If L is the higher-capacity service provider, she is better off not initiating delay announcements because F will respond, which will erode L's initial market share advantage. However, she will give up benefits that will accrue to her in terms of customer delay reduction by opting not to announce delay information. On the other hand, L *should* initiate real-time delay announcements if she is not the higher-capacity service provider, as this gives her a market share advantage even if F responds. This decision also results in a reduction in average delay experienced by her customers.

Our work uncovers an effect of announcing real-time delay information not previously noted: Announcing real-time delay information is a good idea only for service providers who do not have higher service capacity than their competitor. Thus, higher-capacity service providers must exercise caution before providing real-time delay announcements to prevent market share loss: They should only initiate announcements if the resulting delay reduction compensates for the decreased market share. For a lower capacity service provider, announcing delay information can be considered a strategic remedy for capacity shortage.

We also investigate how the delay announcement strategies employed by L and/or F affect social welfare. We define social welfare as the average system-level delay, and we compute it as the market share weighted average of delays at L and F. Social welfare could degrade for some parameters if the higher-capacity L initiates delay announcements, and F does not respond. However, we find that social welfare degradation *never* occurs in equilibrium: There is no change to social welfare when L is the higher-capacity service provider as L does not have an incentive to initiate delay announcements. However, when L is not the higher-capacity service provider, she *does* have an incentive to announce delay information. While F may or may not have an incentive to respond in this case, we find that social welfare *necessarily* improves regardless of F's response. This means that the delay improvement at L is significant enough to offset a possible delay degradation at F. (F can have worse delay and worse market share in Regime 1 than in Regime 0.) In summary, real-time delay announcements result in improved social welfare whenever initiated by L.

We acknowledge that the asymmetry in willingness to adopt new technology, as modeled by our sequential game, is not representative of all duopolistic settings. There are circumstances under which both providers will have identical propensities to being first movers. In this situation, the appropriate modeling choice is a simultaneous game, rather than a sequential one. From our analysis in Section 5, it is clear that it is a dominant strategy for the lower-capacity service provider to announce real-time delay information. Therefore, our findings automatically suggest the outcome of a simultaneous game as well: in equilibrium, the lower-capacity service provider makes real-time delay announcements, while the higher-capacity service provider may or may not do so, depending on the load level.

This work studies a setting with two single-server service providers. We find that even in this setting, obtaining analytical results is non-trivial. A useful extension to our model would be to consider multi-server service providers. Another potentially interesting extension is to model a customer population consisting of dedicated and flexible individuals: Dedicated customers are loyal to one of the service providers regardless of the delay situation, while flexible customers are delay sensitive and use real-time delay information. Such heterogeneity in the customer population is explored in papers such as Dong et al. [2018], but with two providers providing delay information of identical granularity. (It is also explored in a different framework in Chapter 3.) While we investigate a setting with one leader and one follower, investigating settings with one leader and multiple followers could be an interesting avenue for future research. The three conjectures presented in this work also represent interesting avenues for further research: in particular, proving the bounds we conjecture for market shares in Regime 2 (Conjecture 2) would contribute to a deeper understanding of the notoriously difficult Join-the-shortest-queue system.

3

Towards Explicitly Incorporating Competition under Flexible Models of Demand in Dynamic Pricing

1 Introduction

Most revenue management systems assume monopolistic market models to carry out forecasting and optimization for dynamic pricing, even if the firms using such systems are faced with competition [Cooper et al., 2015]. A popular argument put forth by proponents of monopolistic models is that the parameters of these monopolistic models are estimated using *actual* data that comes from the *competitive* market, and that this should offset the effects of model mis-specification. This is called the market response hypothesis. Cooper et al. [2015] characterized the conditions under which the market response hypothesis holds in a one-period game without capacity constraints, with forecast and price optimization cycles. They show that under most general conditions, the hypothesis does not hold.

Advances in technology have made it possible for a firm to receive competitor price updates in real-time. Therefore, it is conceivable that firms could use competitive models for pricing. For example, airlines can track their competitors' fare information by automatically making queries to travel metasearch engines such as Kayak.com. In the retail setting, firms can access their competitors' pricing information by querying aggregators such as Amazon.com. Given the widespread access to competitor price information and Cooper et al. [2015]'s finding that the market response hypothesis does not generally hold, it is important for firms to understand whether, and how, to incorporate their competitors' price information in their own pricing decisions. In this chapter, we consider this question in a duopolistic market, where each seller has a fixed number of units of a product to sell over a finite horizon. Throughout this chapter, we use the motivating example of airlines competing to sell seats on an aircraft. Our findings, however, apply more generally to problems of dynamic price competition with fixed capacities.

The study of how to use competitor fare information relies on understanding what pricing decisions firms take in equilibrium. Related literature that studies these equilibrium decisions assumes that a competitor's current level of remaining inventory can be inferred (in equilibrium) by observing the competitor's price and backing out the inventory level by tracking the equilibrium price path. This chapter also adheres to the same framework. However, we depart from the existing literature in the model of demand that we use. In particular we propose a model of demand that captures three features that the extent literature captures separately, but not together: (i) unknown market size (ii) demand elasticity with respect to price to capture customers' heterogeneous valuations (iii) the presence of loyal customers, who have a preference between the airlines, *and* the presence of flexible customers (also called market priceable) customers who buy down to the lowest fare, leading to a discontinuity in the demand function at the point when an airline's fare is exactly equal to its competition's. We elaborate on the contribution represented by this general model of demand in Section 2.

In line with the existing literature, we focus on exploring the existence and structure of a potential subgame-perfect Nash equilibrium in pure strategies (SPNE-P) in the dynamic pricing game played by the competing airlines. For clarity of exposition, we model a setting with two competing firms, although our approach and main results may readily be extended to any finite number of competing firms. Under our model of demand, we show that in the presence of any customers who exhibit brand loyalty (that is, if an airline stands to capture some customers *even if* she isn't offering the lowest price in the market), an SPNE-P fails to exist. This contribution is significant because this non-existence result encompasses a large set of realistic demand models: reasons for brand loyalty could include frequent-flyer memberships, corporate alliances or proximity to an airline's hub airport.

Next, we study the case when all customers in the market are flexible and buy down to the lowest fare, i.e., the product offerings are perfect substitutes. This setting can be thought of as competition between two low-cost carriers such as *Spirit Airlines* and *Frontier Airlines* [International Civil Aviation Organization, 2019]. We prove the existence of a continuum of SPNE-P and demonstrate how it may be computed. The structure of the equilibrium closely follows that of the equilibrium found in Martínez-de Albéniz and Talluri [2011]. We propose a refinement to the equilibrium concept that yields a practically relevant (and in general, unique) equilibrium. This refinement produces an equilibrium that is analogous to the one found in Martínez-de Albéniz and Talluri [2011]. This is unsurprising: our demand model corresponds to an extension (uncertain valuations) in Martínez-de Albéniz and Talluri [2011]; under their extension, the proof of existence of an equilibrium is left open because the best response function is, in general, not continuous. We overcome this problem by directly analyzing components of the best response function.

Equipped with an equilibrium characterization for the case of competition with fully flexible customers, we explore the strategic question of whether low-cost carriers such as the two above should use competitor fare information. We demonstrate that in general, the unique equilibrium strategy is for both firms to migrate from monopolistic models to competitive models. This leads to questions such as: How does this migration affect industry profits? What about the profits of the individual firms?

Using two standard models of demand: exponential demand and iso-elastic demand, we study how the initial capacities of the firms affect the impact of migration on their profits. Two effects drive our results. The *competition effect* captures the drop in profits attributable to firms undercutting each other's prices, and the *information effect* captures the increase in profits attributable to firms pricing as monopolists when their competitor runs out of inventory.

When total capacity (relative to demand) is very low, the information effect is stronger, and therefore, migrating to competitive models can improve industry profits. When the firms begin with roughly the same capacity, they both share this benefit, otherwise, the benefit is consumed by the firm with the higher capacity. When the total capacity is moderate or high, the competition effect is stronger, and total industry profits are lower. Therefore, when the two firms start off with the same initial capacity, they are both adversely affected. However, when the firms are asymmetric, the industry profits are driven down by the profit erosion of the higher capacity firm. Indeed, the lower capacity firm benefits from this migration because it tends to post lower prices than the higher capacity firm (similar to the effect observed in Dudey [1992] and Martínez-de Albéniz and Talluri [2011]) and sell most of its inventory.

2 Literature Review

Recent developments in dynamic pricing literature are thoroughly reviewed in Chen and Chen [2015]. In their paper, Chen and Chen divide recent contributions into work that incorporates (i) multiple products, (ii) limited demand information, or (iii) competition. We contribute to the ongoing work in the third stream, i.e., studying the role of competition in dynamic pricing. These studies are of particular interest given Cooper et al. [2015]'s finding that the conjectured *market response hypothesis* (see Talluri and Van Ryzin [2004]) does not hold in general.

One stream of work that models competition (with notable papers including Liu and Zhang [2013], Jerath et al. [2010], Levin et al. [2009] and Anton et al. [2014]) focuses on understanding the impact of customers being strategic in their purchase timing. The general finding of this stream of literature is that the presence of strategic customers, as in the case of monopolistic models, erodes firms' profits.

However, most research on modeling competition—including our work—models customers as being myopic in their purchase timing decision. Customers arrive, view the prices available in the market, and make a purchase decision. They leave if they fail to make a purchase and do not return. This stream, again, has two different sub-streams that model firm decisions differently. The first type considers the case of *pre-announced* pricing, wherein the competing firms declare their prices for the entire horizon at the start of the horizon. Some examples of this work are Zhao and Atkins [2002] and Mookherjee and Friesz [2008]. See Chen and Chen [2015] for a detailed review of this literature.

Our work specifically contributes to the stream of literature that models customers as being *myopic* in their purchase timing, and firms as employing *contingent* pricing strategies, where prices are declared dynamically, rather than statically at the start of the horizon.

Some of this work on myopic, contingent pricing models assumes that the products on offer are perfectly substitutable and therefore, customers will only ever purchase from the firm offering the lowest price at any point in time. Dudey [1992] lays the foundation for this stream of work: he is the first to show the existence of an SPNE-P in a dynamic Edgeworth-Bertrand competition, in a simple model with deterministic customer valuations and a deterministic market size. Subsequent work differs in many aspects, including how customer valuations and market size (and therefore demand) are modeled. Dasci and Karakul [2009] model a two-period setting, where there is a fixed and known market size in each period and customers have a fixed willingness-to-pay. They compare dynamic pricing with fixed-ratio pricing (where the second period price is a pre-declared fraction of the first period price). Surprisingly, they find that in most cases, dynamic pricing does not outperform fixed-ratio pricing. Xu and Hopp [2006] model customers as having iso-elastic demand functions. They study firms' pricing and capacity decisions in a continuous time setting, and compare contingent pricing to static pricing. Similar to Dasci and Karakul, they find that contingent pricing is not necessarily helpful. Dudey [2007] extends the work of Dudey [1992] to a situation with random customer valuations (but still a deterministic market size). Similar to Dasci and Karakul [2009], Dudey [2007] uses a specific two-period pricing model wherein the competing firms announce a price in the first period and offer discounts in the second period. In this model, however, the firms also choose their capacity (production quantity) at the start of the horizon. The author finds that in equilibrium each seller produces enough to serve the market. Similar to Dudey [2007], Christou et al. [2007] study a pricing problem over three periods where firms first fix capacities and then compete over prices. Exactly one customer arrives per period (the market size is fixed and known), and this customer's valuation is seen by the firms at the start of each period. Anderson and Schneider [2007] focus on studying the problem of dynamic pricing in the presence of search costs. For an environment with a fixed market size and random customer valuation, they find that in the presence of search costs, prices are higher and firm profits are lower in the competitive setting than the monopolistic setting. Mantin et al. [2011] similarly model search costs implicitly using a Markovian framework and study how the market prices evolve over time as a function of the Markov dynamics. The paper closest to our work in this stream of literature is Martínez-de Albéniz and Talluri [2011], who treat customer valuation as fixed, but the number of arriving customers (the market size) as being unknown. They characterize a unique SPNE-P. In this stream of literature (the case with perfectly substitutable products), we extend Martínez-de Albéniz and Talluri's result to a case with uncertain customer valuations.

Some other papers model the products on offer in the market as *imperfectly* substitutable. The standard demand model assumed in this stream is characterized by continuous functions, i.e., a firms' demand is a continuous function of her price, and does not change discontinuously when she charges more than her competition. Isler and Imhof [2008] and Lin and Sibdari [2009] use a multinomial logit choice model to characterize customer decisions, both in environments with uncertain market size. Isler and Imhof compare the competitive solution to the monopolistic and cooperative solutions and find, unsurprisingly, that the firms have the lowest revenue in the competitive case. Lin and Sibdari compare the complete information case (where each firm knows their competitor's inventory level) with the incomplete information case for which they propose a heuristic solution that works well under various scenarios. Gallego and Hu [2014] also make the assumption of continuity of the demand function. In a continuous time model, they find the structure of the optimal pricing policy in a setting without time-varying demand. They propose a heuristic that handles the time-varying case well, and show that this heuristic results in an asymptotically optimal solution.

We consider a general model of demand that goes beyond deterministic customer valuation and fixed market size, and that unifies both the perfectly substitutable products and the differentiated products cases. We implicitly model some customers as being flexible (always buy the lowest priced product), and some customers as potentially being more loyal to one of the firms, i.e., the products are differentiated. Using our model of demand, we show that the presence of flexible *and* loyal customers precludes the existence of a SPNE-P. (When *only* loyal customers populate the market, the problem collapses to a degenerate one, i.e., one without competition.) When *only* flexible customers populate the market, our setting is equivalent to the setting with perfectly substitutable products. In this case, our demand model collapses to a *still* relevant generalization of the demand model used by Martínez-de Albéniz and Talluri [2011]; in particular, our model allows uncertain customer valuations (as opposed to the fixed customer valuation setting studied in Martínez-de Albéniz and Talluri [2011]). Martínez-de Albéniz and Talluri suggest that proving the existence of a SPNE-P in the uncertain valuations case is not non-trivial. We address this gap in this chapter. Indeed, the equilibrium structure we find is very similar to the one found by Martínez-de Albéniz and Talluri, as conjectured by them. We emphasize the practical relevance of this extension to general willingness-to-pay demand models for real-life applications, for example, in the airline industry. The deterministic willingness-to-pay model employed by papers such as Martínez-de Albéniz and

Talluri [2011] and Dudey [1992], while capable of providing useful managerial insights, falls short of capturing the fact that customers, in reality, have heterogeneous values of willingness-to-pay.

We now discuss some shared features that these models have, that we share. The setting in this stream is either duopolistic or oligopolistic. The typical question of interest is how competing firms should dynamically price in equilibrium. A standard assumption is that in a setting with a unique equilibrium, each firm can track the price path of their competitors and accordingly infer their competitor(s)' current inventory levels. Customer demand is given as some known function of the prices offered by each firm. This demand could be varying over time.

We also contribute to the literature by addressing the strategic question of whether firms should use monopolistic models or competitive models. With the exception of Cooper et al. [2015], the extant literature focuses on simply comparing industry performance in a monopolistic setting (with one firm) to a competitive setting (with more than one firm). In contrast, this chapter and Cooper et al. [2015] study a situation with *two firms* who may employ monopolistic or competitive models in their pricing decisions. Cooper et al. [2015] addresses this by explicitly modeling forecasting and optimization cycles in a sequence of single-period dynamic pricing games without capacity constraints. In contrast, we treat the forecast as being exogenously given, and answer the question in a single multi-period dynamic pricing game with capacity constraints. By doing so, we are able to study the impact of a firm's capacity on their profits if the market migrates from using monopolistic models to using competitive models.

3 The Demand Model

We consider a discrete time model with time indexed by t. The horizon of interest is T periods long, with periods numbered from 0 to $T-1$. There are two firms, me and c, with X_{me} and X_c units of product respectively. These units expire at the end of the horizon and have zero salvage value. A customer arrives in period t with some positive probability (the market size), denoted by λ_t. (We do not impose any restrictions on λ_t; it could, in general, depend on various factors, including the number of units of product me and c have left, as in Martínez-de Albéniz and Talluri [2011].) We assume that there is some strictly positive probability of an arriving customer being flexible as far as choice of an airline, i.e., she will buy the lowest priced product. Some proportion of customers may also be loyal to one of the two airlines, and this proportion may depend on the set of prices on offer. An arriving customer makes a purchase from one of the two airlines (me and c), depending on the prices offered, or decides to leave without purchasing. Arriving customers are myopic, i.e., they are not strategic about their purchase timing. In order to model customer arrival and choice, we specify a probability function $\phi(p_{me}, p_c)$, (where p_{me} represents the price charged by me and p_c the price charged by c) that maps the price system (p_{me}, p_c) to the probability that conditional on a customer arriving, she purchases from me. While the firms may not be symmetric in the amount of capacity they have (number of seats), we assume they are otherwise perfectly symmetric. Therefore, $\phi(p_c, p_{me})$ represents the probability that conditional on a customer arriving, she purchases from

c[1]. As is standard in the literature, we assume the firms face a variable cost of 0 for each customer. As with Martínez-de Albéniz and Talluri [2011], we allow firms to charge negative prices to get rid of capacity. (The interpretation of a negative price is pricing below variable cost.) Therefore, we do not impose any constraints on p_{me} and p_c.

It is convenient to work with the unconditioned demand function $D_t(p_{me}, p_c) = \lambda_t \phi(p_{me}, p_c)$. Function D_t represents the probability that a customer arrives in period t, faces prices p_{me} and p_c, and chooses to purchase one unit of product from *me*.

For the remainder of the chapter, we drop the time subscript t from the demand function D, as it has no impact on our analysis. The demand function D is given by:

$$D(p_{me}, p_c) = \begin{cases} \overline{d}(p_{me}, p_c) & \text{if } p_{me} < p_c \\ \underline{d}(p_{me}, p_c) & \text{if } p_{me} > p_c \\ \dfrac{\underline{d}(p_{me}, p_c) + \overline{d}(p_{me}, p_c)}{2} & \text{if } p_{me} = p_c \end{cases} \tag{3.1}$$

We impose the following conditions on the demand function:

Assumptions:

1. $0 \leq D(p_{me}, p_c), \forall p_{me}, p_c$
2. $0 \leq D(p_{me}, p_c) + D(p_c, p_{me}) \leq 1$
3. $\underline{d}(p_{me}, p_c)$ and $\overline{d}(p_{me}, p_c)$ are defined $\forall p_{me}, p_c$ and are continuous in both arguments.
4. $\underline{d}(p_{me}, p_c)$ and $\overline{d}(p_{me}, p_c)$ are decreasing in own price p_{me}
5. $\underline{d}(p_{me}, p_c)$ and $\overline{d}(p_{me}, p_c)$ are increasing in competitor price p_c
6. $\underline{d}(p_{me}, p_c) < \overline{d}(p_{me}, p_c), \forall p_{me}, p_c$

The first assumption ensures that the demand probabilities are all non-negative. The second assumption ensures that the demand probability function is a valid probability. Note that $1 - D(p_{me}, p_c) - D(p_c, p_{me})$ represents the probability that no customer arrives, or that a customer arrives and chooses not to make a purchase. The third assumption is a simple regularity condition. The fourth assumption ensures that for a fixed value of competitor price, the probability that an arriving customer makes a purchase from *me* decreases as *me*'s price p_{me} increases. The fifth assumption ensures that for a fixed value of p_{me}, the probability that an arriving customer makes a purchase from *me* increases as my competitor's price p_c increases. The fourth and fifth assumptions, taken together, capture customers' price elasticity (which can equivalently be interpreted as

[1]The implication of this assumption is two-fold: (i) There is no asymmetry in the proportion of loyal customers and in the price sensitivities of these loyal customers for a given price system and (ii) flexible customers view both firms' products as perfectly substitutable. We make this assumption for expositional convenience, so that we can focus our study on the role of competition. Our main findings continue to go through without this assumption.

customers having willingness-to-pay values that are uncertain over a continuous support). This feature is present in the demand models such as the ones in Lin and Sibdari [2009] and Isler and Imhof [2008]. Recall that demand probability $\overline{d}(p_{me}, p_c)$ applies when $p_{me} < p_c$ and demand probability $\underline{d}(p_{me}, p_c)$ applies when $p_{me} > p_c$. The sixth assumption, taken together with the continuity imposition in the third assumption, accommodates the behavior of the flexible customers – when price p_{me} exceeds p_c, flexible customers will no longer purchase from *me*, leading to a discontinuous dip in purchase probability. This feature is present in demand models such as the ones in Martínez-de Albéniz and Talluri [2011]. By modeling both price elasticity and demand function discontinuity, we unify existing models of demand in the literature.

This set of assumptions is not restrictive; in fact, it spans a large class of practically relevant demand constructs. For instance, consider a situation with posted prices p_H, p_L, such that $p_H > p_L$. For some $0 < \delta < 1/2$, a fraction $1 - \delta - \delta$ of the market is flexible on choice of firm, a fraction $\delta e^{-(p_H - p_L)}$ is loyal to the firm charging price p_H and a fraction $2\delta - \delta e^{-(p_H - p_L)}$ is loyal to the firm charging price p_L. Note that the extent of loyalty here depends on the difference in prices. Further, assume that all customers have an exponential willingness-to-pay distribution. The following specification is valid for this situation, and is accommodated by our demand model: $\overline{d}(p_{me}, p_c) = (1 - \delta e^{-(p_c - p_{me})})e^{-p_{me}}, \forall p_{me}, p_c > 0$ and $\underline{d}(p_{me}, p_c) = \delta e^{-(p_{me} - p_c)} e^{-p_{me}}, \forall p_{me}, p_c > 0$. (For expositional convenience, we do not specify the demand function for negative prices here.) Under this model of demand, Figure 3.1 graphically demonstrates how the demand probabilities associated with *me* and *c* depend on p_{me}, for $\delta = 0.2$ and $p_c = 1$.

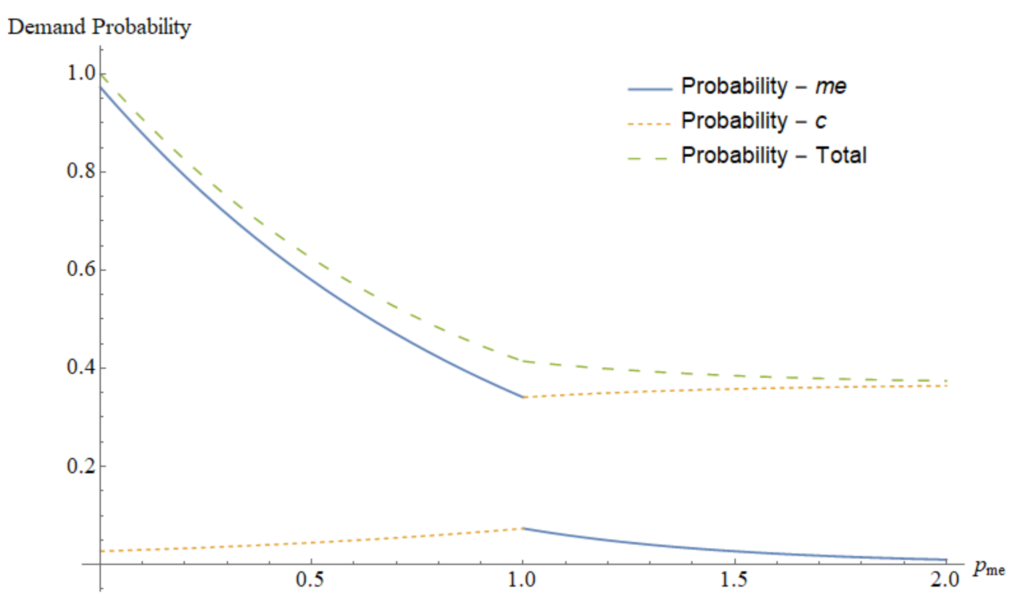


Figure 3.1: Example of a demand model accommodated by our framework

4 Equilibrium Analysis

We analyze the dynamic pricing problem by formulating a firm's (without loss of generality, *me*'s) optimal action in period t using Bellman's equations. The state of the system in period t is described by the ordered pair (x_{me}, x_c), where x_i represents the number of units of inventory remaining with firm i. Let $V_t^*(x_{me}, x_c)$ denote the optimal value-to-go in period t from firm *me*'s perspective. In order to analyze the equilibria resulting from this dynamic pricing game, we write the following best response function for *me*, when the competitor's price is p_c and the current state of the system is given by (x_{me}, x_c).

$$\begin{aligned} & V_t(x_{me}, x_c, p_c) = \\ \max_{p_{me}} \quad & D(p_{me}, p_c)\left(p_{me} + V_{t+1}^*(x_{me} - 1, x_c)\right) + D(p_c, p_{me})\, V_{t+1}^*(x_{me}, x_c - 1) + \\ & (1 - D(p_{me}, p_c) - D(p_c, p_{me}))\, V_{t+1}^*(x_{me}, x_c) \end{aligned} \tag{3.2}$$

We will now look for an SPNE-P using the best response function (3.2). We will be interested in examining two different cases:

1. Not all customers are flexible: the partially flexible case.
2. All customers are flexible: the fully flexible case, i.e., $D(p_{me}, p_c) = 0$ for $p_{me} > p_c$.

We now examine these situations in turn.

4.1 The partially flexible case

When not all customers are flexible, an airline can possibly make a sale even if she offers the higher price. One way to interpret this is as some customers being loyal to one or the other airline. However, the extent of loyalty may depend on the competitor's price. To model this situation, we assume that $D(p_{me}, p_c) > 0$ for $p_{me} > p_c$, that is, the higher priced airline has a strictly positive probability of making a sale.

Proposition 15 *In the partially flexible case, there is no SPNE-P.*

Proof: Presented in Appendix C.1.

Note that the implication of Proposition 15 applies to a variety of underlying customer choice models. Indeed, we did not impose any utility structure on the customers or place any implicit assumption on how many customers are loyal and how many are flexible. In fact, it is possible that the number of loyal customers, or equivalently, the extent of loyalty to a firm, depends on the prices charged by both firms (or even the difference in prices).

4.2 The fully flexible case

When all customers are flexible, we have that $D(p_{me}, p_c) = 0$ for $p_{me} > p_c$. Therefore, we no longer need the notation $\overline{d}(p_{me}, p_c)$ and $\underline{d}(p_{me}, p_c)$. We can simply rewrite $D(p_{me}, p_c)$ as follows:

$$D(p_{me}, p_c) = \begin{cases} 0 & \text{if } p_{me} > p_c \\ d(p_{me}) & \text{if } p_{me} < p_c \\ d(p_{me})/2 & \text{if } p_{me} = p_c \end{cases} \tag{3.3}$$

Note that since all customers are flexible, the only relevant price is the cheaper alternative. We continue to impose the same regularity conditions on this demand function. Under this scenario, the best response function for me can be written as:

$$\begin{aligned} &V_t(x_{me}, x_c, p_c) = \\ &\max_{p_{me}} \left(\mathbb{I}_{p_{me}<p_c} + \frac{\mathbb{I}_{p_{me}=p_c}}{2} \right) d(p_{me}) \left(p_{me} + V^*_{t+1}(x_{me} - 1, x_c) \right) \\ &+ \left(\mathbb{I}_{p_{me}>p_c} + \frac{\mathbb{I}_{p_{me}=p_c}}{2} \right) d(p_c) \, V^*_{t+1}(x_{me}, x_c - 1) \\ &+ \left(\mathbb{I}_{p_{me}<p_c}(1 - d(p_{me})) + \mathbb{I}_{p_{me}>p_c}(1 - d(p_c)) + \mathbb{I}_{p_{me}=p_c}(1 - d(p_{me}) \right) V^*_{t+1}(x_{me}, x_c) \end{aligned} \tag{3.4}$$

Lemma 3 *In the fully flexible case, at each state (x_{me}, x_c), there exist unique thresholds r_{me} and r_c such that:*

- *If $p_c \geq r_{me}$, me's optimal action is to undercut price p_c. Otherwise, she obtains a higher revenue-to-go by not undercutting, that is, by choosing any price $p_{me} > p_c$.*

- *Similarly, if $p_{me} \geq r_c$. c's optimal action is to undercut price p_{me}. Otherwise, she obtains a higher revenue-to-go by not undercutting, that is, by choosing any price $p_c > p_{me}$.*

Proof: Presented in Appendix C.2

The existence of these thresholds is very similar to the equilibrium structure obtained in Martínez-de Albéniz and Talluri [2011]. In their paper, Martínez-de Albéniz and Talluri claim that these thresholds ('reservation values') imply a unique equilibrium. However, we note that the equilibrium is not unique. To see this, consider the case where $r_{me} < r_c$. This implies that me is wiling to undercut to a lower price than c is. Therefore, me undercuts and makes the sale (if the arriving customer has a higher willingness-to-pay than me's price). Observe here that there is a continuum of equilibria: c can choose any price in $[r_{me}, r_c]$ and me will undercut optimally. Neither firm has an incentive to deviate. The $r_c < r_{me}$ case can be treated analogously.

Therefore, we propose a refinement of the subgame-perfect pure strategy Nash equilibrium wherein no firm that settles (i.e., does not undercut) chooses a price below their threshold. Note that such a refinement of the SPNE-P is required to produce a unique equilibrium in the deterministic and known valuation case (as studied in Martínez-de Albéniz and Talluri [2011]). Our proposed refinement, in essence, picks the highest possible sale price among the equilibria; the threshold is thought of as a 'reservation value' below which the 'settling' firm will not price. (In general, the 'undercutting' firm will also not price below their reservation value, but their response cannot be characterized in closed-form for general demand functions.) Under this refinement, we have:

Proposition 16 *In the fully flexible case, a unique refined SPNE-P always exists when the two firms have unique reservation values. When the reservation values are identical, there are two refined SPNE-P.*

Proof: This result is implied by the existence of the reservation values. In particular:

- If $r_{me} < r_c$, the equilibrium involves c pricing at r_c and me undercutting optimally.
- If $r_c < r_{me}$, the equilibrium involves me pricing at r_{me} and c undercutting optimally.
- If $r_c = r_{me} = r$, there are two SPNE-P. Under each, one of the firms prices at r while the other undercuts optimally.

■

In stating Proposition 16, we note that the firm that prices lower does so by 'undercutting optimally'. Without loss of generality, let us consider the case when $r_{me} < r_c$, and therefore, me undercuts c. Formally, undercutting optimally means that me chooses a price in Period t that satisfies:

$$\begin{aligned} p_{me} &= \underset{p_{me}<r_c}{\arg\max} \quad d(p_{me}) \left(p_{me} + \underbrace{V^*_{t+1}(x_{me}-1, x_c) - V^*_{t+1}(x_{me}, x_c)}_{\Delta_1} \right) + V^*_{t+1}(x_{me}, x_c) \\ &= \underset{p_{me}<r_c}{\arg\max} \quad d(p_{me})\,(p_{me} + \Delta_1) + V^*_{t+1}(x_{me}, x_c). \end{aligned} \tag{3.5}$$

We note that this equilibrium cannot be stated as concisely as the equilibrium presented in Martínez-de Albéniz and Talluri [2011] (Lemma 1). This is because of the generality of our willingness-to-pay distribution as opposed to their fixed willingness-to-pay model: In their case, the firm that (potentially) makes the sale simply prices at the reservation price of the competitor. For specific demand functions, we may be able to characterize the equilibrium price at which me prices more sharply. For instance, if the function in (3.5) is concave and is maximized at p^*, we have that $p_{me} = r_c - \epsilon$ for some small $\epsilon > 0$ if $r_c < p^*$, and $p_{me} = p^*$ otherwise.

We also note that in the fully flexible case, our demand model is analogous to an extension (uncertain valuations) in Martínez-de Albéniz and Talluri [2011]. Martínez-de Albéniz and Talluri state that in general, it is difficult to prove the existence of an equilibrium for the case of uncertain valuations because the best response function is, in general, not continuous. We overcome this problem by directly analyzing components of the best response function, allowing us to establish the existence of an SPNE-P.

5 Comparing Monopolistic and Competitive Models

We now study the impact of firms using competitive versus monopolistic models on their revenues.

First, consider the decision at a high level. Each of the firms can choose to use a monopolistic model or a competitive model. Denote the chosen strategy of firm as $s_i \in \{C, M\}$, where M represents using a monopolistic model and C represents using a competitive model. Let us denote a specific strategy profile (a specification of both firms' strategies) by the ordered pair (s_{me}, s_c). Therefore, there are four possible strategy profiles in the strategy space, (M, M), (M, C), (C, M), (C, C). For this model choice problem, we restrict our attention to a one-shot interaction between the competing firms, and investigate which of these strategy profiles (if any) is a Nash equilibrium. (Note that following the choice of model, the two competing firms play a dynamic pricing game over a horizon of length T.)

- (M, M) is not a Nash equilibrium strategy profile. *me* could deviate resulting in the strategy profile (C, M). Naturally, doing so will make her at least as well off as under the strategy (M, M) because she can choose the same prices as she would have chosen under (M, M) and obtain the same profits. Further, it can be shown that strategy profile (C, M) strictly dominates strategy profile (M, M) from *me*'s perspective: Consider all sample paths under which *me* and c are left with some remaining inventory one period from the end of the horizon. Under strategy profile (M, M), both firms select the same price p^* (that maximizes $\frac{d(p)}{2}p$) and obtain an expected profit of $\frac{d(p^*)}{2}p^*$ each, while under the strategy profile (C, M), *me* can undercut c's price by a small $\epsilon > 0$ and obtain a profit arbitrarily close to $d(p^*)p^*$.
- We now argue intuitively that (C, M) (equivalently (M, C)) is not a Nash equilibrium strategy profile either. This is because under strategy profile (C, M), *me* anticipates the price charged by c in every period t, and can undercut c if favorable and not otherwise, without threat of retaliation. This is only a coarse intuition for why c finds (C, M) preferable to (M, M); we confirm this to be the case in all our numerical experiments (described subsequently). This preference does not appear to be analytically provable, however, because the dynamic behavior of the firms under both (M, C) and (C, C) cannot be characterized in closed form.
- Since it is favorable for *me* to deviate from (M, C) to (C, C) and for c to deviate from (C, M) to (C, C), (C, C) is the unique Nash equilibrium strategy profile.

Since (C, C) is the unique Nash equilibrium strategy profile, the natural question to ask is how the payoffs to both firms under strategy profile (C, C) compare to the payoffs under strategy profile (M, M) (the status-quo)? Under what conditions—if any—does the situation resemble the Prisoners' Dilemma?

In order to carry out a consistent comparison of these two strategy profiles, we make the following assumption: if a firm using a competitive model estimates a demand function of $d(p)$ (where p is the price charged by firm with the lower price), the same firm using a monopolistic model will estimate a demand function of $\frac{1}{2}d(p)$. Therefore, if both firms price at p, the total market demand would be $d(p)$ under the cases (C, C) and (M, M). This is analogous to the *demand consistency* property imposed in Cooper et al. [2015]. We review this property in Appendix C.5.

We note that unlike Cooper et al. [2015], we assume that the demand functions are already estimated prior to the dynamic pricing exercise, and that these functions' parameters are not updated from time-to-time based on the outcome of the dynamic pricing exercise. While Cooper et al. [2015] considers cycles of optimization and forecast updates, our focus is restricted to the optimization step; however, the dynamic pricing exercise in our model is T periods long (as opposed to Cooper et al.'s single-period setting) and we model the firms as being capacity constrained. We will demonstrate that capacity plays an important role in determining market outcomes.

We now consider two specific demand models for which the equilibrium prices can be computed simply. The first is a model under which demand decays exponentially with price. The second is a model with iso-elastic (i.e., linear) demand.

Exponential Demand:

Under this model, the demand at a price p is given by:

$$d(p) = \begin{cases} e^{-\alpha p}, & \text{if } p > 0 \\ 1, & \text{otherwise.} \end{cases}$$

Since we are interested in a relative comparison of the revenues obtained by using monopolistic versus competitive models, we can normalize price p (the currency) and set $\alpha = 1$, for convenience. By doing so, we obtain a parameter-free demand function.

Iso-elastic Demand:

Under this model, the demand at a price p is given by:

$$d(p) = \begin{cases} 0, & \text{if } p > 1 \\ 1 - \alpha p, & \text{if } 0 \leq p \leq 1 \\ 1, & \text{otherwise.} \end{cases}$$

Since we are interested in a relative comparison of the revenues obtained by using monopolistic versus competitive models, we can again normalize price p (the currency) and set $\alpha = 1$, for convenience. Again, this normalization gives us a parameter-free demand function.

Note that under both these demand models, a sale will be made with probability 1 if the price offered is low enough. Therefore, these demand models implicitly assume that a customer arrives with probability 1 in each period, i.e., the market size is such that $\lambda_t = 1, \forall t$. With some abuse of notation, this means that $d(p) = \lambda_t \phi(p) = \phi(p)$. Therefore, the only source of uncertainty is the willingness-to-pay of the arriving customer. We reiterate that our general model can accommodate uncertain market size, but we leave this source of uncertainty out of our experiments so as not to detract from the main focus of this section, i.e., the effect of using competitive rather than monopolistic models and the interaction of this effect with capacity. (It is possible to run similar experiments with uncertain and time-varying market sizes because the firms' pricing strategies in each period depend only on the demand D_t in *that* period, and the dependence on future periods is captured through the value function $V^*(\cdot)$. However, doing so would pose an impediment to isolating the effect of capacity on outcomes.)

For each of these demand models, we are interested in computing two quantities: $V_0^{(M,M)}(X_{me}, X_c)$ and $V_0^{(C,C)}(X_{me}, X_c)$. These quantify the net-present value to firm me (for length of horizon T and initial capacities X_{me} and X_c for me and c respectively), under strategy profiles (M, M) and (C, C). We also verify in all cases that (C, C) is indeed the equilibrium strategy, i.e., that me can (weakly) improve profits by deviating from (M, C) to (C, C) and c can (weakly) improve profits by deviating from (C, M) to (C, C).

Computing $V_0^{(C,C)}(X_{me}, X_c)$: To compute $V_0^{(C,C)}(X_{me}, X_c)$, we rely on the equilibrium characterization in Section 4.2. Note that both the exponential and the iso-elastic demand functions are inelastic in the region $p < 0$. We make this simplification for analytical tractability; it can be interpreted as follows: the variable cost is small enough that customer demand is inelastic when the price is set below the variable cost. Inelasticity in this region violates Assumption 4 (that stated that demand was elastic at all prices), and therefore, our existence proof in Proposition 16 does not directly carry over. In Appendices C.3 and C.4, we show that SPNE-P continue to exist, and discuss how they may be computed inexpensively. Solving for the equilibria directly gives us the value of $V_0^{(C,C)}(X_{me}, X_c)$ under both demand models.

Computing $V_0^{(M,M)}(x_{me}, x_c)$: Under strategy profile (M, M), me and c do not track each others' inventory position and prices. Pricing decisions are made based only on their own remaining inventory. To compute $V_0^{(M,M)}(X_{me}, X_c)$, we first find the optimal prices of each of the firms under monopolistic models. To do this, we solve a dynamic programming recursion that relates the period t problem to the value function for period $t+1$ (for a review of these monopolistic models, please see Talluri and Van Ryzin [2004]). We briefly describe the approach here. The dynamic programming formulation is given by:

$$V_t(x_{me}) = \max_{p_{me}} \frac{1}{2} d(p_{me})(p_{me} + V_{t+1}(x_{me} - 1)) + (1 - \frac{1}{2} d(p_{me}))\, V_{t+1}(x_{me}),$$

with the standard boundary conditions $V_T(\cdot) = 0$ and $V_t(x_{me} \leq 0) = 0$. Solving this dynamic program gives us the optimal price $p_t^*(x_{me})$ that a firm would charge in period t with x_{me} units of inventory left. Note that these pricing decisions do not depend on the competitor c's inventory level. We then use these prices to compute $V_0^{(M,M)}(X_{me}, X_c)$ by solving the following recursion for *me*.

$$
\begin{aligned}
&V_t(x_{me}, x_c) = \\
&\qquad \mathbb{I}_{p_t^*(x_{me},x_c)<p_t^*(x_c,x_{me})} \left(d(p_t^*(x_{me}, x_c))(p_t^*(x_{me}, x_c) + V_{t+1}(x_{me} - 1, x_c))\right. \\
&\qquad \left.+(1 - d(p_t^*(x_{me}, x_c)))\, V_{t+1}(x_{me}, x_c)\right) \\
&\qquad + \mathbb{I}_{p_t^*(x_c,x_{me})<p_t^*(x_{me},x_c)} \left(d(p_t^*(x_c, x_{me}))(V_{t+1}(x_{me}, x_c - 1)) + (1 - d(p_t^*(x_c, x_{me})))\, V_{t+1}(x_{me}, x_c)\right) \\
&\qquad + \mathbb{I}_{p_t^*(x_{me},x_c)=p_t^*(x_c,x_{me})} \left(\frac{1}{2} d(p_t^*(x_{me}, x_c))(p_t^*(x_{me}, x_c) + V_{t+1}(x_{me} - 1, x_c))\right. \\
&\qquad \left.+\frac{1}{2} d(p_t^*(x_{me}, x_c))\, V_{t+1}(x_{me}, x_c - 1) + (1 - d(p_t^*(x_{me}, x_c)))\, V_{t+1}(x_{me}, x_c)\right),
\end{aligned}
$$

with the boundary conditions $V_T(x_{me}, x_c) = 0$ and $V_t(x_{me} \leq 0, x_c) = 0$. Since the firms face symmetric demand functions, the value-to-go for firm c in period t is simply $V_t(x_c, x_{me})$.

Now that we can compute $V_0^{(M,M)}(X_{me}, X_c)$ and $V_0^{(C,C)}(X_{me}, X_c)$ under both demand models, we can answer the following questions:

1. **Total capacity net effect:** With symmetric firm capacities, i.e., $X_{me} = X_c = X$, how does the total market capacity ($2X$) in the market affect industry profits, when both firms migrate from monopolistic to competitive models? Under both our demand models, the expected number of arrivals over the horizon is T (as a potential customer arrives in every period). Therefore, the relative capacity level is $\frac{2X}{T}$. We use 'low (high) capacity' to mean 'low (high) capacity' *relative* to total expected number of arrivals T.

2. **Capacity mismatch net effect:** When the firms are asymmetric in their capacity, i.e., $X_{me} \neq X_c$, how is the total industry profit affected by migration to competitive models as a function of the extent of capacity mismatch?

3. **Capacity mismatch individual effect:** When firms are asymmetric, i.e., $X_{me} \neq X_c$, how is the lower capacity firm affected by a migration from monopolistic to competitive models? How is the higher capacity firm affected?

There are two conflicting effects at play that determine the answers to the above questions.

1. Competition Effect: Under the strategy profile (C, C), the presence of competitive forces drives down prices in equilibrium (when both firms have capacity remaining), as observed in prior literature including Martínez-de Albéniz and Talluri [2011]. In fact, these prices can even be negative. This results in a revenue advantage under strategy profile (M, M)—where prices are agnostic to the presence of competition—when both firms have remaining capacity.

2. Information Effect: When one firm (without loss of generality, c) runs out of capacity, the other firm *me* knows this fact under strategy profile (C, C). Therefore, when *me* has remaining capacity x and c has remaining capacity 0, *me* can choose a price to maximize the actual revenue function $d(p_{me})(p_{me} + V_{t+1}(x - 1, 0) + (1 - d(p_{me}))\, V_{t+1}(x, 0)$. This price yields a

higher revenue-to-go for *me* than she would obtain under the strategy profile (M, M), where she chooses a price to maximize the revenue function $\frac{1}{2}d(p_{me})(p_{me} + V_{t+1}(x-1) + (1 - \frac{1}{2}d(p_{me}))\,V_{t+1}(x)$, unaware of the fact that her competitor is *actually* out of capacity and that she is therefore pricing for a demand function $d(p_{me})$, which yields a different value-to-go. Therefore, the information effect gives strategy profile (C, C) an advantage when the competitor has run out of capacity. Note that since the pricing game is dynamic, the possibility of one firm (without loss of generality, c) running out of capacity can be anticipated by *me*, who may initially set high prices to drive c out of the market.

We now examine how these forces interact with each other to answer our three questions.

5.1 Total capacity net effect

To study the total capacity net effect, we consider three different lengths of time horizon: $T = 15$, $T = 75$, and $T = 375$. These correspond to three different values of relative capacity. For each of these three scenarios, we plot the value of $r = \frac{V_0^{(M,M)}(X,X)}{V_0^{(C,C)}(X,X)}$ on a logarithmic scale against values of X from 1 to $T-1$. When $r > 1$, monopolistic models yield higher revenues and when $r < 1$, competitive models yield higher revenues. When $X = T$, $V_0^{(C,C)}(X,X) = 0$, because prices are driven down to zero. $V_0^{(C,C)}(X,X)$ is strictly positive for all other values of X.

The results of our experiments are shown in Figures 3.2 - 3.4. In each of the plots, we see two distinct regions. In Region I, the capacity X is significantly lower than the length of the horizon T. In this region, we observe that $r < 1$ (however, in these cases r is very close to 1 and therefore $\log(r)$ is very close to 0), meaning that the information effect dominates, because there is a high probability that one firm runs out of capacity, leaving the other firm to operate as a monopoly. In Region II, capacity is moderate or high. Here, $r > 1$ and the competition effect dominates, because there is a relatively low probability of one of the firms running out of inventory during the horizon. Therefore, the information effect plays a vanishingly small role, and r increases. We note that the boundary between Region I and Region II differs for the two different demand cases.

In summary, in Region I, the equilibrium strategy profile (C, C) is also Pareto dominant. Therefore, when capacity X is significantly low in comparison to the length of the horizon, a migration to competitive models is both an equilibrium strategy *and* leaves both firms better off. In Region II, however, the equilibrium strategy profile (C, C) is Pareto dominated by strategy (M, M). In other words, the structure of the strategic game resembles that of a Prisoner's dilemma, where both firms are better off playing an off-equilibrium strategy.

5.2 Capacity mismatch net effect

We study the capacity mismatch net effect by tracking how the extent of capacity mismatch affects industry profits in moving from strategy profile (M, M) to strategy profile (C, C). When capacities are matched, we know from Section 5.1 that for relatively low capacity, strategy profile (M, M)

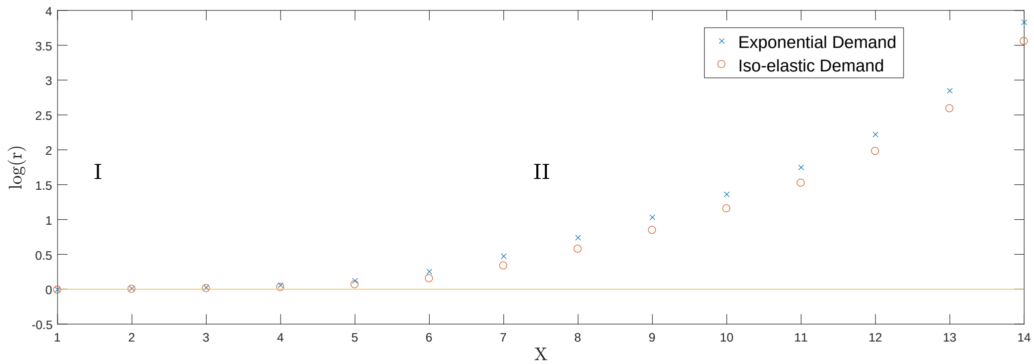


Figure 3.2: $\log(r)$ against X for $T = 15$

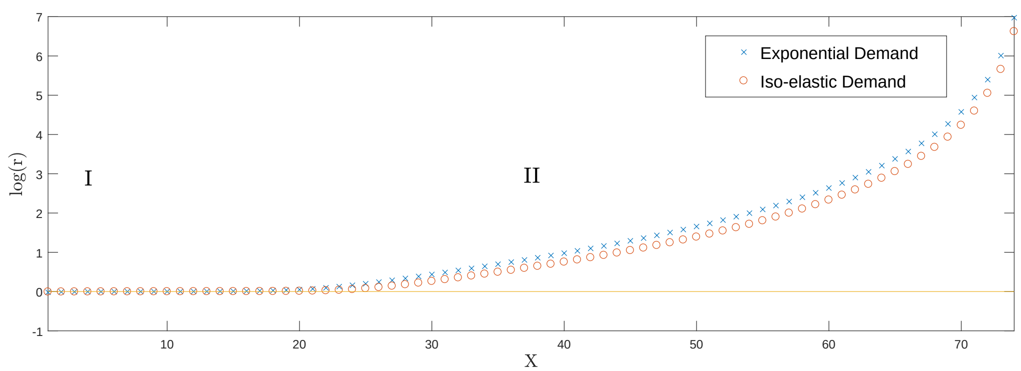


Figure 3.3: $\log(r)$ against X for $T = 75$

results in lower industry profits than strategy profile (C, C). What happens when capacities are mismatched, for the same future demand (in our case, for the same length of horizon T)? Similarly, we know that when capacity is moderate or high, strategy profile (M, M) results in higher industry profits than strategy profile (C, C). Again, what happens when capacities are mismatched? We illustrate this effect for a horizon length of $T = 375$. We start by considering two extremes of total capacities: a very low total capacity of 20 and a very high total capacity of 362. (These specific capacity levels are chosen for expositional convenience, i.e., so that similar behavior can be illustrated under both demand models in a single figure.)

When the total capacity is 20, we have from Section 5.1 that $r < 1$ under both demand models when $X_{me} = X_c = 10$. In Figure 3.5, we plot the logarithm of $r = \frac{V_0^{(M,M)}(X,20-X)+V_0^{(M,M)}(20-X,X)}{V_0^{(C,C)}(X,20-X)+V_0^{(C,C)}(20-X,X)}$ for values of X from 1 to 10.

When the total capacity is 362, we have from Section 5.1 that $r > 1$ under both demand models when $X_{me} = X_c = 181$. In Figure 3.6, we plot the logarithm of $r = \frac{V_0^{(M,M)}(X,362-X)+V_0^{(M,M)}(362-X,X)}{V_0^{(C,C)}(X,362-X)+V_0^{(C,C)}(362-X,X)}$ for values of X from 1 to 181.

In Figures 3.5 - 3.6, we see that under both demand models, as the extent of capacity mismatch decreases (when X increases), the value of r increases, implying that the relative strength of the

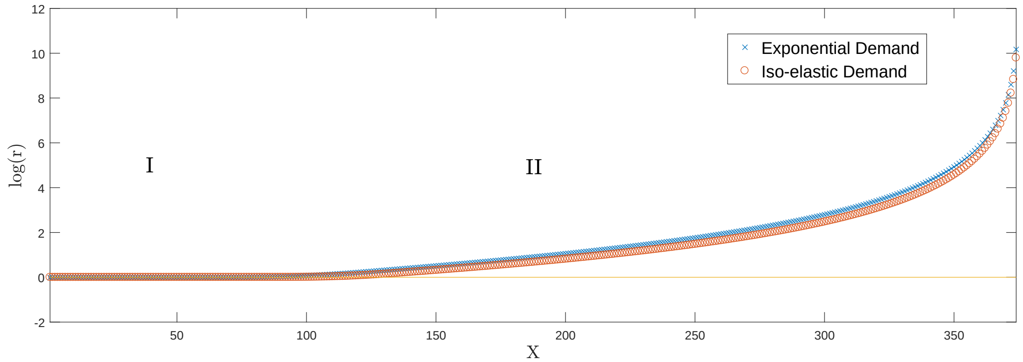


Figure 3.4: $\log(r)$ against X for $T = 375$

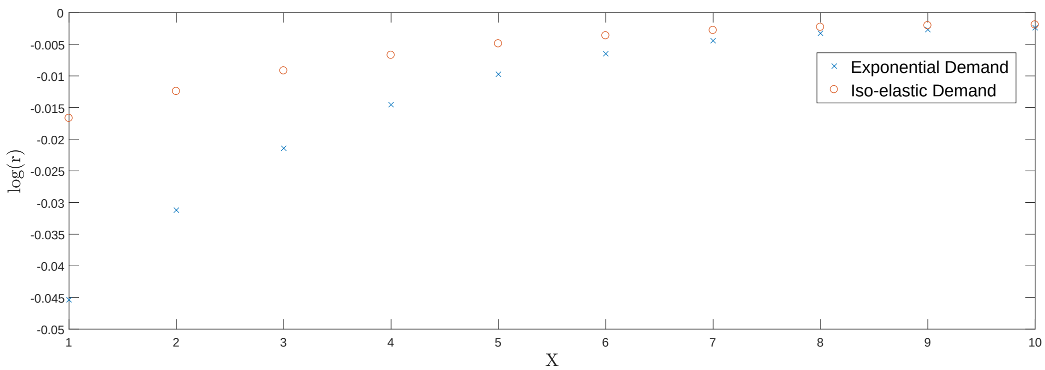


Figure 3.5: $\log(r)$ against X for $T = 375$, very low value of total capacity

information effect increases in the capacity mismatch level. The increase in $\log(r)$ is concave when capacity is low (Figure 3.5), and convex when capacity is very high (Figure 3.6) reflecting the diminishing role of the information effect with increased capacity. This can be attributed to a commensurate decrease in the probability that the lower capacity firm stocks out before the end of the horizon. When capacity is very high (Figure 3.6), we further note that the competition effect is particularly pronounced when capacities are exactly matched, i.e., when $X = 181$.

What happens for intermediate values of capacity? Figure 3.7 presents the plot for exponential demand and a total capacity of $X = 76$. Analogously, Figure 3.8 presents the plot for iso-elastic demand and a total capacity of $X = 98$. In both plots, there are two regions: when capacity mismatch is high (corresponding to lower values of X), the information effect is still strong, and competitive models yield a higher industry profit. As X increases, the value of r increases in a concave fashion until $\log(r)$ becomes 0, and the relative strength of the effects flips. As X is increased further, monopolistic models yield a higher industry profit. Further increasing X causes a convex increase in $\log(r)$.

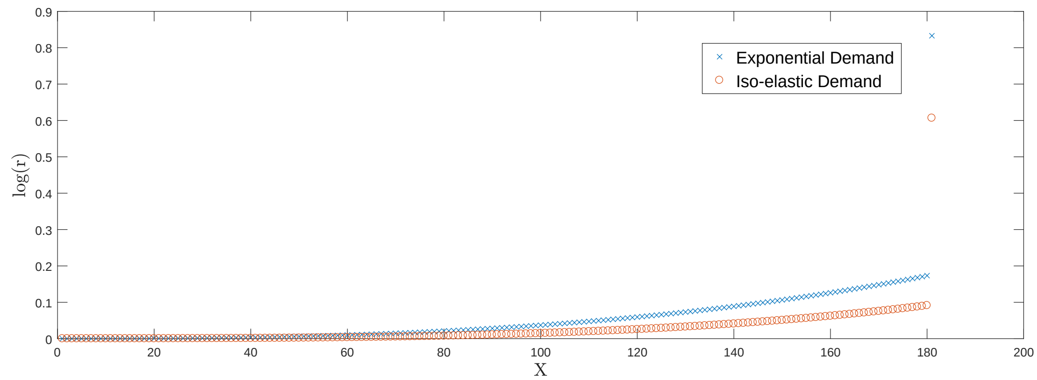


Figure 3.6: $\log(r)$ against X for $T = 375$, very high value of total capacity

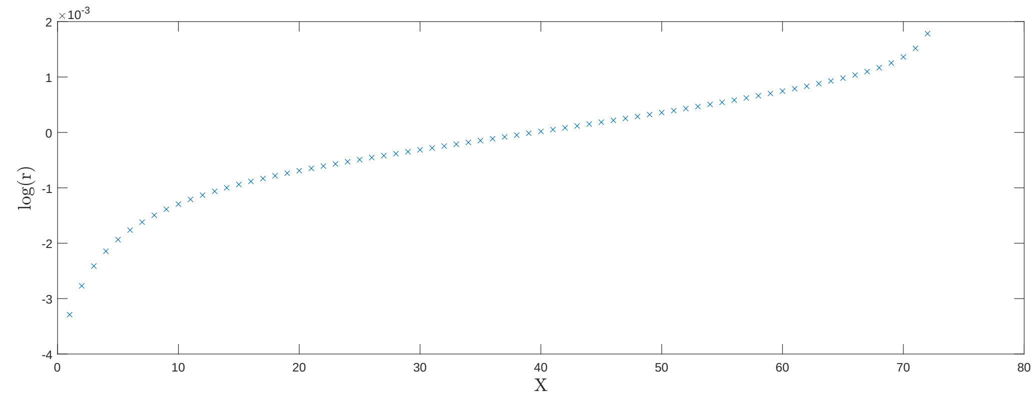


Figure 3.7: $\log(r)$ against X for $T = 375$, Exponential demand, intermediate value of total capacity

5.3 Capacity mismatch individual effect

In Section 5.2, we saw that increased capacity mismatch attenuates the competition effect with respect to total industry profits. But how are these total industry profits split between the two firms? Answering this question will enable us to understand how the equilibrium strategy profile (C, C) compares to the strategy profile (M, M) in terms of payoffs to each firm.

We use the same set of experiments as the ones in Section 5.2 for illustration. However, for brevity, plots are presented only for the exponential demand model. The analogous plots for the iso-elastic demand model are presented in Appendix C.6.

First, we study the case when capacity is very low and total industry profits are higher under strategy profile (C, C). Figure 3.9 plots results for the case of exponential demand. We plot the values of the logarithm of $r = \frac{V_0^{(M,M)}(X,20-X)}{V_0^{(C,C)}(X,20-X)}$ and $r = \frac{V_0^{(M,M)}(20-X,X)}{V_0^{(C,C)}(20-X,X)}$ (the ratio of revenues under monopolistic and competitive models for the lower capacity and higher capacity firm respectively) in adjacent bar graphs, for values of X from 1 to 10. We know that when capacity is low and capacity mismatch is high, competitive models yield larger industry profits (see Figure 3.5). From Figure 3.9, we see that this is attributed to a significant benefit to the higher capacity firm (note

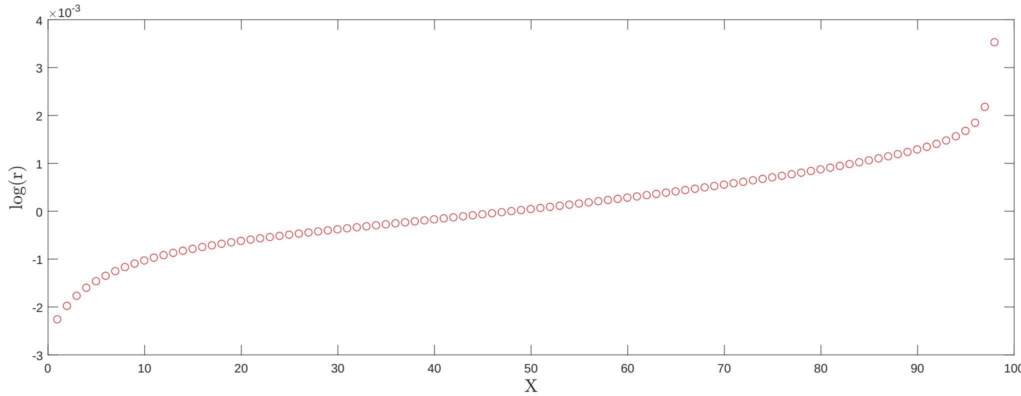


Figure 3.8: $\log(r)$ against X for $T = 375$, Iso-elastic demand, intermediate value of total capacity

that a negative value of $\log(r)$ indicates a preference for competitive models). This outcome is explained by the fact that the information effect favors the higher capacity firm, who benefits from the higher probability that the lower capacity firm runs out of capacity. Again, as the mismatch level decreases (as X increases), the competition effect takes over: the lower capacity firm has less to lose from strategy profile (C, C) and the higher capacity firm has commensurately less to gain from strategy profile (C, C).

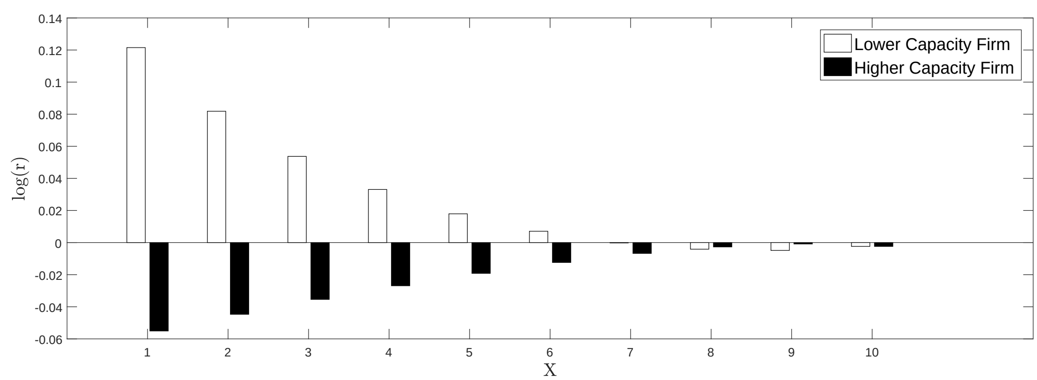


Figure 3.9: $\log(r)$ against X for $T = 375$, total capacity = 20, Exponential demand

Next, we study the case when capacity is very high and total industry profits are better under strategy profile (C, C). In Figure 3.10, we plot the values of the logarithm of $r = \frac{V_0^{(M,M)}(X, 181-X)}{V_0^{(C,C)}(X, 181-X)}$ and $r = \frac{V_0^{(M,M)}(181-X, X)}{V_0^{(C,C)}(181-X, X)}$ (the ratio of revenues under monopolistic and competitive models for the lower capacity and higher capacity firm respectively) in adjacent bar graphs against values of X going from 1 to 181. For visual clarity, we vary X with an interval size of 15 until $X = 166$ and an interval of 1 for $166 < X \leq 181$. The plot indicates that under very high capacity and any amount of capacity mismatch, the high capacity firm prefers monopolistic models while the low capacity firm prefers competitive models. Here, the competition effect is stronger than the information effect, and impacts the higher capacity firm more severely. Furthermore, the stronger competition effect favors the lower capacity firm, because she tends to get rid of capacity first (this same effect has been noted in Martínez-de Albéniz and Talluri [2011] and Dudey [1992] as well); therefore, the lower capacity firm prefers the competitive model. Interestingly, this continues to be true until the capacities are exactly matched, at which point both firms prefer the monopolistic model.

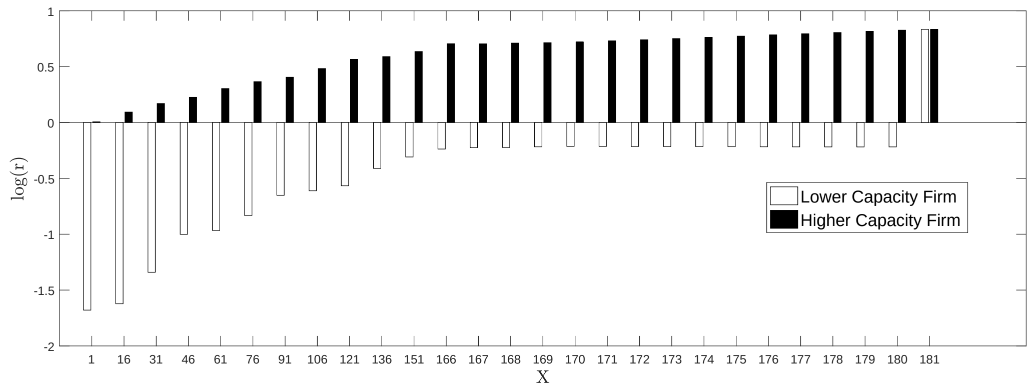


Figure 3.10: $\log(r)$ against X for $T = 375$, total capacity = 362, Exponential demand

What happens for intermediate values of capacity? In Figure 3.11, we consider a total capacity value of $X = 72$; we vary X with an interval of 10 units until $X = 61$ and 1 unit thereafter. (We truncate the vertical axis for visual clarity.) The general observations from the high capacity regime (Figure 3.10) continue to hold: the higher capacity firm prefers monopolistic models while the lower capacity firm prefers competitive models, except when there is no capacity mismatch.

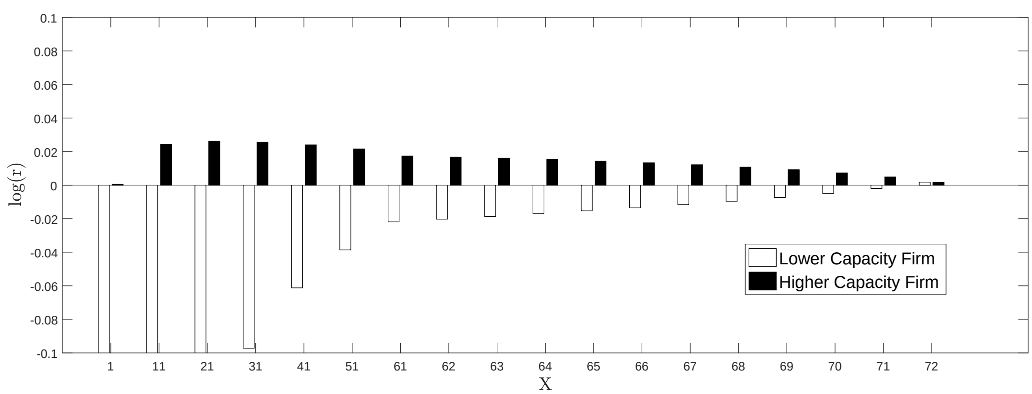


Figure 3.11: $\log(r)$ against X for $T = 375$, total capacity = 76, Exponential demand

In summary, when total capacity is not very low, the lower capacity firm prefers to use competitive models when capacities are mismatched. The higher capacity firm prefers to use monopolistic models, i.e., she would prefer an off-equilibrium strategy. When the total capacity is significantly low, and the level of capacity mismatch is not low, this effect flips. When capacities are symmetric, the equilibrium strategy (C, C) is favorable to both firms only when the total capacity is very low.

6 Extensions

In our base model, we made various simplifications in the interest of expositional convenience. However, some of these assumptions may easily be relaxed without altering the nature of our results.

1. **Asymmetric demand function:** In our base model, we assumed that if $D(p_{me}, p_c)$ represents the probability that me makes a sale under given prices (p_{me}, p_c), $D(p_c, p_{me})$ would represent the probability that c makes a sale under prices (p_{me}, p_c). However, we can make a simple modification to have $D^{me}(p_{me}, p_c)$ represent the probability that me makes a sale and $D^c(p_{me}, p_c)$ represent the probability that c makes a sale. This change continues to allow the existence of reservation prices and hence an equilibrium, as in Section 4.2. To show that an SPNE-P continues to exist as asserted by Proposition 16, the proof of Lemma 3 in Appendix C.2 can be employed directly: this proofs shows, from the perspective of one of the firms, that there exists a threshold price below which that firm will not continue to undercut.

2. **More than two firms:** In our base model, we consider two competing firms. However, this can easily be extended to an oligopolistic situation with more than two firms. The required modifications to our proof closely follow those in Martínez-de Albéniz and Talluri [2011]. The equilibrium characterized in Section 4.2 can be easily extended: each firm continues to have a reservation price below which it will not undercut. In equilibrium, the firm with the lowest reservation price (potentially) makes the sale by undercutting the second lowest reservation price optimally. All other firms choose their reservation prices.

7 Conclusions and Future Work

Prior revenue management and pricing literature has established the need for firms engaging in a dynamic pricing game to explicitly account for the existence of competition in their optimization. This chapter contributes to the growing literature that studies aspects of competition modeling. We posit a general and flexible model of demand that unifies many realistic aspects that determine customer demand in a practical setting: (i) the inherent uncertainty in market size, (ii) uncertainty in customer valuations, and (iii) the existence of both loyal customers (who do not treat the products on offer as perfect substitutes) and flexible customers (who treat the products on offer as perfect substitutes and whose purchases are guided purely by prices). To the best of our knowledge, there is no existing work that unifies all these practically relevant aspects of customer demand. Under our general model of demand, we find that the presence of any loyal customers precludes the existence of a subgame perfect Nash equilibrium in pure strategies (SPNE-P). However, when all customers are flexible, we prove the existence of a continuum of SPNE-P, and propose a refinement that results in a unique equilibrium.

Using this equilibrium, we study the strategic question of whether competing firms should use monopolistic models or competitive models. We argue that in general, it is an equilibrium strategy for both competing firms to use competitive models. We then compare the resulting payoffs—at the industry level and the firm level— to the status-quo payoffs wherein both firms use monopolistic models. The payoffs are driven by two competing effects: the *competition effect* that drives prices and profits down, and the *information effect* that drives profits up. We find that when firms are roughly symmetric in their capacity and capacity is not very low, industry profits are hurt by a migration to competitive models because prices are driven down. However, when firms are highly mismatched in their capacities, this migration can improve industry profits as long as total capacity is not very high. When total capacity is not very low, the lower capacity firm is aided by this migration because she ends up capturing all the customers during the early part of the horizon, at a higher price, as the higher capacity firm seeks to drive her out of the market. The higher capacity firm, though, is ultimately harmed by the migration. On the other hand, when total capacity is very low, the higher capacity firm favors competitive models.

In the general symmetric case (when total capacity is not very low) the firms are *both* hurt by migration to competitive models. Therefore, this situation resembles a prisoners' dilemma: these firms can sustain the strategy of both using monopolistic models in a repeated-game setting under the credible threat that they will deviate if any one firm deviates. Therefore, through tacit collusion, both firms can protect their profitability in this situation.

In the general asymmetric case, i.e., when total capacity is not very low, the lower capacity firm would do well to invest in technology and migrate to using competitive models. The higher-capacity firm will be forced to follow in equilibrium. Therefore, the lower-capacity firm can view these sophisticated revenue management models as a strategic substitute for capacity shortage.

This chapter builds on prior literature, and, for a practically relevant model of demand, provides guidance for firms selling perfectly substitutable products on the effect of migrating to competitive models of dynamic pricing. Under our model of demand, existing equilibrium concepts fail to provide similar guidance to firms selling *imperfectly* substitutable products (i.e., products for which customer loyalty is a salient feature). Finding and analyzing a plausible equilibrium definition for this situation is an interesting avenue for future research. This chapter also assumes that forecasts are given exogenously to the pricing exercise. Another interesting avenue for future research would be to incorporate forecast updates into the dynamic pricing exercise, extending the work of Cooper et al. [2015] to a multi-period setting with capacity constraints.

4

Conclusion

In this dissertation, we study three problems related to the regulation and management of innovative technologies. Using tools from game theory, queueing theory and optimization, and a mixture of analytical modeling and numerical analysis, this dissertation provides guidance and managerial insights for regulators (in Chapter 1) and for firms (in Chapters 2 and 3).

In Chapter 1, we analyze the contentious regulatory issue of how a socially interested utility regulator should design utility tariffs for markets with rooftop solar. We detail the various complexities that the regulator must take into account when designing the tariff: the implications on utility profits, on solar company profits, and on customer welfare and equity, with a view to eliminate unfavorable *cross-subsidization* among customers. We contribute to the policy discourse by creating and analyzing a comprehensive model—cast as an optimization problem—that folds in all these regulatory considerations.

Our analysis reveals that two features are critical to effective tariff design: the ability to discriminate among customer usage tiers and the ability to discriminate between customers with and without rooftop solar. Interestingly, the tariffs used in most states lack at least one of these features. We then present a tariff with both these features and demonstrate, both analytically and numerically, that it outperforms existing tariff structures that lack at least one of the identified critical features. Using our model, utility regulators can compare the implications of using various tariff options, and use our proposed tariff as a basis for designing tariffs for their own states.

In Chapter 2, we take the perspective of a tech-savvy service provider L, and study whether she should initiate announcing real-time delay information when her competitor F could follow. No prior work studies competing firms providing (potentially) heterogeneously-rich delay announcement strategies, chosen endogenously. We fill this gap in the delay-announcement and queueing literature.

To investigate the impact of L initiating delay announcements on her market share, we model queueing systems in three different information regimes. We add to the queueing literature by analyzing the queueing dynamics under these three information regimes, and analytically characterize performance in these regimes employing sophisticated techniques such as coupling and the analysis of dynamical systems. We also conjecture market share bounds that apply to more generally to the Join-the-shortest queue-like system for queues with asymmetric service rates.

Our main finding is that L's announcement decision depends on her service capacity relative to F's: she should initiate delay announcements *only if* she is *not* the higher-capacity service provider. This has an important implication for firms seeking to invest in delay-announcement infrastructure: if they have higher capacity than their competition, initiating real-time delay announcements could cause an erosion in market share. However, if they do not have higher capacity than their competition, real-time delay announcements can be treated as a strategic lever to mitigate capacity shortage.

In Chapter 3, we study the question of whether competing firms (such as airlines) selling a finite number of units of product over a finite horizon should continue with the status-quo of using monopolistic models for their pricing decisions, or whether they should migrate to using competitive models. We contribute to the growing literature that studies revenue management and pricing under competition by unifying the various demand models used in prior literature to capture three realistic considerations: (i) Uncertain market size; (ii) Uncertain customer valuations; and (iii) The presence of loyal and flexible customers. We show that (the popularly used) subgame-perfect Nash equilibrium in pure strategies fails to exist in the presence of any loyal customers. When all customers are flexible (the fully-flexible case), we show the existence of such an equilibrium and provide its structure.

For the fully-flexible case, we then study the strategic question of whether firms who, in the status-quo, use monopolistic models, should migrate to using competitive models. We establish that the unique equilibrium strategy is for both competing firms to migrate to using competitive models. However, in general, this migration only benefits the firm with lower capacity. The implication is that a firm with lower capacity should invest in sophisticated revenue management technology in a bid to improve her profits, even though she can expect that market forces will drive her competitor to also migrate.

There is much scope for further exploration on these and related topics. Related to Chapter 1, there are open policy and tariff design questions associated with other sources of distributed energy, such as community solar arrays. These are seen as viable alternatives to rooftop solar and enjoy economies-of-scale benefits that rooftop solar does not. Chapter 2 studies real-time delay announcements provided by a firm to customers. Various other situations exist in which real-time delay information is disseminated; for instance, ride-sharing platforms such as *Uber* routinely provide real-time delay information to their riders, and *also* give drivers real-time heat maps, as an indication of their time-to-match. From the perspective of a firm like *Uber*, the optimal information disclosure policy is an open problem. Chapter 3 also provides interesting avenues for future research – notably, a plausible equilibrium concept and the associated analysis for a situation with loyal customers is a worthy line of inquiry.

A

Additional Material for Chapter 1

A.1 Proof of Proposition 1

First, notice that R's objective function does not depend directly on $T(\cdot)$, but only on z. Therefore, if she can choose a value of z that maximizes her objective function, and find M and $T(\cdot)$ that satisfy her constraints (and therefore induce adoption outcome z), her ability to maximize her objective is not compromised by this restriction. We now investigate whether, under a fixed value of z, R's ability to satisfy her constraints is compromised by this restriction. Consider a solution that does not follow this restriction. We now show that this solution can be converted to a solution that respects this restriction and continues to respect all R's constraints.

Let O be the set of indices i such that $\exists j : t(i) = t(j)$. Partition O into O_n, the set of indices in O corresponding to non-adopting classes, O_{an}, the set of indices in O corresponding to adopting classes that are not marginal classes, and O_{am}, the set of indices in O corresponding to adopting classes that are marginal. Now, carry out the following steps.

1. Sequentially, for each $i \in O_n$, decrease $t(i)$ by a small value $\epsilon_i > 0$ by appropriately increasing $T(d'_i, e_i, 1)$. Note that doing this does not affect (1.11) because this equation is only affected by $T(d_i, 0, 0)$ for non-adopting classes. Further, since $\epsilon_i > 0$, we can ensure that the tariff restriction is respected, the inequalities in (1.13) continue to be respected, and the inequalities in (1.15) continue to be respected.

2. Sequentially, for each $i \in O_{an}$, decrease $t(i)$ by a small value $\epsilon_i > 0$ by appropriately decreasing

$T(d_i, 0, 0)$. Note that doing this does not affect (1.11) because this equation is only affected by $T(d'_i, e_i, 1)$ for adopting classes. Further, if ϵ_i is small enough and appropriately chosen, we can ensure that the tariff restriction is respected, the inequalities in (1.13) continue to be respected, and the inequalities in (1.14) continue to be respected because $t(i)$ is still larger than $t(m)$.

Now, consider the set O_{am}. Of these, choose exactly one class m to be the marginal class. For all $i \in O_{am} \setminus m$, sequentially increase $t(i)$ by a small value ϵ_i by appropriately increasing $T(d_i, 0, 0)$. Note that doing this does not affect (1.11) because this equation is only affected by $T(d'_i, e_i, 1)$ for adopting classes. Further, if ϵ_i is small enough and appropriately chosen, we can ensure that the tariff restriction is respected. Since these $t(i)$ values are larger than $t(m)$, the incentive compatibility constraints (1.13) corresponding to these classes will now be respected for a small enough ϵ_i, because if S sets a price $t(i)$, since $t(i)$ is now larger than $t(m)$ class m does not adopt, therefore decreasing the quantity in the summation term of (1.13) by at least $h_m g_m$. Naturally, (1.14) continues to hold as $t(i)$ is now larger than $t(m)$. ■

A.2 Proof of Proposition 2

We prove Proposition 2 by showing the following counter-example with three classes of customers. Under this tariff structure, we have

$$t(i) = \frac{d_i r_d + r_0 - ((d_i + e_i - g_i)\, s_d + e_i s_e + s_0)}{g_i}.$$

Let z^* specify that classes 2 and 3 adopt and class 1 does not adopt. Let $h_1 = 4000, h_2 = 250, h_3 = 1000, d_1 = 1000, d_2 = 2000, d_3 = 3000, g_1 = 500, g_2 = 1000, g_3 = 1500, e_1 = 200, e_2 = 400, e_3 = 600$. R now has the choice between two possible orderings: Under ordering o_1, $t(1) < t(2) < t(3)$, and $m(o_1)$ is 2, while under ordering o_2, $t(1) < t(3) < t(2)$, and $m(o_2)$ is 3. We show that $\mathcal{P}_2$ is infeasible under both these orderings by showing that either the ordering constraints or the incentive compatibility constraints fail to hold.

System of constraints under o_1:

$$\frac{1000r_d - 700s_d - 200s_e + r_0 - s_0}{500} < \frac{2000r_d - 1400s_d - 400s_e + r_0 - s_0}{1000} < \frac{3000r_d - 2100s_d - 600s_e + r_0 - s_0}{1500} \tag{A.1}$$

$$1750000\left(\frac{2000r_d - 1400s_d - 400s_e + r_0 - s_0}{1000} - c_s\right) = \Delta_S \tag{A.2}$$

$$3750000\left(\frac{1000r_d - 700s_d - 200s_e + r_0 - s_0}{500} - c_s\right) < \Delta_S \tag{A.3}$$

$$1500000\left(\frac{3000r_d - 2100s_d - 600s_e + r_0 - s_0}{1500} - c_s\right) < \Delta_S \tag{A.4}$$

Solving for $r_0 - s_0$ using equation (A.2), and substituting in (A.3) and (A.4), we obtain:

$$5250000\,(5c_s - 10r_d + 7s_d + 2s_e) + 23\Delta_S < 0$$
$$700000\,(5c_s - 10r_d + 7s_d + 2s_e) + 3\Delta_S > 0$$

Let $\frac{(5c_s - 10r_d + 7s_d)}{2} = v$ The above inequalities simplify to $s_e + v < \frac{-23}{2(5250000)}\Delta_S = -2.19048 \cdot 10^{-6}\Delta_S$ and $s_e + v > \frac{-3}{2(700000)} = -2.14286 \cdot 10^{-6}\Delta_S$, which is not possible, because $\Delta_S > 0$. Therefore, $\mathcal{P}_2$ is infeasible under ordering o_1.

System of constraints under o_2:

$$\frac{1000r_d - 700s_d - 200s_e + r_0 - s_0}{500} < \frac{3000r_d - 2100s_d - 600s_e + r_0 - s_0}{1500} < \frac{2000r_d - 1400s_d - 400s_e + r_0 - s_0}{1000} \tag{A.5}$$

$$1750000\left(\frac{3000r_d - 2100s_d - 600s_e + r_0 - s_0}{1500} - c_s\right) = \Delta_S \tag{A.6}$$

$$3750000\left(\frac{1000r_d - 700s_d - 200s_e + r_0 - s_0}{500} - c_s\right) < \Delta_S \tag{A.7}$$

$$250000\left(\frac{2000r_d - 1400s_d - 400s_e + r_0 - s_0}{1000} - c_s\right) < \Delta_S \tag{A.8}$$

Solving for $r_0 - s_0$ using equation (A.6), and substituting in (A.5), we obtain:

$$3c_s - 4r_d + \frac{14s_d}{5} + \frac{4s_e}{5} + \frac{3\Delta_S}{1750000} < c_s + \frac{\Delta_S}{1750000} < \frac{3c_s}{2} - r_d + \frac{7s_d}{10} + \frac{s_e}{5} + \frac{3\Delta_S}{3500000}$$

However, $c_s + \frac{\Delta_S}{1750000} < \frac{3c_s}{2} - r_d + \frac{7s_d}{10} + \frac{s_e}{5} + \frac{3\Delta_S}{3500000}$ can be rewritten as $350000\,(5c_s - 10r_d + 7s_d + 2s_e) + \Delta_S > 0$ and $3c_s - 4r_d + \frac{14s_d}{5} + \frac{4s_e}{5} + \frac{3\Delta_S}{1750000} < \frac{3c_s}{2} - r_d + \frac{7s_d}{10} + \frac{s_e}{5} + \frac{3\Delta_S}{3500000}$ can be rewritten as $350000\,(5c_s - 10r_d + 7s_d + 2s_e) + \Delta_S < 0$, which contradict each other. Therefore, $\mathcal{P}_2$ is also infeasible under ordering o_2. ■

A.3 Proof of Proposition 3

To prove this property, we choose a number $\epsilon > 0$ and consider a net-metering tariff system (where customers pay a variable charge proportional to their net energy usage). We choose the ordering $t(i) < t(j), \forall i < j$, which is consistent with classes $m, m+1, \ldots, I$ being adopters. Setting $r_d = s_d$, $s_e = -s_d$, and $r_0 = s_0 - \epsilon$, the ordering constraints $t(i) < t(j)$ simplify to $s_d - \frac{\epsilon}{g_i} < s_d - \frac{\epsilon}{g_j}$, which is true, because $g_i < g_j$. This leaves us with the following system of equations.

$$\sum_{i=1}^{I} h_i \left(s_i^{(z^*)}((s_d(d_i - g_i + e_i) + s_e e_i + s_0 - c_u^{z^*}(d_i - g_i)) + (1 - s_i^{(z^*)})((r_d - c_u^{z^*})d_i + r_0) \right)$$
$$- \sum_{i=1}^{I} h_i(p_{r0} - c_u^{E_0})d_i = \Delta_U$$
$$(t(m(o)) - c_s) \sum_{i=1}^{I} s_i^{(z^*)} h_i g_i = \Delta_S$$
$$r_d = s_d$$
$$s_e = -s_d$$
$$r_0 = s_0 - \epsilon$$

This is a linear system of the form $Ax = b$ with five unknowns: r_d, s_d, s_e, r_0, and s_0. Since the rows of A are linearly independent, there exists a solution x to this equation. Therefore $\mathcal{P}_2$ is feasible under this ordering. ■

A.4 Proof of Proposition 4

First, we rewrite the set of constraints that must be satisfied to obtain a feasible solution to $\mathcal{P}_2$.

$$t(i) = \frac{T(d_i, 0, 0) - T(d'_i, e_i, 1)}{g_i}, \forall i \tag{A.9}$$

$$\sum_{i=1}^{I} h_i \left(s_i^{(z^*)}(T(d'_i, e_i, 1) - c_u^{(z^*)}(d_i - g_i)) + (1 - s_i^{(z^*)})(T(d_i, 0, 0) - c_u^{(z^*)} d_i) \right) - \sum_{i=1}^{I} h_i(p_{r0} - c_u^{E_0})d_i = \Delta_U \tag{A.10}$$

$t(i)$ ordering consistent with o (A.11)

$$(t(a(o)) - c_s) \sum_{i=1}^{I} s_i^{(z^*)} h_i g_i = \Delta_S \tag{A.12}$$

$$(t(i) - c_s) \sum_{j=1}^{I} \mathbb{I}_{t(j) > t(i) \text{in ordering } o} h_j g_j < \Delta_S, \forall i \neq a(o) \tag{A.13}$$

This proof proceeds in two parts. In the first part, we will show that by ignoring the set of equations (A.9) and (A.10), we can always find a set of $t(i)$ values that satisfy (A.11)-(A.13). In the second part, we will show that corresponding to the set of $t(i)$ values found, we can find values for $r_c, \forall c \in C$ and f that satisfy (A.9) and (A.10) .

Part 1: First, re-index the usage tiers so the index matches the ordering o. Note that after re-ordering, we no longer have the property that $g_i < g_{i+1} \forall i \in \{1, 2, \ldots, I-1\}$. Now, observe that

(A.11) can be rewritten as $t(i) < t(i+1), \forall i \in \{1, 2, \ldots, I-1\}$. Equation (A.12) can be rewritten as $(t(a(o)) - c_s) \sum_{i=a(o)}^{I} h_i g_i = \Delta_S$ and constraints (A.13) can be rewritten as $(t(i) - c_s) \sum_{j=i}^{I} h_j g_j < \Delta_S \forall i \neq a(o)$. With some manipulation, we obtain:

$$t(i) < t(i+1) \forall i \in \{1, 2, \ldots, I-1\} \tag{A.14}$$

$$t(a(o)) = c_s + \frac{\Delta_S}{\sum_{i=a(o)}^{I} h_i g_i} \tag{A.15}$$

$$t(i) < c_s + \frac{\Delta_S}{\sum_{j=i}^{I} h_j g_j}, \forall i \neq a(o) \tag{A.16}$$

The set of inequalities in (A.16) provide upper bounds on all the $t(i)$ values, and (A.15) pins down the value of $t(a(o))$. Notice that these upper bounds are increasing in i, because $\Delta_S > 0$, and $\sum_{j=i}^{I} h_j g_j$ is decreasing in i. Therefore, there exist values $t(i) \forall i \in \{1, 2, \ldots, I-1\}$ that respect (A.14) and the specified upper bounds.

Part 2: Once a set of values $t(i)$ is found, we need to map them to rate class tariffs r_c. To do this, we examine the set of equations (A.9). Each of these I equations takes the form:

$$t(i) = \frac{r_m d_i + f - (r_n(d_i - g_i) + f)}{g_i} = r_m(\frac{d_i}{g_i}) + r_n(\frac{d_i - g_i}{g_i}), \tag{A.17}$$

where m is the index of the rate class corresponding to net usage level d_i, and n is the rate class corresponding to net usage level $d_i - g_i$. Taken together, these I equations constitute an under-determined linear system of the form $A\vec{r} = \vec{t}$, where r is a $|C|$ dimensional vector of r_c values, $\vec{t}$ is an I dimensional vector of $t(i)$ values, and A is a matrix with I rows and $|C| > I$ columns. If the rows of A are linearly independent, then the system has an infinite number of solutions. For the sake of contradiction, assume that the rows are linearly dependent. Then, there must exist an I dimensional vector $\vec{\lambda} \neq \vec{0}_{I \times 1}$ such that $\vec{\lambda}^T A = \vec{0}_{1 \times |C|}$. Let $m_1 = \arg\max_i \; d_i$. Let c_1 be the column in A corresponding to the rate class into which a tier m_1 customer would fall if they did not adopt solar. This column has exactly one non-zero entry because no other household can fall into this rate class, whether they adopt solar or not. Let w_1 be the index of the row in A that contains this non-zero entry. Then, it must be that the w_1^{th} entry of $\vec{\lambda}$ is 0 in order for the c_1^{th} entry of $\vec{\lambda}^T A$ to be 0. Therefore, the vector $\vec{\lambda}^T A$ is unaltered if we replace all entries in the w_1^{th} row of A by zero. Now, this same argument can be applied repeatedly to assert that all other entries of $\vec{\lambda}$ must also be zero: Choose $m_2 = \arg\max_{i \neq m_1} \; d_i$, find the index c_2 of the column corresponding to the rate class to which a tier m_2 customer would belong if it did not adopt solar, and observe that exactly one row corresponding to this column now has a non-zero entry (recall that we changed all entries in row w_1 to 0). Let w_2 be the index of this row. We can assert that the w_2^{th} entry of $\vec{\lambda} = 0$. By repeating this procedure, we can assert that all entries of $\vec{\lambda}$ are 0. This contradicts our assumption, and therefore, the rows of A *are* linearly independent. Therefore, we can obtain values of r_c consistent with the equations (A.17). These values can then be substituted in (A.10), using which we can find a feasible value of f. ■

A.5 Proof of Proposition 5

We prove Proposition 5 by showing the following counter-example with three classes of customers. Let $h_1 = 445, h_2 = 218, h_3 = 1000, d_1 = 500, d_2 = 681, d_3 = 1024, g_1 = 100, g_2 = 181, g_3 = 343, p_{r0} = 0.1, \Delta_S = 1, c_s = 3/40$. With these parameters, there are four possible values of net demand that a household can have:

1. A class 1 household adopts: Net demand $= d_1 - g_1 = 400$kWh.
2. A class 1 household does not adopt, or a class 2 household adopts: Net Demand $= d_1 = d_2 - g_2 = 500$kWh.
3. A class 2 household does not adopt, or a class 3 household adopts: Net Demand $= d_2 = d_3 - g_3 = 681$ kWh.
4. A class 3 household does not adopt: Net Demand $= d_3 = 1024$ kWh.

Therefore, U's rate schedule must specify 4 different rate classes that apply at each of these net demand levels. Let the rates corresponding to these rate class levels be r_1, r_2, r_3, and r_4 respectively. Accordingly, the set of $t(i)$ values is given by the following expression:

$$t(1) = r_1 + (r_2 - r_1)\frac{d_1}{g_1}$$
$$t(2) = r_2 + (r_3 - r_2)\frac{d_2}{g_2}$$
$$t(3) = r_3 + (r_4 - r_3)\frac{d_3}{g_3}$$

Let Δ_U be chosen so that $\Delta_C = 0$. Therefore, customers as a whole gain exactly 0. We are interested in seeing how close this tariff structure can come to being CS-free. In particular, can customer classes 1, 2, and 3 be financially worse off by an amount arbitrarily close to 0 after solar adoption?

Let every individual household in class 1 be worse off by an amount a_1, and every individual household in class 2 be worse off by an amount a_2. Because $\Delta_C = 0$, we have that every household in class 3 is worse off by exactly $-\frac{a_1 h_1 + a_2 h_2}{h_3}$. Let z^* specify that class 1 adopts, while classes 2 and 3 do not adopt. Since class 1 is the marginal customer, $r_2 d_1 + f = r_1(d_1 - g_1) + f + t(1)g_1$: The class 1 customer is equally worse off whether they adopt solar or not. Accordingly, we can write the following system of equations:

$$r_2 d_1 + f = p_{r0} d_1 + a_1 \tag{A.18}$$

$$r_3 d_2 + f = p_{r0} d_2 + a_2 \tag{A.19}$$

$$r_4 d_3 + f = p_{r0} d_3 - \frac{a_1 h_1 + a_2 h_2}{h_3} \tag{A.20}$$

We solve these equations for r_2, r_3, and r_4. Next, we use the following profit equation to obtain r_1:

$$\Delta_S = h_1 g_1 \left(r_1 + (r_2 - r_1)\frac{d_1}{g_1} - c_s \right) \tag{A.21}$$

Substituting r_2 from equation (A.18) into equation (A.21), we obtain an equation for r_1. Using these expressions for r_1, r_2, r_3, and r_4, we obtain the following expressions for the $t(i)$ values.

$$t(1) = \frac{6677}{89000}$$
$$t(2) = \frac{-10a_1 + 10a_2 + 181}{1810}$$
$$t(3) = -\frac{89a_1}{68600} - \frac{87a_2}{24500} + \frac{1}{10}$$

Now, consider the two possible orderings that R could induce.

1. $t(3) < t(2) < t(1)$: $t(2) < t(1)$ simplifies to the inequality $a_1 > a_2 + \frac{402363}{89000}$. Therefore, a and b cannot both be made arbitrarily close to 0.

2. $t(2) < t(3) < t(1)$: Since this ordering also contains the inequality $t(2) < t(1)$, this ordering also does not allow R to induce a CS-free outcome.

Therefore, this tariff structure does not allow R to induce a CS-free outcome. ■

A.6 Proof of Proposition 7

Corresponding to an adoption outcome z^*, we will choose the ordering yielding the required z^* whereby the non-adopters are arbitrarily arranged to have the lowest indices, the marginal adopter is the adopting class with the lowest value of g_i and all other adopters are arranged in increasing order of g_i[1]. Under this re-indexed system, if m is the index of the marginal adopter, set $t(m) = c_s + \frac{\Delta_S}{\sum_{i=m}^{I} h_i g_i}$. This leads to $f_m = g_m(p_r - t(m))$. For all indices $i < m$, set $t(i) = \epsilon i$ for some small ϵ, and appropriately choose the fixed costs f_i that would lead to this. For all indices $i > m$, set $f_i = \frac{f_m}{g_m} g_i - (i - m)\epsilon g_i$, leading to values $t(i) = p_r - \frac{f_m}{g_m} + (i - m)\epsilon$. Clearly, this schedule of fixed costs respects the ordering that the $t(i)$ values are required to have. Further, the incentive compatibility constraints are respected: For a small enough ϵ, setting a solar price of $t(i), i < m$ brings S negative

[1]Recall that there are $i!j!$ such orderings where $i + j = I$, i classes do not adopt, and j classes do adopt.

profit, and setting a price $t(i), i > m$ increases her margin by an arbitrarily small amount $(i - m)\epsilon$, but brings lower volumes, because $\sum_{j=m}^{I} h_j g_j > \sum_{j=i}^{I} h_j g_j, \forall i > m$.

Now, let us examine what happens to the cash outflow of every tier under this schedule of rates. For a non-adopting tier $i < m$, the decrease in cash outflow (and therefore benefit to a customer in the tier) is $(p_{r0} - p_r)d_i$. For an adopting tier, the decrease in cash outflow is $p_{r0} d_i - (p_r(d_i - g_i) + f_i + p_s g_i) = (p_{r0} - p_r)d_i + p_r g_i - f_i - p_s g_i > (p_{r0} - p_r)d_i$ from Lemma 1.

Accordingly, it is sufficient for us to now show that $p_r < p_{r0}$ under this schedule when $\Delta_C > 0$, and $p_r = p_{r0} + \delta$, for some arbitrarily small $\delta > 0$ when $\Delta_C = 0$.

$$\Delta_C = \sum_1^I h_i(p_{r0} - p_r)d_i + \sum_m^I h_i(-f_i + \frac{f_m}{g_m} g_i) \quad (\textit{From the proof of Lemma 1,} p_r g_i - f_i - p_s g_i = -f_i + \frac{f_m}{g_m} g_i)$$

$$\Leftrightarrow \Delta_C = \sum_1^I h_i(p_{r0} - p_r)d_i + \sum_m^I h_i(i - m)\epsilon g_i$$

$$\Leftrightarrow p_r = p_{r0} - \frac{\Delta_C}{\sum_1^I h_i d_i} + \frac{\epsilon \sum_m^I (i - m) h_i g_i}{\sum_1^I h_i d_i}$$

Therefore, since the term $\frac{\epsilon \sum_m^I (i-m) h_i g_i}{\sum_1^I h_i d_i}$ is vanishingly small when $\Delta_C > 0$, $p_r < p_{r0}$. When $\Delta_C = 0$, p_r can be set arbitrarily close to p_{r0} by reducing ϵ. This completes the proof. ■

A.7 Proof of Proposition 8

When $\Delta_C > 0$, Proposition 7 showed that all tiers can be made to benefit from solar adoption. For tiers i that do not adopt solar, the gain is $(p_{r0} - p_r)d_i$. For tiers that do adopt solar, the gain is $(p_{r0} - p_r)d_i + p_r g_i - f_i - p_s g_i$. From Lemma 1, $p_r g_i - f_i - p_s g_i > 0$. Therefore, the smallest gain accrues to the tier with the lowest d_i value (from the $(p_{r0} - p_r)d_i$ term; since all tiers benefit, $p_r < p_{r0}$). If this d_i value corresponds to a tier l that is a non-adopter, a household of this tier gains the most if p_r is as low as possible. Note that the total gain accruing to households (Δ_C) is fixed, and how it is distributed among the tiers is controlled by the choice of p_r and f_i values for adoptive tiers i. By setting f_i values in order to make the $p_r g_i - f_i - p_s g_i$ values arbitrarily close to 0 for adoptive tiers, we can ensure that all of Δ_C is constituted of the $(p_{r0} - p_r)d_i$ terms. We know exactly how to do this from the proof of Proposition 7. Using the schedule specified in the proof of Proposition 7, adoptive tiers face fixed costs $f_i = \frac{f_m}{g_m} g_i - (i - m)\epsilon g_i$. Therefore, for these tiers, $p_r g_i - f_i - p_s g_i = p_r g_i - \frac{f_m}{g_m} g_i + (i - m)\epsilon g_i - \left(p_r - (\frac{f_m}{g_m})\right) g_i = (i - m)\epsilon g_i$. Since this can be made arbitrarily close to 0 by reducing the value of ϵ, this accomplishes our purpose of ensuring that Δ_C is (almost) entirely constituted of the $(p_{r0} - p_r)d_i$ terms. ■

A.8 Summary of Notation

Table A.1: Summary of Notation

Symbol	Definition
S	Solar Provider
U	Utility Company
R	Regulator
i	Customer tier index
I	Number of tiers
d_i	Annual demand of tier i customer
d_i'	Grid usage of tier i customer, defined as $d_i - g_i + e_i$
g_i	Annual generation of tier i customer
e_i	Annual excess of tier i customer
h_i	Number of households in tier i
p_s	Solar price - price of solar electricity per kWh of generation paid to S
c_s	Solar financial cost - levelized cost of solar electricity paid by S (per kWh of generation)
m_s	Solar environmental cost - monetized cost of environmental externalities imposed by solar electricity (per kWh of generation)
c_u^x	Utility financial cost - levelized cost of electricity to utility paid by U (per kWh of generation) when annual supply is x kWh
m_u^x	Utility environmental cost - monetized cost of environmental externalities imposed by utility's electricity mix (per kWh of generation), when annual supply is x kWh
E_0	Total consumer demand (in kWh) in the base case
Δ_E	Amount of energy dependence migrated to rooftop solar
Δ_U	Increase in utility profits from the base case to the post-solar case
Δ_S	Profit made by S in the post-solar case
Δ_C	Increase in total financial welfare to the customer base, going from the base case to the post-solar case
p_{r0}	Per kWh energy rate paid by customers in the base case
s_i	Solar adoption decision taken by tier i customer. 1 indicates adopt and 0 indicates do not adopt. Superscript * represents optimal decision
$T(\cdot)$	Tariff function
τ	Class of tariff functions
$t(i)$	Threshold solar price that makes class i exactly indifferent between adopting and not adopting solar
z	Index into set of possible adoption outcomes
$A^{(z)}$	Adopters under adoption outcome z
M	Set of marginal adopters, that is, set of tiers i that are indifferent between adopting and not adopting
$\mathcal{P}_1$	Regulator's master optimization problem
$\mathcal{P}_2$	Regulator's optimization subproblem

B

Additional Material for Chapter 2

B.1 Proofs

B.1.1 Proof of Proposition 9:

To prove Proposition 9, we show that when $\mu^{(L)} = 1 < \mu^{(F)}$, $M_1^{(L)} > M_0^{(L)}$. The proof that $M_1^{(L)} \geq M_0^{(L)}$ when $\mu^{(L)} = \mu^{(F)}$ closely follows this proof. We aim to show that $\lambda_1^{(L)} > \lambda_0^{(L)}, \forall \mu^{(F)} > \mu^{(L)}$. Let π_i^0 and π_i^1 denote the probabilities of L being in state i in Periods 0 and 1, respectively. L is an $M/M/1$ queue in Period 0 and an $M/M/1/C_1$ queue in Period 1. From the balance equations and $\mu^{(L)} = 1$, we have $\lambda_1^{(L)} = 1 - \pi_0^1$ and $\lambda_0^{(L)} = 1 - \pi_0^0$. Therefore, to prove $\lambda_1^{(L)} > \lambda_0^{(L)}$, we equivalently prove $\pi_0^0 > \pi_0^1$ by contradiction. We divide the feasible $\lambda_0^{(L)}$ values into two regions: $0 < \lambda_0^{(L)} \leq 0.5$ (Case 1) and $0.5 < \lambda_0^{(L)} < 1$ (Case 2). Cases 1-2 cover the entire set of feasible values for $\lambda_0^{(L)}$ (Assumption (1) and eq. (2.5)). In our proof, we use the following inequality, which holds because $\mu^{(F)} > \mu^{(L)}$:

$$\lambda_0^{(L)} < \frac{\Lambda}{2}. \tag{B.1}$$

<u>Case 1</u> ($0 < \lambda_0^{(L)} \leq 0.5$): Let's assume, for the sake of contradiction, that $\pi_0^1 \geq \pi_0^0$, for some $\mu^{(F)} > \mu^{(L)}$. We prove that this assumption leads to the sum of probabilities of states 0 and 1 in

Period 1 being greater than one, $\pi_0^1 + \pi_1^1 > 1$, resulting in a contradiction.

$$\begin{aligned}
\pi_0^1 + \pi_1^1 &= \pi_0^1 + \pi_0^1 \Lambda && \text{, from the balance equations,} \\
&= \pi_0^1 (1+\Lambda) \geq \pi_0^0 (1+\Lambda) && \text{, from the contradiction assumption,} \\
&> \pi_0^0 \left(1 + 2\lambda_0^{(L)}\right) && \text{, from (B.1),} \\
&= \underbrace{\left(1-\lambda_0^{(L)}\right)\left(1+2\lambda_0^{(L)}\right)}_{LB_1}.
\end{aligned}$$

We now show that $LB_1 \geq 1$, which completes the proof for Case 1. At the two $\lambda_0^{(L)}$ extremes in Case 1, $\lambda_0^{(L)} = 0$ and $\lambda_0^{(L)} = 0.5$, $LB_1 = 1$. For $0 < \lambda_0^{(L)} < 0.5$, the first derivative of LB_1 with respect to $\lambda_0^{(L)}$ confirms that $LB_1 \geq 1$ because,

$$\frac{dLB_1}{d\lambda_0^{(L)}} = 1 - 4\lambda_0^{(L)} \Rightarrow \begin{cases} LB_1 \text{ is increasing in } \lambda_0^{(L)} \text{ when } 0 < \lambda_0^{(L)} < 0.25, \\ LB_1 \text{ is decreasing in } \lambda_0^{(L)} \text{ when } 0.25 \leq \lambda_0^{(L)} \leq 0.5. \end{cases}$$

Case 2 ($0.5 < \lambda_0^{(L)} < 1$): Again, let's assume, for the sake of contradiction, that $\pi_0^1 \geq \pi_0^0$, for some $\mu^{(F)} > \mu^{(L)}$. In this case we prove the contradiction assumption results in $\pi_0^1 + \pi_1^1 + \cdots + \pi_{C_1}^1 > 1$:

$$\begin{aligned}
&\pi_0^1 + \pi_1^1 + \cdots + \pi_{C_1}^1 \\
&= \pi_0^1 \left(1 + \Lambda + \cdots + \Lambda^{C_1}\right) = \pi_0^1 \frac{\Lambda^{C_1+1} - 1}{\Lambda - 1} && \text{, from the balance equations and the geometric summation,} \\
&\geq \pi_0^0 \frac{\Lambda^{C_1+1} - 1}{\Lambda - 1} && \text{, from the contradiction assumption,} \\
&> \pi_0^0 \frac{\left(2\lambda_0^{(L)}\right)^{C_1+1} - 1}{2\lambda_0^{(L)} - 1} && \text{, from (B.1) and the fact that the geometric sum is increasing in } \Lambda, \\
&= \underbrace{\left(1-\lambda_0^{(L)}\right) \frac{\left(2\lambda_0^{(L)}\right)^{C_1+1} - 1}{2\lambda_0^{(L)} - 1}}_{LB_2}.
\end{aligned}$$

We now show that $LB_2 \geq 1$ or

$$\left(2\lambda_0^{(L)}\right)^{C_1+1} \geq \frac{\lambda_0^{(L)}}{1-\lambda_0^{(L)}}. \tag{B.2}$$

In this case $2\lambda_0^{(L)} > 1$. Therefore, we can obtain a lower bound on the left hand side of (B.2) by finding a lower bound for its exponent $C_1 + 1$:

$$C_1 + 1 = \left\lfloor \frac{\lambda_0^{(L)}}{1-\lambda_0^{(L)}} + 1 \right\rfloor + 1 \geq \frac{\lambda_0^{(L)}}{1-\lambda_0^{(L)}} + 1 = \frac{1}{1-\lambda_0^{(L)}}.$$

Therefore to prove (B.2), it suffices that we prove

$$\left(2\lambda_0^{(L)}\right)^{\frac{1}{1-\lambda_0^{(L)}}} \geq \frac{\lambda_0^{(L)}}{1-\lambda_0^{(L)}} \Leftrightarrow \underbrace{\frac{1}{1-\lambda_0^{(L)}} \ln\left(2\lambda_0^{(L)}\right)}_{LHS} \geq \underbrace{\ln\left(\frac{\lambda_0^{(L)}}{1-\lambda_0^{(L)}}\right)}_{RHS}. \tag{B.3}$$

When $\lambda_0^{(L)} \to 0.5$ (the minimum value for $\lambda_0^{(L)}$ in Case 2), $LHS \to 0$ and $RHS \to 0$. Now to prove (B.3) we show that LHS grows at least as fast as RHS as $\lambda_0^{(L)}$ increases; that is, we need to show

$$\frac{dLHS}{d\lambda_0^{(L)}} = \frac{1-\lambda_0^{(L)} + \lambda_0^{(L)} \ln\left(2\lambda_0^{(L)}\right)}{\left(1-\lambda_0^{(L)}\right)^2 \lambda_0^{(L)}} \geq \frac{dRHS}{d\lambda_0^{(L)}} = \frac{1}{\left(1-\lambda_0^{(L)}\right)\lambda_0^{(L)}} \Leftrightarrow$$
$$1-\lambda_0^{(L)} + \lambda_0^{(L)} \ln\left(2\lambda_0^{(L)}\right) \geq 1-\lambda_0^{(L)} \Leftrightarrow \lambda_0^{(L)} \ln\left(2\lambda_0^{(L)}\right) \geq 0,$$

which holds because $2\lambda_0^{(L)} > 1$. This completes the proof for Case 2. ■

B.1.2 Matrix Blocks for Model F and Proof of Proposition 11:

Model F (Fig. 2.3(b)) has the following block-tridiagonal transition matrix where all blocks are square matrices of order $C_t + 1$:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{B} & \mathbf{A}_0 & & & \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & & \\ & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix},$$

where

$$\mathbf{B} = \begin{pmatrix} * & \Lambda & & & \\ 1 & * & \Lambda & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & * & \Lambda \\ & & & 1 & * \end{pmatrix}, \mathbf{A}_0 = \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \Lambda \end{pmatrix}, \mathbf{A}_1 = \begin{pmatrix} * & \Lambda & & & \\ 1 & * & \Lambda & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & * & \Lambda \\ & & & 1 & * \end{pmatrix}, \mathbf{A}_2 = \begin{pmatrix} \mu^{(F)} & & \\ & \ddots & \\ & & \mu^{(F)} \end{pmatrix},$$

where "$*$" represents diagonal elements whose values are set such that $\mathbf{Q}$ has zero row sums.

We now prove Proposition 11. Model F, which is a QBD process, is stable if and only if the following ergodicity condition is satisfied [Latouche and Ramaswami, 1999, Theorem 7.2.4]:

$$\boldsymbol{\nu}\mathbf{A}_0 1 < \boldsymbol{\nu}\mathbf{A}_2 1, \tag{B.4}$$

where the row vector $\boldsymbol{\nu} = (\nu_0, \nu_1, \ldots, \nu_{C_t})$ contains the steady-state probabilities of the Markov chain that corresponds to the generator matrix $\mathbf{A} = \mathbf{A}_0 + \mathbf{A}_1 + \mathbf{A}_2$. Matrix $\mathbf{A}$ corresponds to the generator matrix of Model L (Fig. 2.3(a)). Inequality (B.4) simplifies to $\nu_{C_t}\Lambda < \mu^{(F)}$, where $\nu_{C_t}\Lambda = \lambda_t^{(F)}$, the expected arrival rate to F. ■

B.1.3 Proof of Proposition 12:

When $C_1 = 1$, we use the procedure described in Appendix B.2 to derive

$$C_2 = \left\lfloor \frac{\mu^{(F)} + 2\Lambda + 1 - \sqrt{\left(\mu^{(F)}+1\right)^2 + 4\Lambda}}{\mu^{(F)}\left(\mu^{(F)} - 2\Lambda - 1 + \sqrt{\left(\mu^{(F)}+1\right)^2 + 4\Lambda}\right)} + 1 \right\rfloor. \tag{B.5}$$

Using (2.3) (which specifies the Period 0 delay) and (B.5) and simplifying inequalities $D_0^{(F)} < 1$ and $D_1^{(F)} < 1$, which represent $C_1 = C_2 = 1$, we obtain the required condition. ■

B.1.4 Proof of Short-Term Market Share Limits (eq. (2.5)):

From eq. (2.4), we have:

$$M_1^{(L)} = \frac{\Lambda^{C_1} - 1}{\Lambda^{C_1+1} - 1}. \tag{B.6}$$

When $\Lambda \to 0$, we have from eq. (2.3) and $C_1 = \left\lfloor D_0^{(F)} + 1 \right\rfloor$ that $C_1 = 1$. Accordingly, $M_1^{(L)} = \Lambda - 1/(\Lambda^2 - 1) = 1/(\Lambda + 1) \to 1$ as $\Lambda \to 0$. When $\Lambda \to 1 + \mu^{(F)}$, proving that $M_1^{(L)} \to 1/(1 + \mu^{(F)})$ follows from proving that $\lambda_1^{(L)} \to 1$. From eq. (B.6), we have:

$$\lambda_1^{(L)} = \frac{\Lambda^{C_1+1} - \Lambda}{\Lambda^{C_1+1} - 1} = \frac{1 - \frac{1}{\Lambda^{C_1}}}{1 - \frac{1}{\Lambda^{C_1+1}}}$$

As $\Lambda \to 1 + \mu^{(F)}$, $C_1 \to \infty$, and therefore, $\lambda_1^{(L)} \to 1$. ■

B.1.5 Proof of Lemma 2

Consider systems A and B characterized by thresholds C_t and C_t+1, respectively. In both systems, F alternates between "on" and "off" periods. In an on period, which begins when $n^{(L)}$ is at its threshold, F receives arrivals. In an off period, which begins when $n^{(L)}$ drops below the threshold, arrivals do not join F. We denote the lengths of the i^{th} ($i \geq 1$) off periods by $T_i^{(A)}$ and $T_i^{(B)}$ and the lengths of the i^{th} ($i \geq 1$) on periods by $T_i'^{(A)}$ and $T_i'^{(B)}$.

We denote the customer-average delay[1] at F in system j by $D_t^{(F,j)}$. To show that $D_t^{(F,A)} \geq D_t^{(F,B)}$, it is sufficient to compare the virtual workload processes $W_t^{(F,A)}$ and $W_t^{(F,B)}$ during the on periods as only the customers that arrive during F's on period contribute to her average delay. We make

[1]Not a time-average delay.

this comparison by coupling the arrival times and service times in the two systems in a specific way.

Starting with empty systems, the first few arrivals join L in both systems. The two systems are perfectly coupled till the first moment that the number of L's customers in A reaches the threshold C_t; $n^{(L,A)} = C_t$. At this point, the first on period in A starts while it has not yet started in B as L's threshold in system B is higher. Therefore, it is trivial that $T_1^{(B)} > T_1^{(A)}$. We pause system A at this point and let system B run till it also finishes its first off period by drawing the service times and arrivals for the additional customers arriving to L in B from a separate stack of service times and inter-arrival times.

Now, we ensure that three properties hold using a carefully crafted coupling procedure:

1. $T_i^{(B)} \geq T_i^{(A)}, i \geq 2$: For the next off periods, $i \geq 2$, we redraw the random exponential service times of all current customers in L (if any) for systems A and B at the onset of the i^{th} off period; at this moment $n^{(L,B)} = C_t$ and $n^{(L,A)} = C_t - 1$. We couple the service times of the first $C_t - 1$ customers in the two systems and we redraw an additional random service time for the extra customer in B. Our procedure is permissible due to the fact that service times are i.id. and exponential. During the off period, we couple the arrival times and service times of the new arrivals to L in systems A and B.

2. $T_i'^{(B)} = T_i'^{(A)}, i \geq 1$: The length of an on period is equal to the service times of the threshold customers in L in A and B. By redrawing and coupling the service times of the threshold customers in L at the onset of an on period, the lengths of the on periods in the two systems will be equal.

3. F receives the same amount of workload during $T_i'^{(B)}$ and $T_i'^{(A)}$: This is achieved by using point 2 and coupling the arrival times and service times of new arriving customers to F in A and B in the i^{th} on period.

Based on points 1 and 3 we can conclude that the amount of work in F at the *onset* of the i^{th} on period, $\forall i$, is at least as high in A as it is in B. Furthermore, since we couple the arrivals and service times of customers in systems A and B in the on periods (as explained in points 2 and 3), it follows that the amount of work in F at any time *during* an on period is at least as high in A as it is in B. Therefore, a customer arriving in an on period encounters at least as much delay in A as in B. From the stochastic ordering that this coupling implies, it follows that $D_t^{(F,A)} \geq D_t^{(F,B)}$.

B.1.6 Proof of Proposition 13

In what follows, we assume that F remains stable in all periods (i.e., its expected delay in any period is finite). Since $\lambda_t^{(F)}$ is decreasing in C_t and the stability condition for F is $\lambda_t^{(F)} < \mu^{(F)}$, this means C_t never takes a value low enough that $\lambda_t^{(F)} \geq \mu^{(F)}$. Therefore, $C_t = \left\lfloor D_{t-1}^{(F)} + 1 \right\rfloor, \forall t$ is a finite natural number.

For a specific parameter setting, let $\underline{C} \geq 1$ denote the minimum value that we observe for C_t over all periods and suppose $\underline{C}$ is observed for the first time in Period $\underline{t}$; i.e., $C_{\underline{t}} = \underline{C}$. Due to the stability of F, $\underline{C}$ results in a finite delay $\underline{D}^{(F)}$ that leads to a finite $C_{\underline{t}+1}$. Based on Lemma 1, $C_{\underline{t}+1}$ is the maximum value that C_t can take; we denote this maximum value by $\overline{C}$. Let $\mathbf{S}$ denote the set $\{\underline{C}, \underline{C}+1, \ldots, \overline{C}\}$ with cardinality $|\mathbf{S}|$. From the pigeon-hole principle, if we let the system run for longer than $|\mathbf{S}|$ periods, at least one of the values in $\mathbf{S}$ must be repeated. Since the only determinant of C_t is C_{t-1}, the set of observed C_t values begins to cycle with some finite period T, that is, $C_t = C_{t+T}$ for t large enough.

Let y_i denote the i^{th} observation of C_t after the cycle is initiated; the sequence of C_t values that cycle is $y_1, y_2, \ldots, y_i, \ldots, y_T, y_1, y_2, \ldots$. By definition, all y_i values are distinct. Using contradiction, we argue that either $T = 1$ or $T = 2$. If $T = 1$ we have the convergence pattern, and if $T = 2$ we have the stable oscillation pattern.

Assume for the sake of contradiction that $T \geq 2$. First, observe that

- Case 1: If $y_{i+1} > y_i$, $D_t^{(F)}$ resulting from $C_t = y_i$ must be at least as long as $D_t^{(F)}$ resulting from $C_t = y_{i+1}$ (by Lemma 1); therefore, we have $y_{i+1} > y_{i+2}$.

- Case 2: If $y_{i+1} < y_i$, we have $y_{i+2} > y_{i+1}$ based on a similar discussion as in Case 1.

We can conclude based on Cases 1 and 2 that the direction of the evolution of C_t alternates; an increase is followed by a decrease and a decrease is followed by an increase.

Observation 1. An increase follows a decrease and a decrease follows an increase.

For the remainder, we discuss Case 1. Case 2 follows the same logic.

In Case 1, we have $y_{i+1} > y_i$ and $y_{i+1} > y_{i+2}$. We can consider two different orderings (subcases): either (a) $y_{i+1} > y_{i+2} > y_i$ or (b) $y_{i+1} > y_i > y_{i+2}$. Note that neither (a) nor (b) can themselves form the entire sequence due to the violation of Observation 1[2], which means T cannot be 2.

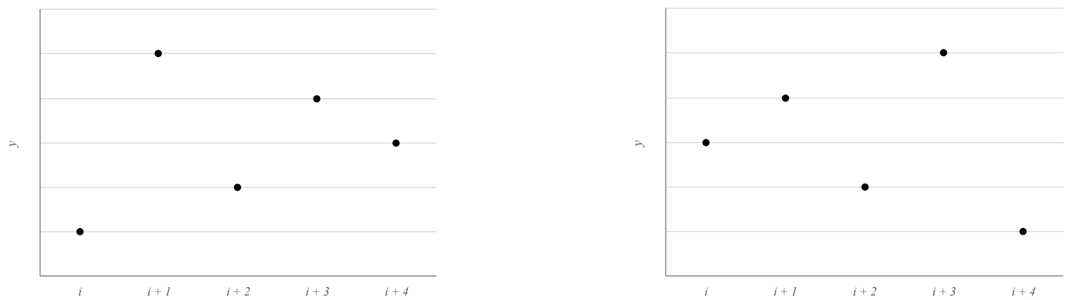
Next, we consider the possibility of $T > 3$. In Case (1), $y_{i+1} > y_{i+2}$; therefore, we have $y_{i+3} > y_{i+2}$ (from Observation 1).

Under (a), we can claim from Lemma 1 that $y_{i+2} < y_{i+3} < y_{i+1}$. Therefore, the ordering should be $y_{i+1} > y_{i+3} > y_{i+2} > y_i$. We call this the "convergent" pattern (Fig. B.1(a)). By continuing this argument for y_{i+4}, we can assert that y_{i+3} and y_{i+4} will be sandwiched between y_{i+1} and y_{i+2}; then, the ordering is $y_{i+1} > y_{i+3} > y_{i+4} > y_{i+2} > y_i$. Therefore, there cannot exist a value T such that $y_{i+T} = y_i$. Hence, subcase (a) does not lead to a feasible sequence.

Under (b), since $y_{i+1} > y_i > y_{i+2}$, it follows from Lemma 1 that $y_{i+2} < y_{i+1} < y_{i+3}$. Therefore, the ordering should be $y_{i+3} > y_{i+1} > y_i > y_{i+2}$. We call this the "divergent" pattern (Fig. B.1(b)). Notice that y_i and y_{i+1} are sandwiched between y_{i+2} and y_3. By continuing this argument for y_{i+1}, we can assert that y_{i+1} and y_{i+2} will be sandwiched between y_{i+3} and y_{i+4}; then, the ordering is

[2] If the sequence was $y_i, y_{i+1}, y_{i+2}, y_i, y_{i+1}, y_{i+2}, \cdots$ then: Under (a), the subsequence y_{i+1}, y_{i+2}, y_i violates Observation 1 due to two consecutive decreases, and under (b) the subsequence y_{i+2}, y_i, y_{i+1} violates Observation 1 due to two consecutive increases.

$y_{i+3} > y_{i+1} > y_i > y_{i+2} > y_{i+4}$. Therefore, there cannot exist a value T such that $y_{i+T} = y_i$. Hence, subcase (b) does not lead to a feasible sequence either.



(a) Convergent pattern **(b)** Divergent pattern

Figure B.1: Illustration of Convergent and Divergent patterns

Therefore, either $T = 1$ or $T = 2$. That is, C_t either stagnates at a single value or alternates between two values.

B.1.7 Proof of Proposition 14

First, observe that $M_{R_1}^{(L)}$ is increasing in C_t (from (2.4)), and therefore, $M_{R_1}^{(F)} = 1 - M_{R_1}^{(L)}$ is decreasing in C_t. Taken together with Conjecture 1, the computation of an appropriate upper (lower) bound to C_∞ will yield a lower (upper) bound to $M_{R_1}^{(F)}$. In particular, we plug in a lower bound for C_∞ to get an upper bound on $M_{R_1}^{(F)}$ and plug in an upper bound for $C_\infty + 1$ to get a lower bound on $M_{R_1}^{(F)}$.

In order to compute bounds on the delay at F when $C_t \to \infty$, we first characterize the distribution of inter-arrival times at F.

Lemma 4 *The first and second moments of F's inter-arrival distribution follow:*

$$E(T) = \frac{\Lambda^{C_t+1} - 1}{(\Lambda - 1)\Lambda^{C_t+1}}, \tag{B.7}$$

$$E(T^2) = \frac{2\Lambda^{-2(C_t+1)}\left(\Lambda^{2C_t+3} - 1 - (2C_t+3)(\Lambda-1)\Lambda^{C_t+1}\right)}{(\Lambda-1)^3}. \tag{B.8}$$

Proof. The arrival process to F is a generalized interrupted Poisson process: the *on*-phase duration is exponentially distributed with rate 1 (the *on*-phase corresponds to the time that L's chain resides in state C_t before transitioning to state $C_t - 1$), and the *off*-phase duration corresponds to the amount of time that elapses until L's chain first visits state C_t, starting at state $C_t - 1$. Let T denote the random variable representing F's inter-arrival times. The Laplace-Stieltjes transform of T for a given C_t follows [Tran-Gia, 1988]:

$$\phi(s) = \frac{\Lambda}{s + \Lambda + 1 - \phi_{off}^{C_t}(s)},$$

where $\phi_{\mathit{off}}^{C_t}(s)$, which is the Laplace-Stieltjes transform of the off-phase duration when the threshold is C_t, follows [Keilson, 1965]:

$$\phi_{\mathit{off}}^{C_t}(s) = \frac{\Lambda}{s + \Lambda + 1 - \phi_{\mathit{off}}^{C_t - 1}(s)},$$

with the initial condition $\phi_{\mathit{off}}^{1}(s) = \Lambda/(\Lambda + s)$. We apply this expression recursively to obtain a general form for $\phi(s)$. Subsequently, taking the first and second derivative (respectively) of $\phi(s)$ with respect to s and setting $s = 0$ result in (B.7) and (B.8). ■

Upper bound: Let S represent F's service time random variable. From Daley et al. [1992, inequality VII], we can first find a lower bound for $D^{(F)}$ as follows:

$$D^{(F)} \geq \frac{\left(\mathrm{E}(S^2) - \mathrm{E}(S)\mathrm{E}(T)\right)^+}{2(1-\rho)\mathrm{E}(T)}, \tag{B.9}$$

where ρ is the probability that F is busy, which follows:

$$\rho = \frac{\Lambda \pi_{C_t}}{\mu^{(F)}} = \frac{(\Lambda - 1)\Lambda^{C_t+1}}{\mu^{(F)}\left(\Lambda^{C_t+1} - 1\right)}.$$

The service time at F follows an exponential distribution with the first and second moments $\mathrm{E}(S) = 1/\mu^{(F)}$ and $\mathrm{E}(S^2) = 2/{\mu^{(F)}}^2$. These along with (B.7)-(B.8) result in:

$$D_{R_1}^{(F)} \geq \left(\underbrace{\frac{\Lambda^{C_t+1}(2\Lambda - \mu^{(F)} - 2) + \mu^{(F)}}{2\mu^{(F)}\left(\Lambda^{C_t+1}(-\Lambda + \mu^{(F)} + 1) - \mu^{(F)}\right)}}_{RHS} \right)^+ .$$

To find a bound based on Conjecture 1, we find the limit of RHS as $C_t \to \infty$ as follows:

$$\lim_{C_t \to \infty} RHS = \begin{cases} -1/(2\mu^{(F)}), & \Lambda \leq 1, \\ \omega = \dfrac{1}{-2\Lambda + 2\mu^{(F)} + 2} - \dfrac{1}{\mu^{(F)}}, & \Lambda > 1. \end{cases}$$

Clearly $\lim_{C \to \infty} RHS < 0$ when $\Lambda \leq 1$. When $\Lambda > 1$, $\lim_{C \to \infty} RHS < 1$ if

$$\Lambda < \frac{2{\mu^{(F)}}^2 + 3\mu^{(F)} + 2}{2\mu^{(F)} + 2}.$$

Note that $(2{\mu^{(F)}}^2 + 3\mu^{(F)} + 2)/(2\mu^{(F)} + 2) \geq 1, \forall \mu^{(F)} > 0$. Also, note again that $C_\infty = \lfloor \inf_{C_t} D_t^{(F)} \rfloor + 1$; therefore, when $\Lambda < (2{\mu^{(F)}}^2 + 3\mu^{(F)} + 2)/(2\mu^{(F)} + 2)$, a lower bound for C_∞ is 1.

Based on $M_{R_1}^{(F)}$ being decreasing in C_∞, Conjecture 1, and the derived lower bound of $D_{R_1}^{(F)}$

corresponding to C_∞, we obtain the following upper bound for $M^{(F)}_{R_1}$:

$$M^{(F)}_{R_1} \leq \overline{M}^{(F)}_{R_1} = \begin{cases} \dfrac{\Lambda^2 - \Lambda}{\Lambda^2 - 1}, & \Lambda < \dfrac{2\mu^{(F)^2} + 3\mu^{(F)} + 2}{2\mu^{(F)} + 2}, \\ \dfrac{\Lambda^{\lfloor \omega+2 \rfloor} - \Lambda^{\lfloor \omega+1 \rfloor}}{\Lambda^{\lfloor \omega+2 \rfloor} - 1}, & \text{otherwise.} \end{cases} \tag{B.10}$$

Lower bound: From Daley et al. [1992, inequality II], we can first find an upper bound for $D^{(F)}$ as follows:

$$D^{(F)} \leq RHS' = \frac{\text{Var}(S) + (1 - (1-\rho)^2)\text{Var}(T)}{2(1-\rho)\text{E}[T]},$$

where $\text{Var}(S) = 1/\mu^{(F)^2}$, $\text{Var}(T) = \text{E}(T^2) - \text{E}(T)^2$ (given in (B.7)-(B.8)). To find a bound based on Conjecture 1, we find the limit of RHS' as $C \to \infty$ as the following:

$$\lim_{C \to \infty} RHS' = \begin{cases} \Omega_1 = \dfrac{1+\Lambda}{\mu^{(F)}(1-\Lambda)}, & \Lambda < 1, \\ \Omega_2 = \dfrac{1 + \Lambda(\mu^{(F)} - 1) + \mu^{(F)}}{\mu^{(F)}(\Lambda - 1)(1 + \mu^{(F)} - \Lambda)}, & \Lambda > 1. \end{cases}$$

The bound is not defined for $\Lambda = 1$.

Based on $M^{(F)}_{R_1}$ being decreasing in C_∞, Conjecture 1, and the derived upper bound of $D^{(F)}_{R_1}$ corresponding to C_∞, we obtain the following lower bound for $M^{(F)}_{R_1}$:

$$M^{(F)}_{R_1} \geq \underline{M}^{(F)}_{R_1} = \begin{cases} \dfrac{\Lambda^{\lfloor \Omega_1+3 \rfloor} - \Lambda^{\lfloor \Omega_1+2 \rfloor}}{\Lambda^{\lfloor \Omega_1+3 \rfloor} - 1}, & \Lambda < 1, \\ \dfrac{\Lambda^{\lfloor \Omega_2+3 \rfloor} - \Lambda^{\lfloor \Omega_2+2 \rfloor}}{\Lambda^{\lfloor \Omega_2+3 \rfloor} - 1}, & \Lambda > 1. \end{cases} \tag{B.11}$$

B.2 Closed-form solutions for $D^{(F)}_1$ when $C_1 = 1$ and 2

Let $\pi_{i,j}$, $i = 0, \ldots, C_1$ and $j \geq 0$, represent the stationary probabilities of Model F in Period 1. Let $\vec{\pi}_j = (\pi_{0,j}, \ldots, \pi_{C_1,j})$. Model F (Fig. 2.3(b)) is a QBD process that is characterized by its square rate matrix $\mathbf{R}$ of order $C_1 + 1$, which satisfies $\vec{\pi}_{j+1} = \vec{\pi}_j \times \mathbf{R}$, $\forall j \geq 0$. In an excursion from Level j to Level $j+1$ initiated by a transition from state (m, j), $m = 0, \ldots, C_1$, element $R_{m+1,n}$ in $\mathbf{R}$, $n = 1, \ldots, C_1+1$, represents the amount of time the excursion spends in state $(n+1, j+1)$ for every unit of time spent in state (m, j) [Latouche and Ramaswami, 1999]. Since transitions from Level j to Level $j+1$ only occur through Phase C_1, $\mathbf{R}$ contains non-zero entries only in its bottom row $C_1 + 1$. Based on this special structure, $\vec{\pi}_{j+1} = \vec{\pi}_j \times \mathbf{R}$ yields,

$$\frac{\pi_{i,j+1}}{\pi_{C_1,j}} = \frac{\pi_{i,j+2}}{\pi_{C_1,j+1}}, \quad 0 \leq i \leq C_1, \forall j \geq 0. \tag{B.12}$$

Let $r_j = \pi_{C_1,j}/(\sum_{i=1}^{C_1}\pi_{i,j})$. From (B.12), $r_j = r_{j+1} = r, \forall j > 0$. From the flow balance equations, we derive,

$$\lambda_1^{(F)} = \mu^{(F)}\,\Pr(F \text{ is busy}) \Longrightarrow \lambda_1^{(F)} = \mu^{(F)}\left(1 - \sum_{i=0}^{C_1}\pi_{i,0}\right) \Rightarrow \sum_{i=0}^{C_1}\pi_{i,0} = 1 - \frac{\lambda_1^{(F)}}{\mu^{(F)}}, \tag{B.13}$$

where $\lambda_1^{(F)}$ follows $\lambda_1^{(F)} = M_1^{(F)}\Lambda = 1 - M_1^{(L)}$ and eq. (2.5). Taking flow balance equations across levels,

$$\pi_{C_1,j}\Lambda = \mu^{(F)}\sum_{i=0}^{C_1}\pi_{i,j+1} \Longrightarrow \sum_{i=0}^{C_1}\pi_{i,j+1} = \frac{\pi_{C_1,j}\Lambda}{\mu^{(F)}}.$$

The expected number of customers in F, which we denote by $L_1^{(F)}$, can be expressed as,

$$\begin{aligned} L_1^{(F)} &= \sum_{j=1}^{\infty} j \sum_{i=0}^{C_1}\pi_{i,j} = \sum_{j=1}^{\infty} j\pi_{C_1,j-1}\frac{\Lambda}{\mu^{(F)}} = \sum_{j=1}^{\infty} j\left(r_j\sum_{i=0}^{C_1}\pi_{i,j-1}\right)\frac{\Lambda}{\mu^{(F)}} = \sum_{j=1}^{\infty} j\left(r_j\pi_{C_1,j-2}\frac{\Lambda}{\mu^{(F)}}\right)\frac{\Lambda}{\mu^{(F)}} \\ &= \cdots = \sum_{j=1}^{\infty} j\pi_{C_1,0}\left(\frac{\Lambda}{\mu^{(F)}}\right)^j r_j^{j-1}. \end{aligned} \tag{B.14}$$

If we can derive the expressions for $\pi_{C_1,0}$ and r_j, we can apply Little's Law and (B.14) to derive $D_1^{(F)}$.

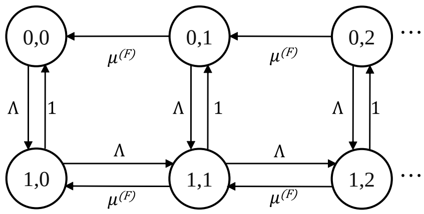
Let y denote the probability of L being in state C_1, which we can easily derive from Model L's balance equations. On the other hand, y can be expressed as a weighted average of r_j values as follows:

$$y = \sum_{j=0}^{\infty}\pi_{C_1,j} = r_0\sum_{i=0}^{C_1}\pi_{i,0} + \sum_{j=1}^{\infty}\sum_{i=0}^{C_1} r_j\pi_{i,j} = r_0\left(1 - \frac{\lambda_1^{(F)}}{\mu^{(F)}}\right) + r\frac{\lambda_1^{(F)}}{\mu^{(F)}}. \tag{B.15}$$

Therefore, (B.15) can be used to find r values based on r_0. We can find r_0 when $C_1 = 1$ and 2 by solving a system of simultaneous non-linear equations governing the stationary probabilities in Levels $j = 0, 1, 2$ of Model F. To find probabilities of states in Levels 1-3 of Model F, we need a system of $3(C_1+1)$ equations. We obtain $2(C_1+1)$ equations from flow balance equations for states in Levels 1 and 2, one equation from the sum of Level $j = 0$ state probabilities from (B.13), and C_1 independent equations from (B.12).

For example, Fig. B.2(a) shows Model F when $C_1 = 1$. The linear equation set (B.16) lists the flow balance equations for states in Levels 1 and 2 (four equations and six unknowns). We then add the linear equation (B.17), using (B.13), and the non-linear equation (B.18), using (B.12), which result in an equation set with six equations and six unknowns.

Solving equations (B.16) and (B.17) simultaneously in terms of $\pi_{1,0}$, we obtain the following



(a) Model F when $C_1 = 1$

$$\begin{aligned}
&\Lambda\pi_{0,0} = \mu^{(F)}\pi_{0,1} + \pi_{1,0},\\
&(1+\Lambda)\,\pi_{1,0} = \Lambda\pi_{0,0} + \mu^{(F)}\pi_{1,1},\\
&\left(\Lambda + \mu^{(F)}\right)\pi_{0,1} = \pi_{1,1} + \mu^{(F)}\pi_{0,2},\\
&\left(1+\Lambda + \mu^{(F)}\right)\pi_{1,1} = \Lambda\pi_{1,0} + \Lambda\pi_{0,1} + \mu^{(F)}\pi_{1,2}.
\end{aligned} \tag{B.16}$$

$$\pi_{0,0} + \pi_{1,0} = 1 - \frac{\lambda_1^{(F)}}{\mu^{(F)}}. \tag{B.17}$$

$$\frac{\pi_{0,1}}{\pi_{1,0}} = \frac{\pi_{0,2}}{\pi_{1,1}}. \tag{B.18}$$

quadratic equation:

$$\pi_{1,0}{}^2\left(\Lambda^2 - (\Lambda+1)\mu^{(F)}\right)\left(\Lambda(\Lambda+1)\mu^{(F)2} + (\Lambda+1)\mu^{(F)2}\right) + \pi_{1,0}\left(\Lambda^2 - (\Lambda+1)\mu^{(F)}\right)\times$$
$$\left(2(\Lambda+1)\Lambda^2\mu^{(F)} - (\Lambda+1)\Lambda\mu^{(F)2} + (\Lambda+1)\Lambda\mu^{(F)}\right) + \left(\Lambda^2 - (\Lambda+1)\mu^{(F)}\right)\left(\Lambda^4 - \Lambda^2(\Lambda+1)\mu^{(F)}\right) = 0,$$

with roots $\Lambda\left(\mu^{(F)} - 2\Lambda - 1 \pm \sqrt{4\Lambda + \mu^{(F)}\left(\mu^{(F)}+2\right)+1}\right)/\left(2(\Lambda+1)\mu^{(F)}\right)$. The second root is not admissible because the expression $\mu^{(F)} - 2\Lambda - 1 - \sqrt{4\Lambda + \mu^{(F)}\left(\mu^{(F)}+2\right)+1} < 0$ as its value is -2 at $\Lambda = 0$, and it is decreasing in Λ. Therefore, $\pi_{1,0} = \Lambda\left(\mu^{(F)} - 2\Lambda - 1 + \sqrt{4\Lambda + \mu^{(F)}\left(\mu^{(F)}+2\right)+1}\right)/\left(2(\Lambda+1)\mu^{(F)}\right)$. We obtain r_0 by dividing $\pi_{1,0}$ by $\sum_{i=0}^{1}\pi_{i,0}$ (given in (B.13)). We obtain the expression for r by plugging r_0 into (B.15). We then obtain $L_1^{(F)}$ using (B.14). Finally, we employ Little's Law to derive (B.5). The expression for the case of $C_1 = 2$ follows a similar procedure.

C

Additional Material for Chapter 3

C.1 Proof of Proposition 15

We will prove this by showing that no equilibrium price pair can be found for state (1,1) one period from the end of the horizon.

First, observe that $V_T^*(x_{me}, x_c) = 0, \forall x_{me}, x_c$. Now, consider the problem in period $T-1$, when $x_{me} = 1$ and $x_c = 1$. The best response function of me simplifies to:

$$V_{T-1}(x_{me}, x_c, p_c) = \max_{p_{me}} \quad \mathbb{I}_{p_{me}<p_c}\overline{d}(p_{me})p_{me} + \mathbb{I}_{p_c<p_{me}}\underline{d}(p_{me})p_{me} + \mathbb{I}_{p_{me}=p_c}\frac{\overline{d}(p_{me}) + \underline{d}(p_{me})}{2}p_{me} \tag{C.1}$$

Since the firms are symmetric, c has the same best response function.

First, note that neither firm will set a price $p < 0$ because this is one period from the end of the horizon: they can achieve a total value-to-go of 0 by simply pricing at 0 which is higher than their revenue-to-go if they set a negative price. Next, note from (C.1) that both firms will not set the same price p because either firm can do strictly better by undercutting by an arbitrarily small $\epsilon > 0$. Moreover, it is not an equilibrium for both firms to set a price $p = 0$. Choosing a price of 0 yields an objective value of 0, and either firm can do strictly better by setting a price $p > 0$ and obtaining an objective value of $D(p, 0)p > 0$.

So, suppose the firms set prices $p_{me} > 0$ and $p_c > 0$ respectively. Without loss of generality, let $p_{me} > p_c$. There are two possibilities:

1. c undercuts me by an arbitrarily small amount $\epsilon > 0$. Since both firms are symmetric and have one unit of inventory left, this must result in the two firms obtaining objective values that can be made arbitrarily close to each other, i.e, $|\underline{d}(p_{me}, p_c)p_{me} - \overline{d}(p_c, p_{me})p_c| < \delta$ for any arbitrarily small value of $\delta > 0$. Observe here that p_{me} can be written as $p_{me} = p_c r$ for a value of r arbitrarily close to 1, approaching 1 from above. In this case, we have:

$$\begin{aligned}&\underline{d}(p_{me}, p_c)p_{me} - \overline{d}(p_c, p_{me})p_c \\ &\qquad = \underline{d}(p_{me}, p_c)p_c r - \overline{d}(p_c, p_{me})p_c \\ &\qquad = p_c \left(\underline{d}(p_{me}, p_c) r - \overline{d}(p_c, p_{me})\right)\end{aligned}$$

Now, observe from the continuity of $\overline{d}$ and $\underline{d}$ and the fact that r is arbitrarily close to 1, the expression $\underline{d}(p_{me}, p_c)r - \overline{d}(p_c, p_{me})$ is arbitrarily close to the expression $\underline{d}(p_c, p_c) - \overline{d}(p_c, p_c)$ which is strictly negative, as implied by Assumption 6. Therefore, p_{me} and p_c cannot form an equilibrium price pair if they are arbitrarily close to each other.

2. c undercuts me by a non-trivial amount. Since both firms are symmetric and have one unit left, this must result in both firms obtaining identical objective values, i.e., $\underline{d}(p_{me}, p_c)p_{me} = \overline{d}(p_c, p_{me})p_c$. Now, observe that $\overline{d}(p_c, p_{me})p_c = \underline{d}(p_{me}, p_c)p_{me} < \overline{d}(p_{me}, p_c)p_{me}$. From continuity, $\overline{d}(p_{me}, p_c)p_{me}$ is arbitrarily close to $\overline{d}(p_{me-\epsilon}, p_c)(p_{me} - \epsilon)$ for a sufficiently small $\epsilon > 0$. Therefore, c can obtain a strictly higher objective value of $\overline{d}(p_{me-\epsilon}, p_c)(p_{me} - \epsilon)$ by setting a price arbitrarily close to p_{me} than her current objective value of $\overline{d}(p_c, p_{me})p_c$. Therefore, (p_{me}, p_c) cannot be an equilibrium price pair.

Since there is no equilibrium price-pair for state (1,1), there can be no SPNE-P. This completes the proof. ■

C.2 Proof of Lemma 3

For any state, consider the best response function of me. Clearly, it is never optimal for me to set a price $p_{me} = p_c$, because me can obtain a strictly higher revenue by setting $p_{me} = p_c - \epsilon$ for some small $\epsilon > 0$.

If me chooses to undercut c, me obtains an expected revenue-to-go of:

$$\begin{aligned} R_{undercut}(p_c) &= \max_{p_{me} < p_c} \quad d(p_{me}) \left(p_{me} + \underbrace{V^*_{t+1}(x_{me} - 1, x_c) - V^*_{t+1}(x_{me}, x_c)}_{\Delta_1} \right) + V^*_{t+1}(x_{me}, x_c) \\ &= \max_{p_{me} < p_c} \quad d(p_{me})\left(p_{me} + \Delta_1\right) + V^*_{t+1}(x_{me}, x_c). \end{aligned} \tag{C.2}$$

Recall that $V^*(\cdot)$, as introduced earlier, represents the value-to-go for me and $d(p)$ is as used in the simplified notation for the fully flexible case, as in (3.3).

If me chooses not to undercut price p_c, she will obtain the same expected revenue-to-go, irrespective of her choice of price p_{me}, as long as $p_{me} > p_c$. This revenue is given by:

$$\begin{aligned} R_{settle}(p_c) &= d(p_c)\left(\underbrace{V^*_{t+1}(x_{me}, x_c - 1) - V^*_{t+1}(x_{me}, x_c)}_{\Delta_2}\right) + V^*_{t+1}(x_{me}, x_c) \\ &= d(p_c)\,(\Delta_2) + V^*_{t+1}(x_{me}, x_c) \end{aligned} \tag{C.3}$$

Our proof is structured as follows. We will show that there exists a point r_{thresh} such that me finds it optimal to undercut when $p_c \geq r_{thresh}$ and finds it optimal to settle otherwise. An analogous argument can be made taking c's perspective.

Both $R_{undercut}$ and R_{settle} can be viewed as functions of p_c. For a given value of p_c, if $R_{settle}(p_c) < R_{undercut}(p_c)$, then me's optimal response is to undercut c. Otherwise, me can settle at any value of $p_{me} > p_c$. Which of these alternatives to choose rests on a comparison of the terms $\underbrace{d(p_{me})\,(p_{me} + \Delta_1)}_{U}$, where p_{me} is constrained to be strictly smaller than p_c, and $\underbrace{d(p_c)\,(\Delta_2)}_{S}$.

Lemma 5 *U is non-decreasing in p_c. Further, U is continuous in p_c*

Proof: Increasing p_c strictly increases the domain over which the optimal value U may be found. Therefore, U must be non-decreasing in p_c. U's continuity follows from the fact that it is the upper envelope of the function $d(p)(p + \Delta_1)$ which is continuous.

Note, therefore, that both U and S are continuous functions of p_c. Consider the following three cases:

1. $\Delta_2 > 0$: In this case, the term S is decreasing in p_c because $d(p_c)$ is decreasing in p_c. Therefore, there is a unique point r_{thresh} such that for $p_c < r_{thresh}$, it is optimal to settle, while for $p_c > r_{thresh}$, it is optimal to undercut. At $p_c = r_{thresh}$, me is indifferent between undercutting and settling (note that these options are both preferable to matching price p_c). Observe that if the functions U and S never intersect $r_{thresh} \to \infty$.

2. $\Delta_2 = 0$: Here, the function U is necessarily larger than 0 when $p_c + \Delta_1 > 0$. So undercutting is preferable when $p_c > -\Delta_1$. When $p_c < \Delta_1$, U is necessarily negative, and it is preferable to settle. Therefore, r_{thresh} is exactly equal to $-\Delta_1$.

3. $\Delta_2 < 0$: In this case, U is non-decreasing in p_c while S is increasing in p_c. Note that S is always negative. We show that in this case, a unique threshold r_{thresh} exists below which U is strictly smaller than S and above which U is strictly larger than S. First, consider the case where $p_c \geq -\Delta_1$. Here $U \geq d(p_c + \Delta_1)(p_c + \Delta_1) - \epsilon \geq 0 - \epsilon$ for any arbitrarily small $\epsilon > 0$. For ϵ small enough, this value is strictly larger than S. Therefore, it is optimal to undercut in this region. Now, we treat the case of $p_c < -\Delta_1 > \Delta_2 - \Delta_1$. Consider $U = d(p_{me})(p_{me} + \Delta_1)$, where $0 \leq p_{me} < p_c$. This function is strictly increasing in p_{me}, because $p_{me} < p_c$ and

$p_{me} + \Delta_1$ becomes more negative as p_{me} decreases (recall that $p_{me} + \Delta_1 < p_c + \Delta_1 < 0$, and $d(p_{me})$ increases as p_{me} decreases. Therefore, under the above conditions, U is maximized at $p_c - \epsilon$ for an arbitrarily small ϵ. Therefore, to decide whether to undercut or settle when $p_c < -\Delta_1$, we simply need to compare $U = d(p_c - \epsilon)(p_c + \Delta_1)$ with $S = d(p_c)\Delta_2$. Upon algebraic manipulation, we see that $U > S$ when $p_c > \Delta_2 - \Delta_1$ and less than S otherwise.

To summarize, when $\Delta_2 \leq 0$, $r_{thresh} = \Delta_2 - \Delta_1$. Below this threshold, settling is optimal, and above this threshold, undercutting is optimal. When $\Delta_2 > 0$, such a threshold exists and can be found (although not in closed form). ■

C.3 Characterizing the equilibrium for the exponential demand case

Characterizing the equilibrium for the case of exponential demand rests on finding the appropriate threshold price (reservation price) below which the firm (without loss of generality, *me*) prefers settling to undercutting. We show such a threshold continues to exist even when the condition that $d(p)$ be decreasing in p is relaxed to $d(p)$ being non-increasing in p when $p < 0$. The key idea to show this is that for prices $p_c < 0$, the function U is strictly increasing. We also show how this threshold may be computed inexpensively. We use the notation defined in Appendix C.2.

Consider the function $f = e^{-p_{me}}(p_{me} + \Delta_1)$, absent any constraints on p_{me}. This function is clearly quasiconcave, because its derivative with respect to p_{me} is negative for $p_{me} > 1 - \Delta_1$ and positive for $p_{me} < 1 - \Delta_1$. Observe that $U = (p_{me} + \Delta_1)$ for $p_{me} < 0$ and $U = f$ for $p_{me} \geq 0$. We now argue that U is quasiconcave: (i) If $1 - \Delta_1 > 0$, then $U = p_{me} + \Delta_1$ is increasing in p_{me} for $p_{me} < 0$, is increasing in p_{me} for $0 \leq p_{me} < 1 - \Delta_1$ and is decreasing in p_{me} for $p_{me} \geq 1 - \Delta_1$. (ii) If $1 - \Delta_1 < 0$, then $U = p_{me} + \Delta_1$ is increasing in p_{me}, for $p_{me} \leq 0$ and $U = e^{-p_{me}}(p_{me} + \Delta_1)$ is decreasing in p_{me} for $p_{me} \geq 0$. Therefore, U has a unique maximum. Note also that $e^{-p}(p + \Delta_1)$ intersects with $e^{-p}\Delta_2$ at exactly one point, i.e., $p = \Delta_2 - \Delta_1$.

First, consider the case of $\Delta_2 = 0$. Here, the argument in Appendix C.2 is unaffected by the form of $d(p)$ because S is 0. Therefore value of the threshold price is $\Delta_2 - \Delta_1$.

Next, consider the case of $\Delta_2 < 0$. The argument that U is maximized at $p_c - \epsilon$ for an arbitrarily small ϵ when $p_c < \Delta_1$ continues to be valid. Therefore, the threshold price in this case, is again $\Delta_2 - \Delta_1$.

Now, consider the case of $\Delta_2 > 0$. First, note that U when not constrained is maximized at $p_{me} = p^* = 1 - \Delta_1$. If $p^* > 0$, three possibilities arise:

1. $\Delta_2 - \Delta_1 > p^*$. This means that at all price points $p_c < p^*$, including points $p_c < 0$, it is preferable for *me* to settle. Therefore, the threshold must be above p^* and below $\Delta_2 - \Delta_1$. Solving $e^{-p^*}(p^* + \Delta_1) = e^{r_{thresh}}(\Delta_2)$, we obtain $r_{thresh} = 1 - \Delta_1 - \log \frac{1}{\Delta_2}$.

2. $0 \leq \Delta_2 - \Delta_1 \leq p^*$. This is a straightforward case to handle. Above $\Delta_2 - \Delta_1$, it is clearly optimal to undercut, and below it, it is optimal to settle. Therefore, $r_{thresh} = \Delta_2 - \Delta_1$.

3. $\Delta_2 - \Delta_1 < 0$. Since $p^* > 0$, it is clearly optimal to undercut for all $p_c > 0$ if the intersection point $\Delta_2 - \Delta_1 < 0$. Therefore, the threshold must be below 0. The value of S in this region is $e^0 \Delta_2$ and the value of U in this region is $e^0(p_{me} + \Delta_1)$. Therefore, the threshold r_{thresh} where these two are equal is $r_{thresh} = \Delta_2 - \Delta_1$.

These three subcases are illustrated in Figure C.1.

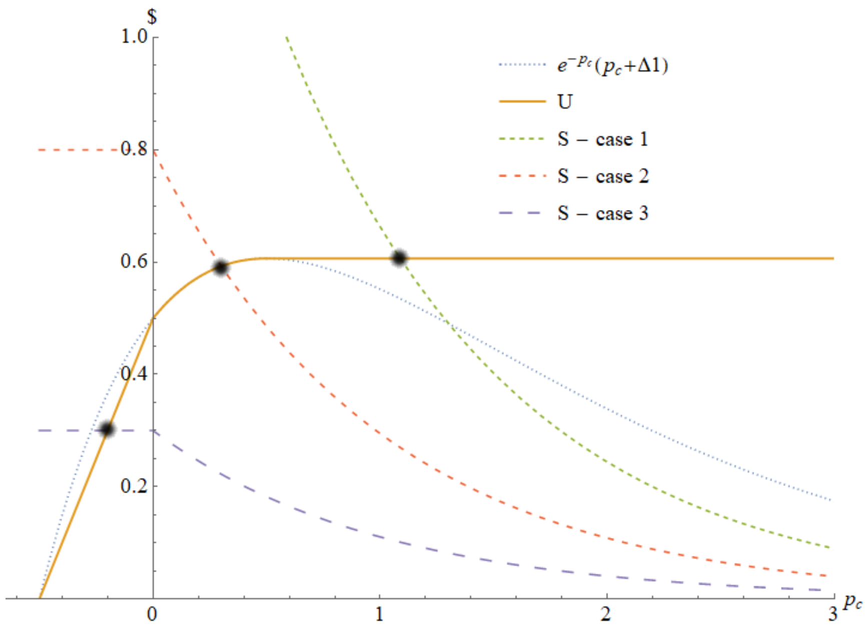


Figure C.1: Possible cases when $p^* > 0$. The relevant thresholds are marked in black.

Now, consider the case where $p^* \leq 0$. Therefore, for $p_c > 0$, the optimal point to undercut to is 0. And for $p_c < 0$, *me* will optimally undercut by some small $\epsilon > 0$. Two sub-cases arise:

1. $\Delta_2 - \Delta_1 > 0$. At price point 0, the revenue from settling is $e^0 \Delta_2 = \Delta_2$, it is optimal to undercut to price 0, and the revenue from undercutting is $e^0 \Delta_1 = \Delta_1$. Since $\Delta_2 > \Delta_1$, the threshold has to lie above 0. We solve for the threshold by solving the equation $e^0(\Delta_1) = e^{-r_{thresh}} \Delta_2$, yielding a solution of $r_{thresh} = \log \frac{\Delta_2}{\Delta_1}$.

2. $\Delta_2 - \Delta_1 \leq 0$. This means that at all price points $p_c > 0$, it is optimal to undercut. Therefore, the threshold must be below 0. The value of S in this region is $e^0 \Delta_2$ and the value of U in this region is $e^0(p_{me} + \Delta_1)$. Therefore, the threshold r_{thresh} where these two are equal is $r_{thresh} = \Delta_2 - \Delta_1$.

These two sub-cases are illustrated in Figure C.2.

Therefore, we can compute the appropriate thresholds analytically in closed form after simply comparing a few quantities.

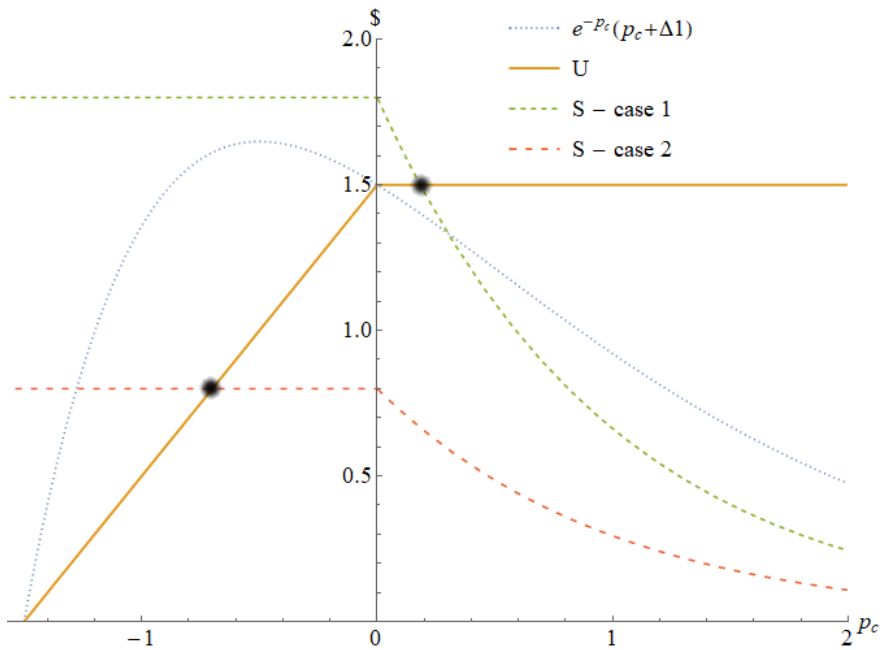


Figure C.2: Possible cases when $p^* < 0$. The relevant thresholds are marked in black.

C.4 Characterizing the equilibrium for the iso-elastic demand case

This characterization closely resembles the characterization in Appendix C.3. As with the exponential case, note that the function $U = (1 - p_{me})\,(p_{me} + \Delta_1)$, absent any constraints on p_{me} is quasiconcave, because its derivative with respect to p_{me} is negative for $p_{me} > \frac{1-\Delta_1}{2}$ and positive for $p_{me} < \frac{1}{1-\Delta_1}$. If $\frac{1-\Delta_1}{2} < 0$, then U remains quasiconcave, because it is decreasing for $p_{me} > 0$ and takes a value of $(p_{me} + \Delta_1)$, which is increasing in p_{me}, for $p_{me} \leq 0$. Therefore, it has a unique maximum. Note also that $(1 - p)\,(p + \Delta_1)$ intersects with $(1 - p)\Delta_2$ at exactly one point, i.e., $p = \Delta_2 - \Delta_1$.

First, consider the case of $\Delta_2 = 0$. Here, the argument in Appendix C.2 is unaffected by the form of $d(p)$ because S is 0. Therefore value of the threshold price is $\Delta_2 - \Delta_1$.

Next, consider the case of $\Delta_2 < 0$. The argument that U is maximized at $p_c - \epsilon$ for an arbitrarily small ϵ when $p_c < \Delta_1$ continues to be valid. Therefore, the threshold price in this case, is again $\Delta_2 - \Delta_1$.

Now, consider the case of $\Delta_2 > 0$. First, note that U when unconstrained is maximized at $p_{me} = p^* = \frac{1-\Delta_1}{2}$. If $p^* > 0$, three possibilities arise:

1. $\Delta_2 - \Delta_1 > p^*$. This means that at all price points $p_c < p^*$, including points $p_c < 0$, it is preferable for *me* to settle. Therefore, the threshold must be above p^* and below $\Delta_2 - \Delta_1$. Solving $(1 - p^*)(p^* + \Delta_1) = (1 - r_{thresh})(\Delta_2)$, we obtain $r_{thresh} = 1 - \frac{(\Delta_1+1)^2}{4\Delta_2}$.

2. $0 \leq \Delta_2 - \Delta_1 \leq p^*$. This is a straightforward case to handle. Above $\Delta_2 - \Delta_1$, it is clearly optimal to undercut, and below it, it is optimal to settle. Therefore, $r_{thresh} = \Delta_2 - \Delta_1$.

3. $\Delta_2 - \Delta_1 < 0$. Since $p^* > 0$, it is clearly optimal to undercut for all $p_c > 0$ if the intersection point $\Delta_2 - \Delta_1 < 0$. Therefore, the threshold must be below 0. The value of S in this region is Δ_2 and the value of U in this region is $(p_{me} + \Delta_1)$. Therefore, the threshold r_{thresh} where these two are equal is $r_{thresh} = \Delta_2 - \Delta_1$.

Now, consider the case where $p^* \leq 0$. Therefore, for $p_c > 0$, the optimal point to undercut to is 0. And for $p_c < 0$, me will optimally undercut by some small $\epsilon > 0$. Two sub-cases arise:

1. $\Delta_2 - \Delta_1 > 0$. At price point 0, the revenue from settling is Δ_2, it is optimal to undercut to price 0, and the revenue from undercutting is Δ_1. Since $\Delta_2 > \Delta_1$, the threshold has to lie above 0. We solve for the threshold by solving the equation $\Delta_1 = (1 - r_{thresh})\Delta_2$, yielding a solution of $r_{thresh} = 1 - \frac{\Delta_1}{\Delta_2}$.

2. $\Delta_2 - \Delta_1 \leq 0$. This means that at all price points $p_c > 0$, it is optimal to undercut. Therefore, the threshold must be below 0. The value of S in this region is $1\Delta_2$ and the value of U in this region is $(p_{me} + \Delta_1)$. Therefore, the threshold r_{thresh} where these two are equal is $r_{thresh} = \Delta_2 - \Delta_1$.

Again, for the iso-elastic case, we can compute the appropriate thresholds analytically in closed form after simply comparing a few quantities.

C.5 Demand Consistency Property imposed in Cooper et al. [2015]

We briefly describe the demand consistency property imposed in Cooper et al. [2015] for the following specific with two players, denoted as i and $-i$.

- Player i's demand when she prices at p_i and her competition prices at p_{-i} is given by $d_i(p_i, p_{-i}) = \beta_{i,0} + \beta_{i,i}p_i + \beta_{i,-i}p_{-i}$. A similar function holds for player $-i$.

- Player i and $-i$ use monopolistic models and assume that demand is only a function of their own price. However, they know the correct intercept. Therefore, player i's demand function is specified as: $\delta_i(p_i) = \beta_{i,0} + \alpha_i p_i$. A similar function holds for player $-i$.

Under the demand consistency property, "if the chosen prices converge, the estimated expected demand at the chosen prices converges to the actual demand at the limit prices." In this case, demand consistency requires that i's estimate of α_i follows:

$$\alpha_i = \frac{\beta_{i,0}(\beta_{-i,i}\beta_{i,-i} - \beta_{-i,-i}\beta_{i,i})}{\beta_{-i,0}\beta_{i,-i} - \beta_{i,0}\beta_{-i,-i}}.$$

Further, the demand consistency property in Cooper et al. [2015] imposes a restriction on the convergent equilibrium price. This imposition cannot be applied to our dynamic (multi-period) setting.

C.6 Section 5.3 plots for Iso-elastic demand

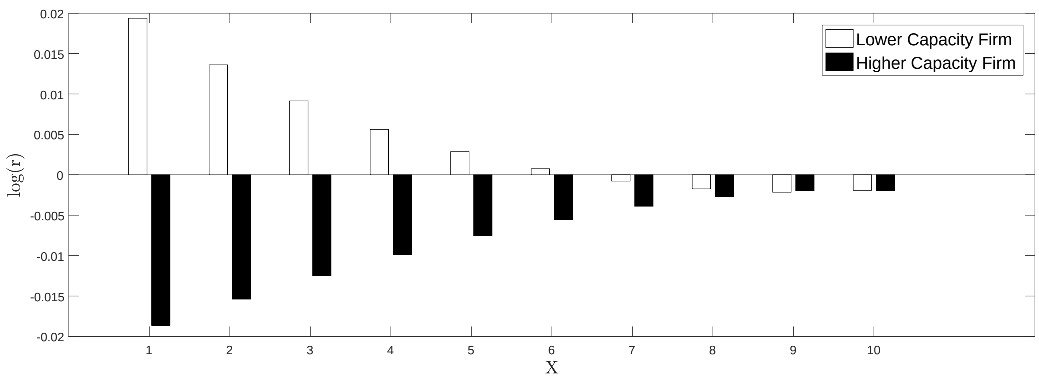


Figure C.3: $\log(r)$ against X for $T = 375$, total capacity $= 20$, Iso-elastic demand

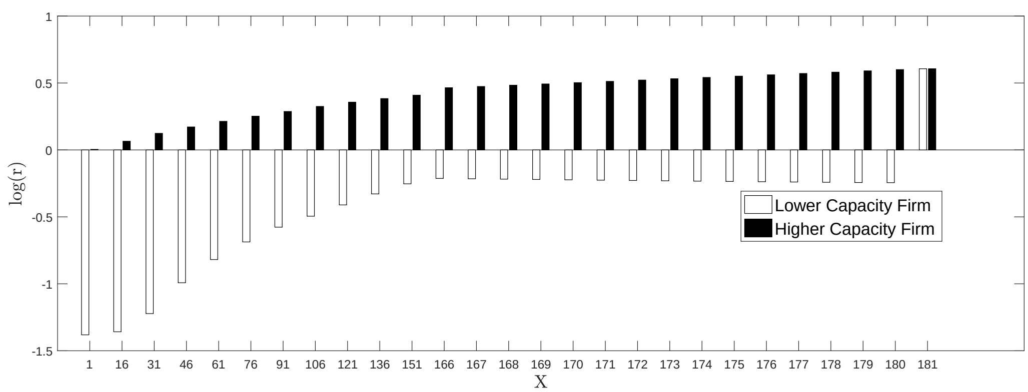


Figure C.4: $\log(r)$ against X for $T = 375$, total capacity $= 362$, Iso-elastic demand

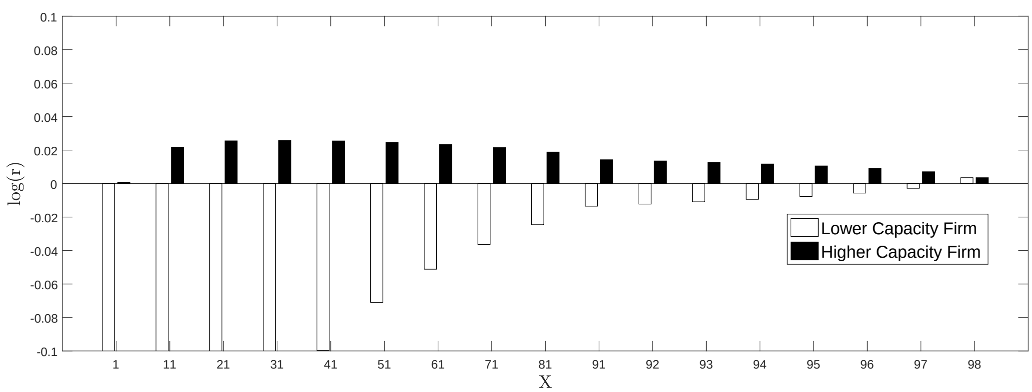


Figure C.5: $\log(r)$ against X for $T = 375$, total capacity $= 98$, Iso-elastic demand

References

Adan, I., Wessels, J., and Zijm, W. H. M. (1990). *Analysis of the asymmetric shortest queue problem*. Department of Mathematics and Computing Science, University of Technology.

Aflaki, S. and Netessine, S. (2017). Strategic investment in renewable energy sources: The effect of supply intermittency. *Manufacturing & Service Operations Management*, 19(3):489–507.

Agarwal, A., Cai, D., Shah, S., Chandy, M., and Sherick, R. (2015). A Model for Residential Adoption of Photovoltaic Systems. In *2015 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE.

Akşin, Z., Ata, B., Emadi, S. M., and Su, C.-L. (2016). Impact of delay announcements in call centers: An empirical approach. *Operations Research*, 65(1):242–265.

Al-Gwaiz, M., Chao, X., and Wu, O. Q. (2016). Understanding How Generation Flexibility and Renewable Energy Affect Power Market Competition. *Manufacturing & Service Operations Management*.

Alizamir, S., de Véricourt, F., and Sun, P. (2016). Efficient Feed-in-Tariff Policies for Renewable Energy Technologies. *Operations Research*, 64(1):52–66.

Allon, G., Bassamboo, A., and Gurvich, I. (2011). “We will be right with you”: Managing customer expectations with vague promises and cheap talk. *Operations Research*, 59(6):1382–1394.

Altman, E., Jiménez, T., Núñez-Queija, R., and Yechiali, U. (2004). Optimal routing among ./M/1 queues with partial information. *Stochastic Models*, 20.

Anderson, S. P. and Schneider, Y. (2007). Price competition with revenue management and costly consumer search. Technical report, Technical report, University of Virginia, Department of Economics ....

Ang, E., Kwasnick, S., Bayati, M., Plambeck, E. L., and Aratow, M. (2015). Accurate emergency department wait time prediction. *Manufacturing & Service Operations Management*, 18(1):141–156.

Anton, J. J., Biglaiser, G., and Vettas, N. (2014). Dynamic price competition with capacity constraints and a strategic buyer. *International Economic Review*, 55(3):943–958.

Armony, M. and Maglaras, C. (2004). Contact centers with a call-back option and real-time delay information. *Operations Research*, 52(4):527–545.

Armony, M., Shimkin, N., and Whitt, W. (2009). The impact of delay announcements in many-server queues with abandonment. *Operations Research*, 57(1):66–81.

Babich, V., Lobel, R., and Yucel, S. (2017). Promoting solar panel investments: Feed-in-tariff versus tax-rebate policies.

Bauner, C. and Crago, C. L. (2015). Adoption of Residential Solar Power Under Uncertainty: Implications for Renewable Energy Incentives. *Energy Policy*, 86:27–35.

Bird, L., McLaren, J., Heeter, J., Linvill, C., Shenot, J., Sedano, R., and Migden-Ostrander, J. (2013). Regulatory Considerations Associated with the Expanded Adoption of Distributed Solar.

Blackburn, G., Magee, C., and Rai, V. (2014). Solar Valuation and the Modern Utility's Expansion into Distributed Generation. *The Electricity Journal*, 27(1):18–32.

Bollinger, B. and Gillingham, K. (2012). Peer Effects in the Diffusion of Solar Photovoltaic Panels. *Marketing Science*, 31(6):900–912.

Borlick, R. and Wood, L. (2014). Net Energy Metering: Subsidy Issues and Regulatory Solutions. *Issue brief september*.

Brown, A. and Bunyan, J. (2014). Valuation of Distributed Solar: a Qualitative View. *The Electricity Journal*, 27(10):27–48.

Buhayar, N. (2016). Who Owns The Sun? [accessed on October 24, 2016].

Cai, D. W., Adlakha, S., Low, S. H., De Martini, P., and Chandy, K. M. (2013). Impact of Residential PV Adoption on Retail Electricity Rates. *Energy Policy*, 62:830–843.

California Distributed Generation Statistics (2017). Statistics and Charts. [accessed on February 2, 2018].

California Public Utilities Commission (2017). California Public Utilities Commission. [accessed on November 29, 2017].

Chediak, M. and Buhayar, N. (2015). Buffett's Utility Wins as Sun Sets on Nevada Home Solar Subsidy. [accessed on April 11, 2017].

Chen, M. and Chen, Z.-L. (2015). Recent developments in dynamic pricing research: multiple products, competition, and limited demand information. *Production and Operations Management*, 24(5):704–731.

Choudhary, V., Ghose, A., Mukhopadhyay, T., and Rajan, U. (2005). Personalized pricing and quality differentiation. *Management Science*, 51(7):1120–1130.

Christou, C., Kotseva, R., and Vettas, N. (2007). Pricing, investments and mergers with intertemporal capacity constraints.

Cooper, W. L., Homem-de Mello, T., and Kleywegt, A. J. (2015). Learning and pricing with models that do not explicitly incorporate competition. *Operations research*, 63(1):86–103.

Costello, K. W. and Hemphill, R. C. (2014). Electric Utilities' 'Death Spiral': Hyperbole or Reality? *The Electricity Journal*, 27(10):7–26.

Daley, D., Kreinin, A. Y., and Trengove, C. (1992). Inequalities concerning the waiting time in single-server queues: A survey. *OXFORD STATISTICAL SCIENCE SERIES*, pages 177–177.

Darghouth, N. R., Barbose, G., and Wiser, R. (2011). The Impact of Rate Design and Net Metering on the Bill Savings from Distributed PV for Residential Customers in California. *Energy Policy*, 39(9):5243–5253.

Darghouth, N. R., Wiser, R. H., Barbose, G., and Mills, A. D. (2016). Net Metering and Market Feedback Loops: Exploring the Impact of Retail Rate Design on Distributed PV Deployment. *Applied Energy*, 162:713–722.

Dasci, A. and Karakul, M. (2009). Two-period dynamic versus fixed-ratio pricing in a capacity constrained duopoly. *European Journal of Operational Research*, 197(3):945–968.

Denholm, P., Margolis, R. M., and Drury, E. (2009). *The Solar Deployment System (SolarDS) Model: Documentation and Sample Results*. National Renewable Energy Laboratory.

Deo, S. and Gurvich, I. (2011). Centralized vs. decentralized ambulance diversion: A network perspective. *Management Science*, 57(7):1300–1319.

Do, H. and Shunko, M. (2015). Pareto-improving coordination policies in queueing systems with independent service agents. *Available at SSRN 2351965*.

Dong, J., Yom-Tov, E., and Yom-Tov, G. B. (2018). The impact of delay announcements on hospital network coordination and waiting times. *Management Science*.

Drury, E., Denholm, P., and Margolis, R. (2013). Sensitivity of Rooftop PV Projections in the SunShot Vision Study to Market Assumptions. *Contract*, 303:275–3000.

Dudey, M. (1992). Dynamic edgeworth-bertrand competition. *The Quarterly Journal of Economics*, 107(4):1461–1477.

Dudey, M. (2007). Quantity pre-commitment and price competition yield bertrand outcomes. Technical report, Working Paper, Rice University.

Enders, P. (2010). Applications of Stochastic and Queueing Models to Operational Decision Making. *Carnegie Mellon University*. [PhD Dissertation].

Energy.gov (2017). Residential Renewable Energy Tax Credit. [accessed on February 2, 2018].

Environmental Protection Agency (2017). The Scoial Cost of Carbon. [accessed on February 2, 2018].

Florida Hospital (2016). Know ER wait times before you go. [accessed on August 8, 2017].

Fu, R., Chung, D., Lowder, T., Feldman, D., Ardani, K., and Margolis, R. (2016). US Solar Photovoltaic System Cost Benchmark Q1 2016. Technical report, NREL/PR-6A20-66532. Golden, CO: National Renewable Energy Laboratory (NREL).

Gagnon, P. and Sigrin, B. (2015). Distributed PV Adoption in Maine through 2021. Technical report, NREL (National Renewable Energy Laboratory (NREL), Golden, CO (United States)).

Gallego, G. and Hu, M. (2014). Dynamic pricing of perishable assets under competition. *Management Science*, 60(5):1241–1259.

Goodarzi, S., Aflaki, S., and Masini, A. (2018). Optimal feed-in tariff policies: The impact of market structure and technology characteristics. *Production and Operations Management*.

Google (2018). Google Project Sunroof. [accessed on January 10, 2018].

Groeger, L., Tigas, M., and Wei, S. (2014). ER wait watcher - which emergency room will see you the fastest? [accessed on August 8, 2017].

Guo, P. and Zipkin, P. (2007). Analysis and comparison of queues with different levels of delay information. *Management Science*, 53(6):962–970.

Halfin, S. (1985). The shortest queue problem. *Journal of Applied Probability*, 22(4):865–878.

Hassin, R. (1986). Consumer information in markets with random product quality: The case of queues and balking. *Econometrica: Journal of the Econometric Society*, pages 1185–1195.

Hassin, R. (1996). On the advantage of being the first server. *Management Science*, 42(4):618–623.

Hassin, R. (2016). *Rational queueing*. CRC press.

He, Y.-T. and Down, D. G. (2009). On accommodating customer flexibility in service systems. *INFOR: Information Systems and Operational Research*, 47(4):289–295.

Hordijk, A. and Koole, G. (1990). On the optimality of the generalized shortest queue policy. *Probability in the Engineering and Informational Sciences*, 4(4):477–487.

HospitalStats (2017). ER Wait Time in San Jose Hospitals. [accessed on August 8, 2017].

Hu, S., Souza, G. C., Ferguson, M. E., and Wang, W. (2015). Capacity Investment in Renewable Energy Technology with Supply Intermittency: Data Granularity Matters! *Manufacturing & Service Operations Management*, 17(4):480–494.

Ibrahim, R. (2018). Sharing delay information in service systems: a literature survey. *Queueing Systems*, 89(1-2):49–79.

Ibrahim, R., Armony, M., and Bassamboo, A. (2015). Does the past predict the future? The case of delay announcements in service systems.

Ibrahim, R. and Whitt, W. (2009a). Real-time delay estimation based on delay history. *Manufacturing & Service Operations Management*, 11(3):397–415.

Ibrahim, R. and Whitt, W. (2009b). Real-time delay estimation in overloaded multiserver queues with abandonments. *Management Science*, 55(10):1729–1742.

Ibrahim, R. and Whitt, W. (2011). Wait-time predictors for customer service systems with time-varying demand and capacity. *Operations Research*, 59(5):1106–1118.

Intergovernmental Panel on Climate Change (2014). Technology-specific Cost and Performance Parameters. [accessed on February 2, 2018].

International Civil Aviation Organization (2019). Low Cost Carriers. [accessed on January 29, 2019].

Isler, K. and Imhof, H. (2008). A game theoretic model for airline revenue management and competitive pricing. *Journal of Revenue & Pricing Management*, 7(4):384–396.

Ivan Penn (2017). California Invested Heavily in solar power. [accessed on February 2, 2018].

Jeff, St. John (2016). The California Duck Curve Is Real, and Bigger Than Expected. [accessed on December 4, 2017].

Jerath, K., Netessine, S., and Veeraraghavan, S. K. (2010). Revenue management with strategic customers: Last-minute selling and opaque selling. *Management Science*, 56(3):430–448.

Jordan, D. C., Smith, R., Osterwald, C., Gelak, E., and Kurtz, S. R. (2010). Outdoor PV degradation comparison. In *Photovoltaic Specialists Conference (PVSC), 2010 35th IEEE*, pages 002694–002697. IEEE.

Jouini, O., Akşin, O. Z., Karaesmen, F., Aguir, M. S., and Dallery, Y. (2015). Call center delay announcement using a newsvendor-like performance criterion. *Production and Operations Management*, 24(4):587–604.

Jouini, O., Aksin, Z., and Dallery, Y. (2011). Call centers with delay information: Models and insights. *Manufacturing & Service Operations Management*, 13(4):534–548.

Jouini, O., Dallery, Y., and Akşin, Z. (2009). Queueing models for full-flexible multi-class call centers with real-time anticipated delays. *International Journal of Production Economics*, 120(2):389–399.

Keilson, J. (1965). A review of transient behavior in regular diffusion and birth-death processes. part ii. *Journal of Applied Probability*, 2(2):405–428.

Keyes, J. and Rábago, K. (2013). A Regulator's Guidebook: Calculating the Benefits and Costs of Distributed Solar Generation. *Latham, NY: Interstate Renewable Energy Council*.

Kök, A. G., Shang, K., and Yücel, Ş. (2016). Impact of electricity pricing policies on renewable energy investments and carbon emissions. *Management Science*, 64(1):131–148.

Krysti Shallenberger (2017a). Is Rooftop Solar Just a Toy For the Wealthy? [accessed on November 29, 2017].

Krysti Shallenberger (2017b). Maine Lawmakers Fail to Override Governor's Veto of Solar Bill. [accessed on December 4, 2017].

Latouche, G. and Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*, volume 5. Siam.

Lazard (2017). Lazard's levelized cost of energy analysis.

Lehr, R. L. (2013). New Utility Business Models: Utility and Regulatory Models for the Modern Era. *The Electricity Journal*, 26(8):35–53.

Levin, Y., McGill, J., and Nediak, M. (2009). Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *Management science*, 55(1):32–46.

Lin, K. Y. and Sibdari, S. Y. (2009). Dynamic price competition with discrete customer choices. *European Journal of Operational Research*, 197(3):969–980.

Linvill, C., Shenot, J., and Lazar, J. (2013). Designing Distributed Generation Tariffs Well. *Montpelier, VT: Regulatory Assistance Project*.

Liu, Q. and Zhang, D. (2013). Dynamic pricing competition with strategic customers under vertical product differentiation. *Management Science*, 59(1):84–101.

Lobel, R. and Perakis, G. (2011). Consumer Choice Model for Forecasting Demand and Designing Incentives for Solar Technology.

Mantin, B., Granot, D., and Granot, F. (2011). Dynamic pricing under first order markovian competition. *Naval Research Logistics (NRL)*, 58(6):608–617.

Martínez-de Albéniz, V. and Talluri, K. (2011). Dynamic price competition with fixed capacities. *Management Science*, 57(6):1078–1093.

Mike Munsell (2017). US Solar Market Grows 95% in 2016, Smashes Records. [accessed on December 4, 2017].

Mookherjee, R. and Friesz, T. L. (2008). Pricing, allocation, and overbooking in dynamic service network competition when demand is uncertain. *Production and Operations Management*, 17(4):455–474.

Moore, J., Keen, J., and Apt, J. (2016). The Sustainability of Net Metering. *Electric Markets Research Foundation*.

NC Clean Energy Technology Center (2017a). Net Metering Program Overview. [accessed on December 4, 2017].

NC Clean Energy Technology Center (2017b). Net Metering Program Overview. [accessed on December 4, 2017].

NC Clean Energy Technology Center and Meister Consultants Group (2015). The 50 States of Solar.

NV Energy (2015). Net Metering Overview. [accessed on January 6, 2017].

NV Energy (2016a). Consoidated Financial Statements. [accessed on February 2, 2018].

NV Energy (2016b). Renewable Portfolio Standard Annual Report. [accessed on February 2, 2018].

NV Energy (2017). Rates and Regulatory. [accessed on February 2, 2018].

O'Donnell, K. (2014). Head of troubled Phoenix VA hospital removed. [accessed on June 26, 2016].

Ong, S., Denholm, P., and Doris, E. (2010). *The Impacts of Commercial Electric Utility Rate Structure Elements on the Economics of Photovoltaic Systems*. National Renewable Energy Laboratory.

Pender, J., Rand, R. H., and Wesson, E. (2018). An analysis of queues with delayed information and time-varying arrival rates. *Nonlinear Dynamics*, 91(4):2411–2427.

Perez, S. (2014). Nowait, the app that lets you join a restaurant wait list from your phone, goes nationwide. [accessed on August 8, 2017].

PNM Energy (2017). 2017 Form 10-K. [accessed on February 2, 2018].

PNM Energy (2018). Rates. [accessed on February 2, 2018].

Pyper, J. (2016). Nevada Regulators Restore Retail-Rate Net Metering in Sierra Pacific Territory. [accessed on January 6, 2017].

Rai, V. and Sigrin, B. (2013). Diffusion of Environmentally-Friendly Energy Technologies: Buy Versus Lease Differences in Residential PV Markets. *Environmental Research Letters*, 8(1):014022.

Ramirez-Nafarrate, A., Hafizoglu, A. B., Gel, E. S., and Fowler, J. W. (2014). Optimal control policies for ambulance diversion. *European Journal of Operational Research*, 236(1):298–312.

Rawls, J. (1974). Some reasons for the maximin criterion. *The American Economic Review*, 64(2):141–146.

Ritzenhofen, I., Birge, J. R., and Spinler, S. (2016). The Structural Impact of Renewable Portfolio Standards and Feed-in Tariffs on Electricity Markets. *European Journal of Operational Research*, 255(1):224–242.

Robert Walton (2016). New Mexico Settlement Blocks Solar Fees Proposed by Xcel Energy. [accessed on March 15, 2018].

Roselund, C. (2015). SolarCity Installs More Than 1/3 of All U.S. Residential Solar in 2015. [accessed on October 24, 2016].

Sadick, B. (2012). No wait at the ER. [accessed on August 8, 2017].

Satchwell, A., Mills, A., and Barbose, G. (2015). Quantifying the Financial Impacts of Net-Metered PV on Utilities and Ratepayers. *Energy Policy*, 80:133–144.

Sawers, P. (2017). Yelp acquires restaurant waiting list tech startup Nowait in a $40 million all-cash deal. [accessed on March 17, 2019].

Selen, J., Adan, I., Kapodistria, S., and van Leeuwaarden, J. (2016). Steady-state analysis of shortest expected delay routing. *Queueing Systems*, 84(3-4):309–354.

Shahan, Z. (2016). Cost Of Solar Systems From SolarCity vs Sunrun & Vivint Solar, and Cost Per Watt Of Solar Sales/Acquisition. [accessed on October 24, 2016].

Solar-Estimate (2012). Solar and Wind Calculations: The Basics. [accessed on January 6, 2017].

SolarCity (2016a). Cost Calculation Methodology. [accessed on January 6, 2017].

SolarCity (2016b). Solar Energy Production. [accessed on January 6, 2017].

SolarCity (2016c). Solar Panel Lifespan. [accessed on January 6, 2017].

SolarCity (2016d). SolarCity's SolarPPA. [accessed on October 24, 2016].

SolarDirect (2016). Solar Electric System Sizing - Determine the Sun Hours Available Per Day. [accessed on January 6, 2017].

Sunar, N. and Swaminathan, J. M. (2018). Net-metered distributed renewable energy: A peril for utilities? *Kenan Institute of Private Enterprise Research Paper*, (18-24).

Sundararajan, A. (2004). Nonlinear pricing of information goods. *Management Science*, 50(12):1660–1673.

Talluri, K. T. and Van Ryzin, G. J. (2004). *The theory and practice of revenue management*, volume 68. Springer Science & Business Media.

TechCrunch (2015). Google Search Now Shows You When Local Businesses Are Busiest. [accessed on March 20, 2019].

Tran-Gia, P. (1988). A class of renewal interrupted poisson processes and applications to queueing systems. *Zeitschrift für Operations-Research*, 32(3-4):231–250.

US Department of Energy (2013). Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States. [accessed on January 6, 2017].

U.S. Energy Information Administration (2009). 2009 RECS Survey Data. [accessed on October 24, 2016].

U.S. Energy Information Administration (2015). 2015 RECS Survey Data. [accessed on January 31, 2018].

U.S. Energy Information Administration (2016a). Nevada State Profile and Energy Estimates. [accessed on February 2, 2018].

U.S. Energy Information Administration (2016b). New Mexico State Profile and Energy Estimates. [accessed on February 2, 2018].

U.S. Energy Information Administration (2017). Wholesale Electricity and Natural Gas Market Data. [accessed on January 11, 2017].

Varian, H. R. (1989). Price discrimination. *Handbook of industrial organization*, 1:597–654.

Villa, D. L., Reno-Trujillo, M. D., and Passell, H. D. (2012). Sun City Progress Report: Policy Effects on Photovoltaic Adoption for City Planning. Technical report, Sandia National Laboratories.

Whitt, W. (1999). Improving service by informing customers about anticipated delays. *Management Science*, 45(2):192–207.

Wu, O. Q. and Kapuscinski, R. (2013). Curtailing Intermittent Generation in Electrical Systems. *Manufacturing & Service Operations Management*, 15(4):578–595.

Xu, X. and Hopp, W. J. (2006). A monopolistic and oligopolistic stochastic flow revenue management model. *Operations Research*, 54(6):1098–1109.

Zhao, X. and Atkins, D. (2002). Strategic revenue management under price and seat inventory competition. Technical report, Working Paper, University of British Columbia.

Zhou, Y., Scheller-Wolf, A., Secomandi, N., and Smith, S. (2014). Managing wind-based electricity generation in the presence of storage and transmission capacity. *Production and Operations Management*.