From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2017

Essays on Market Structure and Product Quality

Goetz, Daniel Thomas

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp013t945t42c

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS ON MARKET STRUCTURE AND PRODUCT QUALITY

Daniel T. Goetz

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ECONOMICS
Advisers: Jan de Loecker and Bo Honoré

September 2017

©Copyright by Daniel Thomas Goetz, 2017. All rights reserved.

Abstract

In three essays, this dissertation examines market structure, firm investments in product quality, and how the interplay between them can affect households.

Chapter 1 measures the effect of horizontal mergers in the broadband industry when content providers can invest in product quality. I estimate a structural model of dynamic multilateral bargaining between U.S. broadband internet service providers (ISPs) and Netflix over how to split the surplus from investments Netflix makes to improve the quality of its streaming video. Unlike previous studies of firm bargaining, I find ISP scale—independent of ISP market power—influences how much surplus ISPs extract. Allowing the two largest U.S. cable internet providers to merge would have increased the magnitude of aggregate consumer welfare loss by 4.3% due to a longer period of degraded Netflix quality, and reduced the probability of Netflix making the quality investments by between 0% and 5.1%.

In Chapter 2, I investigate how broadband ISPs provide quality in the form of faster or slower download speeds. Using a dataset of actual broadband internet speeds and the speeds advertised, I find that an additional broadband competitor raises the ratio of actual to advertised speeds for incumbents by between 23% and 32%, although the increase is temporary. This increase is due to improvements in the actual speed, and not just reductions in the claimed speed. I provide suggestive evidence that the reduction in firms' strategic misrepresentation of their products led to reduced misallocation of consumers across internet plans.

Finally, in Chapter 3, my co-author Alexander Rodnyansky and I focus on how firm cost shocks affect consumer welfare when firms can adjust the composition of qualities of their products. Using a uniquely granular online retail dataset that spans Russia's currency depreciation in late 2014, we document that firms lower product qualities in response to a positive cost shock instead of raising prices. A structural model of the apparel industry that endogenizes the distribution of costs and qualities can rationalize the quality reductions. Compared to a model that features only cost heterogeneity, our model predicts larger drops in welfare after a positive cost shock.

Acknowledgments

I first and foremost wish to acknowledge my primary adviser, Jan de Loecker. He has helped me through countless iterations of my research, and his generosity with his time and insights about big picture questions has been invaluable during my journey to become a researcher and economist. I also thank my second adviser, Bo Honoré. No one has done more to teach me the value of thoroughness in research, and Bo's enthusiasm and humour has been an inspiration throughout my Ph.D. For their active influence in my development, I would also like to thank Jakub Kastl, Myrto Kalouptsidi, Nicholas Buchholz, and Eduardo Morales from the economics department at Princeton, as well as Anindya Sen from the University of Waterloo.

I owe much thanks to participants in the graduate IO/Applied Micro lunch seminar groups; in particular, C.J. Verbeck, Giulia Brancaccio, Molly Schnell and Moshe Katzwer. I would also like to thank Gabby Herrera, without whose pro-active administrative help the lunch seminar would not have been possible, and Laura Hedden, for her tireless administrative support and compassionate commitment to the prospects of each and every Ph.D. student under her care.

To my co-author Alexander Rodnyansky I would like to express my gratitude; this dissertation would have been much the lesser if not for his persistence, insights, and commitment to our joint work.

The Griswold Center for Economic Policy and the NET Institute at NYU both gave a vote of confidence in this research, which meant far more to me than the financial help alone. I would also like to thank Princeton University, the International Economics Section at Princeton University, and the Dean's Fund for Scholarly Travel for assistance.

Without friends to share it with the Ph.D. would have been a far more difficult journey, and I thank Justin Weider, Alex Jakobsen and Cristian Alonso for all the times spent together. I also want to thank Jeff Chan and Andrew Dickens for reminding me that Princeton was not the beginning or end of economics research, and Sharon Traiberman for never doubting in my ability to deliver.

I owe a deep and lifelong debt of thanks to my parents, Laurie and Wendel, for instilling

in me a love of learning and the confidence to pursue my ideas—their support and sacrifice enabled everything. For his honesty, advice, and unconditional willingness to stand beside me against any and all comers I am forever grateful to my brother Nickolas. To my uncle Neil, thank you for your friendship and for providing a port in any storm. Finally I would like to thank Megan Steffen for always being her loving, forgiving and perceptive self, and for showing me how it's done—dissertation-wise not least of all.

Contents

1 Competition and Dynamic Bargaining in the Broadband Industry 1

1.1 Introduction . . . 1

1.1.1 Related literature . . . 6

1.2 The Broadband Industry . . . 7

1.2.1 Structure of the Internet . . . 7

1.2.2 Netflix Bargaining Event . . . 8

1.3 Data . . . 10

1.3.1 Demand Data . . . 10

1.3.2 Supply Data . . . 11

1.4 Patterns of Consumer Substitution and Agreement Timing . . . 12

1.4.1 Agreement timing patterns . . . 12

1.4.2 Consumers switch ISPs . . . 14

1.4.3 Netflix Demand is Inelastic . . . 15

1.5 Model . . . 16

1.5.1 Upstream Bargaining . . . 17

1.5.2 ISP Plan Pricing and Profits . . . 22

1.5.3 Demand . . . 23

1.6 Estimation and Identification . . . 27

1.6.1 Demand . . . 27

1.6.2 Upstream Bargaining . . . 33

1.7 Empirical Results . . . 37

1.7.1 Demand estimates . . . 37

1.7.2 Upstream Bargaining Estimates . . . 39
1.8 Counterfactuals . . . 42
1.8.1 Comcast-TimeWarner merger . . . 42
1.8.2 Prohibiting Quality Degradation During Bargaining . . . 46
1.9 Conclusion . . . 47
1.10 Appendices . . . 49
1.10.1 Appendix A: Tables . . . 49
1.10.2 Appendix B: Figures . . . 58
1.10.3 Appendix C: Data Construction . . . 64

2 Competition and Product Misrepresentation 70
2.1 Introduction . . . 70
2.2 Literature and Research Questions . . . 73
2.3 Data . . . 75
2.4 Modelling and Estimation . . . 77
2.4.1 Identification . . . 78
2.5 Results . . . 81
2.5.1 Promise Ratio . . . 81
2.5.2 Extensions . . . 82
2.5.3 Robustness . . . 84
2.6 Conclusion . . . 84
2.7 Appendices . . . 86
2.7.1 Appendix A: Tables . . . 86
2.7.2 Appendix B: Figures . . . 97

3 Endogenous Quality and Exchange Rate Pass-Through 99
3.1 Introduction . . . 99
3.1.1 Relation to the Literature . . . 102
3.2 Data . . . 104
3.2.1 Data Sources . . . 104
3.2.2 Institutional Background . . . 105

3.3 Empirical Motivation . . . 106
3.3.1 Quality . . . 107
3.3.2 Demand Channel . . . 108
3.3.3 Prices . . . 110
3.4 Model . . . 113
3.4.1 Overview . . . 113
3.4.2 Notation and timing . . . 115
3.4.3 Model primitives and assumptions . . . 116
3.4.4 Equilibrium . . . 117
3.5 Estimation . . . 118
3.5.1 Moments . . . 120
3.6 Constructing moments from the data . . . 120
3.6.1 Demand . . . 121
3.6.2 Supply . . . 122
3.6.3 Estimation and identification of the moment-generating model . . . 123
3.6.4 Moment-generating model results . . . 123
3.7 Results . . . 124
3.7.1 Model Fit . . . 124
3.7.2 Counterfactual . . . 125
3.8 Conclusion . . . 126
3.9 Appendices . . . 128
3.9.1 Appendix A: Tables . . . 128
3.9.2 Appendix B: Figures . . . 130

List of Tables

1.1 Cross-sectional ISP disagreement durations 49
1.2 Panel of correlated ISP agreement timings 50
1.3 Panel of household ISP switching 51
1.4 Panel of ISP subscriber share changes 52
1.5 Panel of household speed upgrading 53
1.6 Demand Model Parameter Estimates 54
1.7 Quality Degradation Induced Consumer Substitution 55
1.8 Supply model estimates 56
1.9 Choice set stats 57
1.10 Conditional means of demographics from the ACS 58
1.11 Complete Durations Data 69

2.1 Means by Date 86
2.2 Ex Ante Differences Between Counties 86
2.3 Effect of Thunderstorms on Internet Quality 87
2.4 Effect of Thunderstorms on Internet Speeds 88
2.5 Main Results 89
2.6 Main Results by Initial Competition 90
2.7 Promise Ratio Decomposition 91
2.8 Effect of treatment on peak consumption 92
2.9 Effect of treatment on switching 93
2.10 Robustness: only ratios ≤ 1 94
2.11 Robustness: different sat. share cutoffs 95

2.12 Robustness: different terrestrial competitors 96

3.1 Within-SKU Pass-through 128
3.2 Parameters and moments 128
3.3 Structural demand parameter estimates 129
3.4 Reallocation towards low cost goods 129

List of Figures

1.1 Average Netflix throughput to 21 U.S. ISPs 58
1.2 Time Variation in Internet Data 59
1.3 Histogram of Disagreement Durations 60
1.4 Residual ISP subscriber growth rates by whether negotiation occurred 60
1.5 Netflix subscriber growth rates 61
1.6 Joint distribution of agreement timings under different profit interactions . . 61
1.7 Supply Model Fit 62
1.8 Bargaining Delay Changes 62
1.9 Distributions of Netflix Surplus Net of Transfers 63
1.10 Distributions of Netflix Surplus Net of Transfers 64
1.11 ISP-Netflix Pairwise Throughput in NYC, LA and DC 66

2.1 Promise Ratio ECDF at 4 ISPs 97
2.2 Ex Ante Satellite Subscription Shares 98

3.1 Overlapping Generations of Goods 130
3.2 Cost of Goods Sold 131
3.3 Price Indexes 131
3.4 Price Adjustments 132
3.5 Fabrics & Quality Index 133
3.6 Demand Channel 134
3.7 Small, positive relationship between cost and quality 135

Chapter 1

Competition and Dynamic Bargaining in the Broadband Industry

1.1 Introduction

Two-sided markets, where intermediaries provide consumers access to content supplied by third parties, are increasingly common.[1] Recent empirical work including Dafny, Ho and Lee (2016) and Gowrinsankaran, Nevo and Town (2015) has sought to understand how market power in two-sided markets affects consumer welfare. In this literature, the quality of available content is taken as given; as a result, greater intermediary market power against the content side of the market can only help consumers, as larger intermediaries negotiate better terms with content providers and then pass those efficiencies on to consumers (as in Crawford and Yurukoglu (2012)) without any reduction in content quality.

This paper evaluates the consequences of market structure in the market for broadband internet when the quality of content is not fixed. In this market, content streaming quality—for instance, how quickly and at what resolution online videos load—is an impor-

[0]Thanks to Jan de Loecker, Bo Honoré, Jakub Kastl, and Myrto Kalouptsidi for substantial guidance on this project. I am also thankful to participants at the 2016 CEA Annual Conference where this paper was presented, as well as participants in the Applied Micreconomics lunch at Princeton University. This project received financial support from the Griswold Center for Economic Policy at Princeton University and the NET Institute at New York University.

[1]Older industries with two-sided characteristics such as cable television, medical insurance, and broadband internet have recently been joined by online platforms providing access to vacation homes, ride sharing services, and contract work.

tant component of internet users' experience. Motivated by data on bargaining between U.S. internet service providers (ISPs) and Netflix—the leading provider of streaming video content—over the terms of interconnection, I build a structural model where streaming quality can be degraded in two ways. First, bargaining frictions between ISPs and Netflix will imply that negotiations take time, and Netflix streaming quality is degraded at an ISP until negotiations resolve. Second, bargaining is over how to split the surplus from a Netflix capacity investment that will maintain streaming quality; if too much of Netflix's surplus from the investment is extracted by ISPs then Netflix may choose not to invest. I structurally estimate the model and simulate the effects of counterfactual policies, including a merger and a prohibition on quality reductions during bargaining, on both bargaining durations and on Netflix's incentive to invest.

There are two novel empirical findings from this paper, as well as a methodological contribution. The first empirical finding is that both ISP size and ISP market power matter for how much surplus ISPs extract from Netflix's ISP-specific investments. An ISP with more market power—whose subscribers cannot switch to another provider when Netflix quality is degraded during bargaining—extracts more surplus; however, an ISP that is present in many markets—even if its consumers have multiple internet options in all the markets the ISP is present in—will also extract relatively more surplus. The intuition behind the result that ISP size matters independently from market power is that Netflix has economies of scale in its surplus which weakens its bargaining position against ISPs that serve more subscribers. The second empirical finding is that in this two-sided market with quality investment, the ISPs find it optimal to extract Netflix's investment surplus even though doing so disincentivizes Netflix from making the investment. This result need not be true in theory, as in a two-sided market the intermediary may find it optimal to incentivize investment in content quality to charge a higher markup to consumers who value quality. Finally, the paper makes a methodological contribution with a model of dynamic bargaining that takes data on bargaining durations and consumer substitution in response to reduced quality and recovers information about the magnitude and split of Netflix's investment surplus.

The empirical core of this paper is a cross-section of bargaining durations between U.S. ISPs and Netflix over how to split the surpluses from Netflix's ISP-specific capacity invest-

ments. In reduced form regressions, an ISPs' market footprint is positively and significantly associated with bargaining length, while the number of competitors an ISP faces in the markets it serves has a negative but insignificant effect. This cross-sectional correlation forms the basis for the insight that ISP scale, in addition to market power, matters for bargaining. Panel data regressions of ISP subscriber numbers on Netflix quality—which is reduced during bargaining—show that a small but significant number of subscribers value Netflix enough to leave the affected ISPs, suggesting a welfare cost to longer bargaining, as well as a cost to ISPs.

The supply model is a structural joint durations model that rationalizes the delays in bargaining between Netflix and the ISPs, and which maps delays and subscriber substitution into the split of Netflix's investment surplus. As in the data, in the model Netflix can make ISP-specific investments that generate surplus in two ways: an investment improves transmission of Netflix content to its subscribers at an ISP, which is a quality improvement whose value to consumers I identify from demand data, and an investment also saves Netflix money on transmission costs, the value of which only Netflix knows. I assume that ISPs do not know the exact value of the surplus they face, but they do know its distribution, and that they make take-it-or-leave-it offers of interconnection fees in each bargaining period. Netflix will reject a high fee offer to signal that its investment surplus with that ISP is low, leading to positive probabilities of delay. As in the data, in the model Netflix streaming quality remains degraded at an ISP until the capacity investment is made.

The model maps bargaining durations to surplus splits by assuming that longer periods of bargaining reflect strategic choices by the ISPs to screen Netflix's surplus more finely. ISPs' strategic tradeoff in the model is between making a low initial fee offer to Netflix, in which case the offer will be accepted quickly and the ISP will have avoided losing subscribers but will not extract much surplus from Netflix, versus making high initial offers that slowly decrease, which may cost the ISP subscribers during periods of quality degradation but which will extract more of Netflix's surplus in expectation. Consumers' willingness and ability to substitute away from ISPs with degraded Netflix quality is a measure of ISP market power, and lower market power will make a finer screening strategy less profitable. Consumer substitution also generates interactions between ISPs' screening strategies: by offering a low

fee to Netflix that is likely to be accepted an ISP can poach subscribers from rivals whose quality remains degraded. How Netflix's ISP-specific distribution of surplus varies with ISP characteristics—such as the physical size of the ISP's network—captures whether Netflix experiences economies of scale in its cost-savings.

The model is estimated in a two-step procedure: I first recover the demand for internet access as a function of Netflix quality degradation and plan features; I then assume ISPs take the demand curve as given and estimate the unobserved surplus distributions—and associated optimal screening strategies—that maximize the model's probability of matching the bargaining delays in the data.

I estimate consumer preferences for internet access and for Netflix streaming quality using individual level data on internet plan choices, combined with nationwide ISP subscriber numbers and data on Netflix-ISP pairwise quality of service degradation. While ISPs charge higher monthly subscription fees for plans with more features, this pricing and feature schedule varies little nationwide, making it difficult to estimate how consumers value internet service with market share data alone. I use fine geographic variation in the set of available ISPs and plans to identify the distribution of preferences for price, download speed, access technology, and Netflix quality of service in a mixed-logit framework with random choice sets.

I assume that the observed data on bargaining durations is a perfect Bayesian equilibrium of the multilateral dynamic screening model. Given parameters, I solve for ISPs' optimal state-contingent fee offers, which imply a joint distribution of bargaining durations for all ISPs with which I can perform MLE. Intuitively, if ISPs with certain characteristics—i.e., larger networks—take longer to bargain on average, then conditional on subscriber loss they must be facing surplus distributions with higher means to make the finer screening strategy worthwhile.

From the demand curve results, I estimate that during periods of disagreement, ISP market shares decrease by about 0.5 percent on average with substantial heterogeneity across ISPs. For instance, Comcast is predicted to lose more, and higher-margin, subscribers than AT&T during bargaining, which will help match the fact that Comcast agrees more quickly in the data. The median price elasticity among all ISP-plan-quarter combinations is 2.667,

substantially higher than the 0.7 estimated in Dutz, Orszag and Willig (2012).

In the bargaining game, I estimate that there are increasing returns to scale in Netflix's cost-saving infrastructure investment; that is, the mean and variance of the distribution of surplus are convex in the number of housing units ISPs are connected to. These estimates are driven by a key reduced form fact in the data: network size is strongly positively associated with bargaining durations even after controlling for measures of ISP market power. The surplus distribution mean being convex in size makes sense of this correlation by showing Netflix brings relatively more surplus to a larger ISP, so that ISP finds it worthwhile to trade off greater expected lost subscriber revenue in order to screen Netflix's surplus more finely. That returns to scale in investment can be a source of bargaining power in two-sided markets is a key insight of the paper. I find that Netflix retains only 29% of the surplus against the four largest ISPs, but 34% against the smaller ISPs, implying that ISP size worsens a hold-up problem in Netflix's quality investment.

With the estimates in hand, I examine two regulatory interventions: a merger between Comcast and TimeWarner Cable—the two largest providers of cable internet in 2013 in the U.S.—that was blocked by the FCC in 2015, and a policy that prohibits content quality degradation during bargaining.

My main counterfactual allows Comcast and TimeWarner to merge before the bargaining event begins. The model predicts that the magnitude of the aggregate consumer welfare loss due to degraded quality during bargaining increases by 4.3 percentage points after the merger, and Netflix's share of the upstream surplus from its investment decreases by 9.4 percentage points. In reality, since Netflix must pay an upfront cost to make the investment, Netflix therefore faces a holdup problem that is exacerbated by the merger. I find that Netflix's probability of making the investment in interconnection infrastructure that precipitated bargaining would be reduced by 5.1 percent under the merger. Intuitively, the merged firm has more to gain from screening than Comcast or TimeWarner combined due to economies of scale in Netflix's interconnection surplus, but the merged firm's marginal loss in subscribers from disagreement increases only linearly. The tradeoff between losing subscribers and extracting investment surplus tips towards extracting surplus, giving the merged firm the incentive to prolong negotiations. Moreover, in the estimated model the

marginal gain in subscribers to an ISP from concluding bargaining is larger when rival ISPs have not yet agreed; thus, the merged firm's finer screening induces slightly more aggressive offers on average by competing ISPs, decreasing disagreement lengths and improving welfare for consumers at those ISPs.

In my second counterfactual, I prohibit the degradation of Netflix quality during bargaining. This policy is related to recent Federal Communication Commission rules on network neutrality that ban ISPs from selectively degrading content providers' connection quality to the end consumer; my policy is slightly more general, in that it bans both content providers and ISPs from degrading the connection quality. I find consumer welfare increases, with the largest increases at large ISPs like AT&T and Comcast whose bargaining took the longest to resolve. However, bargaining times actually increase and Netflix's surplus share decreases, making Netflix 8.7 percentage points less likely to invest in interconnection infrastructure. This follows from the fact that in the data, marginal consumers are more willing to switch ISPs than cancel Netflix in response to reduced quality, so prohibiting quality degradation actually removes Netflix's advantage in bargaining.

1.1.1 Related literature

This paper contributes to several literatures within industrial organization. To begin, I add to a growing literature that analyzes bargaining between firms along the supply chain. Starting with Grennan (2013), recent industry analyses have endogenized firm costs by posing a model of upstream bargaining over inputs. Papers that allow for substitution by downstream consumers in response to upstream bargaining include Crawford and Yurukoglu (2012), Gowrinsankaran, Nevo and Town (2015) and Dafny, Ho and Lee (2016). Consumer substitution is a key source of upstream bargaining incentives, and I extend these papers' insights to a dynamic setting.

A literature including Evans (2003), Evans (2010) and Evans and Schmalensee (2013) has recognized that two-sided markets, like the market for internet service, face a unique set of antitrust issues, since market power may only be exerted against the content side of the market. Empirical work including Argentesi and Filistrucchi (2007) and Chandra and Collard-Wexler (2009) focuses on the consequences of intermediary market power on content-

side prices, and on exit and entry of content providers. I contribute to these analyses by allowing content providers to negotiate with intermediaries, and by allowing the margin of adjustment on the content side to be investment by an existing content provider, as opposed to entry and exit of content.

There is a small but growing set of tools to empirically analyze dynamic bargaining. A complete information framework for multilateral dynamic bargaining is developed in Merlo and Wilson (1995) and Merlo and Tang (2012), where delays arise if the value of surplus to be split is stochastic and may rise over time. My model complements their analysis by allowing for delay even with constant surplus and introducing downstream competition among bargaining parties. Ambrus, Chaney and Salitsky (2016) have structurally estimated a dynamic bargaining game with incomplete information, and I extend their framework to the multilateral setting. Bargaining durations resemble joint optimal stopping problems, notable examples of which include Berry and Tamer (2006),Honoré and de Paula (2010) and Björkegren (2015). I draw on insights from this literature, especially with regards to how payoff interactions affect the set of equilibira.

Finally, the paper contributes to a theory literature on whether allowing content quality degradation is welfare improving. Lee and Wu (2009), Becker, Carlton and Sider (2010), Economides and Hermalin (2012) Gans (2015), and Peitz and Schuett (2016) have analyzed whether a policy of network neutrality—that is, prohibiting ISPs from selectively reducing content quality—is welfare improving, and what types of distributional impacts different neutrality policies will have. I add to this discussion the idea that temporary violations of neutrality may actually be to the benefit of content providers if consumers are willing to substitute ISPs to improve content quality.

1.2 The Broadband Industry

1.2.1 Structure of the Internet

The internet is a two-sided market. On one side are consumers, who purchase access to the internet in order to consume online services and content such as email and streaming video.

On the other side are the providers of services and content, such as Google and Netflix, who charge consumers either indirectly, via advertisements, or directly, via subscription fees, for using services and viewing content. In the middle are layers of firms that intermediate the relationship between consumers and content providers. In what is to follow I refer to the service/content side of the market as content providers.

Consumers and small businesses interact with "last-mile" or "edge" internet service providers like Comcast and Verizon. A consumer's choice set for wired internet service depends on which ISPs have infrastructure connected to her house, since last-mile ISPs have the exclusive right to sell service on infrastructure they own. Service is differentiated by infrastructure technology (cable, fiber optic, etc.) across providers, and by tiered menus of plans varying by monthly price and download speed in megabits per second (MB/s) within providers.[2] By 2013, 70% of households had access to two or more wired providers offering maximum download speeds greater than 10MB/s. However, the industry is concentrated: for 91% of those consumers, at least one alternative was provided by the four largest last-mile ISPs: AT&T, Comcast, TimeWarner, and Verizon.

Netflix and other large content providers seek to connect to last-mile ISPs, and have several options to do so. The largest, like Google or Microsoft, incur a large fixed cost to install infrastructure that allows them to connect directly with last-mile ISPs at low variable cost. Others buy access from "transit" ISPs like Level3 and Cogent, who connect with last-mile providers to transmit content to consumers. Using third parties to transmit content comes with a higher variable cost, and content providers must ensure they purchase sufficient access to meet consumer demand. To avoid purchasing enough transit access to meet demand at peak times, content companies can also pay to upload content to so-called "content delivery networks" (CDNs)—caches of servers distributed around the country that ensure no consumer is far from a content source.

[2]Upload speed in MB/s, caps on how much content can be consumed in a month, and contract length are also plan features, but these are much less important: 92.5% of respondents in the 2013 Current Population Survey Internet Supplement list price, download speed, or reliability as the most important feature of service, from a list of choices that also includes upload speed, data usage caps, mobility, and bundling options.

1.2.2 Netflix Bargaining Event

Starting in mid-2012, Netflix developed a strategy to transition from using mainly third parties to disseminate content, to using its own infrastructure. They developed a custom CDN, called Open Connect, and in so doing incurred a large fixed development and deployment cost.[3] Open Connect would save Netflix money in two ways. First, it would allow them to save on the variable cost of using third party CDNs. As the largest online video distributor, Netflix not only paid transit ISPs for connections and the CDNs for servers, but also pursued a policy of paying the fees that last-mile ISPs charged CDNs and transit ISPs carrying Netflix content.[4] Second, by locating the servers inside last-mile ISPs' own networks, Netflix would no longer need to ensure that it paid for sufficient bandwidth from transit ISPs to accommodate demand at peak times. With Open Connect servers located in, for instance, Comcast's network, Netflix could update the servers slowly and during off-peak times when Comcast consumers were not streaming, and therefore save on transit costs.[5] Open Connect would allow Netflix to deliver service reliably and at lower cost.

By mid-2013, Netflix had not installed Open Connect in the vast majority of last-mile ISP networks, and had begun to report degraded quality of service to a number of U.S. ISPs. I emphasize the quality degradation for the largest two U.S. ISPs by subscriber count, AT&T and Comcast, who in 2013 collectively accounted for 43% of all U.S. broadband subscribers, in Figure 1.1. Starting in mid-2013, the average transmission rate of Netflix data to subscribers at these ISPs dips far below trend, and is restored after varying amounts of time. ISPs including TimeWarner (13% of subscribers) and Verizon (10.5%) also experience degradation, while Cox (5.5%) and Cablevision (3.3%) do not.

I argue that these slowdowns and their resolutions correspond to periods of bargaining disagreement over the negotiated fees for installation of Open Connect. In Figure 1.1 Comcast service quality is fully restored during the first quarter of 2014, which corresponds to Netflix FCC filings indicating that by January, 2014, Netflix and Comcast had reached a

[3]Netflix Petition to Deny, pg. 49, paragraph 1. Fixed investment in R&D and deployment on the order of $100 000 000.

[4]Paragraph 12, Statement of Ken Florance, Vice President of Content Delivery at Netflix since 2012.

[5]Netflix petition to deny, pg. 49, paragraph 2. "Open Connect...uses a 'proactive caching' method to conduct daily content updates during periods when networks are least used, such as early in the morning, to avoid congesting the network."

deal on interconnection fees.[6] AT&T service quality is only restored later: in Netflix's April 2014 Q-10 filing, they state that AT&T still has not agreed to Open Connect interconnection,[7] but data from the Center for Applied Internet Data Analysis (CAIDA) indicates that AT&T began interconnecting with Netflix in August 2014—around the time AT&T service quality is restored. When describing the event in FCC filings at the end of 2014, Netflix notes that "none of the U.S.'s four major ISPs [had] agreed to partner with Open Connect without payment", implying that the parties were indeed negotiating over explicit transfers from Netflix to the ISPs.[8]

1.3 Data

1.3.1 Demand Data

Demand data is collected from several sources. An overview of the time trends between in market shares, choice sets, and plan characteristics between 2010 and 2014 is presented in Figure 1.2.

Market shares: Data on market shares are gathered from ISPs' quarterly and yearly earnings reports (10-Q and 10-K) which are available for all publicly traded companies in the U.S. Total internet subscriber numbers are given every quarter; combined with auxiliary data on market sizes detailed below, these numbers imply nationwide market shares. The reports also contain ancillary data on mergers, which provide a source of variation in available plans. Some ISPs are privately held—e.g. Cox and RCN—in which case I use estimates of the subscription base from Leichtman Research Group. Market share movements are dominated by trends and mergers.

Plan characteristics: The menu of prices and download speeds each ISP offers are gathered primarily from the FCC Urban Rate Survey and Open Connectivity Database.[9] Where prices are missing, I collect them by hand from stored ISP frontpages on the Internet Archive Project. When the Internet Archive is unable to recover the prices—for instance,

[6] Petition to Deny, pg. 57, paragraph 2 — pg. 58 paragraph 2.
[7] Netflix 2014Q1 letter to investors, pg. 5 paragraph 3.
[8] Petition to deny, pg. 49, paragraph 2.
[9] No relationship with Netflix's Open Connect CDN.

due to prices being hidden behind a localization layer—I comb ISP-specific consumer reviews on DSLreports.com. ISPs add or drop plans from their menu across different regions, but conditional on offering a plan it is advertised at the same price everywhere during the sample period. The price per megabit averaged across offered plans drops by 58% over the sample period.

Choice sets: Most consumers have access to only one or two wired internet options. Data on what choices are available to consumers comes from the National Broadband Map (NBBM), a government initiative with data available from 2010 through 2014 which collects information at half-yearly intervals on ISP connections at the census block level. For each census block, ISPs report whether they provide service to that block and their maximum advertised speed. The maximum advertised speed truncates ISP plan offerings in that block, generating geographic variation in menus. Combined with census data on exact counts of households within each block, this data gives the weight of households across choice sets for any level of geographic aggregation. I assume that all consumers have access to satellite internet as part of their choice set. The share of consumers with access to two or more high speed ($\geq$ 25 Mb/s) providers increases from below 20% to almost 70% during the sample.

Plan microdata: I construct a time series of within-ISP plan shares using data from the FCC's Measuring Broadband America (MBBA) program. The program consists of high frequency testing data for an unbalanced panel of roughly 10 000 households from 2012 through 2014. I observe a household's ISP and their tested download speeds, which I use to back out which plan within an ISP's menu each household subscribes to in each quarter. The distribution of customers across plans changes over time, but substantial upgrading only occurs later in 2014.

Demographic microdata: The 2013 and 2014 waves of the American Community Survey (ACS) link household choices of access technology to their demographics within over 2000 Public Use Microdata Areas (PUMAs). The 2011, 2013 and 2015 waves Current Population Survey (CPS) link demographics to a variety of questions about internet usage, including use of streaming video and ISP switching behaviour.

Netflix data: I recover Netflix's quarterly subscribers, as well as the share of paid subscribers, from their Q-10 and K-10 filings. Netflix prices do not change over this time

period. Whether and to what degree Netflix streaming quality is degraded is linked to bargaining disagreement durations, which I describe in Section 1.3.2 below.

1.3.2 Supply Data

Bargaining Delays: I gather data on the Netflix quality degradation event from several sources, including the Netflix data in Figure 1.1, data from an independent measurement company MLab, and CAIDA. In addition I draw extensively from business filings; the full data construction description, as well as the list of ISPs and associated bargaining times, is provided in Section 1.10.3. Bargaining begins simultaneously for all U.S. ISPs in the third quarter of 2013, and lasts for between 0 and 4 quarters. I present the histogram of disagreement lengths in Section 1.10.2. Disagreement between an ISP and Netflix affects that ISP's customers nationwide.

1.4 Patterns of Consumer Substitution and Agreement Timing

In this section, I document two key reduced form facts in the data that will inform the results in the structural model.

The first fact is that ISPs with more households in their network take longer to bargain than smaller ISPs. This fact is surprising because network size is not necessarily an indication of market power: an ISP with many households in its network may face many competitors at the households its serves, while a smaller ISP may be a monopolist to the houses it serves. Why larger ISPs find it worthwhile to take longer and extract more surplus from Netflix at the cost of subscriber revenues is a key question for the model.

The second fact is that marginal consumers substitute away from ISPs that take longer to bargain. Robustly demonstrating this relationship is key to the identification of the surplus distribution, which relies on the intuition that bargaining ISPs expect to extract at least as much surplus from a screening strategy as subscribers they expect to lose. Moreover, that marginal subscribers substitute implies that delaying agreement over how to split the surplus

from the capacity investment has welfare costs for consumers.

1.4.1 Agreement timing patterns

To document how disagreement timings vary across ISPs, I first run simple cross-sectional regressions of disagreement durations on features of an ISP's network:

$$duration_f = \beta X_f + \epsilon_f.$$

The dependent variable is how long bargaining lasts in quarters. X_f contains a measure of ISP size, as well as other covariates including what technology the ISP uses to serve consumers (cable, DSL or fiber), the degree of geographic connectedness of an ISP's network, and measures of the intensity of competition the ISP faces.[10]

Results from several specifications are presented in Table 1.1. The two measures of size, footprint—which measures how many housing units are in the ISP's network—and its national market share, are included separately. A one standard deviation increase in size is associated with between a 0.5 and 0.8 standard deviation increase in disagreement duration; other covariates are insignificant.

Note that size is not equivalent to market power since regardless of geographic scale, most ISPs are duopolists to the households they serve. However, if Netflix has relatively more to gain from making its capacity investment in a larger ISP compared to a smaller ISP, then larger ISPs have a relatively stronger bargaining position and can extract a greater share of Netflix's surplus.

If ISPs tend to agree quickly after their competitors it may be an indication that market power matters for bargaining. I estimate the following specification using the panel of ISPs

[10] An ISP whose network is geographically contiguous will have a high degree of connectedness. Taking each census block as a vertex and whether that block is physically adjacent to another block as an edge, for each ISP I sum the total number of edges and divide it by the product of the number of vertices for that ISP and the average number of edges per vertex in the network given by all census blocks in the continental United States.

that were present at the beginning of bargaining in mid-2013:

$$agree_{ft} = \sum_{k=0}^{1} \gamma_k comp_agree_{f,t-k} + \beta X_{ft} + \bar{\alpha}_f + \bar{\alpha}_t + \epsilon_{ft},$$

where $agree_{ft}$ takes a value of 1 if ISP f concludes bargaining at time t. $comp_agree_{ft}$ takes a value of 1 if the ISP with the greatest share of overlap in f's markets concludes negotiations with Netflix at time t. An ISP exits the panel the period after it concludes negotiations.

Results presented in Table 1.2 indicate a positive relationship between when an ISP agrees and when its primary competitor agrees; that is, agreement timings are correlated. However, this relationship is insignificant after time fixed effects are included. The data cannot distinguish between whether ISP interaction or correlated shocks are driving the agreement timing patterns. In my structural model I will allow both mechanisms to operate in a restricted way.

1.4.2 Consumers switch ISPs

In this section I show that there exist marginal consumers who leave ISPs that do not reach an agreement with Netflix quickly, implying that degraded Netflix quality is costly for consumers. I further show that ISPs that take time to negotiate with Netflix experience reductions in the rate of subscriber growth during negotiations, implying that lengthy negotiations are also costly for ISPs.

Using the individual level panel of individual ISP and speed choices from the MBBA, I estimate the following equation:

$$P_{it}(f'|f) = \gamma disagree_{ft} + \beta \Delta \log p_{ft} + \bar{\alpha}_i + \bar{\alpha}_t,$$

where $P_{it}(f'|f)$ is the probability that consumer i leaves ISP f for any other f' at time t, including the outside option, $disagree_{ft}$ is a dummy indicating whether ISP f is still negotiating, and p_{ft} is the price of f's entry level plan. Results are presented in Table 1.3 and indicate a robust correlation between the probability of exiting an ISP and whether that ISP is currently experiencing degraded Netflix quality.

Suggestive evidence that Netflix users drive the increased switching during the periods of degraded streaming quality comes from cross-sectional data from the mid-2013 CPS Internet Use Supplement. 18.8% of households in the CPS report having switched ISPs in the past 3 years, and households with a TV-based internet streaming device were 36% more likely to switch than households without such a device.[11] Moreover, consumers who switched and own a TV-based streaming device were 19% more likely to report that reliability was their reason for switching than the average switcher.

Longer periods of disagreement will only be costly for ISPs if Netflix quality reductions generate a net outflow of subscribers or reduce the rate of subscriber growth. I use movements in aggregate subscriber shares to directly evaluate whether the quality reduction cost ISPs subscribers. Figure 1.4 presents median residual ISP growth rates for two groups of ISPs: those that experienced a delay in negotiation with Netflix, and those that did not. The first vertical line indicates the beginning of bargaining, the dashed vertical line indicates the conclusion of the first wave of bargaining, and the final line indicates when the last ISPs agreed. The median ISP that did not experience disagreement and a quality degradation grows 0.5% faster during the slowdown than ISPs that did. To construct these series I purge ISP specific time trends, time dummies, and control for mergers and price movements.

To formalize the relationship in the graph, I run the following regression:

$$\Delta \log subscribers_{ft} = \gamma disagree_{ft} + \beta \Delta \log p_{ft} + \bar{\alpha}_f + \bar{\alpha}_t + \epsilon_{ft},$$

where the dependent variable is the log change in ISP f's reported number of internet subscribers, $disagree_{ft}$ is a dummy variable indicating whether an ISP is currently in disagreement with Netflix, and p_{ft} is the price of the ISP's entry tier plan.

Results are reported in Table 1.4. The estimated reduction in subscriber growth for ISPs that delay the resolution of negotiations is between 0.4 and 0.8 percent during periods of delay. The negative relationship is robust across specifications that include ISP and time fixed effects and that control for price changes and mergers. Price also has the expected negative

[11] This may be due to a difference in switching costs: streaming video users were 4.7% less likely to bundle internet with cable. It may also be an extreme manifestation of a prevalent preference for reliability: 36.7% of respondents list reliability as most important feature of a plan, greater than price (30.5%) or download speed (25.3%)

coefficient, although it is not significant, potentially due to endogeneity with unobserved demand shocks.

The results suggest an economically significant role for subscriber substitution between ISPs in response to Netflix quality degradation. That is, I find that marginal consumers do substitute away from ISPs that are in a state of disagreement with Netflix. This suggests that not only do these consumers value Netflix streaming quality, they value it enough (and are informed enough) to leave ISPs (or at least, not sign up with ISPs) whose streaming quality is degraded.

1.4.3 Netflix Demand is Inelastic

I argue that Netflix demand during 2013-2014 is inelastic. Since I lack microdata on switching behaviour of individual Netflix subscribers and Netflix does not provide a measure of churn, I analyze the residual growth rates in Netflix aggregates from Q-10 and K-10 filings.

The leftmost image in Figure 1.5 gives the residual growth rate in Netflix streaming subscribers after controlling for a time trend and seasonality. The series begins in 2012, since prior to this date Netflix aggregated its DVD rental and streaming customers. Even with controls for quarter, the growth rate exhibits periodicity; however, the pattern of subscription growth around 2014 looks very similar to that in 2013.

One possible explanation for why subscriber growth does not drop is that Netflix made more free trial memberships available during the slowdown. That is, while Netflix does not change the sticker price of its service during this time, they may be reacting to the (endogenous) negative demand shock by offering more free trials. In the rightmost image of Figure 1.5 gives the residual growth rate in the fraction of streaming customers that pay for service. In 2014 the growth rate in paid subscribers is lowest during the bargaining event, unlike in 2013 and 2015 when the low point happens midway through the year. The series is very noisy, and only weakly suggests that relatively more of the growth in Netflix's subscriber base were free trial offers during the slowdown.

The evidence for substitution away from Netflix is weak. This agrees with the results in the previous section in that it points to a high consumer valuation for Netflix. Speculatively, the inelastic demand may come from the lack of similarly priced alternatives for on-demand

TV and movies,[12] as well as strong consumer sentiment that ISPs such as Comcast are solely responsible for network slowdowns.[13] In my base model specification, I will assume that demand for Netflix is completely inelastic, but that slowdowns affect consumer valuations for ISPs. I relax this condition in the online appendix.

1.5 Model

There are three types of agents in the model: consumers, internet service providers, and Netflix. I assume that time is discrete, with each time period t representing three months (one quarter).

Consumers demand internet access and purchase it from the ISPs, according to their heterogeneous valuations of plan characteristics. This feature of the model predicts ISP plan shares in each period as a function of parameters, whether Netflix quality has been restored, and data on household and plan characteristics. I do not model households' choice of whether to purchase Netflix or not, but allow disutility for reductions in ISP-specific Netflix throughput.

ISPs earn profits from selling subscriptions, and may lose subscriber profits from a quality reduction in Netflix. Using estimates of the demand curve and assumptions on ISP competition and price setting, the model will predict profit as a function of the complete vector of which ISPs have restored Netflix quality of service.

On the bargaining side, Netflix and ISPs bargain multilaterally. The model predicts agreement time and fee offer probabilities as a function of supply parameters and the estimated profit elasticities. That is, the model endogenizes agreement timings, and allows for strategic interaction between ISPs in their offers as they trade off higher fee offers against potentially losing subscribers to a rival.

I begin by describing the dynamic bargaining framework, taking the ISP subscriber profit function as given. I then describe the model of price setting and consumer demand that will

[12]By the beginning of 2013, DISH—the purchasers of Blockbuster—had shut down 1100 of 1500 stores, and shuttered 1450 of 1500 by 2015. A monthly Netflix subscription granting unlimited streaming was $7.99 per month in 2013, while pay-per-view movies were anywhere from $2.99 to $5.99 for a one week rental. F

[13]Consumer Reports

serve as an input to the bargaining model.

1.5.1 Upstream Bargaining

Upstream bargaining is a dynamic game played between all downstream internet service providers indexed by $f = 1, \ldots, F$, and the upstream content provider Netflix, indexed by N.

Time is discrete and runs from t_0 to a terminal period T.[14] Bargaining is exogenously and simultaneously initiated with all ISPs by Netflix at time t_0. At t_0, N draws conditionally independent, ISP-specific types μ_f from a distribution $F(\mu_f|w_f, \theta_s)$ where w_f are observables.[15]

Netflix's vector of draws $(\mu_1, \ldots, \mu_F)$ is its private information. In this setting, these draws correspond to Netflix's ISP-specific marginal increase in surplus from installing their CDN servers in that ISP's network. w_f includes functions of observable ISP network characteristics such as their total footprint and technology, which are plausibly informative about Netflix's benefit. The eventual goal will be to estimate the parameters θ_s and use them as primitives in counterfactuals.

Actions and Timing

Starting from t_0, within each period t:

1. Any ISP whose prior offers have not been accepted observes the history of past agreement timings, its own private information about the vector (μ_f), and information on ISP demand shifters, and proposes a lump sum interconnection fee $\tau_{ft} \in \mathbb{R}$

2. Netflix accepts or rejects each f's offer, $\mathbf{a}_t = (a_{1t}, \ldots, a_{Ft})$ where $a_{ft} \in \{0, 1\}$. If N accepts f's offer, N pays f τ_{ft} and realizes surplus μ_f.

3. ISPs observe the vector of acceptance/rejections $\mathbf{a}_t$ and compete for consumers. ISPs receive flow payoffs and ISPs whose offer is accepted exit the bargaining game.

[14] I postpone a discussion for why I assume a terminal period to the estimation and identification section.

[15] Note that the vector μ_f does not depend on the number of subscribers at the ISP. Therefore, if consumers substitute between ISPs in response to slowdowns, it does not change the size of the ISP-specific investment surplus to be split.

At the beginning of each period, all remaining ISPs make offers simultaneously. That is, ISPs experience both incomplete information vis-à-vis Netflix's marginal valuation, and imperfect information regarding each others' actions. Moreover, I assume that in addition to the draws (μ_f) which neither the econometrician nor f observes, there is a vector of demand shocks (ξ_{ft}) which the firms observe but which the econometrician and Netflix do not.

Netflix's problem

After exogeneously initiating bargaining at t_0, Netflix chooses its strategy of acceptances and rejections in each period to maximize its profits. Netflix period profits are

$$\pi_{Nt}(\mathbf{a}_t, \boldsymbol{\tau}_t) = \tilde{\pi}_{Nt} + \sum_f (\mu_f - \tau_{ft}) a_{ft}.$$

$\tilde{\pi}_{Nt}$ is profit from subscriptions. In my base model I assume Netflix demand is perfectly inelastic during this time period, so $\tilde{\pi}_{Nt}$ does not depend on the vector of disagreements.

If Netflix is dynamic, their problem is to choose a vector of acceptances in each period as a function of ISP offers, the observable history of acceptances and rejections, and whatever they can infer about ISPs' information sets given histories and ISP strategies. For a history $h_t = \{\mathbf{a}_{t_0}, \ldots, \mathbf{a}_{t-1}\}$, Netflix's continuation value is:

$$V(h_t, \boldsymbol{\tau}_t) = \max_{\mathbf{a}_t, \ldots, \mathbf{a}_T} E_{\boldsymbol{\xi}_t, \ldots} \left[\sum_{t' \geq t} \beta^{t'-t} \pi_{Nt}(\mathbf{a}_t) \Big| h_t, \boldsymbol{\tau}_t \right]$$

Netflix's problem is difficult for informational and computational reasons. To know whether they should accept f's offer in a given period, Netflix must be able to forecast the next period's offer, which requires understanding how a rejection will affect the evolution of f's beliefs about the distribution of μ_f. This issue is made complicated by the demand shock $\boldsymbol{\xi}$, since the unobserved shock will affect f's learning in each period in a way that is difficult for Netflix to infer. Computationally, Netflix must optimally choose a vector of actions $\boldsymbol{a}$ from an action space of size 2^{23} each period (although this space shrinks as firms exit.)

To manage this complexity, in my base model I assume that Netflix is myopic with $\beta = 0$.

Combined with the assumption on inelastic demand, this implies that N's optimal strategy is separable in offers:

$$a_{ft}(\boldsymbol{b}, \boldsymbol{\tau}) = a_{ft}(\tau_f) = \begin{cases} 1 & \text{if } \mu_f \geq \tau_f \\ 0 & \text{else.} \end{cases}$$

Before receiving offers, a myopic N would like to adopt a more sophisticated strategy, but after receiving offers they can do no better than the above.

Assuming Netflix is myopic and that demand is inelastic simplifies Netflix's strategy greatly. It also makes it trivial to invert their strategy as a function of their unobserved type, which is important for the ISPs—who will be dynamic—to be able to learn from rejected offers. Because I assume Netflix and the ISPs are bargaining over a lump sum transfer paid in the same period, there is no sense in which I am forcing Netflix to irrationally accept high offers by making them ignore a future stream of payments to ISPs. I show how to allow for elastic Netflix demand in robustness exercises in an online appendix.

ISPs' problem

At the end of time t, after f has observed the vector of demand shocks $\boldsymbol{\xi}$, made offers, and Netflix has accepted or rejected each offer, firm f flow profits are written as follows:

$$\pi_{ft}(\mathbf{a}_t, \tau_{ft}) = \tilde{\pi}_{ft}(\mathbf{a}_t) + a_{ft}\tau_{ft}$$

where $\tilde{\pi}_{ft}(\mathbf{a}_t)$ is the profit earned from f subscribers at time t, which depends on $\mathbf{a}_t$ as marginal consumers may substitute between ISPs in response to slowdowns. $\tilde{\pi}_{ft}$ will be recovered from optimal prices and the demand curve, estimated in the next section. The second component of profits $a_{ft}\tau_{ft}$ is equal to the lump sum transfer f stands to receive if its offer is accepted.

Each ISP f seeks to maximize expected profits by choosing a best response sequence of offers. Expectations are taken over the probability of each realization of the agreement vector in each period induced by Netflix and $-f$'s strategies. f is best responding both to

other ISPs' offers, as well as Netflix's optimal strategy of accepting/rejecting offers:

$$V(h_t, \boldsymbol{\xi}) = \max_{\tau_{ft},\ldots,\tau_{fT}} E_{\mathbf{a}_t,\ldots}\left[\sum_{t'\geq t} \beta^{t'-t}\pi_{ft}(\mathbf{a}_t, \tau_{ft})\middle| h_t, \boldsymbol{\xi}\right]$$

where $\boldsymbol{\xi}$ is the complete vector of demand shock realizations for all firms.

To formulate the ISPs' problem recursively, I show that given Netflix's optimal strategies and a set of ISP strategies, each ISP's information set will be a vector of (Bayesian) beliefs about the upper bound on the distributions every ISP faces. That is, f's information will be a vector of real numbers $(b_1, \ldots, b_F)$ such that f knows $\mu_{f'} < b_{f'}$ for every f'.

First, given a set of ISP strategies, if any ISPs remain in the bargaining game at time t it must mean that their lowest offer in prior periods was rejected; given Netflix's strategy vis-à-vis f to accept any offer less than or equal to μ_f, μ_f must therefore be lower than that lowest offer. Second, notice that given ISP strategies and histories, it is possible for each ISP to construct the full information set since ISPs have no unobserved heterogeneity with respect to each other. I restrict ISPs to strategies such that if two histories of acceptance and rejection timings lead to the same information set, the optimal strategies going forward from that information set are identical. This restriction implies equivalence between histories and the information state, so that ISP optimal strategies can be solved as functions of the information state alone.

I formulate ISP f's problem recursively as a function of the complete information state:[16]

$$V_{ft}(b_f, \mathbf{b}_{-f}) = \max_{\tau} \mathbb{E}_{\mathbf{a},\boldsymbol{\tau}}\left[\pi_{ft}(\mathbf{a}, \tau) + \beta V_{ft+1}(\tau, \boldsymbol{\tau}_{-f})\middle| \mu_f < b_f, \boldsymbol{\mu}_{-f} < \mathbf{b}_{-f}\right]$$

I assume all demand shocks ξ_{ft} are perfectly forecasted, and therefore do not include it as a

[16]Either the full information state, or the complete history of agreement times is necessary. The current vector of commonly observed agreement plus a firm's own information about μ_f is not a sufficient state when there are more than two ISPs. For instance, with three firms, suppose the state is such that one firm has agreed but the other two have not. Without knowing how long ago the first firm agreed, the second and third firms cannot infer each others' current strategies, which are necessary to determine the transition probabilities to next period's agreement vector. At any given time, if the first firm agreed very early, then the rate of screening by the other two firms would have been different than if the first firm agreed only recently. Each firm must be able to accurately infer what the other knows about its distribution to predict the probability of agreement in this period.

state variable since value functions are indexed by ft.[17] Moreover, I index the value functions by f and t, so that each named ISP will have its own value function to solve. Since ISPs are highly heterogeneous in their cross sectional plan offerings, market coverage, and demand shocks, as well as in the evolution of those characteristics over time, I index value functions to avoid introducing a high dimensional state space of characteristics.

The goal in the estimation section will be to solve each ISP's optimal policies to recover probabilities of agreement timings, then use MLE to estimate the parameters of the surplus distribution, θ_s.

Alternate Sources of Asymmetric Information

To induce delay in agreement, one can introduce (1) asymmetric information as in Fudenberg and Tirole (1983), (2) irrational optimism as in Yildiz (2004), or (3) stochastic payoffs as in Merlo and Wilson (1995). I rule out optimism to keep agents rational in this first attempt at using dynamic bargaining in an industry setting. I rule out the stochastic payoff model as it requires that payoffs might increase if there is delay, which is difficult to square with this environment where quality degradation persists, and possibly grows worse, as long as there is disagreement.

An alternative to an unknown additive supply-side surplus might be unknown consumer elasticities. If ISPs and/or Netflix do not know how the marginal subscriber will substitute in response to a slowdown and they are bargaining over subscriber surplus, a delay might arise. However, if the purpose of the bargaining delay was for market participants to learn about their own subscriber elasticities, it begs the question of why ISPs and Netflix chose to learn through a costly, multi-month long quality degradation as opposed to smaller, cheaper experiments.

[17] In literature using mixed logit demand systems, firms are either presumed to know each others' demand shocks as in Berry, Levinsohn and Pakes (1995), or demand shocks are assumed away as in Bajari, Benkard and Levin (2007).

1.5.2 ISP Plan Pricing and Profits

In this section I discuss how I recover $\tilde{\pi}_{ft}(\mathbf{a}_t)$ as a function of the demand curve. I assume that ISPs take the menus of their offerings and the period's vector of agreements $\mathbf{a}_t$ as given.

If the price elasticities from the demand curve are not greater than one, then I assume a marginal cost of zero and calculate profits as revenues from the demand curve. If the price elasticities are greater than one, then f chooses a vector of prices for its plans to maximize profits, conditional on the best response pricing of other firms. That is, f solves

$$\tilde{\pi}_{ft}(\mathbf{a}_t) = \max_{\{p_{jft}\}} M_t \sum_{j|j\in\mathcal{J}_{ft}} s_{jft}(\mathbf{p}_t(\mathbf{a}_t), \mathbf{a}_t)(p_{jft}(\mathbf{a}_t) - mc_{jft}),$$

where j is an ISP plan (price-download speed combination), M_t is the size of the market at time t, and $\mathcal{J}_{ft}$ is the set of plans offered by f at t. s_{jft} is the nationwide demand for plan j offered by f at time t.

Taking first order conditions and inverting the system of equations to solve for marginal costs yields the standard equation:

$$\mathbf{mc}_t = \mathbf{p}_t + \Delta_t^{-1} s_t,$$

I assume a plan has the same price nationwide. This restriction is in keeping with the data, where plan prices do not vary across geographic markets. Moreover, since I lack share data at the market level, it is impossible to recover market specific unobservable shocks which would be necessary for taking the derivative of the market specific demand curve.

1.5.3 Demand

This section of the model predicts demand curves as a function of parameters, Netflix quality degradation, and other data to feed into $\tilde{\pi}_{ft}$.

Flow utility

The ISPs available to a household vary broadly by geographic market m and by time t. A consumer in market m will have an individual specific choice set, reflecting the fact that not all ISPs in that market m may have infrastructure connected to that individual's dwelling. Markets are relevant to the consumer as the same ISP operating in two markets may offer different plans.

An individual i in market m at time t chooses among internet service providers f that belong to that individual's choice set, $\mathcal{F}_{mt}$. Each available firm f offers a menu of vertically differentiated plans $j \in \mathcal{J}_{fmt}$ which vary by market. Conditional on choosing a firm, a consumer chooses among the available j offered by f in m at time t. The indirect utility to i from choosing firm $f \in \mathcal{F}_{imt}$ is[18]

$$u_{ifmt} = \delta_{ft} + \lambda_{ifmt} + \epsilon_{ifmt},$$

where δ_{ft} is the mean utility each consumer derives from consuming f at time t. Mean utility depends on ISP fixed effects, time, whether the ISP has reached agreement with Netflix or not, and an unobserved firm specific shock ξ_{ft}:

$$\delta_{ft} = \bar{\gamma}(1 - a_{ft}) + \bar{\alpha}_f ISP_f + \bar{\alpha}_{ft} ISP_f \times t + \xi_{ft}, \tag{1.1}$$

where a_{ft} is a dummy indicating whether Netflix and ISP f have reached agreement and restored quality of service. Based on the reduced form analysis, I expect that $\bar{\gamma} < 0$. Each ISP f has an unobserved dimension of heterogeneity, ξ_{ft}. As in similar papers on demand for telecommunications services, this time-varying heterogeneity may reflect quality of customer service and the quality of bundled services—for example, whether a cable internet provider adds or drops channels from its TV service.

Firms are not uniquely identified with a price or download speed, so these variables do not enter δ_{ft}. Individuals choose among f's menu of offered plans, and the effect on indirect

[18] I assume that switching costs between vertically differentiated products within a provider, as well as switching costs between providers or to the outside option, are all zero. Prohibiting switching costs between ISPs is done to reduce the computational burden in the supply model, since otherwise ISPs would need to keep track of consumer states.

utility is captured by λ_{ifmt}:

$$\lambda_{ifmt} = \max_{j \in \mathcal{J}_{fmt}} \{\alpha_{ip} p_{jft} + \alpha_{iq} q_{jft}\} + \alpha_{if} + \gamma_i (1 - a_{ft}).$$

p_{jfmt} and q_{jfmt} are the price and download speed associated with each plan j in the set of plans offered by f in m at time t, $\mathcal{J}_{fmt}$. $(\alpha_{ip}, \alpha_{iq}, \alpha_{if}, \gamma_i)$ are individual specific coefficients on price, download speed, firm technology, and whether or not there is a slowdown, respectively.

I model consumer heterogeneity $(\alpha_{ip}, \alpha_{iq}, \alpha_{if}, \gamma_i)$ as being comprised of observed and unobserved characteristics. In particular,

$$\begin{aligned} \alpha_{ip} &= -\exp\left(-\left(\boldsymbol{x}_i' \boldsymbol{\alpha}_p^o + \sigma_p \nu_{ip}\right)\right) \\ \alpha_{iq} &= \boldsymbol{x}_i' \boldsymbol{\alpha}_q^o + \sigma_q \nu_{iq} \\ \alpha_{if} &= \boldsymbol{x}_i' \boldsymbol{\alpha}_g^o + \sigma_g \nu_{ig} \\ \gamma_i &= \boldsymbol{x}_i' \boldsymbol{\gamma}^o + \sigma_s \nu_{is} \end{aligned} \tag{1.2}$$

where $\boldsymbol{x}_i$ denotes a vector of consumer characteristics including functions of income, household size, etc. and $\{\nu_{ip}, \nu_{iq}, \nu_{ig}, \nu_{is}\}$ denote unobserved (by the econometrician) consumer tastes. Define the complete vector of characteristics for a household as $\omega_i \equiv (\boldsymbol{x}_i, \boldsymbol{\nu}_i)$, and the complete vector of demand heterogeneity parameters $\theta_d^h \equiv (\boldsymbol{\alpha}, \boldsymbol{\sigma})$. The full vector of demand parameters θ_d also includes the mean effect of Netflix quality degradation and ISP fixed effects, $\theta_d \equiv (\bar{\gamma}, \bar{\alpha}, \theta_d^h)$.

I assume that the unobserved tastes are distributed independent standard normal across and within consumers. As usual, ϵ_{ift} is assumed to be distributed according to the type I extreme value distribution. I borrow the functional form for the heterogeneous price coefficient from Berry, Levinsohn and Pakes (2004).

Consumers have an outside option regardless of their choice set, which reflects either non-purchase or purchase from a dialup internet provider. Indirect utility from the outside option is

$$u_{i0mt} = \boldsymbol{x}_i' \boldsymbol{\alpha}_o^o + \epsilon_{i0mt},$$

where ξ_{0t} is normalized to zero to fix relative utility levels, which are otherwise unidentified.

Market shares

As usual, the conditional choice probabilities for a given consumer i in market m at time t have an analytic form due to the assumption of the type I extreme value error. Given a choice set $\mathcal{F}_{mt}$, i's probability of choosing f is:

$$\Pr(f_{imt} = f|\mathcal{F}_{mt}, \omega_i) = \frac{\exp(\delta_{ft} + \lambda_{ifmt})}{\sum_{f' \in \mathcal{F}_{mt}} \exp(\delta_{f't} + \lambda_{if'mt})},$$

where δ_{ft} incorporates the mean utility from choosing f in market m at time t, and λ_{ifmt} incorporates individual and market specific deviations from the mean. I assume the outside option is contained in every choice set.

Within a market m, households have a vector of probabilities of being assigned to each choice set in the market. For intuition, a market will be a geographic area defined by the census (a Public Use Microdata Area) while choice sets are observed combinations of ISPs operating in the market. The assignment probabilities potentially depend on households' observable characteristics, the moments of the distribution of observable characteristics for individuals in each choice set, and the fraction of households each choice set covers within the market. Define i's probability of being assigned to $\mathcal{F}_{mt}$ as $\phi_{mt}(\mathcal{F}_{mt}|\boldsymbol{x}_i)$, where indexing ϕ by mt incorporates the potential for the probability to depend on the distribution of observed characteristics and household weights across choice sets within a market.

To construct market shares at the national level, I aggregate across choice sets for a given i, then across i within a market, and finally across markets using potentially time-varying market weights w_{mt}:

$$s_{ft} = \sum_m w_{mt} \int \left(\sum_{\mathcal{F}_{mt}} \Pr(f_{imt} = f|\mathcal{F}_{mt}, \omega_i) \phi_{mt}(\mathcal{F}_{mt}|\boldsymbol{x}_i) \right) dG(\omega_i|\theta_d). \tag{1.3}$$

The equation Equation (1.3) is the model's prediction of the distribution of ISP market shares in a given t given the distribution of consumer characteristics, plan characteristics, choice sets and a vector of parameters θ_d. Matching these predicted shares to the observed

shares from quarterly filings will form the basis of my estimation strategy as in Berry, Levinsohn and Pakes (2004).

The model also generates the distribution of consumer plan purchases within each firm for given parameter values. In this model, consumer heterogeneity in the valuation of prices and download speeds $(\alpha_{ip}, \alpha_{iq})$ is the only reason why within-firm plan shares are non-degenerate. Matching conditional purchase moments from the MBBA with the model's predictions will be informative about these consumer heterogeneity parameters. Because the heterogeneity parameters govern the disutility of price, these moments are essential to identify the price elasticity of demand. The share of individuals purchasing plan j from firm f at time t is given by:

$$s_{j|ft} = \sum_{m|f\in m} \tilde{w}_{mt} \times \frac{\int \left(\sum_{\mathcal{F}_{mt}} \mathbf{1}\left[j = \underset{j'\in\mathcal{J}_{fmt}}{\operatorname{argmax}} \left\{ \alpha_{ip} p_{jft} + \alpha_{iq} q_{jft} \right\} \right] \Pr(f|\mathcal{F}_{mt}, \omega_i) \phi_{mt}(\mathcal{F}_{mt}|\boldsymbol{x}_i) dG(\omega_i) \right)}{\int \left(\sum_{\mathcal{F}_{mt}} \Pr(f|\mathcal{F}_{mt}, \omega_i) \phi_{mt}(\mathcal{F}_{mt}|\boldsymbol{x}_i) dG(\omega_i) \right)}, \quad (1.4)$$

where $\tilde{w}_{mt} = w_{mt} / \left(\sum_{m'|f\in m'} w_{m't} \right)$.

1.6 Estimation and Identification

1.6.1 Demand

I use the Berry, Levinsohn and Pakes (1995) approach to estimate demand parameters, with the addition of micro moments as in Berry, Levinsohn and Pakes (2004) to help identify heterogeneity in parameters across the population and provision for choice sets that vary across individuals as in Goeree (2008). I first recover estimates of the unobserved firm (and time) specific utility δ_{ft} as a function of demand heterogeneity parameters θ_d^h. Next, for choice of instruments Z_{fmt} I form a GMM objective function using micro moments only to identify the heterogeneity parameters. At the θ_d^h that minimizes the GMM criterion, I recover $\delta_{ft}(\theta_d^h)$, and use it to estimate the mean effect of Netflix quality degradation and ISP fixed effects under various assumptions on the joint distribution of the exogenous component

of $\delta_{ft}(\theta_d^h)$ and $\xi_{ft}(\theta_d^h)$.

From Equation (1.2), the demand parameters θ_d to estimate are:

$$\begin{aligned} \text{Mean shifters: } & \{\bar{\gamma}, \bar{\boldsymbol{\alpha}}\} \\ \text{Observed heterogeneity: } & \left\{\boldsymbol{\alpha}_p^o, \boldsymbol{\alpha}_q^o, \boldsymbol{\alpha}_g^o, \boldsymbol{\gamma}^o\right\} \\ \text{Unobserved heterogeneity: } & \{\sigma_p, \sigma_q, \sigma_g, \sigma_s\}. \end{aligned}$$

To estimate these parameters, I match three sets of predicted moments to their sample analogues: (1) the covariance between unobserved firm-level heterogeneity and a set of instruments that shift firm markups; (2) the covariances of the observed technology type with observed consumer characteristics; and (3) the covariance of the set of instruments with the difference between the predicted conditional plan shares and observed conditional plan shares.

The first set of moments are useful for identifying the mean consumer valuation for each ISP and for the mean response to the Netflix quality degradation. Given parameter guesses, ξ_{ft} will allow the model to exactly fit the observed market shares by shifting the mean utility for each firm at each date. Depending on the joint distribution of the ISP dummies, a_{ft} and ξ_{ft}, I can recover the coefficients $(\bar{\gamma}, \bar{\alpha})$ by OLS, FE or IV. In my base specification, interacting these ξ_{ft} with a set of contemporaneous and lagged instruments will help identify the mean shifters even in the presence of latent switching costs (see Scherbakov (2015)).

The second set of moments match observed consumer attributes from the 2013 and 2014 waves of the ACS to each consumer's chosen technology within a geographic market. These moments will be particularly useful in identifying the observed heterogeneity parameters $\boldsymbol{\alpha}$. If, for instance, large households choose the cable option relatively more in a market where cable download speeds are higher compared to large households in a region with poor cable options, the estimation will attribute a positive coefficient to the interaction of download speed with household size. There is substantial variation in the set of plans and technologies available across markets, mostly due to plausibly exogenous geographic variation in ISP service areas. Cross sectional and time series variation in the sets of ISPs that are still negotiating with Netflix in 2013 and 2014 identifies heterogeneous consumer valuations to

the slowdown.

The final set of moments will help identify unobserved heterogeneity in the valuation of price and download speed, (σ_p, σ_q). I match the model's prediction for the vector of nationwide conditional plan shares for each ISP observed in the MBBA data with the observed conditional shares.[19] Variation in the menus of available plans over time and across ISPs will provide the identifying power for these parameters.

Recovering $\boldsymbol{\delta}$

I recover the δ_{ft} that rationalize the observed nationwide ISP shares according to the standard BLP algorithm. That is, given a guess for θ_d and an initial guess of the vector of $\boldsymbol{\delta}^0 \equiv (\delta^0_{ft})$, I iterate the following equation until it converges for each t:

$$\boldsymbol{\delta}_t^{(a+1)}(\theta_d^h) = \log(\mathbf{s}_t) - \log(\hat{\mathbf{s}}_t(\theta_d^h, \boldsymbol{\delta}^{(a)}(\theta_d^h))) + \boldsymbol{\delta}_t^{(a)}(\theta_d^h), \tag{1.5}$$

where $\mathbf{s}_t$ is the vector of data on national market shares at time t and $\hat{\mathbf{s}}_t(\theta_d^h, \boldsymbol{\delta})$ is the model's prediction of shares. Denote the recovered parameters by $\hat{\delta}_{ft}$, where dependence on θ_d^h is implicit.

To predict $\hat{s}_t(\theta, \boldsymbol{\delta})$ requires integrating individual conditional choice probabilities across consumers and choice sets within markets. I integrate by simulation. In each m, I sample $r = 1, \ldots, R$ households (with replacement) from the empirical distribution of consumer (and characteristics $\boldsymbol{x}$) in that market, using census provided weights as sampling weights. I draw unobserved heterogeneity $\boldsymbol{\nu}$ from a multivariate standard normal.

Individuals must still be assigned choice sets to construct the integral. In my baseline case, I take $\phi_{mt}(\mathcal{F}_{mt})$ to be the empirical share of households in PUMA m who are in choice set $\mathcal{F}_{mt}$ from the National Broadband Map. For each sampled household r, I assign them a single choice set, with probabilities of each choice set dictated by $\phi_{mt}(\mathcal{F}_{mt})$. That is, if PUMA m has 40% of households in a Comcast-Verizon choice set, I will assign sampled households r to that choice set with 40% probability regardless of household characteristics. Denote r in market m at time t's choice set as $\hat{\mathcal{F}}_{rmt}$.

[19] I drop the highest plan for each provider to have an excluded category, since otherwise the moment will be flat in the parameter value.

The simulated market shares for firm f at time t:

$$\hat{s}_{ft}(\theta_d^h, \boldsymbol{\delta}) = \sum_m w_{mt} \frac{1}{R} \sum_r \Pr(f|\hat{\mathcal{F}}_{rmt}, \omega_r, \theta_d^h, \boldsymbol{\delta}_t)$$

The simulated conditional share for plan j offered by f at time t:

$$\hat{s}_{j|ft}(\theta_d^h, \boldsymbol{\delta}) = \sum_{m|f\in m} \tilde{w}_{mt} \frac{\sum_r \mathbf{1}\left[j = \underset{j'\in\mathcal{J}_{fmt}}{\operatorname{argmax}} \{\alpha_{ip} p_{jft} + \alpha_{iq} q_{jft}\}\right] \Pr(f|\hat{\mathcal{F}}_{rmt}, \omega_r, \theta_d^h, \boldsymbol{\delta}_t)}{\sum_r \Pr(f|\hat{\mathcal{F}}_{rmt}, \omega_r, \theta_d^h, \boldsymbol{\delta}_t)},$$

where $\tilde{w}_{mt} = w_{mt} / \left(\sum_{m'|f\in m'} w_{m't}\right)$.

Estimating the Heterogeneity Parameters

Having recovered $\boldsymbol{\delta}$, I estimate the heterogeneity parameters using the second and third sets of moments. The second moment requires that within a market, interacting instruments with the difference in the mean characteristics of households predicted by the model to choose technology g and the mean observed characteristics of consumers who actually choose g is minimized:

$$G_R^2(\theta_d^h, \boldsymbol{\delta}) = \frac{1}{NM_t} \sum_m Z_{gmt} \left[\frac{1}{n_{gmt}} \sum_{i_{gmt}=1}^{n_{gmt}} \boldsymbol{x}_{i_{gmt}} - E[\boldsymbol{x}|g(f_{imt}) = g, \theta_d^h, \boldsymbol{\delta}] \right], \tag{1.6}$$

where n_{gm} is the number of ACS respondents reporting they use technology g in market m, and NM_t is the number of markets at time t. $g(f_{imt})$ is a function that returns the technology of firm f in market m at time t.[20]

The conditional mean term in the square brackets is the model's prediction of the mean characteristics of an individual choosing technology g. I estimate the conditional mean using the simulated individuals.

$$E[\boldsymbol{x}|g(f_{imt}) = g, \theta_d^h, \boldsymbol{\delta}] \approx \frac{\sum_r \boldsymbol{x}_r \Pr(g(f_{rmt}) = g|\boldsymbol{x}_r, \nu_r, \theta_d^h, \boldsymbol{\delta})}{\sum_r \Pr(g(f_{rmt}) = g|\theta_d^h, \boldsymbol{\delta})}$$

[20] A firm's technology may vary, e.g. Verizon serves markets using both DSL and fiber optic cable.

A sample of the conditional variation in characteristics across technologies in the data is presented in Table 1.10.

The third set of moments use data from the MBBA testing program. These moments are similar in spirit to the second set of moments, interacting the difference between actual and predicted conditional shares with an instrument. The moments are:

$$G_R^3(\theta_d^h, \boldsymbol{\delta}) = \frac{1}{\tilde{F}_t} \sum_f \frac{1}{|\mathcal{J}_{ft}|} \sum_{j \in \tilde{\mathcal{J}}_{ft}} Z_{ft} \left[\frac{n_{j|ft}}{n_{ft}} - \hat{s}_{j|ft}(\theta_d^h, \boldsymbol{\delta}) \right], \tag{1.7}$$

where $\tilde{F}_t$ is the number of firms present in the MBBA data at time t (with $\tilde{F}_t \leq F_t$), $n_{j|fmt}$ is the number of individuals in the data who choose j as their plan—given that they choose f at time t—and n_{mt} is the number of individuals who choose firm f at time t. $\tilde{\mathcal{J}}_{ft}$ is the set of plans offered by f at t dropping the top plan, and $|\mathcal{J}_{ft}|$ is the total number of plans offered by f at t.

Potential instruments Z_{ft} and Z_{fmt} include a firm's own prices, technology dummies, and functions of the menu of download speeds, as well as the equivalent quantities for a firm's competitors. The main concern is that price is endogenous to the unobserved demand shock ξ_{ft}.

Since demand shocks are firm wide, they correspond intuitively to changes in the quality of (non-Netflix related) bundled services, system-wide disruptions in reliability, or the quality of customer service. Since the shocks are observed by f, they are endogenous to contemporaneously set inputs in the firm's profit maximization problem. I assume that only price is set contemporaneously.

The instruments Z_{ft} for G^3 that I use are variables that affect f's price at time t but do not depend on ξ_{ft}. I assume that the menu of download speeds evolves exogenously, according to background technological progress that enables faster speeds over the same infrastructure. Functions of competitors' contemporaneous menu of download speeds, as well as f's own speeds, are valid instruments under this assumption. I also use lagged functions of the instruments to separately identify preferences from latent switching costs, as in Scherbakov (2015).

The instruments Z_{gmt} for G^2 I use are similar to the above: I include a constant, the

weighted average of ISP minimum, mean, and maximum offered download speeds across ISPs of the same technology g, as well as a weighted average over all competitors' minimum, average and maximum download speeds within each market m.

I stack $G^2(\cdot)$ and $G^3(\cdot)$ and use two-step GMM to recover $\hat{\theta}_d^h$ as the parameter that minimizes the objective function. As Hansen (1982) shows, provided that $R \to \infty$ and the ACS sample sizes go to infinity, the estimator will be consistent.

Estimating the Mean Parameters

Given an estimate of θ_d^h that minimizes the GMM objective function, I use the recovered $\hat{\delta}_{ft}$ to estimate the mean parameters $(\bar{\gamma}, \bar{\alpha})$. From Equation (1.1), $\hat{\delta}_{ft}$ is comprised of ISP-specific fixed effects and time trends, as well as the mean effect of the slowdown.

In my baseline case, I estimate Equation (1.1) by a fixed effects regression with instruments for the slowdown dummy. My timing assumption is that firms observe ξ_{ft} before they optimize over the offer they make to Netflix. Thus, an anticipated demand shock will change the marginal value of agreement, inducing different screening and a correlation in the distributions of ξ_{ft} and a_{ft}. I instrument for a_{ft} using a subset of the markup-shifting instruments Z_{ft} described above. If f's opponents exhibit changes in their menu of offered speeds, this change also affects f's marginal returns to agreement with Netflix but is plausibly exogenous to ξ_{ft}.

Falsifying an Alternative Demand Model

A concern that arises from my formulation of the demand model is that it does not directly allow the Netflix slowdown to affect consumers based on their chosen download speed. One can imagine that if the slowdown is a proportional reduction in throughput to Netflix, higher download speed consumers will suffer less since they can absorb greater absolute throughput reductions before streaming quality becomes a problem.

If this concern is valid, then it introduces another reason for ISPs to delay bargaining: ISPs may throttle Netflix traffic to force consumers to purchase faster plans to upgrade their way out of the slowdown. Since faster plans are higher margin, ISPs stand to benefit, especially if consumers experience a ratcheting effect in their preferences for speed.

I examine whether consumers change their speed upgrading behaviour at affected ISPs during periods of disagreement. I use the panel of MBBA consumers to show using individual level data that consumers do not disproportionately increase their rate of upgrading at ISPs affected by the disagreement.

I run LPM specifications of the following form:

$$P_{it}(j'|j,f) = \gamma disagree_{ft} + \beta \Delta \log p_{ft} + \bar{\alpha}_i + \bar{\alpha}_t,$$

where $P_{it}(j'|j,f)$ is the probability that consumer i changes their plan at ISP f to any j' with speeds greater than those of their current plan j, $disagree_{ft}$ is a dummy indicating whether ISP f is still negotiating, and p_{ft} is the price of f's entry level plan. Results are presented in Table 1.5 and indicate a robust correlation between the probability of exiting an ISP and whether that ISP is currently experiencing degraded Netflix quality. The coefficient on the $disagree_{ft}$ variable is consistently negative and significant. Far from upgrading their speed during the slowdown, consumers in fact slowed down their rate of upgrading, which provides evidence against the "forced upgrading" rationale for prolonging negotiations with Netflix.

1.6.2 Upstream Bargaining

The supply side parameters to estimate are θ_s, a vector that governs the distribution of μ_f, $F(\mu_f|w_f,\theta_s)$. I assume that μ_f is the product of an observable function of ISP attributes w_f, and an ISP-specific shock ζ_f that is observed by Netflix but not by the ISP or the econometrician:

$$\mu_f = \exp(\mu' w_f)\exp(\sigma_\zeta \zeta_f),$$

I assume that ζ_f is distributed normal but truncated from below, and that each draw ζ_f is independent and identically distributed, which implies a truncated lognormal distribution for μ_f.

To estimate $\theta_s \equiv (\mu, \sigma_\zeta)$, I will rely on maxmimum likelihood. Given a guess of the

parameters, I solve the optimal state contingent policy functions for ISPs, which generates a joint probability distribution of agreement timings and offers. From the data section, I only have observed agreement/disagreement timings $\{T_f\}$ so the likelihood function will have the form:

$$\mathcal{L}(\{T\}_f ; \mu, \sigma_\zeta) = P(\hat{T}_1(\mu, \sigma_\zeta) = T_1, \ldots, \hat{T}_F(\mu, \sigma_\zeta) = T_F),$$

where $\hat{T}_f(\mu, \sigma_\zeta)$ are agreement times predicted by the model for a given parameter vector. I suppress the implicit dependence of predicted agreement times on demand side parameters.

For the likelihood to be well-identified, one of two situations must be true: the model may give unique predictions for agreement times for a given set of parameter values. Alternatively, the model may have multiple equilibria for given θ_s, but as long as the data is only generated by one equilibrium and I select the correct predicted equilibrium when forming the likelihood I will recover the correct parameter values.

To construct the likelihood I must find the predicted agreement times as a function of parameters, which will follow from recovering the optimal policies. I first detail what variation in the data will identify the parameters of the distribution, then describe how I recover optimal policies and in doing so, select an equilibrium to input into the likelihood.

Identification of (μ, σ_ζ)

To identify the supply side parameters, I rely on (1) cross-sectional variation in the disagreement durations, and (2) the cross sectional correlation in covariates and durations, conditional on the marginal increase in flow profits from agreement.

Estimating σ_ζ is straightforward: only $\sigma_\zeta > 0$ allows the model can generate positive disagreement durations, since w_f is observed. Variation in disagreement times across ISPs that look similar in terms of covariates and flow profits will identify the dispersion parameter σ_ζ.

Recovering μ is more challenging: in the standard screening model with linearly transferable utility and no outside stream of profits, the rate of screening is independent of the parameter μ. Increasing μ does not affect the *relative* cost of postponing agreement, so the

probability of an agreement time appearing in the data remains the same for different μ. In other words, the duration based likelihood is flat in μ. However, with side payments, if high w_f ISPs that face large dips in profits if agreement is not reached have long disagreement lengths in the data, then μ must be positive and significant. This follows because μ increases both the mean and the variance of μ_f: if increasing μ lead to a pure mean shift, then μ would still be unidentified even with side payments.

Solving for Policy Functions

To form the likelihood, I must solve for optimal policies. If firms care about the states and strategies of all other firms the model quickly becomes intractable. I adopt a cutoff rule to mitigate this curse of dimensionality. First, for f consider all $-f$ such that each $-f$ overlaps at least 15% of f's footprint. Starting with Comcast, I determine all such $-f$—Comcast's primary competitors—and then determine the $-f$ of those primary competitors. Following this procedure until no new firms are added will form a minimal group that cares about the actions of all other members of that group, even if some members' footprints do not overlap.

The minimal group constructed in such a way is Comcast, AT&T, TimeWarner and Verizon; call this set of ISPs $\mathcal{F}_L$. I assume when solving these firms' optimal policies that they pay attention to the markets that are exclusively served by members of the group. Other ISPs react to the policies of the biggest four, but do not affect the four in turn, which keeps the state space of the large ISPs manageable.

Small ISPs primarily compete with large ones (see Table 1.9), so I assume small ISPs do not pay attention to each other. I also assume that small ISPs play strategies such that large ISPs with which they do not overlap behave deterministically according to their durations in the data. For instance, the small cable internet provider Mediacom's state space includes the information state of AT&T and Verizon, but Mediacom assumes Comcast—a large cable provider that AT&T and Verizon care about—deterministically reaches agreement in the third quarter of negotiations.

The equilibrium I select gives Comcast the first mover role. As in Jia (2008) or Björkegren (2015), I begin by specifying that all other $f \in \mathcal{F}_L$ agree immediately. After solving Comcast's optimal response, I perform a round robin algorithm, updating the optimal strategies

of each of the four large ISPs in turn until their policy functions converge.[21] I then recover the policy functions of the smaller ISPs given the optimal behaviour of $f \in \mathcal{F}_L$.

The full set of policy functions gives the joint distribution over agreement timings for the chosen θ_s. Since large ISPs do not pay attention to small ISPs and small ISPs only pay attention to large ISPs, the likelihood function can now be written as

$$\mathcal{L}(\{T\}_f ; \mu, \sigma_\zeta) = P\left(\left\{\hat{T}_f(\mu, \sigma_\zeta) = T_f\right\}_{f \in \mathcal{F}_L}\right) \times \prod_{f \in \mathcal{F}_S} P\left(\hat{T}_f(\mu, \sigma_\zeta) = T_f \middle| \left\{\hat{T}_f(\mu, \sigma_\zeta) = T_f\right\}_{f \in \mathcal{F}_L}\right),$$

where $\mathcal{F}_S$ is the set of smaller ISPs.

Terminal Dates

Since ISP value functions depend on t and not states alone, policy functions will be solved backwards from a terminal condition. My condition is dictated by data limitations: I have price data until the end of 2014, and the National Broadband Map extends only through the middle of 2014. I assume that bargaining terminates in December 2014 with the minimal offer, which is always accepted.[22]

I predict the evolution of profits after 2014 with a flexible time series. This is necessary since firms that exit bargaining enjoy their discounted payoff into the infinite future. Given the terminal condition, I solve policy functions by iterating backwards.

Shape of Joint Density of Agreement

Each ISP will offer a strictly decreasing sequence of fees to screen Netflix's unobserved ISP specific type. Indeed, even without flow profits it is never optimal to make an offer outside the updated conditional support of μ_f, and the fact that slowdowns always hurt ISP flow profits intensifies the incentives to reduce offered fees over time.

[21] Jia (2008) uses this approach for estimation as I do, while Björkegren (2015) uses it for computing counterfactual equilibria. By choosing the first mover and the order of responses, this procedure selects an equilibrium; recent papers that mention this feature of the algorithm include Wollman (2016) and Lee and Pakes (2009).

[22] I will evaluate different assumptions on the terminal date T in future robustness checks.

The question remains whether an ISP will decrease its offers more quickly or more slowly once a rival has concluded agreement. This feature of the model will be important for the counterfactual merger analysis. If a rival's agreement increases the marginal benefit (from flow profits) to own agreement (supermodularity in flow payoffs), then if the merged firm screens faster so too will the other firms. If a rival's agreement decreases the marginal benefit to own agreement (submodularity in payoffs), then faster screening by the merged firm will lead to slower agreement times on average by the other firms. Figure 1.6 gives the shape of the joint density of agreement timings for two ISPs in the case where flow profits are submodular and supermodular. *Ex ante*, analytic derivations of the flow profits imply submodular profits as in Jia (2008). Full tests of the degree of submodularity are presented in the online appendix.

1.7 Empirical Results

This section presents estimation results of the demand model and dynamic bargaining framework.

1.7.1 Demand estimates

Table 1.6 summarizes the key estimated parameters for the demand model. The first column refers to the ISP characteristic, and the second the household characteristic that it is being interacted with. The mean effect of disagreement is negative and significant, implying a baseline level of disutility from disagreement as reflected in the aggregate market share movements. Unobserved consumer preferences against the slowdown are large and significant while observed heterogeneity is mostly insignificant, suggesting that latent Netflix subscribership patterns are not captured by the observable covariates.

The coefficients on price and download interactions are mostly significant, and unobserved heterogeneity is very important. Recall that the price coefficient is of the form

$$\alpha_{ip} = -\exp(-(\alpha^o_{p1} + \sigma_p \nu_{ip} + \alpha^o_{p2} Inc + \alpha^o_{p3} Inc \times q_{25,Inc})).$$

Coefficient signs are intuitive: increases in income reduce the disutility of price increases, although these effects are dampened for income earners in the first quartile. For download speed, larger households have preferences for higher speeds, but this sharply drops off for households larger than 4 members.

The dispersion in unobserved preference for the price will help recover a non-degenerate predicted distribution of plan shares. This feature of the model is important not only to match the data, but to recover meaningful price elasticities. With degenerate plan shares, the price elasticity for plans with no share will be zero or infinite. Moreover, the elasticity for plans with a positive share will be too small if price changes mostly induce switching across ISPs and not switching within an ISP's own menu.

I estimate a large set of heterogeneity parameters for the cable technology dummy. Since bundling opportunities are most prevalent with cable, allowing valuation of different technologies to vary by household characteristics helps recover credible estimates of the price coefficient. I find that larger households and households headed by men draw greater utility from having a cable subscription.

The demand curve's main purpose is as an input to the bargaining game. Consumer substitution in response to a slowdown makes lengthy bargaining costly, and will identify the parameters of the distribution of unobserved surplus in combination with disagreement durations. I present the substitution patterns for the four largest ISPs in table Table 1.7. The experiment is to turn on the disagreement dummy for each ISP in turn, assuming that all other ISPs have finished bargaining. Since demand is static, this experiment captures the marginal consumers an ISP has to lose if it does not conclude negotiations with Netflix if all other ISPs have already agreed.

In response to the slowdown, consumers at cable internet providers (Comcast, Time-Warner) mostly switch to AT&T, while consumers at non-cable providers switch to Comcast. These patterns reflect the overwhelming market presence of Comcast and AT&T as the providers with the largest footprints. This also reflects my assumptions on the functional form of utility: the best alternative to 1.50Mb/s AT&T is not 6 Mb/s AT&T because the disutility of lower streaming quality does not decrease in the purchased download speed.

Switching rates increase mostly uniformly by download speed: individuals with higher

speeds switch more in response to the slowdown. The increase is especially steep for Comcast and Verizon, while at AT&T higher speed subscribers do not switch much more than lower speed ones. Since different speeds of internet are provided at roughly the same marginal cost, high speed customers are also high margin customers. Results from this table imply that Comcast and Verizon have an especially strong incentive to conclude negotiations quickly, whereas the marginal loss to AT&T from prolonged disagreement is lower.

The demand curve must be combined with a method for deriving the price-cost margin to recover profits. I compute the price elasticity for every plan in every quarter, and find that the median elasticity across offered plans in all quarters is 2.667. However, there is large variation, with an interquartile range of over 10 and almost 30% of plan price elasticities falling below one. In theory all elasticities should be above one, although elasticities for broadband below one are consistent with recent estimates in Dutz, Orszag and Willig (2012). The substantial fraction of elasticities falling below one precludes using a demand system inversion to recover marginal costs, so I assume a marginal cost of zero and take the evolution of prices over time as exogenous in my base case for estimating the supply parameters.

To asses model fit, I regress the model predicted conditional plan shares against actual plan shares from the MBBA data. I recover recover an R^2 of 0.76, and across the entire sample I cannot reject the null that the slope is one. I also find that the model predicts degenerate within-ISP plan shares in only 6% of quarter-ISP combinations, and never for any of the largest four ISPs. To indirectly validate the predictions in Table 1.7 and verify that the model does not do poorly during the periods of the slowdowns, I calculate the quarter by quarter R^2 and slope estimates for the regression of predictions on data.[23] I find that the R^2 has a slight inverted u-shape, with the highest values during 2013. The model's good performance even during the slowdown is encouraging, although it also likely reflects the fact that unobserved price and download heterogeneity in the model does not vary over time but plan characteristics are slowly improving.

[23]Unfortunately the MBBA does not track well consumers who switch ISPs, so there is no way to directly verify in the data if the substitution patterns in response to a slowdown predicted in hold.

1.7.2 Upstream Bargaining Estimates

The demand estimates from the previous section are combined with the dynamic multilateral bargaining assumptions and disagreement durations to estimate the distribution of Netflix's marginal value of agreement vis-à-vis each ISP. Estimates of the coefficients of the mean shifting variables w_f, as well as the standard error of the dispersion of information σ_ζ, are presented in Table 1.8.

The most parsimonious version of the model in column (1) features σ_ζ only, so that $\mu_f = \exp(\sigma_\zeta \zeta_f)$, with ζ_f unobserved by the ISPs. With no shifters w_f, the ISPs whose marginal loss in subscribers is largest would be predicted to agree most quickly. In the data the largest ISPs have the greatest marginal loss in subscribers but also take the longest time to agree, which will require the presence of the shifters w_f to rationalize.

Since the number of ISP that bargain is limited ($N = 22$), I restrict the number of covariates to functions of the footprint—the number of housing units in each ISP's physical network—and the ISP's connection technology. My preferred estimates are column (3), where I allow a quadratic term in pass-by. The quadratic term allows for increasing or decreasing returns to scale for Netflix in installing the interconnection technology; the positive estimated coefficient in specification (3) suggests increasing returns to scale, so that Netflix's cost savings from making the investment increase in the size of the ISP's network. This agrees with Netflix's own statements that there are returns to scale in interconnection.[24] All else equal, the larger a provider's network, the greater the value in interconnection: evaluated at mean pass-by, increasing the number of households passed by 10 million increases the mean of the surplus distribution by 3.2 million.

Combining firm optimal screening strategies with the supply model estimates yields a key prediction of the model: the expected share of bilateral surplus that accrues to Netflix and each ISP. Understanding how this share varies with ISP size is crucial to understanding whether how Netflix's incentive to make its quality investment changes with the size of a

[24] "...our actual costs rise with every additional point of interconnection...Moreover, interconnecting at more locations can be less efficient, and thus more expensive, for Netflix. With each additional location, the costs associated with placing and maintain each location—e.g., servers, routers, rent, cabling costs—are spread across a smaller amount of capacity." Paragraph 20, Statement of Ken Florance, Vice President of Content Delivery at Netflix since 2012.

negotiating ISP's network.

To see how the model predicts shares, suppose that ISP f makes offer τ_{ft} that is accepted by Netflix. The model predicts that the surplus is distributed $F(\mu_f|\mu_f < \tau_{ft-1}; w_f, \theta_s)$, so that Netflix's expected surplus share in that period is $E_{\mu_f}[(\mu_f - \tau_{ft})/\mu_f|\mu_f < \tau_{ft-1}]$. Integrating over the marginal distribution of agreement timing offers for a given ISP will give Netflix's expected share for that ISP.

Netflix's *ex ante* expected share against the four largest ISPs ranges between 28% against AT&T and 33% against Comcast. On average, Netflix retains 34% of its surplus against smaller firms. That Netflix's share is low even though its subscribers are inelastic is partially driven by modelling assumptions: myopia substantially increases Netflix's willingness to accept high fee offers, decreasing their share of the surplus. However, variation in shares across ISPs will incorporate differences in ISP subscriber elasticities, which reflects ISP's consumer-facing market power, as well as the size and dispersion of Netflix's surplus for that ISP, which reflects the how the model rationalizes systematic differences in agreement times.

Large ISPs are predicted to take their time in bargaining because the surplus they stand to extract from Netflix is relatively large compared to their subscriber loss. Since it is more worthwhile to try and extract surplus with a fine screening strategy for these ISPs, they will naturally appropriate a greater share in equilibrium.

Since the surplus is only divided after a delay, from the *ex ante* perspective a certain fraction of surplus will be lost by pushing its realization into the future. On average lost surplus is roughly 12%, with a high of 14% against AT&T. Recall that this inefficient delay is necessary in order for Netflix to signal its true valuation: the only credible way Netflix can claim to have a low surplus from investing vis-a-vis a particular ISP is by rejecting high fee offers, which will push bargaining and the realization of the surplus into the future.

I asses the model fit by plotting the predicted time to a successful offer against the actual time in Figure 1.7. The correlation is positive and significant and while an unweighted regression is nowhere near the 45 degree line, the model does perform better for the larger ISPs. The ISPs that agree immediately in the data are predicted to take just over a quarter longer to agree in the model. Since their pass-by is so small, their average time to agreement mostly reflects the estimate of σ_ζ. Recall that the observable component on μ_f can increase

dispersion and time to agreement, but cannot ensure positive probabilities of delay on its own. Since the estimate of σ_ζ must also rationalize the presence of small ISPs who still take time to bargain, it will tend to overpredict delays for ISPs that agree quickly in the data and underpredict delays for ISPs that take more time. This is not a problem with the model however, as differences in agreement times for similar ISPs simply reflects variation in the draws of μ_f.

Note also that the model fit improves with the additional of the quadratic surplus, validating the inclusion of this term as a mean shifter in the surplus distribution. The positive estimated quadratic term helps with fit, and will be especially important for the merger counterfactual, since it implies that interconnecting with Comcast and TimeWarner's networks separately is not as valuable as interconnecting with the single, merged provider whose network is equal in size to the sum of the other two.

Finally, as a byproduct of estimation, the model predicts the exact offers that are made in the period where Netflix agrees. These fees are potentially of interest to other large content providers if the practice of paying terminating fees to last-mile ISPs becomes more prevalent in the future. I find lump sum transfers of 19.0 million USD to AT&T, 12.64 million to Comcast, and 3.89 million and 4.90 million to TimeWarner and Verizon, respectively. Smaller ISPs recoup fees around 1 million USD, with the lowest fees going to Mediacom and Brighthouse—two small ISPs whose bargaining only resolved after a 4 quarter delay. The largest ISPs and Netflix earn billions of dollars yearly, so these interconnection fees are comparatively small.

1.8 Counterfactuals

1.8.1 Comcast-TimeWarner merger

With both the supply and demand estimates in hand, I turn to the paper's main counterfactual—the effect of a merger between Comcast and TimeWarner.[25] There are three dimensions on which to evaluate the merger. First, does the hold-up problem for Netflix's capacity invest-

[25]Since there are many price elasticities that are less than one, in my base case I assume that the merged firm simply takes on the plan characteristics and pricing of Comcast instead of doing a competitive repricing.

ment change? That is, do the *ex ante* predicted splits of the upstream surplus change, both at the merged firm and its competitors? Second, does consumer welfare change? Third, and related to the second point, how do bargaining times—and the fraction of ex ante surplus lost due to delays—change?

The first criterion speaks to the increased degree of extractive power held by the larger merged firm which affects the split of surplus, and disincentivizes Netflix from making the quality investment. The last two criteria speak directly to dead weight loss and consumer welfare, which are the typical targets for judging the value of a merger. In this case the delay in agreement generates a deadweight loss, both because streaming quality degradation directly reduces welfare and because it pushes realization of surplus to the (discounted) future.

I first examine how bargaining times vary across all ISPs. I plot the change in the *ex ante* expected bargaining times in Figure 1.8 for ISPs that experience a non-negligible change. The merged ISP has an *ex ante* expected disagreement length of 3.69; while this is not a huge percentage increase in duration length for customers who were previously with Comcast (3.43 quarters), it is quite large for former TimeWarner customers (2.84 quarters). Other ISPs that experience changes in bargaining duration due to the merger uniformly experience slight decreases. From Table 1.9, notice that these ISPs are all ones that have a great deal of market overlap with Comcast and/or TimeWarner.

The quadratic term in the mean-shifting component of the surplus distribution drives the result that bargaining takes more time. The merged firm has the combined footprint of Comcast and TimeWarner, but the mean and dispersion in μ_f increase more than linearly. Meanwhile, since there is 0% market overlap between these ISPs, the merged firm's marginal loss in subscriber revenue as a result of disagreement is simply the sum of Comcast and TimeWarner's marginal losses. However, the surplus they stand to gain from finer screening is greater than the sum of the surpluses that Comcast and TimeWarner stood to get individually. To reiterate, this intuitively represents that for Netflix, there are more cost-savings associated with investing in one large network than two smaller ones.

To understand why direct competitors of the merged firm take slightly less time, recall that since profits are submodular in the agreement vector, ISPs have more subscriber revenue

to (re)gain on the margin by being the first to agree. In equilibrium the merged firms will screen slightly faster on average to capture the greater expected marginal gain of being the first to agree.

I estimate aggregate consumer surplus under three scenarios: without any quality degradation, with the joint distribution of quality degradation predicted by the model, and with the counterfactual joint distribution of degradation when Comcast and TimeWarner are allowed to merge. I use the standard closed-form logit formula to derive individual surplus and then aggregate across consumers and markets.[26]

The *ex ante* aggregate consumer surplus decreases by 0.51 percent moving from no quality degradation to the expected degradation lengths produced by the fitted model in Figure 1.7. Aggregate surplus decreases by 0.54 percent moving from no quality degradation to the expected degradation lengths produced in the new competitive environment after the merger, a roughly 4.3 percentage point increase in magnitude. The aggregate decrease masks heterogeneity in responses across choice sets. The 20 percent of households facing choice sets that previously included TimeWarner see the magnitude of their aggregate welfare loss increase by 17.6 percentage points (to 0.60 percent), while the 34 percent of households in choice sets that include Comcast see their aggregate welfare loss increase by 6.5 percentage points. Households facing choice sets that include a DSL or Fiber provider that share a large footprint with the merged firm, but that do not include the firm itself actually enjoy an increase in aggregate surplus of 0.3 percentage points due to slightly shorter expected slowdown durations.

The merger affects the amount of surplus that is lost due to delays. The substantial increase in *ex ante* predicted disagreement times compared to the base case implies that 16 percent of the upstream surplus is expected to be lost in the bargaining game between Netflix and the merged ISP, an increase of almost 30 percentage points from the case with no merger. Reductions in *ex ante* disagreement times for other ISPs imply that the share of surplus that is lost in bargaining decreases slightly from 12 to 11 percent on average.

[26]

$$\hat{W}_{rmt} = \frac{1}{\hat{\alpha}_r} \log\Big(\sum_{f \in \mathcal{F}_{imt}} \exp(\hat{\delta}_{ft} + \hat{\lambda}_{rfmt}) \Big), \qquad \hat{W}_t = \sum_m w_{mt} \frac{1}{R} \sum_r \hat{W}_{imt}$$

Netflix's *ex ante* expected share of the surplus with respect to the merged firm decreases to 29 percent, from an aggregate (combined Comcast and TimeWarner) share of 32 percent, a decrease of just over 9 percentage points.

The split and size of upstream surplus matters for Netflix's decision to invest in the CDN infrastructure. In particular, if the bargaining environment grows more adverse then there will be a hold-up problem, as either bargaining delays or greater ISP bargaining power prevents Netflix from realizing all of the surplus from its investment. I investigate the effect of the merger on Netflix's incentive to invest by analyzing how the merger changes the *ex ante* distribution of aggregate surplus net of transfers to ISPs. I will compare these distributions of aggregate surplus to an estimate of Netflix's upfront investment cost to develop the CDN.

I assume that Netflix first draws ISP specific surpluses, then pays the R&D cost.[27] From the model, I can recover the *ex ante* distribution of aggregate surplus net of transfers accruing to Netflix given optimal firm strategies in both the original and merger cases:

$$\sum_f \beta^{\hat{T}_f}\left(\mu_f - \tau_{f,\hat{T}_f(\mu_f)}\right) \sim G, \qquad \sum_f \beta^{\hat{T}'_f}\left(\mu'_f - \tau'_{f,\hat{T}'(\mu'_f)}\right) \sim G',$$

where μ_f is random, $\hat{T}_f(\mu_F)$ is the random stopping time predicted from optimal strategies that depends on the draw μ_f, and the variables with primes denote the same quantities drawn under the counterfactual industry structure where Comcast and TimeWarner merged.

I do not observe the initial fixed cost Netflix must expend on research and development before it can begin offering to install the CDN in ISP networks. I recover an estimate of the maximum possible fixed costs: given the actual observed agreement times, the model predicts the actual offer and an upper bound on each ISP-specific surplus as the previous period's rejected offer.[28] The sum of the upper bounds minus the actual offer (appropriately discounted) gives Netflix's maximum possible surplus from negotiation, which is an upper bound on its fixed cost.I estimate this quantity as 9.03 million USD. The lower bound on the fixed cost is zero: if Netflix's surplus is exactly equal to ISPs' fee in the periods Netflix

[27]Intuitively, this reflects the idea that Netflix does a cost benefit analysis before committing to the project.

[28]Since the surplus distribution is unbounded from above, for ISPs that agree in the first period of bargaining I assume that maximum surplus is the 97.5% quantile of their surplus distribution. Moreover, to recover this estimate requires that Netflix is able to forecast ISPs' demand shocks to know ISPs' exact offers in each period.

accepts, then Netflix retains no surplus, and may therefore have been indifferent about initiating bargaining:

$$0 \leq FC \leq \sum_f \beta^{T_f} \left(\tau_{f,T_f-1} - \tau_{f,T_f} \right)$$

I plot the histograms of G and G' in Section 1.10.2. The distribution of net surpluses without the merger first order stochastically dominates the distribution of surpluses with the merger. Given the fixed cost, the leftward shift in the distribution of surpluses reduces the probability that Netflix invests in the CDN by 5.1 percentage points, from 19.4 percent to 18.4 percent. These probabilities of investment are low because the fixed cost is an upper bound.

A fixed cost of zero implies there is no disincentive effect from not retaining additional surplus. However, for any non-zero fixed cost, that the original Netflix retained surplus distribution first order stochastically dominates implies that the hold-up problem grows worse with the new market structure after the merger.

1.8.2 Prohibiting Quality Degradation During Bargaining

In this section I analyze how outcomes would change if bargaining was not linked to quality degradation of the content provider's content.[29] Insofar as degradation is a choice made by one or both bargaining parties to gain leverage in the bargaining game, prohibiting quality degradation can be implemented.

This constraint is a more general form of what is referred to as "network neutrality." Network neutrality is a policy—recently affirmed by the FCC—that ISPs cannot differentially throttle or privilege the tranmission of different content, and is partly intended to prevent them from exercising their market power to extract payments from content providers. In my model, the marginal consumer will switch ISPs to prevent Netflix quality degradation, so it is Netflix that has the incentive to degrade quality to induce faster screening by ISPs. The idea that a content provider with market power may degrade quality to extract concessions

[29] Explicitly regulation the rules of bargaining is not an unusual counterfactual. In Gowrinsankaran, Nevo and Town (2015), the authors analyze the effect of a counterfactual hospital merger if the constituents are still forced to bargain separately.

in bargaining with ISPs is a new one in the network neutrality literature, since in typical theory models ISPs are assumed to be monopolists.[30]

Preventing quality degradation during bargaining increases aggregate welfare by 0.54 percent (the reverse of the 0.51 percent decrease in the previous section). ISPs with longer delays such as AT&T and Comcast see consumers in their choice sets have the largest increases in welfare of 0.58 percent and 0.56 percent respectively.

Netflix's share of the *ex ante* bargaining surplus goes down and the delay to agreement increases. Without the marginal loss in subscribers affecting the screening tradeoff between finer offers and a higher payout, and coarser offers and a quicker payout, the ISPs screen much more slowly. I summarize the effects in the graph of *ex ante* distributions of consumer surplus versus the fixed investment cost Section 1.10.2.

The probability of investment declines by 8.7 percentage points under the regime where quality degradation is prohibited during bargaining. In general, there is a two-sided holdup problem since infrastructure investment will only come on line if both sides in a bilateral monopoly agree to its installation. My setup analyzes a major investment by Netflix, and shows that the holdup problem grows more severe when quality degradation is prohibited since the marginal consumer is more elastic with respect to ISPs than Netflix subscriptions.

1.9 Conclusion

This paper combines a model of dynamic multilateral bargaining with demand for broadband internet to evaluate a counterfactual merger and a change in bargaining regime policy. I estimate an industry model of demand for internet access, plan choice, pricing and interconnection bargaining using data on plan prices, consumer choice sets and bargaining delays between U.S ISPs and the leading purveyor of streaming video content, Netflix. The model rationalizes a cross-section of bargaining delays between Netflix and U.S. ISPs, where the bargaining was over how to split the surplus from investments that Netflix wanted to make at each ISP that would save Netflix operating costs and improve video streaming quality.

I find that a proposed merger between TimeWarner and Comcast that was challenged

[30]See, for instance, Economides and Hermalin (2012)

by the Federal Communications Commission would have both raised the fees Netflix paid to make its investment and increased bargaining length, increasing the magnitude of aggregate consumer welfare loss and disincentivizing Netflix from making the investment. A regime where quality degradation is prohibited during bargaining improves consumer welfare compared to the base bargaining case, but Netflix has a lower share of surplus in this regime and is less likely to invest in the infrastructure project whose surplus is being bargaining over. This result follows because allowing quality degradation during bargaining actually improves Netflix's bargaining position, since marginal consumers in the data would rather switch ISPs to get better Netflix than cancel Netflix.

There are two portable insights from this paper. The first is that increasing returns to scale from a content provider's intermediary-specific investment can be a source of bargaining power for larger intermediaries. Even if a large intermediary does not have more market power against the consumers it serves, if it serves many consumers then the content provider benefits especially from any investment (with a fixed cost component) that it makes in that intermediary, implying a lower outside option for the intermediary. The content provider will retain a smaller share of a larger surplus; whether the net effect is to disincentive investment is an empirical question.

The second insight is that adding an investment margin for content providers implies that intermediaries that negotiate better prices on content may actually harm consumers by doing so, since retaining a lower share of the surplus from the content they deliver may disincentivize content providers from making investments in the quality of their content. Since content provider investment in quality in a salient feature in most two-sided markets, adding this margin is potentially important for welfare and merger analysis in a broad range of industries.

1.10 Appendices

1.10.1 Appendix A: Tables

Table 1.1: Cross-sectional ISP disagreement durations

	Dependent variable:					
	Duration				Duration*	
	(1)	(2)	(3)	(4)	(5)	(6)
log(footprint)	0.757***	0.733***	0.726***		0.823***	
	(0.211)	(0.220)	(0.230)		(0.205)	
share				0.478**		0.558**
				(0.220)		(0.205)
log(connect)	0.289	0.278	0.283	0.161	0.334	0.196
	(0.322)	(0.329)	(0.341)	(0.380)	(0.303)	(0.354)
avg.num.comp.		−1.217	−1.326	−2.384	−1.085	−2.260
		(2.218)	(2.366)	(2.610)	(2.107)	(2.437)
comp.price			0.007	0.014	0.004	0.010
			(0.039)	(0.044)	(0.035)	(0.041)
Tech. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23	23	23	23	23	23
R^2	0.450	0.460	0.461	0.317	0.573	0.405

The dependent variables are the disagreement durations. Duration* provides an alternate construction of disagreement durations where durations are increased for some ISPs. log(footprint) gives the number of housing units the ISP network includes, log(connect) is a measure of the connectedness of the ISP physical network, avg.num.comp. gives the average number of competitors the ISP faces, comp. price gives the average minimum price offered by competitors. share gives the share of subscribers in mid-2013, and is an alternative measure of size. Tech. controls indicates whether technology (cable, DSL, fiber) dummies are included. *p<0.1; **p<0.05; ***p<0.01.

Table 1.2: Panel of correlated ISP agreement timings

	Dependent variable:					
	agree_t				agree^*_t	
	(1)	(2)	(3)	(4)	(5)	(6)
comp.agree_t	0.385^{***}	0.046	0.257^{**}	0.046		
	(0.116)	(0.111)	(0.101)	(0.112)		
comp.agree_{t-1}			−0.195	0.058		
			(0.174)	(0.209)		
comp.agree^*_t					0.141	−0.021
					(0.120)	(0.122)
$\text{comp.agree}^*_{t-1}$					0.412^{**}	−0.018
					(0.161)	(0.194)
ISP FE	✓	✓		✓		✓
Time FE		✓		✓		✓
Observations	70	70	70	70	70	70
R^2	0.270	0.714	0.110	0.715	0.109	0.794

The dependent variables are dummies that take a value of 1 in the period when an ISP concludes negotiations with Netflix. The covariates are dummies equal to one when an ISP's primary competitor concludes negotiations, or lags thereof. Stars on a variable indicate the alternative duration construction. *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the ISP level.

Table 1.3: Panel of household ISP switching

	(1)	(2)	(3)	(4)	(5)	(6)
disagree	0.010***	0.011**	0.007***	0.009*		
	(0.003)	(0.004)	(0.003)	(0.004)		
disagree*					0.008***	0.010**
					(0.003)	(0.005)
Δp_{jt}			−0.02***	−0.01***	−0.02***	−0.01***
			(0.001)	(0.002)	(0.001)	(0.002)
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Balanced		✓		✓		✓
Observations	33,560	12,126	33,560	12,126	33,560	12,126
R^2	0.235	0.159	0.242	0.164	0.242	0.164

The dependent variable is a dummy that equals one if a consumer exits ISP j. disagree and disagree* are dummies that equal one if ISP j is negotiating in that quarter with Netflix. Δp_{jt} is the change in price of ISP j's entry level plan (in tens of dollars). *p<0.1; **p<0.05; ***p<0.01. Standard errors at clustered at the household level.

Table 1.4: Panel of ISP subscriber share changes

	OLS		FE			
	(1)	(2)	(3)	(4)	(5)	(6)
disagree	-0.007*	-0.06***	-0.004**		-0.004**	
	(0.002)	(0.002)	(0.002)		(0.002)	
disagree*				-0.009		-0.009
				(0.006)		(0.006)
Δp_{jt}^{med}					-0.02	-0.02
					(0.04)	(0.04)
ISP FE			✓	✓	✓	✓
Time FE			✓	✓	✓	✓
Observations	541	526	526	526	526	526
R^2	0.006	0.019	0.641	0.638	0.641	0.638

The dependent variable in all specifications is the log difference in ISP subscriber shares between t and $t-1$. (1) includes all data, including time periods when a merger happens; (2) removes mergers—as these are large, noisy events, removing them increases significance even as it reduces the point estimates; (3) and (4) add in ISP and time fixed effects and perform the exercise for the alternative specification of disagreement times; (5) and (6) also control for ISPs' median price (in tens of dollars). ***, ** and *: 0.1%, 1% and 5% significance. Standard errors are clustered at the ISP level.

Table 1.5: Panel of household speed upgrading

	Dependent variable:					
	Speed Upgrade Dummy					
	(1)	(2)	(3)	(4)	(5)	(6)
disagree	-0.029^{***}	-0.017	-0.033^{***}	-0.020		
	(0.008)	(0.014)	(0.008)	(0.014)		
disagree*					-0.030^{***}	-0.021
					(0.009)	(0.015)
Δp_{jt}			-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.02^{***}
			(0.004)	(0.01)	(0.004)	(0.01)
Unit FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Balanced		✓		✓		✓
Observations	33,560	12,126	33,560	12,126	33,560	12,126
R^2	0.138	0.089	0.139	0.090	0.139	0.090

The dependent variables is a dummy that equals one if a consumer upgrades their plan speed. disagree and disagree* are dummies that equal one if ISP j is negotiating in that quarter with Netflix. Δp_{jt} is the change in price (in tens of dollars) of ISP j's entry level plan. *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the household level.

Table 1.6: Demand Model Parameter Estimates

ISP char.	HH char.	Notation	Parameter	S.E.
Disagree	Constant	$\bar{\gamma}$	-0.011^{*}	0.004^{*}
	HH size	α^{o}_{s1}	-0.040	0.029
	HH size $\times \mathbf{1}[\text{size} \geq 5]$	α^{o}_{s2}	-0.076	0.040
	Speak English	α^{o}_{s3}	0.294	0.152
	Male head	α^{o}_{s4}	-0.002	0.001
	Unobs. het	σ_s	0.544^{*}	0.247
Price	Constant	α^{o}_{p1}	3.698^{***}	1.099
	Income	α^{o}_{p2}	0.093^{**}	0.034
	Income $\times q_{25,Inc}$	α^{o}_{p3}	-0.020	0.012
	Unobs. het	σ_p	0.686^{***}	0.203
Download	Constant	α^{o}_{q1}	0.027^{***}	0.008
	HH size	α^{o}_{q2}	0.762^{***}	0.202
	HH size $\times \mathbf{1}[\text{size} \geq 5]$	α^{o}_{q3}	-0.365^{**}	0.117
	Speak English	α^{o}_{q4}	0.559	0.455
	Male head	α^{o}_{q5}	0.489^{***}	0.145
	Age	α^{o}_{q6}	0.010^{*}	0.005
	Unobs. het	σ_q	1.157^{***}	0.330
Cable Dummy	HH size	α^{o}_{g1}	0.061^{*}	0.029
	HH size $\times \mathbf{1}[\text{size} \geq 5]$	α^{o}_{g2}	0.002	0.002
	Speak English	α^{o}_{g3}	0.034	0.018
	Male head	α^{o}_{g4}	0.107^{*}	0.043
	Age	α^{o}_{g5}	0.041	0.033
	Unobs. het.	σ_g	0.239	0.162

This table presents estimates of the non-linear consumer preference parameters, as well as the linear effect of disagreement $\bar{\gamma}$. Standard errors for the non-linera parameters are computed from the numerical derivatives of the GMM moments at the estimated parameters. The linear parameter estimate standard errors come from the linear IV with fixed effects regression of $\delta(\theta^{h}_{d})$ on firm dummies and time trends, and the disagreement dummy.

Table 1.7: Quality Degradation Induced Consumer Substitution

ISP	Plan (Mb/s)	% Switch	Best Alt.	Best Plan (Mb/s)	% Switch cond.	% Out cond.
AT&T	1.50	0.31	Comcast	3	22.27	46.88
AT&T	6	0.29	Comcast	3	23.20	40.30
AT&T	12	0.32	Comcast	3	22.87	40.38
AT&T	18	0.34	Comcast	50	27.08	36.15
AT&T	24	0.34	Comcast	50	26.46	22.00
AT&T	45	0.36	Comcast	50	21.95	9.28
Comcast	3	0.38	AT&T	1.50	16.27	46.90
Comcast	50	0.52	AT&T	24	20.41	25.14
Comcast	105	0.65	AT&T	45	30.35	6.43
TimeWarner	1	0.43	AT&T	1.50	20.56	49.06
TimeWarner	15	0.43	AT&T	24	21.18	37.28
TimeWarner	20	0.45	AT&T	24	40.14	10.64
TimeWarner	50	0.53	AT&T	45	20.72	16.04
Verizon	0.50	0.29	Comcast	3	18.41	46.68
Verizon	7.10	0.25	Comcast	3	11.76	31.30
Verizon	50	0.35	Comcast	50	25.62	29.29
Verizon	75	0.43	Comcast	50	21.66	10.28
Verizon	150	0.47	Comcast	105	37.68	2.52

The experiment in this table is to turn on the disagreement dummy for each ISP in turn, assuming that all other ISPs have agreed already. I use data from 2013-06-01. The Best Alt. and Best Plan columns gives the ISP and plan to which the most subscribers switch conditional on switching. The last two columns give the percent of individuals who switch to the next most popular plan, and the percentage who switch to the outside option, conditional on switching.

Table 1.8: Supply model estimates

		MLE	
	(1)	(2)	(3)
σ_ζ	1.538	0.204	0.171
footprint	—	7.311	6.618
footprint squared	—	—	0.709
DSL dummy	—	0.456	−0.130

The DSL dummy takes a value of 1 if the ISP's primary transmission technology in the last-mile is DSL. Pass-by is the number of houses connected to the internet by the ISP in hundreds of millions. Standard errors coming soon.

Table 1.9: Choice set stats

Market	Fraction	Cumulative Fraction
AT&T-COMCAST	0.154	0.154
AT&T-TIMEWARNER	0.098	0.252
COMCAST-VERIZON	0.091	0.343
CENTURYLINK-COMCAST	0.067	0.410
TIMEWARNER-VERIZON	0.053	0.463
AT&T-CHARTER	0.051	0.514
CABLEVISION-VERIZON	0.044	0.558
AT&T-COX	0.030	0.588
HUGHES	0.026	0.614
CENTURYLINK-COX	0.026	0.639
RCN-VERIZON	0.019	0.658
COMCAST-FRONTIER	0.017	0.675
CENTURYLINK-CHARTER	0.016	0.691
FRONTIER-TIMEWARNER	0.016	0.707
AT&T-SUDDENLINK	0.013	0.721
AT&T-BRIGHTHOUSE	0.013	0.734
CENTURYLINK-TIMEWARNER	0.013	0.746
COX-VERIZON	0.012	0.759
AT&T-WIDEOPENWEST	0.012	0.771
BRIGHTHOUSE-VERIZON	0.011	0.782

Table 1.10: Conditional means of demographics from the ACS

	Tech.	Income	HH Size	Speak English	Male Head	Age of Head
Uncond.	—	89.220	2.193	0.970	0.591	51.860
Internet	No	43.774	1.789	0.939	0.512	61.543
	Yes	97.801	2.489	0.979	0.614	51.098
Tech	Cable	100.704	2.460	0.979	0.606	50.446
	Dialup	63.112	2.067	0.975	0.590	61.992
	DSL	90.060	2.531	0.979	0.625	52.297
	Fiber	121.480	2.607	0.984	0.636	51.684
	Mobile	79.529	2.495	0.970	0.602	48.401
	Satellite	77.424	2.440	0.981	0.631	54.976

1.10.2 Appendix B: Figures

Figure 1.1: Average Netflix throughput to 21 U.S. ISPs

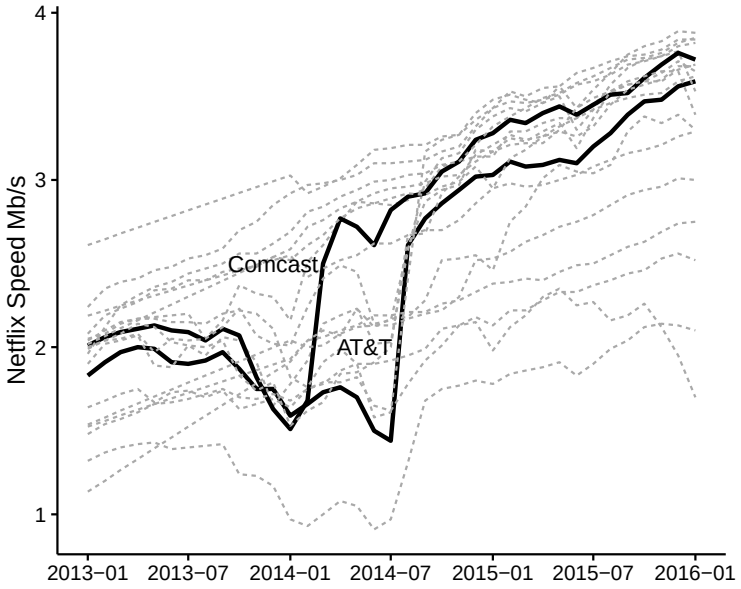
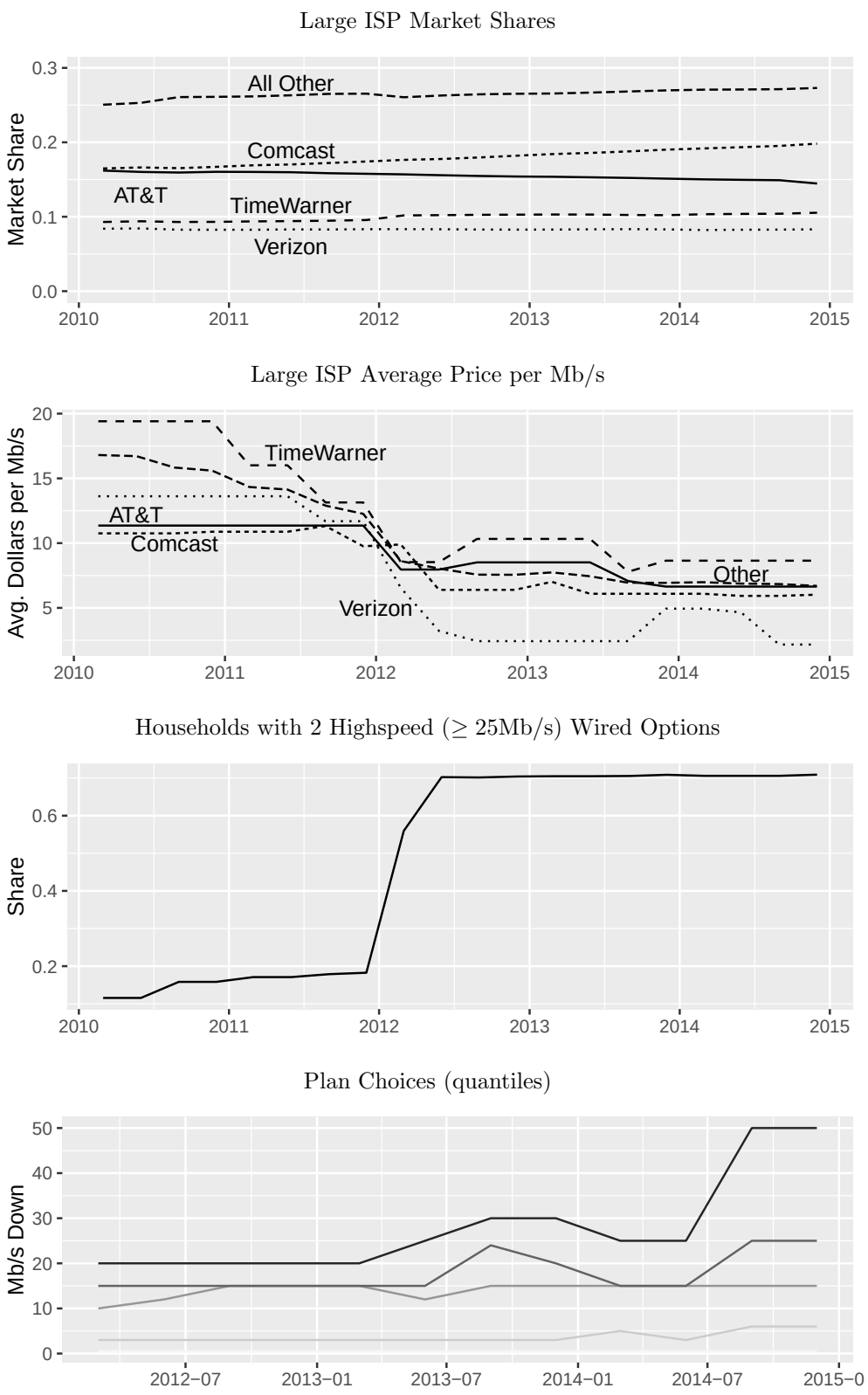


Figure 1.2: Time Variation in Internet Data



The quantile graph shows the 20,40,60 and 80% quantiles for a balanced subpanel of 1913 MBBA participants. The unbalanced panel has substantial attrition, and large additions that change the distribution.

Figure 1.3: Histogram of Disagreement Durations

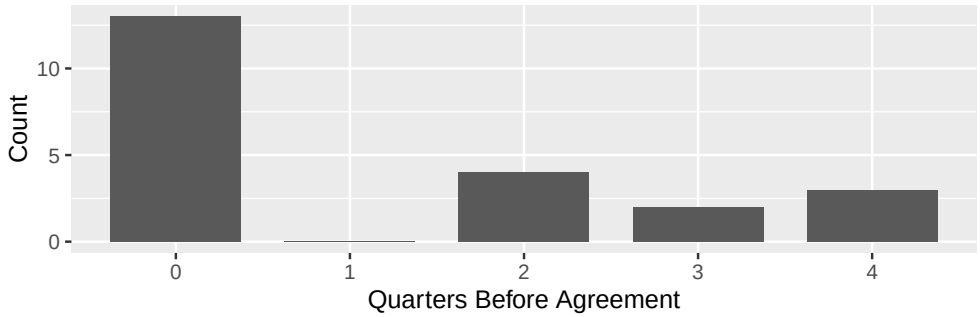
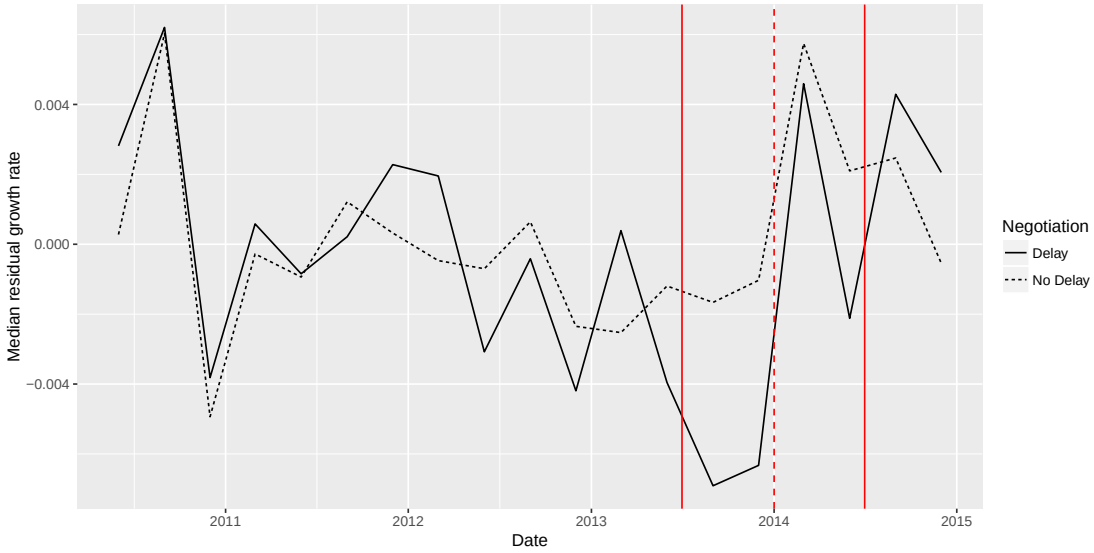
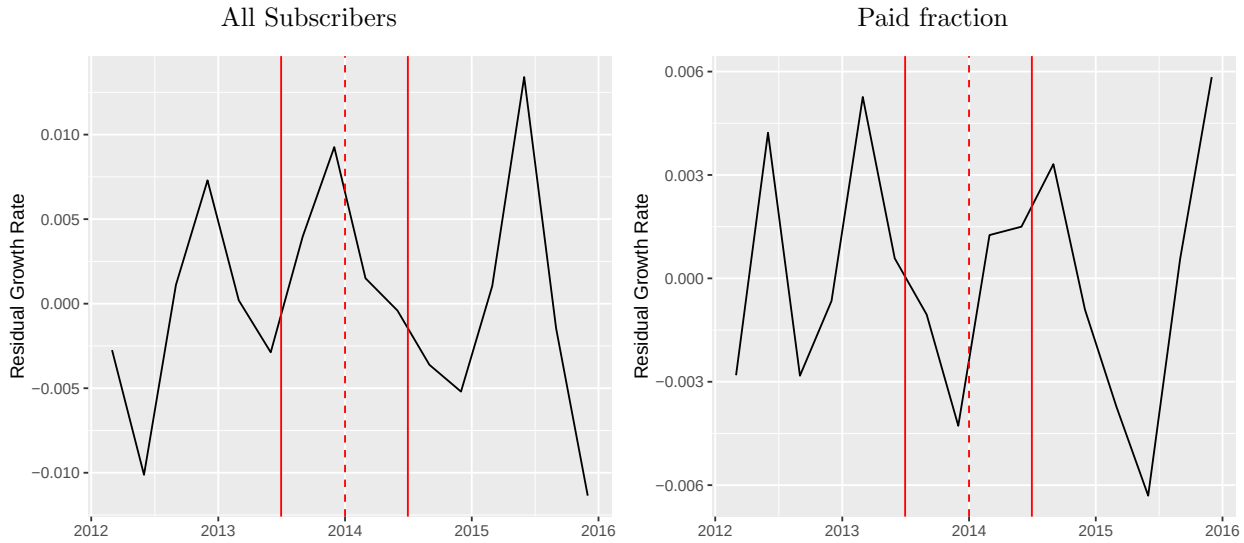


Figure 1.4: Residual ISP subscriber growth rates by whether negotiation occurred



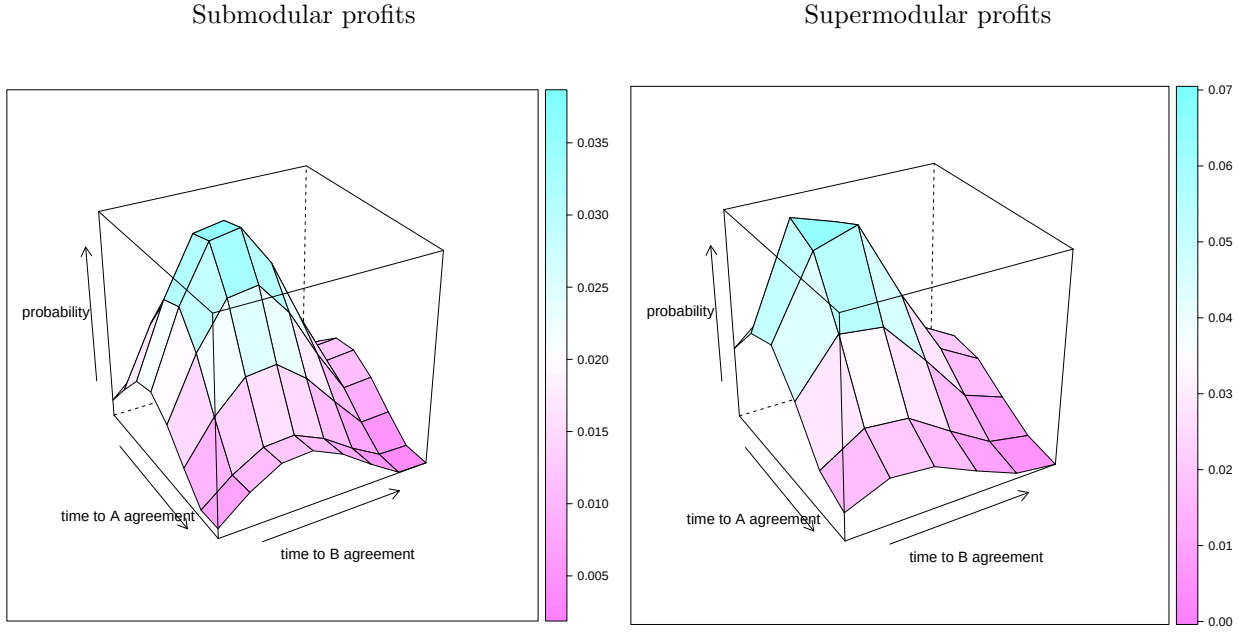
Growth rates are purged of variation due to price changes, mergers, and ISP specific time trends before the median is taken within the group that bargained and the group that did not. The dashed vertical line is the time by which most ISPs agreed, and the vertical lines indicate the beginning and end for all ISPs.

Figure 1.5: Netflix subscriber growth rates



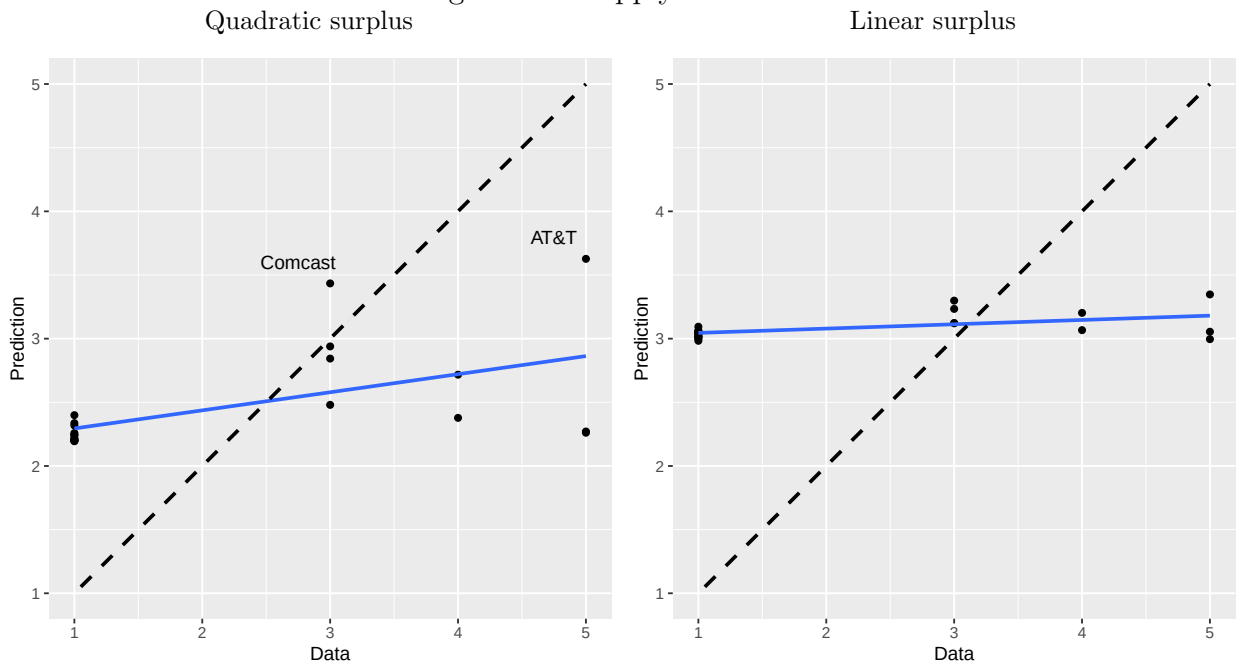
The growth rate in both graphs is detrended. There is no substantial reduction in subscriber growth during the quality slowdown. There is a slight reduction in growth in the paid fraction of subscribers corresponding to the slowdown, suggesting Netflix may be offering more free trials during this period.

Figure 1.6: Joint distribution of agreement timings under different profit interactions



Whether the profits of competing ISPs are sub or supermodular in the agreement of their competitors has a larger effect on the shape of the joint distribution of agreement timings. With supermodular profits, the probability of a joint agreement timing is much higher compared to the case with submodular profits.

Figure 1.7: Supply Model Fit



Disagreement times in the data are on the x axis, while predicted average disagreement times from the model are on the y axis. If the model were performing perfectly, then the points would still not fall along the 45 degree line due to the random error component in the actual surplus. The leftmost graph includes a term in the surplus distribution that increases in the quadratic of subscribers and improves the fit of the model, especially for the largest ISPs.

Figure 1.8: Bargaining Delay Changes

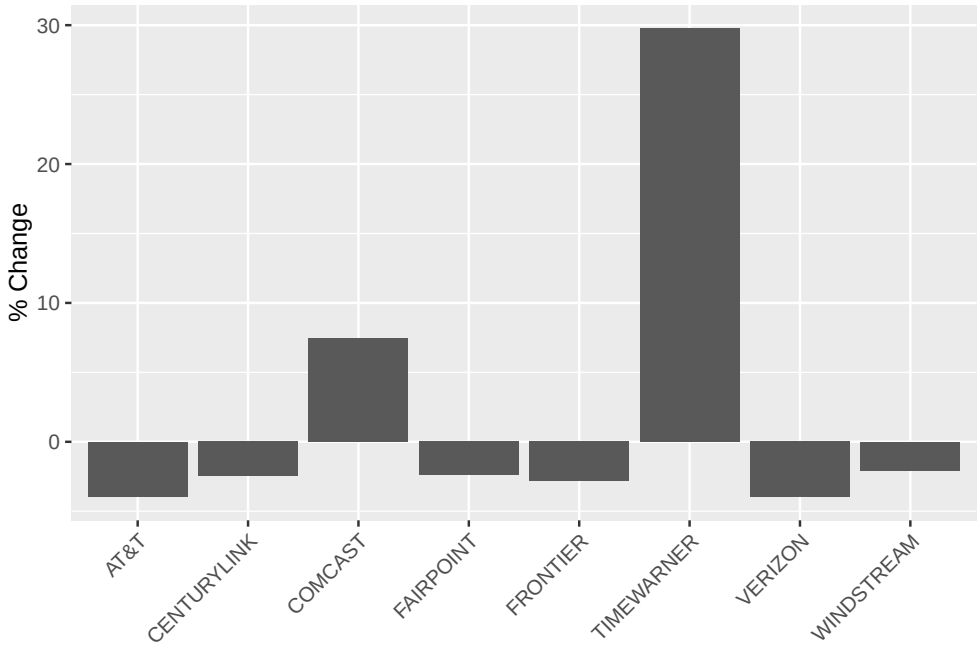
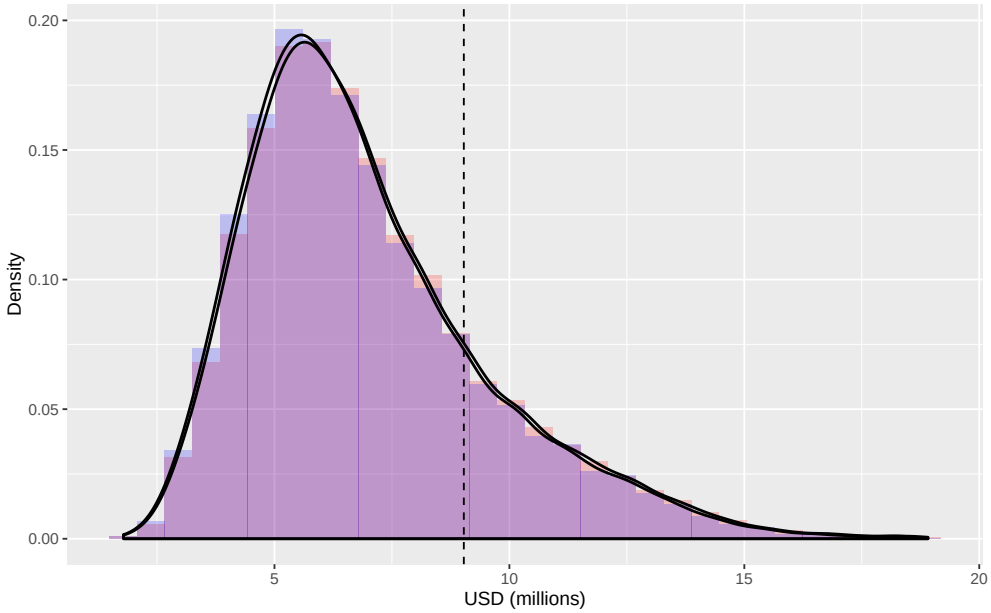
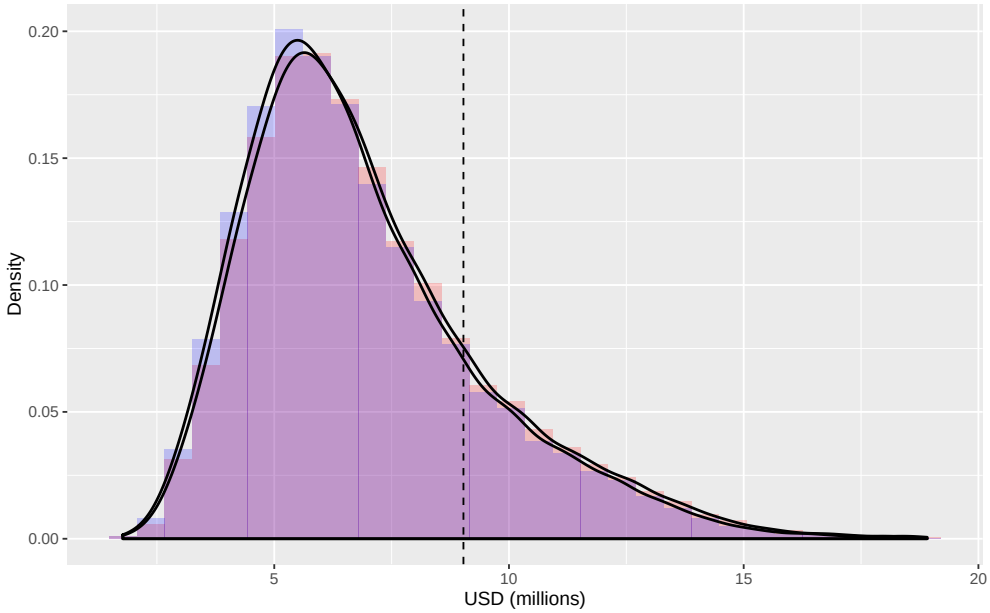


Figure 1.9: Distributions of Netflix Surplus Net of Transfers



The blue histogram is the distribution of aggregate net surplus accruing to Netflix under the merged regime, the red histogram is the distribution under the base regime. Distributions are recovered by sampling the vector ζ_f $R = 10000$ times and deriving the surpluses net of transfers from ISP optimal strategies under each regime.

Figure 1.10: Distributions of Netflix Surplus Net of Transfers



The blue histogram is the distribution of aggregate net surplus accruing to Netflix under the regime with no quality degradation during bargaining, the red histogram is the distribution under the base regime. Distributions are recovered by sampling the vector ζ_f $R = 10000$ times and deriving the surpluses net of transfers from ISP optimal strategies under each regime.

1.10.3 Appendix C: Data Construction

I draw on Netflix's throughput data, as well as MLab measurement data, CAIDA data on interconnection, and Netflix qualitative business filings data to argue that (1) quality reductions were for business reasons, not technical ones and (2) the duration of quality reductions corresponds to the duration of bargaining over interconnection.

A key document in my analysis is the public (partially redacted) version of Netflix's "Petition to Deny". The Petition to Deny is a legal document filed by Netlix to the Federal Communications Commission to argue against the application for merger of Comcast and Time Warner that was announced in mid 2014. In making its case for why these entities should not merge, Netflix detailed the difficulties it had during the installation of its Open Connect servers in 2013. Perhaps because its goal in the Petition to Deny was to argue the dangers of too much market power, only the four largest ISPs are explicitly named in the document. While I take statements in the document about agreement timings at face value,

statements that the ISPs are to blame for the slowdown will need to be evaluated.

Ruling Out ISP Technical Difficulties

From the Netflix data it is clear that some ISPs experience a slowdown and others do not. However, it is possible that affected ISPs were suffering from general network problems and not Netflix-specific problems. I falsify this hypothesis with data from MLab, an independent research group that measures the rate of transmission of data between last-mile ISPs and transit ISPs. Since Netflix primarily uses the service of the transit ISP Cogent during this time, if slowdowns occur between Cogent and the affected ISPs but not between other transit providers and the affected ISPs, then it will be evidence of a Netflix-ISP specific problem. Figure 1.11 illustrates that starting in July 2013, there was a sharp drop in Netflix streaming quality to Comcast, Verizon and TimeWarner. Throughput of other content to these ISPs was not affected, and throughput of Netflix to ISPs such as Cox—a large provider with roughly 5.5% share of subscribers nationwide—was also not affected. Although the MLab data is not comprehensive in its coverage of ISPs, it strongly suggests that slowdowns occurred pairwise. Combined with the qualitative data cited in Section 1.2.2, it is clear that the slowdowns were business driven.

Assigning Responsibility for the Slowdown

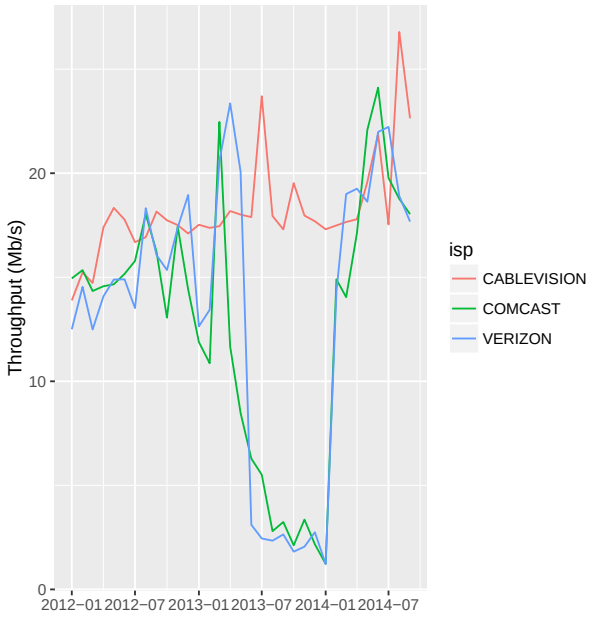
Although Netflix's standing offer to install Open Connect servers was not taken up by the largest ISPs until 2013, transmission of Netflix content to the end users remained reliable until mid-2013. At that point, Netflix, the last-mile ISPs, or both Netflix and last-mile ISPs either actively precipitated a collapse in reliable transmission, or failed to proceed with status quo network upgrades.[31] The end result was that quality of service plummeted for the largest ISPs.

From Netflix's point of view, it is the ISPs that were at fault: Netflix argues that Comcast, Verizon, Time Warner and AT&T "presumably made the business decision that the present discounted value of benefits from degrading the quality of the Netflix video stream to [their]

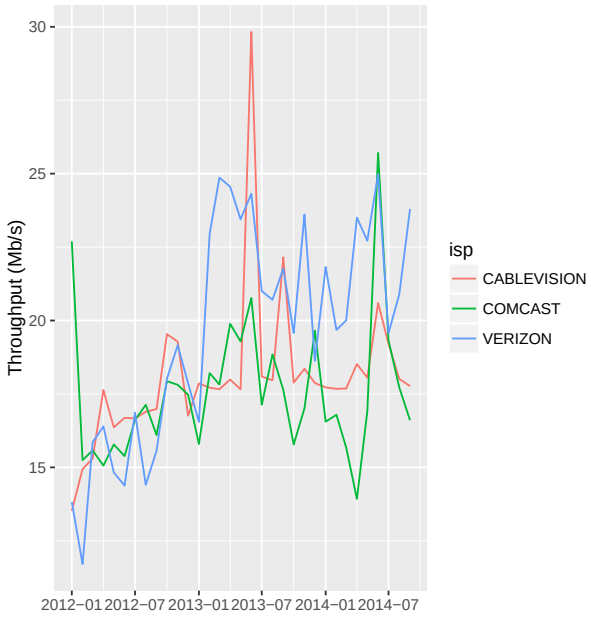
[31] Harvard Business School case N9-616-007.

Figure 1.11: ISP-Netflix Pairwise Throughput in NYC, LA and DC

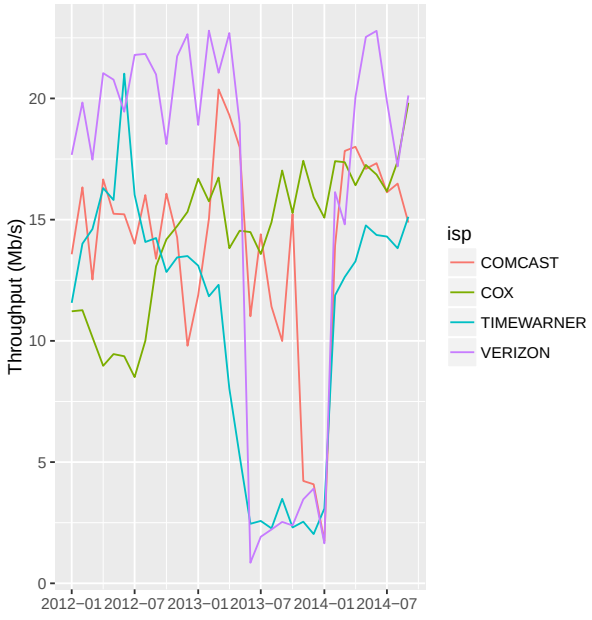
Netflix Throughput, NYC



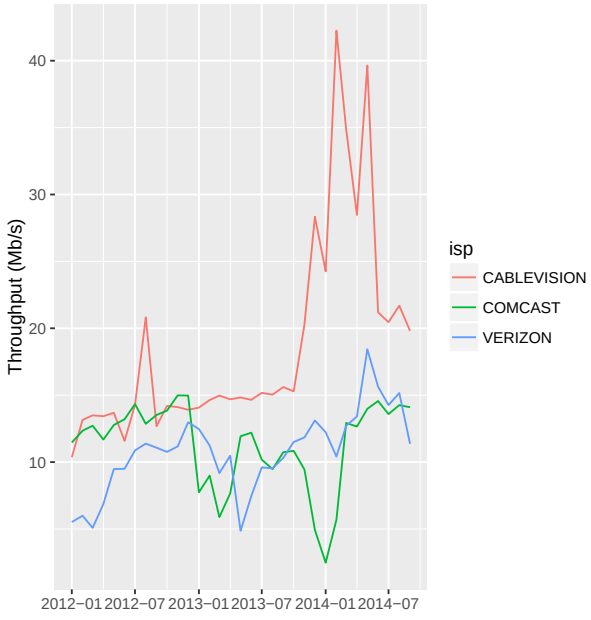
Non-Netflix Throughput, NYC



Netflix Throughput, LA



Non-Netflix Throughput, DC



subscribers was greater than the present discounted value of the costs." [32]

However, from CAIDA data, Netflix cancelled service with a crucial third party in late 2013. Limelight was a large CDN that Netflix had relied on to smooth delivery of its services, but after November 2013—the low point of quality degradation from Figure 1.1—they no longer interconnected. Meanwhile, Comcast claimed that Netflix's throughput problems could be solved if they purchased more bandwidth from transit providers—a statement reported and dismissed by Netflix.[33]

In the model, I show that Netflix had the incentive to degrade quality of service to induce faster agreement times from ISPs. I also show that some ISPs had positive *ex ante* expected payoffs from the dispute even with quality degradation. However, since the marginal consumer appears to be more elastic with respect to switching ISPs than canceling Netflix, there is no incentive for ISPs to slow down traffic.

Constructing Durations

I assume that bargaining begins for all ISPs in 2013Q3, corresponding to the sharp drop off in quality in Figure 1.11.

The duration of disagreement is constructed in the following steps:

1. Using the Netflix data in Figure 1.1, for each ISP for which there is data, I regress throughput on a linear time trend and code a "slowdown" dummy as 1 if throughput falls below 80% of its predicted value. This is a necessary condition for ISPs to be considered as having a lengthy slowdown, and provides candidate disagreement durations.

2. If an ISP is explicitly mentioned in Netflix's Q-10 or Petition to Deny Filings as having reached agreement or not by a certain time, I adjust the disagreement durations to reflect this (Comcast, TimeWarner, and Verizon reach agreement in 2014Q1, while AT&T does not. No other ISPs are mentioned.)[34]

[32]Petition to deny, pg. 52, paragraph 2.

[33]Statement of Ken Florance, Vice President of Content Delivery at Netflix since 2012.

[34]In its first quarter letter to investors issued on April 21, 2014, pg. 5 paragraph 3, Netflix notes "now nearly all cable Internet households receive great quality Internet video", implying that Time Warner Cable also concluded negotiations by the first quarter of 2014. From the same document, Netflix mentions the

3. If any remaining ISP appears in the CAIDA data as interconnecting with OpenConnect at a certain time, I adjust the disagreement data to reflect this timing

4. All remaining ISPs are those that did not experience a quality degradation. Netflix pursued a policy of installing Open Connect infrastructure in the networks of even small ISPs, so I assume that these ISPs reached agreement with Netflix immediately.[35] CAIDA data indicates that some small ISPs reached agreement earlier—for instance, RCN interconnects with Open Connect in late 2012. I assume that ISPs that agree immediately do so in the first period of bargaining in 2013Q3.

The full set of derived durations is displayed in Table 1.11. I include Hughes and Wildblue, the two satellite operators in the sample, but I do not assume that these providers bargain with Netflix over interconnection since their infrastructure is not amenable to Open Connect. There is no slowdown vis-a-vis these providers however, so their bargaining duration is set to zero and they are not used in the supply estimation.

extremely poor streaming quality that AT&T U-Verse customers receive, and argues that "[it] is free and easy for AT&T to interconnect directly with Netflix and quickly improve their customers' experience, should AT&T so desire.", implying that the slowdown could be alleviated as soon as Netflix and AT&T could settle on a price. From this report it is clear that AT&T actually took longer to resolve negotiations, so that the slowdown truly indicates negotiation time and not just greater technological difficulties implementing interconnection with Open Connect.

[35]"if an ISP has an individual market area serving a population of at least 100,000 subscribers, Netflix will install Open Connect appliances at that location at no charge to the ISP.", pg.49, paragraph 2, Netflix Petition to Deny.

Table 1.11: Complete Durations Data

ISP	Duration	ISP	Duration
AT&T	4	FRONTIER	3
BRIGHTHOUSE	4	HAWAIIAN TELCOM	0
CABLE ONE	0	HUGHES	0
CABLEVISION	0	MEDIACOM	4
CENTURYLINK	3	RCN	0
CHARTER	2	SUDDENLINK	0
CINCINNATI BELL	0	TDS TELECOM	0
COMCAST	2	TIMEWARNER	2
CONSOLIDATED	0	VERIZON	2
COX	0	WIDEOPENWEST	0
EARTHLINK	0	WINDSTREAM	0
FAIRPOINT	0	FRONTIER	3

Chapter 2

Competition and Product Misrepresentation

2.1 Introduction

When making consumption decisions, consumers often do not perceive the true characteristics of a good until they actually purchase it. Ex ante, consumers rely on descriptions of product attributes supplied by firms. Recent work on cheap talk models of advertising has shown that when a product's features are unknown before consumption, firms may have an incentive to claim their product is higher quality than it actually is to attract consumers (Gardete, 2013). This incentive to misrepresent product quality is particularly strong in industries where there is little chance of repeat business or where switching costs are high (Cai and Obara, 2009). Two natural empirical questions arise from the theory: first, do firms actually strategically misrepresent their product characteristics? And if so, in the absence of third-party regulation or monitoring, will competitive forces in the market discipline firms' product claims?

To understand the relationship between claimed and actual product quality, I use a panel dataset of consumer purchases of broadband internet service. The download speed of an internet plan is one of its most important features, but whether an internet service

[0]Thanks to participants at the 2017 CEA Annual Conference for valuable feedback on this project, where it was presented as "Competition and Unobserved Product Quality."

provider (ISP) actually delivers the speed that is paid for is not known until the consumer actually signs the contract and begins service, at which point the cost to cancel service can be substantial. The panel data provides both the nominal speed the ISP claims to be providing as well as the measured speed that is actually being provided, which allows me to directly measure how the difference in claimed and actual product quality responds to covariates.

The paper's main empirical contribution is to recover the effects of competition on the ratio of consumers' actual to claimed download speed. This ratio, henceforth referred to as the *promise ratio*, exhibits substantial variation: across consumers and within an ISP-month, the consumer at the 10th percentile of the ratio receives approximately 77% of the speed they pay for, while the 90th percentile consumer receives 117%. In my preferred specification, I show that competition has a substantial and statistically significant effect on the ratio of actual to claimed speed; the presence of an additional broadband internet competitor offering high speed plans (up to 15Mb/s) increases the promise ratio by 23-32%, although with a full set of ISP-month fixed effects the coefficient is significant at only the 10% level. I show that this result is not driven by reductions in claimed speed.

I next turn to the question of whether strategic misrepresentation matters for economic outcomes. In particular, I look at whether ISPs' strategic noisiness in their plan descriptions induces consumers to pick sub-optimal plans, or whether misrepresentation is simply cheap talk and consumers accurately perceive the true qualities. I use two empirical tests of whether consumers were misallocated across plans: whether consumers on plans that were strategically misrepresented change their average consumption of content after the qualities are improved in response to competition, and whether the same consumers change their plan search behaviour. I find that both tests provide suggestive evidence for the hypothesis that firms' strategic misrepresentation led to sub-optimal plan choice.

My empirical strategy to understand the effect of competition on the promise ratio is a difference-in-differences specification. The main threat to identification is that the entry of competitors in an area might be endogenous to the promise ratio. Since a fall in an ISP's promise ratio may lead to an increased willingness of consumers to switch internet providers, competitors are more likely to enter, which would generate a negative relationship between changes in the promise ratio and the number of competitors. To resolve this issue and

establish causation from number of competitors to the ratio, I use the fact that a provider of satellite internet brought a broadband-capable satellite online during 2012. The launch provides an exogenous increase in the number of competitors. While nominally the launch may have provided an additional competitor for every household, some areas may not be suitable for satellite internet, due to geography, weather, or characteristics of local housing stock. I use the presence of existing satellite customers in a county in 2010 and 2011 to construct cross-sectional variation in which counties are affected by the launch; the intuition is that counties with relatively large ex ante share of satellite customers are the only areas suitable for satellite internet.

An area's suitability for satellite internet depends partly on covariates that are exogenous to short run shocks to local firms' promise ratios. I leverage two sources of cross-sectional variation orthogonal to short run shocks to instrument for whether a county is treated by the satellite launch: variations at the county level in severe weather events, and variations in the housing stock—in particular the fraction of manufactured/mobile homes—across census block groups (Boik, 2017). The instruments are predictive and have intuitively correct signs in the first stage, and increase the estimates of the effect of the satellite launch on the promise ratio by an order of magnitude. I also validate the weather instrument by showing that thunderstorms indeed have an effect on internet quality using high-frequency testing data.

I use the same instrumented difference-in-differences empirical methodology as in the main regression to evaluate the implication of firm misrepresentation for consumer outcomes. In my first specification, I find that consumers at ISPs that improve promise ratios after the competitor enters substantially increase their consumption during peak times. If consumers did not increase consumption in response to the increase in real speeds it would suggest that the initial misrepresentation did not prevent consumers from achieving their optimal consumption profile. Increasing consumption after the treatment is a necessary, but not sufficient, condition for there to have been misallocation, so the evidence is suggestive that ex ante misrepresentation induced misallocation that was ameliorated by competition. In the second specification, I find that treated consumers switched plan speeds less frequently after treatment; intuitively, after controlling for payoff-relevant changes in observable plan

characteristics, the remainder of switching partly reflects costly experimentation. Substantial reductions in switching provide evidence that consumers more easily identified their optimal plan ex post.

The remainder of the paper proceeds as follows. I first discuss the industry background, the literature, and the research questions I will take to the data in Section 2.2. I then explain the industry background, the data and the construction of my variables in Section 3.2. Section 3.5 covers the estimation and analysis, and Section 2.6 concludes.

2.2 Literature and Research Questions

Policy makers have long worried whether firms sufficiently inform consumers about payoff-relevant product attributes. In the U.S. especially, federal agencies such as the Federal Department of Agriculture and Federal Aviation Authority have broad mandates to enforce mandatory disclosure of product attributes ranging from calorie content to airline on-time performance. In the absence of regulation, whether firms will disclose information about their products, what they will choose to disclose if they do, and how the incentives to disclose interact with market structure, have all generated substantial debate among economists but only limited validation in the data, necessitating the empirical approach taken in this paper.

Theoretical work on how competition affects the incentive to honestly signal product attributes offers mixed predictions. There is a substantial literature on quality disclosure, where firms can choose to either honestly and credibly announce their quality or else not provide any information at all. This scenario is a subset of the case I examine, where firms can also choose to misrepresent the attribute. A stark result from the early literature focusing on quality disclosure in monopoly (see Grossman (1981) and Milgrom (1981), as well as Jovanovic (1982) and Stiglitz (1975)) is that in the absence of disclosure costs any firm, regardless of quality, will choose to honestly disclose for strategic reasons. The argument is simple: a firm with quality greater than the mean quality of firm types that choose not to disclose will find it improves profit to disclose; iterating on this logic, only a firm with the very worst quality will then be indifferent between disclosing and not. Recent papers such as Board (2009) have shown that competition can increase or reduce the incentive to disclose;

intuitively, disclosure may allow firms to benefit from differentiation. When the unknown attribute is the price, Ellison and Wolitzky (2012) show that strategic obfuscation of prices is optimal in oligopolistic competition with search costs, while in the model of Gabaix and Laibson (2006) the incentive to suppress pricing information about product add-ons is non-monotonic in the level of competition.

The empirical literature on quality disclosure has typically found that firms are strategic in disclosure, and that competition actually reduces the incentive to disclose.[1] Mathios (2000) finds that lower quality firms profitably obfuscate their attributes: prior to mandatory nutrition labelling, low-fat salad dressings advertised as such while high-fat dressings do not advertise, and after labeling high-fat dressings experienced a relative loss in sales. Zhe Jin (2005) shows that greater competition between HMOs reduces voluntary disclosure and leads to greater obfuscation of qualities. Misrepresentation of pricing is profitable even in a fairly competitive online environment: Brown et al. (2010) empirically explore the conditions under which shrouding a particular attribute of a good—the shipping price, when the good is sold through eBay—is profitable for sellers; Ellison and Ellison (2009) find evidence that obfuscation of prices by an online retailer raise profits by making consumers less informed.

There is a recent strand of empirical work closely related to this paper that examines how product misrepresentation changes with market structure and the regulatory environment. This work focuses on the airline industry, where the true quality of a flight is its actual flight time, while the quality the firm represents to consumers is the predicted flight time. Forbes et al. (2011) show that misrepresentation is strategic in the airline industry: airlines with the incentive to do so adjust on-time performance for flights whose late time would exceed a threshold, in order to reduce the proportion of late flights and increase their ranking. In a more recent paper, Forbes et al. (2017) show that airlines increase reported performance by lengthening scheduled flight times on routes where they compete with entrants who also make quality reports. In other words, in response to competition airlines do not decrease flight length, but only increase *projected* flight length, resulting in more on-time or ahead of schedule flights. This idea that competition can lead firms to costlessly claim lower quality and thereby better the "optics" of their actual service finds a counterpoint in Zhe Jin

[1] See Dranove and Zhe Jin (2010) for an overview of the literature.

(2003), where mandatory disclosure in competitive markets leads restaurants to actually improve their hygiene scores. This motivates the central research question of this paper: does competition lead firms to disclose download speed more honestly, or to misrepresent by claiming even faster speeds to grab consumer attention? And if disclosure becomes more honest, do firms revise their claimed download speed downwards or invest to increase the actual speeds?

The second research question this project focuses on is whether product misrepresentation leads to consumer misallocation across internet plans. A recent strand of empirical literature on consumer choice has focused on whether consumers make *ex ante* correct choices of subscription plans—be they health plans, cellphone plans Björkegren (2015), etc.—given their observed usage. The idea is that since most plans entail both an upfront tariff and a usage fee, one can back out whether a consumer has chosen the wrong plan by observing if they would have spent less money by choosing a different plan given their observed usage of plan services. Nevo et al. (2016) find that within the ISP they have data from, most consumers choose the lowest-cost internet plan given their observed data consumption. However, their research is not focused on explaining variation in the fraction of consumers making the "wrong" choice. Whether or not changes in misrepresentation due to competition change the degree of misallocation is an important question for policymakers interested in non-standard welfare effects of mergers.

2.3 Data

Since 2012, the Federal Communications Commission (FCC) has been collecting information at a household level to track ISP service quality as part of its Measuring Broadband America (MBA) program. The most important number the FCC provides statistics on is what I have termed the *promise ratio*: the average ratio of consumers' actual to purchased speeds across households at the same ISP.[2] The FCC updates information on the distribution of promise ratios at each ISP every year to help consumers make more informed choices about their internet options. The FCC emphasizes the promise ratio in its releases partly because

[2]See `https://www.fcc.gov/general/measuring-broadband-america`

the majority of consumers rate both reliability and download speed as the most important features of an internet plan. Prior to this program, there was no government data for tracking whether ISPs were providing the speeds they claimed; however, there were third-party testing sites that measure households' speeds, allowing a household to verify its actual internet quality but only after signing up.

The main dataset is an unbalanced panel of MBA testing data. Units are observed for a maximum of four months spread over 3 years: April 2012, and September 2012, 2013 and 2014. Within each month, the plan that the consumer is subscribed to is verified with the ISP, and measurements are taken every two hours of the actual download speed, upload speeds, latency, video streaming rates and connection speeds to various popular websites (in later samples). The ratio of the measured download speed to the ISP-verified download speed is the main dependent variable of interest.

Using the FCC testing data, I first divide actual speed by claimed speed for each unit in each hour during which measurements are taken. I then trim the 1% and 99% quantiles to remove outliers. Finally, since the measurements are all taken for residential broadband subscribers and most usage occurs between 7pm and 11pm, I average the promise ratios across these peak hours within a month. This average forms the main dependent variable, and varies over time within a household across months during which the household is part of the sample. In Figure 2.1 I plot the empirical distribution of the promise ratio in the first month of the sample (April 2012) for 4 different ISPs: each ISP has a substantial fraction of households receiving less than the promised speed, but this fraction varies across ISPs.

I use the Current Population Survey's (CPS) Internet Use Supplement to recover the fraction of individuals subscribing to satellite internet in 2010 and 2011. This fraction will form the basis for constructing the cross-sectional dummy variable indicating which counties are affected by the new satellite launch. While satellite internet was available prior to the launch of the new satellite in mid-2012, the maximum advertised speed was 1.5Mb/s and actual speeds were closer to 0.8Mb/s—not high enough to stream video smoothly.[3] To construct the treatment dummy, I classify treated counties as those in the 75th percentile or higher of satellite subscribers in 2010 and 2011. Figure 2.2 shows the distribution of the

[3]Measuring Broadband America report 2012

fraction of satellite subscribers across counties, in counties for which I have data.

To the panel data and treatment data I add supplementary data from the National Broadband Map (NBM). The NBM provides geographic data, measured half-yearly between 2010 and the middle of 2014, on which ISPs are available in each census block and on what speeds those ISPs offer. By matching units in the FCC panel to their census block, I can control for time-varying measures of terrestrial competition for each unit.

There will be two instruments for the treatment dummy: the fraction of houses that are prefabricated in a census block group, and the frequency of severe weather events. Information on the housing stock, including the fraction of mobile homes in a census block group, is recovered from the American Community Survey (ACS) 5-year sample. Information on the history of severe weather events at a county level comes from the National Oceanic and Atmosphere Administration (NOAA). The severe weather instrument is then constructed as the fraction of days during which the county is affected by thunderstorm winds. Finally, data on the daily quality of internet providers at a city level for 2010-2014 comes from Ookla, a large, free, online internet speed testing website. I use this data to verify that severe weather does indeed have an effect on the quality of wireless internet.

Summaries of variables for each of the four months that units are observed is available in Table 2.1. Note that there is a steady increase over time in both the promise ratio and claimed download speeds. The maximum advertised download speed is also increasing but not by as much as the actual download speeds individuals subscribe to. The average number of terrestrial competitors for a household barely moves over this time period—most individuals have a choice between only two providers—but there is a roughly 10% increase in the average number of competitors offering very fast internet by the end of the sample. Sample means of quantities such as housing density, the fraction of housing comprised of mobile homes, and the median house age that only vary cross-sectionally, will vary over time in Table 2.1 due to selection in the set of units that are in the sample as units drop out and the panel is refreshed.

2.4 Modelling and Estimation

The main specification for understanding the effect of the entry of the satellite competitor on the promise ratio is

$$y_{ijct} = \sum_{t>1} \delta_t \text{Treatment}_c \times D_t + \beta' X_{ijct} + \alpha_i + \alpha_{jt} + \epsilon_{ijct}, \tag{2.1}$$

where i indexes a household, j an ISP, c a county and $t = 1, \ldots, 4$ one of the four months a household can be present in the sample.

The coefficients of interest are the δ_t, which measures the effect of being in a county exposed to the new satellite competitor at different time windows after the launch. y_{ijct} is the log of the promise ratio, and Treatment_c is a dummy taking a value of 1 if the consumer is in a county where satellite is a plausible option. The specification allows for time varying, ISP-specific changes in the promise ratio and variation in the average promise ratio across units. X_{ijct} includes the number of terrestrial competitors a unit has access to besides the ISP it is subscribed to, as well as the claimed download speed a unit is paying to receive.

As the literature makes clear, the sign of δ_t is an empirical question. In Forbes et al. (2017), airlines increase their predicted flight time in response to competition while not improving the actual flight time; in my setting, the equivalent response would be to reduce the claimed speed consumers are receiving while maintaining the actual speed, leading to an apparent improvement in the promise ratio from competition but no actual change in quality. If some consumers are "naive" and take the claimed download speed to be true, then ISPs may have even greater incentives in the presence of competition to exaggerate and induce consumers to switch. Finally, ISPs may also have the incentive to upgrade the quality of their service, improving the actual speed while maintaining claims. I will therefore run an auxiliary regression of the same form as Equation (2.1) where the dependent variable is the claimed download speed.

2.4.1 Identification

Insofar as the differential change in the mean promise ratio in treated counties can be attributed to a different composition of ISPs—who may be upgrading their national network differentially—the α_{jt} dummy will control for it. Moreover, since only within-household variation in the promise ratio is used, the effect within a treated county of treatment on the promise ratio cannot be attributed to consumer-side selection (e.g., if only lower promise ratio consumers were dropping service to switch to satellite). That is, neither firm-side or consumer-side differences in ex ante distribution across treated and untreated counties should pose a threat to identification of the competition effect.

In the background of the reduced form specification is an entry game: ISPs enter a small geographic region (a census block, or even a housing unit), choose how much to invest to provide a certain quality, and supply information—potentially false—about what quality can be provided to consumers. Firms also experience local shocks in their ability to maintain their promise ratio. Meanwhile, consumers can switch between providers, upgrade their nominal speed by paying more, or both.

The satellite-treatment methodology is intended to address a particular source of endogeneity: that of reverse causality from changes in the promise ratio to entry where entrants react to negative shocks to a competitor's ability to provide service by entering the market. The satellite launch is not endogenous to any local shocks. However, the ex ante fraction of satellite internet subscribers, which is used to construct the treatment dummy, may be endogenous to county level shocks. Ex ante difference of means tests across treated and untreated counties in the log promise ratio, number of competitors, number of competitors offering more than 50 Mb/s, and the maximum advertised download speed are presented in Table 2.2. The table suggests there are no ex ante difference in key observables between treated and untreated counties; however, the tests do not rule out differences in unobservable local shocks across counties, which I address below.

If there were ex ante unobserved negative shocks to competitors' promise ratios in some counties, then individuals may have substituted into satellite, leading these counties to be classified as treatment counties. There are then two options: if the shocks are persistent,

then it implies treated counties' non-satellite providers are on a downwards trajectory. Any improvement in their promise ratios after treatment would therefore be a lower bound on the actual treatment effect. If the shocks are not persistent, then any improvement in treated counties' promise ratios may simply be mean reversion. To address this possibility, I instrument for the treatment dummy using variables that are likely not correlated with ex ante negative shocks to competitors' promise ratios, but which determine the viability of satellite in a county. The instrumentation strategy also has the benefit of dealing with attenuation bias, which may arise because of mismeasurement/random non-response in the CPS.

The first instrument is the historic frequency of severe thunderstorms in a county. The location and duration (in days and hours) of severe weather events are reported by NOAA and come in many forms: tornadoes, thunderstorm wind, hail, etc. Cloud cover and heavy, prolonged rain are considered by experts to interfere substantively with wireless internet signals. In Table 2.3, I use data from Ookla Speedtest.net on to show that thunderstorms have a differential negative effect on several measures of internet quality for wireless versus fixed internet.[4] In particular, there are significantly more tests run during thunderstorms for wireless internet subscribers and latency—a measure of the responsiveness of real time interactions including gaming and voice calling—also increases differentially for wireless internet. In Table 2.4 I use testing data that does not provide measures of responsiveness, but which has more observations, download and upload speeds, and includes satellite internet subscribers. From this table it is clear that during thunderstorms, wireless and satellite internet both suffer a differential hit to both their download and upload speeds, with a larger magnitude decrease for satellite internet (although the coefficient is not significant.)

Counties with historically high storm activity are therefore less suitable for satellite simply because they will experience longer, more frequent service disruptions. Since the reliability of internet service is reported by consumers to be the most important feature of a plan, severe weather helps determine which counties are affected by the entry of the broadband satellite, while also being exogenous to transient shocks to fixed broadband providers'

[4]A disclaimer for this table is that since there are very few observations in the testing data for satellite internet, I rely on the effect of weather on wireless broadband (e.g., 3G and 4G) as a proxy.

promise ratios.

The second instrument is the fraction of housing stock comprised of mobile and manufactured homes. In Boik (2017), the author finds that "households dwelling in manufactured/modular homes, conditional on the value of the home, are 9% less likely to adopt wired broadband ... [this] result cannot be explained by these households having lower disposable income to spend on internet access because such households are 20-67% *more* likely to adopt satellite broadband compared to households in non-manufactured homes, even in regions where there is no wired broadband provider." The result that households in mobile homes are much more likely to subscribe to satellite due to a higher fixed cost of setting up wired broadband provides cross-county variation in satellite suitability. Characteristics of the housing stock do not change quickly, and so are likely exogenous to the types of transient shocks to competitors' promise ratios that have the potential to bias the treatment estimate.

2.5 Results

2.5.1 Promise Ratio

Regression results from the main specification Equation (2.1) are provided in Table 2.5. The first three columns provide results from standard fixed effects regressions without instrumenting for the treatment; these models suggest that the effect the satellite competitor is to raise the promise ratio of incumbents by about 1.7%. When the treatment dummy is instrumented for by the weather and housing stock instruments the estimate of the treatment effect increases by an order of magnitude. With date fixed effects, the entry of the satellite competitor increases the promise ratio of households at incumbent ISPs by 23.0%, while with an ISP-date fixed effect the treatment effect is 31.7% although it is significant only at the 10% level with standard errors clustered at the county level. Effects are significant for the time period immediately following the launch in September 2012, but attenuate by September 2013 and 2014.

The regression specification takes care of selection at the household and ISP level, and the instrumentation strategy should deal with unobserved shocks to the promise ratio that may

be correlated with the treatment dummy. These estimates provide an unbiased estimate of the effect of competition on the promise ratio; as competitors enter, firms reduce the negative misrepresentation of their products, but only in the short run.

I provide additional support for the idea that the improvement in the promise ratio is due to competitive pressures by running the regression separately for households exposed to more and less ex ante competitive environments. In Table 2.6, I show in the third and fourth columns that for households who already have an outside option offering speeds greater than 50 Mb/s, the introduction of the satellite competitor does not significantly affect the promise ratio of the incumbent; meanwhile, for households that do not have a high-speed alternative, there is a 26.0% improvement in the promise ratio.

Since the promise ratio can be improved by either reducing the claimed speed or by increasing the actual speed, I provide two auxiliary regressions where claimed download speed and actual download speeds are the dependent variables in Table 2.7. I find that by September 2012, treated areas did not experience any significant differential reduction in claimed download speeds compared to untreated areas; indeed, the sign of the treatment dummy is positive. Treated areas did, however, experience a relative decrease in claimed download speed by September 2013. Meanwhile, treated areas did experience a significant differential increase in actual download speeds compared to untreated areas, and the magnitude of 47.9% is almost triple the size of the magnitude of 17.1% for the increase in claimed speeds by September 2012 in the specification with full fixed effects.

Taken as a whole, the results suggest that the mechanism present in Forbes et al. (2017)—where firms react to competition by changing their claims but not their actual performance—is not operational here. Instead, firms increase their actual performance and do not reduce the claimed speed. This behaviour suggests both that misrepresentation is strategic insofar as it changes in response to market structure, and that changing only product claims in response to competition, while inexpensive, does not dominate actually improving service as a strategy.

2.5.2 Extensions

In this section I turn to the question of whether misrepresentation leads to consumer misallocation across internet plans. I construct two tests: whether treated consumers changed their data consumption after the reduction in misrepresentation, and whether treated consumers switched plans less frequently after the reduction in misrepresentation.

The basic intuition behind the first test is that plan choice entails an extensive and intensive margin: consumers first pick a plan, and then choose how many Gb of data to consume. If a household does not change their consumption after the increase in actual speeds, then they were already consuming at their optimal level, suggesting that they perceived the actual plan speed accurately enough even with the misrepresentation to make the utility maximizing choice. Observing no differential change in consumption for treated households would therefore suggest there was no misallocation. However, if a household does change their consumption ex post, it need not be evidence of misallocation: perhaps the increase in speed simply led to an increase in consumption through substitution effects. A positive, significant effect on consumption for treated households is therefore a necessary but not sufficient condition to conclude that there was misallocation.

In Table 2.8, I find that after instrumentation, consumers in the treatment group exhibit a large, positive effect on consumption compared to untreated consumers in the first observed period after treatment. As discussed above, this provides evidence in favour of ex ante misallocation insofar as we cannot rule it out. However, the effect of treatment does attenuate over the next two periods; combined with the results from Table 2.7 which show that actual speeds decreased in 2013 and 2014, the results do seem to suggest that substitution effects may be responsible for at least some of the changes in data consumption.

The intuition underlying the second test is that for goods with unknown attributes, consumers may need to experiment to find their best internet plan match. If treated consumers experience a larger drop in switching ex post it may suggest the reduction in misrepresentation has allowed consumers to find their optimal plan more quickly. That is, if consumers were able to deduce their optimal plan ex ante even with misrepresentation, then we would expect no difference in their switching frequencies once misrepresentation was reduced be-

cause there was no greater need to experiment ex ante.

To recover switching frequencies, I download auxiliary testing data from the MBA records; in particular, test records from 2 months before and after each verified month. The auxiliary data does not have the verified speeds, so I infer them from actual speeds and extrapolating the average promise ratio in the closest verified month. A household is then coded as having switched plans if their verified download speed changes at all in the 5 month window around the original verified months.

Results are presented in Table 2.9. The regressions do not provide much support for rejecting the null that switching frequencies declined after treatment. While the signs of the treatment coefficients are correct, they are not significant.

Taken as a whole, the results provide weak evidence that households were misallocated ex ante across internet plans. For a consumer wishing to stream video during peak hours, buying a plan with a claimed speed that is sufficient for streaming but an actual speed that is insufficient represents a misallocation. Households where misrepresentation was reduced due to improvements in the speed quickly increased their consumption of data, suggesting that they may have been misallocated in their initial plan choice.

2.5.3 Robustness

In this section I provide a number of robustness checks on the main results having to do with the promise ratio.

First, as shown in Figure 2.1, some ISPs actually provide speeds greater than what they promise to some measure of households. In Table 2.10, I run the same baseline regression but restrict the sample to only households with promise ratios less than 1 for the entire sample. The positive, significant coefficient remains, although the magnitude is decreased to about 20.5%.

Next, to construct the satellite treatment dummy I code regions in the 75th percentile or higher of ex ante satellite subscriber shares as being treatment groups. In Table 2.11, I experiment with changing the cutoff level. The sign remains the same in most specifications, although for lower cutoffs the significance disappears, partly due to lower estimated coefficient magnitudes. This result is not unexpected given that by lowering the cutoff, more and

more counties that were actually not treated are included in the treatment group, so that the average improvement in the promise ratio after treatment is diluted.

Finally, I experiment with different controls for the terrestrial competitors. As shown in Table 2.6, having a high-speed competitor is an important determinant of the promise ratio and has important interactions with the treatment effect. I include various controls in Table 2.12 including the number of competitors that are national brands, and the number of competitors with medium, fast, and super fast download speeds. The inclusion of the different controls does not significantly move the coefficient on the treatment effects.

2.6 Conclusion

In this paper I analyze whether competition leads firms in the market for internet service provision to disclose product attributes more accurately. The paper connects to a rich literature on how competition affects firms' incentives to disclose product attributes; I expand on this literature by considering the case where firms can misrepresent their products by disclosing inaccurate information. I find that the effect of an additional high speed competitor is to raise the ratio of actual to claimed download speed by between 23 and 31 percent. Furthermore, I find that this effect is driven by improvements in the actual speed, and not simply reductions in the claimed speed.

I also provide preliminary evidence on whether firms' choice to misrepresent their products leads to misallocation of consumers across plans. If consumers correctly perceive true attributes even when firms misrepresent then misrepresentation is simply an efficient, cheap-talk equilibrium. I find weak support for the idea that misrepresentation led to misallocation of consumers across plans by showing that their data usage increased in response to the reduction in misallocation and increase in actual speeds.

Substantial work remains to be done on the topic of how market structure affects efficiency through the channel of misrepresentation. Given that Forbes et al. (2017) finds that competition leads to weaker claims but no improvement in service, it is necessary to write a theory model that can accommodate both behaviours—weakening claims or improving service—as equilibria, perhaps depending on parameters. Understanding how strategic mis-

representation can lead to lower investment in service quality is understudied even in the disclosure literature, which typically does not model an endogenous investment in quality.

2.7 Appendices

2.7.1 Appendix A: Tables

Table 2.1: Means by Date

	Apr2012	Sept2012	Sept2013	Sept2014
Number of units	6790	6102	5245	4493
Promise ratio	0.9945845	1.0036044	1.0428606	1.0377242
Promise ratio 7pm-11pm	0.9696025	0.9722008	1.0098240	1.0018899
Download	15.09796	16.93083	23.33167	31.79436
Upload	4.196571	4.447210	5.864187	8.561023
Max adv. down	46.90078	48.08522	55.61642	64.21735
Max adv. up	4.691047	4.706396	7.028915	10.730917
Num. competitors	1.185095	1.189382	1.204331	1.196963
Num. comp. down > 50Mb/s	0.4344138	0.4250468	0.4572835	0.4705476
Housing Density	893.5948	871.8765	748.2968	713.7691
Frac. of Housing Mobile	0.04078837	0.04268021	0.04700761	0.04748228
Median House Age	39.66746	39.58704	38.55341	38.70599

Table 2.2: Ex Ante Differences Between Counties

Variable	Untreated	Treated	T test	p value
Promise Ratio	0.959	0.967	-0.898	0.370
Number of Competitors	1.122	1.136	-0.221	0.825
Num. Comps. >50Mb/s	0.348	0.372	-0.564	0.574
Max Adv. Down	43.178	47.779	-1.364	0.174

This table tests for ex ante differences in observable variables between treatment and non-treatment counties. The t test and p value columns come from the standard difference of means statistic with unequal variance.

Table 2.3: Effect of Thunderstorms on Internet Quality

	Dependent variable:			
	Total Tests (log)	VOIP Quality	Jitter	Latency
	(1)	(2)	(3)	(4)
Thunder ×	0.003	-0.396^{***}	0.478^{***}	-0.388^{***}
Fixed	(0.002)	(0.036)	(0.073)	(0.133)
Thunder ×	0.104^{***}	0.650^{***}	0.573^{**}	3.140^{***}
Wireless	(0.007)	(0.139)	(0.278)	(0.506)
ISP×City FE	✓	✓	✓	✓
Date FE	✓	✓	✓	✓
Observations	184,961	184,961	184,961	184,961
R^2	0.689	0.756	0.760	0.795

Since testing speed represents an endogenous response to lowered quality, the first column is a measure of whether individuals are responding to the adverse weather; a positive coefficient indicates a greater response. Columns 2 and 3 measure the quality of voice over internet and the variation in data transmission delays; negative and positive coefficients represent an adverse impact, respectively. Column 4 measures the "round trip time" to fetch data from servers and is important for real time applications (video chat, gaming), and a positive coefficient measures an adverse impact of weather. Standard errors are clustered at the level of fixed effects. *p<0.1; **p<0.05; ***p<0.01

Table 2.4: Effect of Thunderstorms on Internet Speeds

	Dependent variable:		
	log(Total Tests)	MB/s down	MB/s up
	(1)	(2)	(3)
Thunder×Fixed	-0.009^{***}	0.151^{***}	0.172^{***}
	(0.0005)	(0.007)	(0.005)
Thunder×Satellite	-0.007	-0.194	-0.200
	(0.021)	(0.296)	(0.216)
Thunder×Wireless	-0.009^{***}	-0.123^{***}	-0.072^{***}
	(0.001)	(0.015)	(0.011)
ISP×City FE	✓	✓	✓
Date FE	✓	✓	✓
Observations	11,292,923	11,292,923	11,292,923
R^2	0.886	0.784	0.791

Since testing speed represents an endogenous response to lowered quality, the first column is a measure of whether individuals are responding to the adverse weather; a positive coefficient indicates a greater response. Columns 2 and 3 measure the upload and download speed; a negative coefficient indicates a greater adverse impact of weather. Errors are clustered at the level of the fixed effects, $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 2.5: Main Results

	Dependent variable:			
	log(Promise Ratio)			
	FE		*IV*	
	(1)	(2)	(3)	(4)
Treatment $\times t_1$	0.017	0.017**	0.230**	0.317*
	(0.012)	(0.008)	(0.116)	(0.182)
Treatment $\times t_2$	0.035**	0.020**	0.162	0.096
	(0.014)	(0.009)	(0.121)	(0.175)
Treatment $\times t_3$	0.026	0.004	0.037	0.085
	(0.017)	(0.010)	(0.133)	(0.192)
log(down)	−0.028***	−0.060***	−0.028***	−0.056***
	(0.008)	(0.004)	(0.005)	(0.005)
Ncomp	−0.002	−0.006	−0.002	−0.0001
	(0.012)	(0.007)	(0.010)	(0.010)
Unit FE	✓	✓	✓	✓
Date FE	✓		✓	
ISP × Date FE		✓		✓
Observations	16,627	16,627	10,940	10,940
R^2	0.647	0.680	0.628	0.646

The coefficients can be interpreted as percentage changes when multiplied by 100. The treatment coefficients show the effect of adding an additional high speed competitor after 6, 18, and 30 months respectively. log(down) is the logged claimed download speed, and Ncomp is the number of terrestrial competitors. Standard errors are clustered at the level of the fixed effect for the OLS columns, and at the level of the instrument (county level) for the IV columns, *p<0.1; **p<0.05; ***p<0.01.

Table 2.6: Main Results by Initial Competition

	Dependent variable:			
	log(Promise Ratio)			
	OLS		*IV*	
	(1)	(2)	(3)	(4)
Treatment$\times t_1$	-0.058	0.293^{**}	-0.096	0.260^{*}
	(0.137)	(0.145)	(0.195)	(0.239)
Treatment$\times t_2$	0.125	0.230	0.167	0.064
	(0.167)	(0.151)	(0.207)	(0.243)
Treatment$\times t_3$	-0.234	0.109	0.025	0.020
	(0.215)	(0.156)	(0.341)	(0.222)
log(down)	-0.012	-0.043^{***}	-0.046^{***}	-0.057^{***}
	(0.008)	(0.008)	(0.008)	(0.008)
Ncomp	-0.012	0.023	-0.011	0.007
	(0.014)	(0.015)	(0.018)	(0.015)
Init.Comp.$\geq$ 50 Mbit/sec	✓		✓	
Unit FE	✓	✓	✓	✓
Date FE	✓	✓		
ISP $\times$ Date FE			✓	✓
Observations	4,293	5,285	4,293	5,285
R^2	0.618	0.549	0.671	0.595

See table notes for Table 2.5 for interpretation. The sample is split into groups where there is the presence of an initial competitor offering speeds greater than 50 Mb/s and where there is not. Standard errors are clustered at the level of the fixed effect for the OLS columns, and at the level of the instrument (county level) for the IV columns, *p<0.1; **p<0.05; ***p<0.01.

Table 2.7: Promise Ratio Decomposition

	Dependent variable:			
	log(Claimed Mbit/sec down)		log(Actual Mbit/sec down)	
	(1)	(2)	(3)	(4)
Treatment $\times t_1$	0.386	0.171	0.606**	0.479*
	(0.330)	(0.468)	(0.345)	(0.487)
Treatment $\times t_2$	−0.632*	−0.770*	−0.452	−0.631
	(0.346)	(0.450)	(0.363)	(0.468)
Treatment $\times t_3$	−0.529	−0.324	−0.478	−0.221
	(0.381)	(0.497)	(0.400)	(0.516)
Ncomp	0.021	0.015	0.018	0.014
	(0.028)	(0.027)	(0.029)	(0.028)
Unit FE	✓	✓	✓	✓
Date FE	✓		✓	
ISP × Date FE		✓		✓
Observations	10,940	10,940	10,940	10,940
R^2	0.887	0.913	0.889	0.916

This table decomposes the effect of competition on the promise ratio into its effect on the denominator (claimed speed) and the numerator (actual speed.) log(down) is dropped as a control since it appears as a dependent variable. Standard errors the level of the instrument (county level) for all columns, *p<0.1; **p<0.05; ***p<0.01.

Table 2.8: Effect of treatment on peak consumption

	Dependent variable:			
	log(Gb Consumption)			
	OLS		*IV*	
	(1)	(2)	(3)	(4)
Treatment $\times t_1$	-0.037	-0.062	1.397^{**}	1.423^{*}
	(0.090)	(0.092)	(0.746)	(0.787)
Treatment $\times t_2$	-0.061	-0.143	0.949	0.303
	(0.099)	(0.102)	(1.174)	(1.730)
Treatment $\times t_3$	-0.049	-0.125	0.462	0.459
	(0.111)	(0.114)	(1.369)	(1.883)
log(down)	0.209^{***}	0.134^{***}	0.199^{***}	0.120^{***}
	(0.035)	(0.039)	(0.038)	(0.042)
Ncomp	0.040	0.033	0.040	0.039
	(0.070)	(0.073)	(0.078)	(0.088)
Unit FE	✓	✓	✓	✓
Date FE	✓		✓	
ISP × Date FE		✓		✓
Observations	7,181	7,181	7,168	7,168
R^2	0.770	0.775	0.755	0.760

The coefficients can be interpreted as percentage changes when multiplied by 100. Standard errors are clustered at the level of the fixed effect for the OLS columns, and at the level of the instrument (county level) for the IV columns, *p<0.1; **p<0.05; ***p<0.01.

Table 2.9: Effect of treatment on switching

	Dependent variable:			
	Switched Speeds Dummy			
	OLS		*IV*	
	(1)	(2)	(3)	(4)
Treatment $\times t_1$	-0.025	-0.035^{*}	-0.221	-0.100
	(0.019)	(0.020)	(0.282)	(0.679)
Treatment $\times t_2$	-0.047^{**}	-0.036	-0.171	0.046
	(0.021)	(0.022)	(0.290)	(0.537)
Treatment $\times t_3$	-0.051^{**}	-0.077^{***}	0.015	0.383
	(0.024)	(0.025)	(0.330)	(0.743)
ldown	0.068^{***}	0.071^{***}	0.070^{***}	0.071^{***}
	(0.008)	(0.009)	(0.009)	(0.011)
Ncomp	-0.046^{***}	-0.042^{**}	-0.050^{***}	-0.055^{**}
	(0.016)	(0.016)	(0.017)	(0.024)
Unit FE	✓	✓	✓	✓
Date FE	✓		✓	
ISP × Date FE		✓		✓
Observations	9,972	9,972	9,954	9,954
R^2	0.391	0.407	0.372	0.359

The coefficients on the treatment variables can be interpreted as the effect of treatment on the probability of switching speeds within your provider after 6, 18 and 30 months respectively. Standard errors are clustered at the level of the fixed effect for the OLS columns, and at the level of the instrument (county level) for the IV columns, *p<0.1; **p<0.05; ***p<0.01.

Table 2.10: Robustness: only ratios ≤ 1

	Dependent variable:	
	log(Promise Ratio)	
	(1)	(2)
Treatment$\times t_1$	0.171	0.205**
	(0.146)	(0.100)
Treatment$\times t_2$	0.149	0.199
	(0.158)	(0.221)
Treatment$\times t_3$	−0.212	−0.158
	(0.168)	(0.214)
ldown	−0.044***	−0.061***
	(0.011)	(0.012)
Ncomp	−0.003	−0.006
	(0.019)	(0.020)
Unit FE	✓	✓
Date FE	✓	
ISP × Date FE		✓
Observations	5,133	5,133
R^2	0.693	0.705

This table drops all observations with promise ratios greater than one. Standard errors are clustered at the level of the instrument, *p<0.1; **p<0.05; ***p<0.01.

Table 2.11: Robustness: different sat. share cutoffs

	Dependent variable:			
	log(Promise Ratio)			
	(1)	(2)	(3)	(4)
Treatment $\times t_1$	0.100	0.202	0.138	0.440*
	(0.125)	(0.179)	(0.105)	(0.289)
Treatment $\times t_2$	0.138	0.176	0.086	0.144
	(0.146)	(0.216)	(0.114)	(0.246)
Treatment $\times t_3$	0.099	0.120	0.067	0.143
	(0.136)	(0.210)	(0.116)	(0.304)
ldown	−0.053***	−0.054***	−0.054***	−0.056***
	(0.005)	(0.006)	(0.005)	(0.006)
Ncomp	0.001	−0.003	−0.001	−0.001
	(0.010)	(0.011)	(0.009)	(0.010)
Cutoff	0.5	0.6	0.7	0.8
Unit FE	✓	✓	✓	✓
ISP × Date FE	✓	✓	✓	✓
Observations	10,940	10,940	10,940	10,940
R^2	0.676	0.662	0.680	0.629

This table varies the cutoff that is used to create the treatment dummy; the leftmost column categorizes a treatment county as one in the 50th percentile or greater of ex ante satellite shares, and the rightmost categorizes a treatment counties as being in the 80th percentile or greater of the same. Standard errors are clustered at the level of the instrument, *p<0.1; **p<0.05; ***p<0.01.

Table 2.12: Robustness: different terrestrial competitors

	Dependent variable:			
	log(Promise Ratio)			
	(1)	(2)	(3)	(4)
Treatment$\times t_1$	0.317*	0.320*	0.317*	0.316*
	(0.182)	(0.182)	(0.182)	(0.182)
Treatment$\times t_2$	0.094	0.092	0.093	0.093
	(0.174)	(0.175)	(0.176)	(0.175)
Treatment$\times t_3$	0.086	0.094	0.086	0.078
	(0.193)	(0.192)	(0.192)	(0.191)
log(down)	−0.056***	−0.056***	−0.056***	−0.056***
	(0.005)	(0.005)	(0.005)	(0.005)
Ncomp National	−0.011			
	(0.012)			
Ncomp $> 10Mb/s$		0.009		
		(0.008)		
Ncomp $> 50Mb/s$			0.026*	
			(0.015)	
Ncomp $> 1Gb/s$				−0.030
				(0.029)
Unit FE	✓	✓	✓	✓
ISP × Date FE	✓	✓	✓	✓
Observations	10,940	10,940	10,940	10,940
R^2	0.646	0.645	0.646	0.646

This table varies how the terrestrial competitor control is defined; in the first through fourth columns, the competitors are only national brands, or ones that satisfy a certain minimum speed respectively. Standard errors are clustered at the level of the instrument, *p<0.1; **p<0.05; ***p<0.01.

2.7.2 Appendix B: Figures

Figure 2.1: Promise Ratio ECDF at 4 ISPs

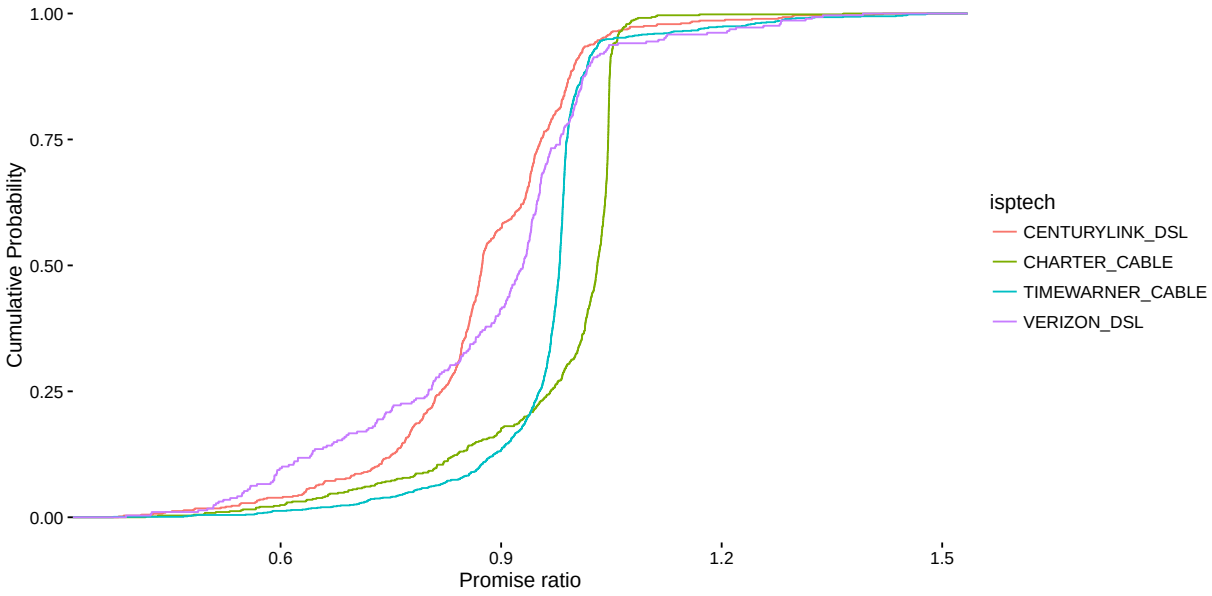
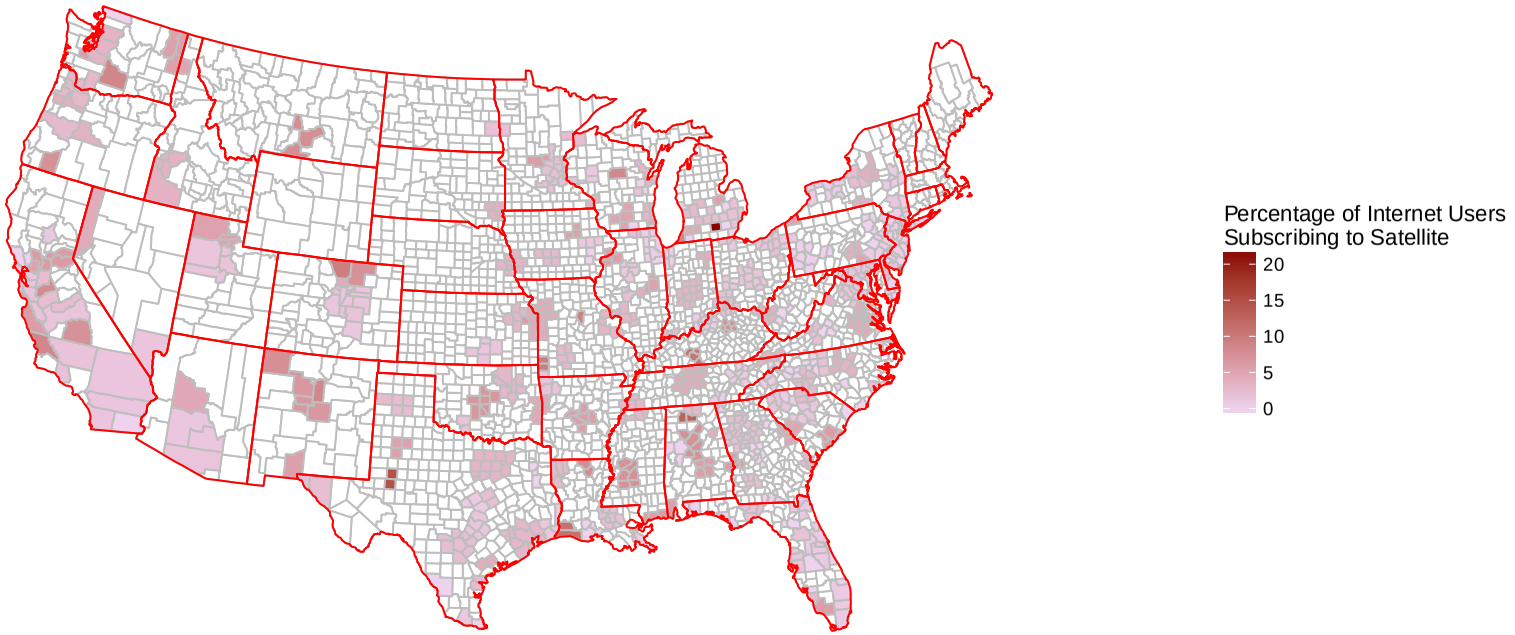


Figure 2.2: Ex Ante Satellite Subscription Shares



Chapter 3

Endogenous Quality and Exchange Rate Pass-Through

3.1 Introduction

A large body of empirical work documents a small response of traded good price indices to exchange rate fluctuations. These facts are usually described by saying that exchange rate pass-through is slow and incomplete. Surprisingly, very large devaluations are similarly associated with declines in the real exchange rate and correspondingly low inflation rates. In this paper, we use a uniquely granular micro-dataset spanning Russia's huge devaluation episode in late 2014 to argue that compositional shifts in product offerings are a pertinent force inducing seemingly low price levels after a cost shock.

Basic international macro and trade models with firm heterogeneity predict low aggregate price level changes due to compositional shifts in traded goods. During a devaluation, as costs rise foreign firms that are least efficient—those with the highest marginal costs—no longer find it profitable to trade. The rise in the local price index is therefore dampened by the exit of high cost, high price goods. We document in the data that there is indeed a reallocation towards relatively low cost goods after a currency devaluation. However, we

[0]This chapter is co-authored with Alexander Rodnyansky. Thanks to participants in the Bendheim Center for Finance student lunch, the Macroeconomics student lunch, and the Applied Micreconomics student lunch at Princeton University for their valuable feedback.

further show that there is a reallocation towards relatively low *quality* goods as well. We view this as a puzzle: if a cost shock precipitated by an exchange rate movement induces the least efficient foreign firms to exit, why is there a reallocation towards lower quality goods? Moreover, how does this quality reallocation interact with the cost reallocation to affect inflation and consumer welfare?

We address these questions by constructing a structural model of demand and firm entry and exit, where consumers value both low prices and high quality. Unlike in models built with constant elasticity of substitution utility, marginal cost and quality will not be isomorphic for firm pricing in this model. Therefore, goods in the model truly vary on two dimensions. We make the intuitive argument that if the joint distribution of marginal cost and quality has a positive covariance, then quality downgrading can be rationalized as high cost, high quality firms drop out. Although high quality has an insulating effect for high cost firms, if the *ex ante* variance of costs is high relative to quality then a proportional cost-shock will nonetheless induce these firms to exit in greater numbers than low cost firms.

Using a calibrated version of the model, we find that a proportional cost-shock equal in magnitude to the Russian ruble devaluation in 2014 leads to a substantial reallocation towards low cost, low quality goods. The lowest quartile of the cost distribution increases its relative share of firms 10 percent respectively after the cost shock, while high cost quartiles reduce their shares by between 17 and 43 percent. Because of a positive empirical covariance between quality and costs this also represents a reallocation towards lower quality goods. By way of comparison, in a model that only features cost heterogeneity the lowest quartile increases its share by only 6 percent. This results from a subtle feature of including quality heterogeneity: forcing high quality firms to exit reduces the average markup by making the environment more competitive even in the absence of a cost shock. This especially affects medium cost, medium quality firms, leading more of these firms to exit than in the case with only cost heterogeneity.

Including the quality dimension also affects how consumer welfare changes after the depreciation. In the model with quality heterogeneity, welfare decreases by 21 percent in response to the shock. In contrast, in the model with only cost heterogeneity, the welfare decrease is only 14 percent. Relative to the 50 percent cost shock we impose these welfare

reductions may seem modest; that the reduction is not higher reflects both a decrease in markups across the board as well as consumer willingness to substitute to domestic goods which do not experience a cost shock.

For our mechanism, we considerd the Russian ruble devaluation purely as a cost shock. There may be some worry that the reallocation towards lower quality goods stems from household substitution due to a corresponding real income shock. This phenomenon has been called the "flight from quality", and has been observed in the aftermath of large contractionary devaluations (Burstein et al. (2005)), potentially contaminating the measurement of CPI. We present several pieces of evidence that for the online retailer under consideration, the consumer response to the exchange rate shock was negligible in the short run. Firstly, since we observe both inventories and sales, we show that inventory reallocates towards low cost and low quality goods and that this is not driven by mass purchases of expensive items in an effort to move consumption forward. Firm choices, rather than consumer behavior, drives the reallocation towards low cost goods.[1] Secondly, we find in a robust difference-in-differences analysis that regions in Russia whose economies were heterogeneously exposed to the devaluation experienced no differential change in demand patterns. Thirdly, the Russian apparel market data used in this paper is an ideal laboratory to study how textile fabrics, which can be seen as a direct proxy of quality, change following large cost shocks. Restricting attention to one category of goods, shirts, we find that high quality materials disappear entirely from the composition during the first post-devaluation season.

We take care to show that although our data is from an online retailer, it exhibits similar features to data used in other studies of currency devaluations. In this way we hope to convince the reader that there is nothing "unusual" about the devaluation transmission mechanism we observe, and therefore that the results are externally valid. In particular, we document a relatively subdued aggregate price reaction across the store, and show it mainly to be the result of the low frequency of price changes, with only 8% of all goods being repriced each month. Conditional on price adjustments, however, we find that exchange rate

[1]Several other industries are likely to have been affected by the same quality downgrading mechanism during the course of the Russian currency crisis. In particular, a study by PwC found that domestic food producers reacted to the economic collapse by offering new and cheaper products, economizing on packaging, as well as switching to lower quality ingredients when manufacturing their produce. Source: "Food is changing in taste and weight.", *Vedomosti*, July 19, 2015.

pass-through is almost immediate and complete for imported clothing items. Both of these features of the data are present in studies using offline sales, which we discuss below.

3.1.1 Relation to the Literature

This paper is closely related to studies that have focused on large exchange rate devaluations and pass-through into prices, relying on aggregate industry level data (Burstein et al. (2005), Goldberg and Campa (2010)) as well as micro studies using product-level prices: Broda and Weinstein (2008) examine retail price data for individual products (UPCs) across different retail establishments, Gopinath et al. (2011) use data from a single retail chain to decompose variation in prices across the US and Canada into costs and markups. Burstein and Jaimovich (2012) use the same data to analyze patterns of wholesale prices to shed light on pricing-to-market (PMT) in aggregate and micro-level real exchange rate movements. Fitzgerald and Haller (2014) also provide very direct evidence on PTM by comparing the exchange rate response of prices of the same item sold to both the domestic and the international market. The lack of data on domestic product prices at the firm-level matched with international data shifted the focus of analysis from the response of domestic prices broadly to the response of prices of exporters and importers. For example, Gopinath and Itskhoki (2010) provide indirect evidence that is consistent with the presence of strategic complementarities in pricing. Amiti et al. (2014) decompose the variation across exporters in the exchange rates pass-through into markup and marginal cost channels, while Amiti et al. (2016) extend the analysis to provide direct evidence on strategic complementarities in price setting across industries. Recently, Cravino and Levchenko (2016) have looked into the distributional consequences of large devaluations and showed the cost of the consumption basket of people in the bottom decile of the income distribution to rise a lot more than of individuals in the first decile.

The paper also builds on the literature emstimating pass-through and markup variability in specific industries such as cars (Feenstra et al. (1996)), coffee (Nakamura and Zerom (2010)), and beer (Goldberg and Hellerstein (2012)). These industry studies typically rely on structural estimation by adopting a specific model of demand and market structure, which is tailored to the industry in question. We take a similar approach by considering

a framework that is similar in spirit to most dynamic oligopoly models in the applied IO literature, starting with Ericson and Pakes (1995). In notation and exposition we closely follow Weintraub et al. (2008). When estimating the model, we take the Bajari et al. (2007) methodology one step further by using it in a simulated method of moments approach to estimate parameters. Following Khandelwal (2010), we also exploit both price and quantity information to estimate the quality of products offered by the online store over time.

This study is also linked to a body of existing work quantifying how the treatment of sales influences measured degrees of price rigidity in posted prices. Bils and Klenow (2004), Eichenbaum et al. (2011), and Kehoe and Midrigan (2015) provide evidence on the extent to which discount prices influence effective prices paid by households in reality. Coibion et al. (2015) build on this literature by investigating the cyclicality in both posted retail and effective prices actually paid by consumers. Rather than discount sales, the authors find that consumer expenditure reallocation across stores in response to economic conditions drives a wedge between cyclical changes in posted and effective prices. Our key results are complementary to these findings as they document a separate amplification channel on the part of firms wishing to adjust the quality (and cost) of their products in response to aggregate economic shocks.

The focus of our paper is not on the pass-through of cost shocks at the good level, however a number of important factors explaining incomplete exchange rate pass-through have been studied over the past decades. Firstly, short-run nominal rigidities with sticky local currency prices (LCP) lead to zero pass-through for firms that do not respond to cost shocks immediately (Devereux and Engel (2002), Gopinath and Rigobon (2008)). Secondly, pricing-to-market (PTM) arises in a framework with variable markups where firms choose different prices for various destination markets depending on local conditions. Atkeson and Burstein (2008) provide a quantitative exploration of the PTM channel and its implication for aggregate prices. And thirdly, local costs may play a key role in determining pass-through by driving a wedge between prices and import costs and lowering firms' responsiveness to exchange rate fluctuations (Burstein et al. (2003) and Goldberg and Campa (2010)).

Finally, the paper is related to a number of studies on "flight from quality" and international trade. In a recent contribution, Medina (2017) shows that both quality upgrades

and increases in export activity can be direct outcomes of import competition, quantifying the importance of both factors. Chen and Juvenal (2015) show that higher quality goods enjoyed a stronger growth in exports before the crisis, but this trend reversed during the Great Recession, largely due to income shocks. Chen and Juvenal (2016) develop a model with more pricing-to-market and a smaller response of export volumes to a real depreciation for higher quality products, while Bems and Di Giovanni (2014) show that income effects can drive expenditure switching between domestic and imported goods using Latvian scanner-level data during the 2008 crisis. In contrast to these papers, we uncover a novel endogenous amplification channel on the firm side driving quality adjustments in product offerings.

The paper proceeds as follows. Section 3.2 provides an overview of the data and institutional background used in the paper. Section 3.3 presents stylized facts about price adjustments, potential demand (substitution and income) channels, as well as direct evidence on quality downgrading in the Russian online apparel industry. Section 3.4 describes the demand model. Section 3.5 outlines the estimation procedure. Section 3.6 explains how each moment from the data is obtained. Section 3.7 presents results and counterfactual simulations. Section 3.8 concludes.

3.2 Data

3.2.1 Data Sources

The main data is sourced from one of two most prominent online apparel marketplaces and distributors in Russia. The store specializes in clothing but offers certain semi-durable products as well. The company is a subsidiary of a publicly traded German enterprise, listed on the Frankfurt Stock Exchange. The data are available at daily frequencies and the main period of analysis is from January 1, 2012 until September 18, 2015. As of today, the online retailer operates in four Commonwealth of Independent States (CIS) countries (Belarus, Kazakhstan, Russia, and Ukraine), although the present study focuses exclusively on the largest market, which is Russia.

The following set of variables are available at a stock keeping unit (SKU) level, sorted

by orders over time: article id, SKU, product characteristics (such as size, color, fabric composition, premium status, brand, date of introduction, season, gender, target group or category etc.), order time and date, shipping method, city and region of the order, regular retail and discount prices, average wholesale costs, customer id, customer OS, inventory count on an almost daily level, and invoice currencies. This structure of the data allows to extract the precise number of quantities ordered and sold across thousands of destination cities within Russia—a central advantage relative to extant studies with no observations on the quantity dimension.

We further merge this dataset with aggregate macro and regional variables, such as the EURRUB and USDRUB exchange rates, regional GDP growth in 2015, as well as regional unemployment and income levels that are sourced from Global Financial Data and Russian Federal State Statistical Service Agency, respectively.

3.2.2 Institutional Background

Before the main analysis, it is important to highlight some key institutional features of the e-commerce apparel market in Russia as well as the huge devaluation episode in late 2014. Starting with the latter, the Russian currency crisis of 2014 broadly refers to a major decline in confidence that caused investors to sell off their Russian assets and led to a fall in the value of the Russian ruble. The lack of confidence in the Russian economy stemmed from two major sources: first and foremost, the price of crude oil, a key export of Russia, declined by nearly 50% between its yearly high in June 2014 and December 2014. The second factor was related to international economic sanctions imposed on Russia following the annexation of Crimea and a military intervention in Ukraine.[2] As shown in Figure 3.2, part of the results were a ruble depreciation of about 60% against the US dollar between July and December 2014.

From the vantage point of an individual firm, these abrupt exchange movements represent exogenous cost shocks to the values of imported inputs or consumer products. We later use this variation together with invoice currency information at the SKU-level to instrument

[2]It should be noted that Russian counter-sanctions against the EU were targeted exclusively at the food and agricultural sector and had no direct effects on apparel imports from South-East Asia.

for prices when estimating demand. As far as the apparel industry is concerned, its most distinctive trademark feature pertains to seasonality: most of the goods in our sample have a very short lifespan (the median is 182 days) and are put on sale as soon as their native season ends. Figure 3.1 provides a depiction of the percentage revenues attributable to each season over time. And since, by definition, any given SKU remains perpetually assigned to the same season, it is clear that hardly any goods live beyond three quarters of a year. Product-level entry and exit margins of adjustments will therefore be the central mechanism by which quality changes might come about. This industry structure necessitates brands to come up with novel products on a half-yearly basis, while strongly negative economic perturbations will influence the types of goods each firm chooses to supply. Yet since most apparel manufactures aim to retain their customer loyalty, the bulk of any quality downgrading should be expected to occur across rather than within brands over time.

A very first glance at the data manages to reveal a number of striking regularities supportive of potential compositional shifts in product offerings. Figure 3.2 shows that the steep ruble depreciation at the end of 2014 led to a notable increase in the deseasonalized average wholesale cost of goods sold (mean COGs). However, one can also discern a fall in the deseasonalized average *inventory-weighted* wholesale cost (i-weighted mean COGs) over the same time frame. This pattern is not driven by a large scale removal of high cost goods from the retailer's warehouses (which could perhaps be rationalized with consumers moving forward consumption), but rather by a disproportionate amount of stocking-up on low cost goods: the closely associated movement of the deseasonalized average *quantity-weighted* wholesale cost (q-weighted mean COGs) confirms this interpretation. In fact, it is worth pointing out that simple mean wholesale costs are consistently below both quantity- and inventory-weighted mean wholesale costs until the onset of the currency crisis in late 2014, when the order is reversed and the simple mean reaches its maximum level.

3.3 Empirical Motivation

We have three objectives in this section: first, we present direct evidence on quality downgrading by looking at the composition of textile fabrics within a narrow category of goods

(shirts). Second, we justify our exclusive focus on the devaluation as a cost shock by showing that demand effects of the shock were negligible. Finally, we provide evidence that the pattern of price adjustment for our data is not unusual. The concern that conclusions derived from the pricing behavior of an online retailer might not be externally valid must be addressed, as the literature has found that mechanisms that are not present online—such as menu costs—are important for pass-through. We show that both the price index of goods sold by the firm as well as the pattern and frequency of individual goods price adjustments are not unusual.

3.3.1 Quality

Due to the highly seasonal nature of the apparel industry, and independent of within-SKU pass-through, a large chunk of the response to the currency crisis and resulting cost shock could naturally have occurred on the extensive margin of goods entry and exit. The objective of this section is therefore to construct a quality index for products in the shirts category offered by the online store using a similar approach as the one employed by the BLS in calculating the CPI.

Initially, quality weights have to be constructed for each fabric type. We choose to base these weights on relative textile prices from China, the main production country for almost every apparel item in the dataset. Further, the vast majority of all fabrics happen to be either cotton, viscose or polyester fibers, and are ranked according to the highly stable relative cost difference of these materials over time. Once these time-invariant fabric rankings are obtained, each product can be assigned a quality measure, $\lambda_{i,t} = \sum_k \omega_k \cdot \lambda_{i,k,t}$, for each season t, SKU i, fabric type k and its respective quality ranking ω_k. The cross-seasonal quality index is then computed as follows:

$$\mathcal{Q}_t = \mathcal{Q}_{t-1} \cdot \frac{\bar{\lambda}_t}{\bar{\lambda}_{t-1}} \tag{3.1}$$

where $\mathcal{Q}_t$ is the quality index for season t, $\mathcal{Q}_{SS2012}$ is normalized to 100, and $\bar{\lambda}_t$ is the simple average SKU quality measure for season t.

Panel (a) of Figure 3.5 shows the available shirt sole material fabric composition over time,

distinguishing between fall-winter (F) and spring-summer (S) seasons. Panel (b) plots the time-series for the quality index in equation 3.1, splitting goods by whether they are imported or locally produced. The results show striking patterns: the more expensive higher quality materials disappear entirely from the composition during the first post-crisis season (15S). In particular, the percentage of natural leather and agriculturally grown gum fabrics abruptly drops to zero, leaving all existing shirts composed of the artificial and cheaper polyurethane material. Looking at the quality index, it is obvious that all of these adjustments are driven by imported goods rather than the locally produced items. And even though the quality index for Russian goods displays some volatility due to the much smaller fraction of apparel manufacturing in Russia, the quality of these products does not react in any noticeable way to the massive ruble devaluation while the fabrics based quality of imported shirts drops by around 30% relative to the previous spring-summer season.

Taken together these findings suggest a substantial compositional reallocation on the part of fashion manufacturers from higher towards lower quality goods.

3.3.2 Demand Channel

One might suspect the observed compositional changes stem from a large demand shift towards cheaper or lower quality goods as a result of an income shock to consumers, rather than a cost shock to apparel manufacturers. In this section we assess the quantitative importance of this mechanism by looking at regions that were more adversely affected during the crisis and comparing their demand patterns to regions that had higher economic growth. We find little evidence of demand non-homotheticities or consumption reallocation towards cheaper goods in Russian regions (oblasts) suffering from extremely low or even negative economic growth in 2015. The basic approach entails a difference-in-differences (DinD) estimation strategy of the following form:

$$\log(Y_{it}) = \alpha_i + \sum_t \gamma_t \mathbf{D}_t + \sum_t \delta_t \left(\mathbf{D}_t \cdot Growth_i\right) + X'_{it}\theta + \sum_t \psi_t \left(\mathbf{D}_t \cdot X_{it}\right) + \epsilon_{it} \qquad (3.2)$$
$$\forall i,\ \forall t \in \{2012m1, \ldots, 2015m9\} \setminus \{2014m12\}$$

where Y_{it} is either the i) median regular price, ii) mean (weighted or unweighted by sales) regular price, or iii) fraction of premium sales in region i at time t, α_i are region fixed effects, $Growth_i$ is the nominal regional GDP growth in 2015, $\mathbf{D}_t$ is an indicator for the time period (year-month), with $2014m12$ taken as the omitted category, $(\mathbf{D}_t \cdot Growth_i)$ represents an interaction term between the time indicators and a region's economic performance in 2015, and X_{it} is a matrix of control variables that includes total regional sales (in logs), as well as regional unemployment and income levels. All standard errors are clustered at the region-level to allow for serial correlation across time.

The Russian currency crisis had a vastly differential impact on various regions of the country. This provides for a clean distinction between exposed (low growth) and unexposed (high growth) oblasts that can be utilized when estimating specification 3.2. Panel (a) of Figure 3.6 shows a map with geographic regions that grew relatively fast (in dark colors) as well as slowly (in light colors) in 2015. Exclusively devoting attention to oblasts with positive retail sales, the steepest contraction saw regional GDP growth of -10.1% while the oblast with the highest growth expanded by 16.1%. The standard deviation of income growth was 3.26 over this period.

As would be necessary with any DinD estimation approach, this specification also provides evidence on the parallel trends assumption in all outcome variables. That is, in the absence of treatment the unobserved disparities between high- and low-growth regions should be constant over time – the validity of the estimation procedure relies on outcome variables that would have continued to develop as they did before the economic shock in all regions. Unless this assumption is valid, the estimated treatment effects would be biased versions of the true impact. As an additional robustness check on the identification strategy, all control variables are interacted with the $\mathbf{D}_t$ indicators to allow for possible heterogeneous responses to negative economic shocks across very distinct regions (e.g. poor vs. richer oblasts might react differently to the crisis).

The main parameters of interest are the δ_t since they capture the difference between crisis exposed and relatively unscathed regions over time. The estimated fixed-effects model includes leads going back to early 2012 and lags reaching the last available month, September 2015. The specification allows for any causal direction of the findings and assesses if the

effects grow or fade over time.

One may also entertain a causal interpretation of the δ_t estimates in equation 3.2 for other important reasons. Firstly, about 93% of goods sold by the retailer are not produced in Russia, and even when the good is home made it is almost never manufactured in the region under consideration. Hence the specification will not suffer from endogeneity issues typically associated with regressions of prices on economic activity. For instance, unobserved productivity innovations for a specific SKU are unlikely to be correlated with local growth rates. In principle, aggregate shocks could lead to simultaneous movements in prices of goods and local economic growth. But since time fixed effects are included, they should eliminate this endogeneity issue too. Finally, the retailer does not discriminate in price setting across geographic regions within Russia and thus any observed divergence in regional (sales weighted) median or mean prices can only be explained by movements in quantities (purchases).

The findings are summarized in Figure 3.6, which plots the key estimated parameters of interest, $\boldsymbol{\delta}_t$, with 95% confidence intervals around them. As would be consistent with the parallel trends assumption, the estimates in Panel (b) show no robust differences between the positively exposed (high growth) and negatively hit (low growth) regions in the months prior to the onset of Russia's currency crisis. Then, starting around mid-2014, volatility in the treatment effects for all three outcome variables starts to appear. However, the results are insignificant and hardly moving in the expected positive direction. Together with unreported but also highly robust evidence suggesting no differential effects on total regional sales, this leads us to conclude that income shocks across Russian geographies played a marginal role in the observed compositional shifts in the affordable fashion industry and that endogenous amplification channels on the firm-side must be driving most of the quality downgrading.

3.3.3 Prices

Price Indexes

The first objective is to construct a price index for the online store using a similar approach as the one employed by the BLS in calculating the CPI. Initially, the total quantity sold is

computed for each good (SKU) in every month. The index is aggregated to the monthly level to reduce the number of missing values. The following category-level Laspeyres price index is used:

$$P_{j,t}^{L} = P_{j,t-1}^{L} \cdot \frac{\sum_{i \in j} p_{i,t} \bar{q}_{i,t}}{\sum_{i \in j} p_{i,t-1} \bar{q}_{i,t}} \tag{3.3}$$

where $P_{j,t}^{L}$ is the Laspeyres price index for category j in month t, $P_{j,2013m1}^{L}$ is set to equal 100, the price of a specific good i at time t is given by $p_{i,t}$, and $\bar{q}_{i,t}$ is the average monthly quantity sold of good i across months *t-1* and *t*. By fixing quantities in this way, household consumption patterns are held fixed as prices change. The basket of goods is updated in two-month intervals (e.g. Jan–Feb, Feb–March, etc.), and the resulting indexes are chained to produce one index for each category, denoted by $P_{j,t}^{L}$.

A key issue when computing monthly price indexes is how to deal with missing values from period-to-period. For example, a product may show up in month m but not in month $m+1$, making it impossible to compute the price change for that good between two months. Missing values may be due to new products entering the market, old products withdrawing from the market, and seasonality in sales. Given the highly ephemeral nature of products in the apparel industry, we build a month-to-month chained price index to ensure the presence of a large enough fraction of goods at every step of the calculation. This way, more than 50% of SKUs are present in each period.

The results for the aggregate price index based on all goods listed by the online marketplace reveal a striking pattern: Figure 3.3 suggest that as the ruble devalued by about 60% against the US dollar, the ruble prices of imported apparel items rose only by approximately 4%. Locally produced (Russian) goods faced a similar albeit somewhat smaller price inflation during this period. After about four months into 2015, prices appear to have stabilized at a new and slightly higher level.

Puzzling as they may seem, these low inflation numbers are mainly explained by the low frequency of price adjustment during the whole period of operations. Figure 3.4 shows the sales-weighted and unweighted percentages of goods that faced a price change in any

given month relative to the previous month. The average percentage of price adjustments is roughly 8% throughout all years for which data is available. Price adjustments seem to become slightly more frequent just before each winter sale period, except for the currency crisis year of 2014 when there were fewer price decreases (discounts) than normally.

Besides explaining the limited reaction of aggregate price indexes, these pattern once again underscore the importance of the extensive margin of product entry and exit. In particular, all goods are put on sale in the latter half of their lifespan, and on average after 60% of the weeks a good is available for purchase have been passed. This mechanism of phasing out older generations of products before the arrival of new ones defines transitions between seasons in many industries.

Micro-dynamics of Price Adjustments

Once a conditioning on price adjustments is introduced, however, inflation and within-SKU pass-through become a lot higher, and turn out to be almost immediate and complete for imported goods.

At the SKU-level, we estimate pass-through into prices of exchange rate shocks realized during the most recent period of price non-adjustment and of those that were realized prior to the previous price adjustment. As discussed in the literature (Gopinath and Itskhoki (2010)), in the absence of real rigidities, all adjustment should take place at the first instance of price change and hence the coefficient on the exchange rate change prior to the previous price adjustment should be zero. More precisely, the following regression is estimated:

$$\Delta p_{i,t} = \beta_1 \Delta_{\tau_1} e_t + \beta_2 \Delta_{\tau_2} e_{t-\tau_1} + \eta_i + \epsilon_{i,t} \tag{3.4}$$

where i indexes the SKU, t stands for the date, the outcome variable, $\Delta \bar{p}_{i,t}$, is the change in the log ruble price of a good, *conditional on price adjustment*, and $\Delta_{\tau_1} e_t \equiv e_t - e_{t-\tau_1}$ is the the cumulative change in the log of the bilateral nominal exchange rate over the duration when the previous price was in effect (denoted as τ_1). Analogously, τ_2 denotes the duration of the previous price of the firm so that $\Delta_{\tau_2} e_{t-\tau_1} \equiv e_{t-\tau_1} - e_{t-\tau_1-\tau_2}$ is the cumulative exchange rate change over the previous period of non-adjustment, i.e. the period prior to the previous

price change. Finally, solely within-SKU variation is exploited via the inclusion of good-specific fixed effects, η_i, and standard errors are clustered at the SKU-level to allow for serial correlation across time.

Table 3.1 reports the results from estimations of regression 3.4, with evidence for various sub-samples of the data and restricting attention to goods that generate above 90% of all revenues, i.e. with a lifespan of less than 300 days. In contrast to some findings in previous studies, exchange rate shocks that took place prior to the current period of non-adjustment have a much weaker and hardly robust effect on current price changes. This is reflected in the estimated β_2 coefficient being statistically indistinguishable from zero in across specifications.

When the regression model is restricted to imported goods only in columns (1) through (4), pass-through occurs immediately after the cost shock, which is indicative of no substantial real rigidities in pricing. In fact, pooling all categories of goods (target groups), pass-through seems almost complete, with the estimated β_1 coefficient almost equal to one and significant at the 1% level. Later, when developing the structural framework for counterfactual analysis in section 3.4, we focus exclusively on a single product category, shirts, to ensure meaningful quality comparisons. For that purpose, columns (2) and (4) list very similar results when the sub-sample is restricted to shirts. Remarkably, when only locally produced goods (made in Russia) are considered in columns (5) through (8), the pass-through coefficients are almost everywhere insignificant and hardly point in one robust direction.

Together with the fact that exchange rate movements are exogenous cost shocks from the vantage point of an individual retailer, these findings lend additional support to the causal interpretation of these results: consistent with Table 3.1, and apart from potential strategic complementarity considerations, the prices of Russian goods should be much less sensitive to exchange rate fluctuations. If all of our main results were driven by some aggregate income shock story, we should not expect to see such different outcomes for Russian and imported items.

3.4 Model

3.4.1 Overview

Having documented that (1) there is incomplete pass through in the aggregate price index, and (2) there is a reallocation towards lower cost goods, and (3) there is a reallocation towards lower quality goods, we formulate a theoretic framework that speaks to all of these stylized facts.

The basic intuition of the model must rationalize the fact that there is a reallocation towards goods that are low cost, but also low quality. In the apparel industry, a low cost good can be inexpensive to produce because it is produced by an efficient firm in a low cost location, because it uses cheaper materials, or both. In a competitive industry, it is reasonable that on average a low cost good must be cheaper or lower quality in some way on average, even if there is considerable dispersion for any given cost quantile.

If low cost goods also tend to be low quality goods, then these goods may not account for a large share of sales *ex ante* because they are less profitable. However, after a proportional exchange rate shock, high cost goods experience a much larger level increase in cost and therefore may experience a disproportionately larger decrease in profits and market share. High cost goods should be more insulated from the shock by their higher average quality. Yet if the variance of costs is relatively high and the variance of quality relatively low, then their higher quality will not be enough to keep them profitable. Low cost, low quality goods will increase their market share, especially relatively to the low *ex ante* share they enjoyed pre shock.

To formalize this intuition, we construct a simple model of oligopolistic competition with logit demand and match it to key moments of the data covering the periods prior to the shock. In the model, firms (individual products) enter and sell at wholesale cost to the monopolist. The monopolist charges a constant proportional markup over wholesale cost, so that firms actually face the residual demand curve of the consumers. On entry, firms draw a cost quality pair for their product from a fixed distribution; the steady state distribution of firms over cost quality states is determined by endogenous exit choices, as firms are faced with idiosyncratic payoffs to exit.

It should be made clear that individual SKUs, and not brands or the monopolist online retailer, are the main economic actors in this model. Based on the reduced form evidence on pass-through, the monopolist simply prices at a constant markup above wholesale cost. For this reason, as well as the fact that the monopolist's problem is computationally difficult, we do not focus on the monopolist. In this version of the paper we refrain from focusing on brands as the economic actors simply because they exhibit huge variation in the quality of goods they offer: to model a brand as a product would drop key information about the cost-quality distribution, while allowing brands to chose the set of goods they offer would create an intractably large state space.

We model the exchange rate shock as increasing the entire cost distribution proportionally to the depreciation in the ruble, with no effect on income as suggested by our reduced form evidence. In the new long run steady state, we find substantial reallocation towards low cost goods: the lowest quartile increases its share of firms by 10 percent, while higher quartiles decrease their shares by between 17 and 43 percent. We also construct the model's predicted aggregate price index as a measure of model fit. The fixed-quantity price index (with quantity shares fixed to pre-shock levels) increases by only about 4%, which is roughly equal to the price index constructed from data alone.

Adding in the quality dimension better explains the pattern of increasing low cost goods share after the shock than a model with cost heterogeneity alone. Consider a model with cost and quality heterogeneity: the high quality firms are able to charge higher markups since they have larger positive demand shifters, which in turn makes other firms' residual demand curves more inelastic, allowing them to charge higher markups. If high quality firms exit, then the competitive environment induces a general reduction in the markups of the remaining firms. In the model with only cost heterogeneity, the absence of this second order effect implies that the lowest quartile increases its share by only 6 percent. Evidently, medium cost, medium quality firms are most strongly affected by the second order effect, so that they lose relatively more—and low cost, low quality firms gain more—in the model with quality heterogeneity.

The model also generates new insights about the true effects of the exchange rate depreciation on consumer welfare. The model without quality heterogeneity implies a roughly 14%

decrease in welfare following the shock, while the model with quality heterogeneity implies an almost 21% decrease in welfare. The model we specify is similar in spirit to most dynamic oligopoly models in the applied IO literature, starting with Ericson and Pakes (1995). In notation and exposition we closely follow Weintraub et al. (2008).

3.4.2 Notation and timing

The industry evolves over discrete time periods $t \in \mathbf{N}$ and an infinite horizon. Brands that enter draw a state x from an invariant, discrete distribution F over all possible states A, and keep that state until they exit. Firm j's state is a cost quality pair for the single good it sells, $x_j = (c_j, \lambda_j)$.

Define the industry state s_t to be a vector enumerating how many firms there are at each state $x \in A$ in each period, so that the aggregate state space is $\bar{S} = \left\{ s \in N^{\infty} | \sum_{x \in A} s(x) < \infty \right\}$. For each $j \in S_t$, define $s_{-j,t} \in S$ to be the state of the competitors of firm j, so that $s_{-j,t}(x) = s_t(x) - 1$ if $x_{jt} = x$ and $s_{-j,t}(x) = s_t(x)$ otherwise.

Firms earn profits each period on a spot market, with single period expected profit denoted by $\pi(x_{jt}, s_{-j,t}, \theta)$. Profits will derived from demand and marginal cost primitives, indexed by parameter θ. New firms choose whether to enter and pay a fixed cost of entry κ; incumbent firms observe a firm-time specific shock ϕ_{jt} and decide whether to exit.

In each period, timing is as follows:

1. Incumbent firms observe their sell-off values and make exit decisions
2. The number of entrants is determined; entrants pay the fixed cost κ
3. Incumbent firms (even those that have committed to exit) compete in the spot market and receive profits
4. Exiting firms exit and receive scrap values, and new entrants enter and draw states from the *ex ante* distribution.

3.4.3 Model primitives and assumptions

We assume that demand for goods is logit, marginal costs are constant and perfectly observed, and that prices are the Nash equilibrium of a Cournot pricing game. We further assume that when brands wholesale to the monopolist, the simple heuristic that firm uses to set retail prices is to set a constant markup γ over the wholesale cost w_{jt}, so $p_{jt} = \gamma w_{jt}$.

As usual, utility to consumer i from consuming firm j's product is:

$$U_{ijt} = \alpha p_{jt} + \lambda_j + \epsilon_{ijt}$$
$$U_{i0t} = \lambda_0 + \epsilon_{ijt}$$

Where the errors ϵ_{ijt} are distributed type II extreme value, giving the standard expression for market shares:

$$\sigma_{jt} = \frac{\exp(\alpha p_{jt} + \lambda_j)}{\exp(\lambda_0) + \sum_k \exp(\alpha p_{kt} + \lambda_k)}$$

Where σ_{jt} denotes the firm j's share of the market at time t. On entry, brands draw a constant marginal cost and quality pair. They observe the aggregate industry state and set wholesale prices accordingly, choosing their wholesale price w_{jt} to solve:

$$\max_{w_{jt}} \sigma_{jt}(\gamma w_{jt})(w_{jt} - c_j)$$

Profits for j are given as $\pi_{jt}(p_t^*)$, where p_t^* are the prices that satisfy a static Nash equilibrium.

We assume that the scrap values ϕ_{it} are iid, with finite expectation and well-defined density functions over R_+. When matching the model to the data, we will pick the exponential distribution.

Finally, we assume that there are an asymptotically large number of potential entrants who play a symmetric mixed entry strategy, as in Weintraub et al. (2008). They prove that this results in a Poisson-distributed number of entrants. The expected number of firms entering at industry state s_t is denoted by δ_{s_t}; the expected entry rate will be endogenously determined and satisfy a zero expected discounted profits condition.

3.4.4 Equilibrium

The data is assumed to generated by a pure strategy Markov perfect equilibrium, where firms play symmetric strategies (that is, conditional on state, firms' strategies are identical.) Continuing to follow the notation of Weintraub et al. (2008), firms follow an cutoff-rule exit strategy: there is a function ρ such that j exits at time t if and only if the randomly drawn scrap value exceeds this cutoff, $\phi_{jt} \geq \rho(x_{jt}, s_{-j,t})$. Crucially, because the state is two-dimensional, cutoffs are not monotonically decreasing in cost (although they are conditional on quality.) Finally, the entry rate function $\delta(S_t)$ determines the Poisson average entry rate conditional on the aggregate state S_t.

The value function $V(x, s|\rho', \rho, \delta)$ is the expected NPV for a firm at state x when its competitors' state is s, given that its competitors each follow common strategy ρ:

$$V(x, s|\rho', \rho, \delta) = E_{\rho', \rho, \delta}\left[\sum_{k=t}^{\tau_j} \beta^{k-t}\pi(x_{jk,s_{-j,k}}) + \beta^{\tau_j - t}\phi_{j,\tau_j}|x_{jt} = x, s_{-j,t} = s\right]$$

j is the index of a firm at state $x = (c, \lambda)$ at time t, τ_j is a random variable giving the optimal stopping time, and the expectation is over the firm's own strategy as well as the strategy of its competitors.

We assume a symmetric MPE with positive entry rates. The full MPE comprises strategies ρ and an entry rate δ such that

1. Incumbent firm strategies represent an MPE:

$$\sup_{\rho'} V(x, s|\rho', \rho, \delta) = V(x, s|\rho, \rho, \delta)$$

2. At each state, entrants have zero expected discounted profits and $\lambda > 0$:

$$\beta E_{\rho, \delta, F}\left[V(x, s_{-j,t}|\rho, \delta)|s_t = s\right] - \kappa \leq 0$$

3.5 Estimation

Our data contains information on hundreds of brands and thousands of products. Since even fitting the MPE is intractable in such an environment, we take an alternative approach. We assume that the data pre-shock is generated by firms playing Oblivious Equilibrium (OE) strategies. We then pick parameters and simulate data from the model, and match key simulated moments to moments in the data. Once we have matched moments, we can do experiments by altering parameters and using the same simulation mechanisms.

Oblivious Equilibrium is a solution concept pioneered by Weintraub et al. (2008) to deal with the problem of simulating data from industries with dozens or even hundreds of firms that are still plausibly competing oligopolistically. In their seminal work on solving MPE models, Bajari et al. (2007) demonstrate that simulating from an MPE model even with only 3 firms induces enormous computational burdens; moreover, problems of multiple equilibria grow in severity as the number of firms increases.

Oblivious Equilibrium assumes that all firms treat the aggregate state vector as being in its long run steady state. This simplifies the computational burden enormously: instead of needing to keep track of the state of every firm in the industry when solving its value function, a firm needs only pay attention to its own state. We take the Bajari et al. (2007) methodology one step further by using it in a simulated method of moments approach to estimate the model's parameters.

Define $\rho(x)$ to be symmetric, oblivious equilibrium strategy, δ to be the oblivious entry rate, and the expected number of firms at state x at time t as $\tilde{s}_t(x) = E[s_t(x)]$. Given the *ex ante* distribution function F, there is a simple expression for the expected number of firms in each state:

$$\lim_{t\to\infty} \tilde{s}_t(x) = \delta \int_{x\in A} \sum_{k>0} P(\phi > \rho(x))^k dF$$

Let $\tilde{s}_{\rho,\delta} = \lim_{t\to\infty} \tilde{s}_t(x)$. The oblivious value function is the expected NPV of a firm at state x following oblivious exit strategy ρ under the assumption that its competitors' states are

fixed at $\tilde{s}_{\rho,\delta}$. It is given by: d

$$\tilde{V}(x|\rho', \rho, \delta) = E_{\rho'}\left[\sum_{k=t}^{\tau_j} \beta^{k-t}\pi(x_{jk}, \tilde{s}_{\rho,\delta}) + \beta^{\tau_j - k}\phi_{j,\tau_j}|x_{jt} = x\right]$$

The oblivious value function does not directly depend on its competitors' dynamic strategies; however, since the aggregate state $\tilde{s}_{\rho,\delta}$ does, the dependence is indirect. As in Benkard et al. (2008), we define an oblivious equilibrium as:

1. Incumbent firms optimize an oblivious value function:

$$\sup_{\rho'} \tilde{V}(x|\rho', \rho, \delta) = \tilde{V}(x|\rho, \rho, \delta)$$

2. At each state, entrants have zero expected discounted profits and $\lambda > 0$:

$$\beta E_{\rho,\delta,F}\left[\tilde{V}(x|\rho, \delta)\right] - \kappa \leq 0$$

3.5.1 Moments

Define the distribution of scrap values by Φ. For a given set of parameters

$$\{\alpha, \gamma, \lambda_0, F, \kappa, \Phi\},$$

we can generate data from the Oblivious Equilibrium, using Bellman's equation and value function iteration to solve the firm's optimal stopping problem. There are two infinite dimensional probability distributions to be identified: the *ex ante* distribution of costs and qualities F, and the distribution of scrap values Φ.

To make the problem tractable, assume that the scrap value distribution is exponential with mean parameter K. Further assume that F is distributed lognormal, with $\log(X) \sim N(\mu, \Sigma)$. We discretize the distribution (at the marginal quantiles of the original conditional distribution) to make the state space discrete, with 25 cost quality combinations total. This parameterization leaves us 10 objects (2 means in μ and 3 unique terms in Σ) to identify. γ and λ_0 are identified directly from the data, while for the remainder we minimize the

unweighted sum of the difference between the model's simulated moments and the empirical moments in 3.2.

3.6 Constructing moments from the data

The key moment we need to match in the data is the conditional joint distribution of cost and qualities, where the conditioning is on firm entry and exit. Although we have a measure of quality that we directly observe in the data and use to motivate our approach, better fabric composition is just one small part of why a good's demand might be more inelastic. We estimate quality as a SKU specific fixed effect in a logit demand system. In particular, if an SKU has higher sales conditional on an identical price, we impute that good as having a high quality. Moreover, while we observe prices and average wholesale costs, we do not observe the marginal costs of the wholesalers themselves; this will also be recovered from the logit demand system.

Recovering quality from the data requires taking into account features of the environment that are not in our simple model. First, most brands are multiproduct; although substitution patterns may differ slightly within and between brands' goods, we still use the simple (non-nested) logit to map closely to the structural model. However, when inverting the demand system to recover the brands' marginal costs, we take account of the fact that firms selling multiple goods will incorporate cross-elasticities of substitution across their own products when setting prices.

Second, as our data comes from an apparel retailer there is a strong seasonal component to some product lines; we therefore include month dummies to control for aggregate time-varying shocks. Finally, in order to use the Berry (1994) inversion and linear regression to estimate the qualities and costs, we must add in a product time specific demand shock to consumers' utility. As usual, this shock is not observed by the econometrician but is observable to wholesalers when they make their pricing decisions, which leads to endogeneity, which we deal with using a number of instruments.

3.6.1 Demand

The monopolist firm sells J_t differentiated goods in month t, which can be partitioned into B_t different brands. Let J_{bt} be the set of brand $b \in B_t$ goods offered at time t.

The upstream brands—the full set of which is denoted by B—are, as before, the principal decision makers in this model. Brands observe the full set of characteristics (qualities) for each good they can offer. They decide at what wholesale price (in rubles) to sell to the monopolist. Similarly to Nakamura and Zerom (2010), the downstream retailer will operate according to a heuristic rule; in this model, they charge a fixed percentage markup over the marginal cost, which we will recover from the data.

The value to consumer i of consuming good j at time t is given by:

$$U_{ijt} = \lambda_j + \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

We assume ϵ_{ijt} is distributed Type II extreme value to give a logit structure for demand. ξ_{jt} is a demand shock observable at the time brands choose the wholesale price, and is therefore endogenous with p_{jt} (which will simply be a constant multiplied by the wholesale price). λ_j will be recovered as a product SKU specific quality.

The value of the outside good is allowed to vary over time to capture time-varying aggregate shocks and seasonality:

$$U_{i0t} = \lambda_t + \epsilon_{i0t}$$

Nested logit demand leads to the standard expression for the share of each product j. Taking log differences (as in Berry (1994)) gives a simple linear expression:

$$\ln \sigma_{jt}(\theta) - \ln \sigma_{0t}(\theta) = \alpha p_{jt} + \lambda_j - \lambda_t + \xi_{jt}$$

Where $\sigma_{jt}(\theta)$ denotes the share of good j at time t given parameters θ.

3.6.2 Supply

Let the monopolist's constant markup over the wholesale cost be denoted γ, so $p_{jt} = \gamma w_{jt}$. This assumption is identical to the one made by Nakamura and Zerom (2010) in their paper on cost-shock pass-through for coffee manufacturers, except that their retail firms set prices a constant amount above wholesale cost, whereas our firm sets a constant percentage markup.

As in Berry et al. (1995), conditional on the set of products it offers in a season and the set of competitors it faces, each brand statically solves the profit maximization problem in every month:

$$\max_{\{w_{jt}\}_{j \in J_{bt}}} \sum_{j \in J_{bt}} (w_{jt} - mc_{jt}) M_{st} \sigma_{jt}(\gamma w_t, \xi_t, \theta)$$

Brand b will have a first order condition for each $j \in J_b$:

$$\sigma_{jt}(\gamma w_t, \xi_t) + \sum_{k \in J_{bt}} (w_{kt} - mc_{kt}) \frac{\partial \sigma_{kt}(\gamma w_t, \xi_t, \theta)}{\partial p_t} = 0$$

Define the J_t by J_t matrix entry by entry as $\Delta_{jk} = \frac{-\partial \sigma_k}{\partial p_j} \mathbf{1}[j, k \text{ owned by same } b]$. Then the vector of marginal costs of all wholesalers at time t, mc_t is given as

$$mc_t(\theta) = w_t - \Delta(\gamma w_t, \xi_t, \theta)^{-1} \sigma(\gamma w_t, \xi_t, \theta) \tag{3.5}$$

3.6.3 Estimation and identification of the moment-generating model

Given the assumption $p_{jt} = \gamma w_{jt}$, if we observed time-varying wholesale costs then γ could be identified by $\sum p_{jt} / \sum w_{jt}$. Instead, the data includes a weighted sum $w_j = \frac{1}{T_j} \sum_t q_{jt} w_{jt}$ where T_j is the total number of time periods j is observed. Thus we estimate γ by:

$$\hat{\gamma} = \frac{\sum_j \frac{1}{T_j} \sum_t q_{jt} p_{jt}}{\sum_j w_j} \tag{3.6}$$

IV on the linear demand equation is used to identify the parameters $\theta \equiv \{\lambda_t, \lambda_j, \alpha\}$.

Once again, the estimating equation for the demand curve is:

$$\ln \sigma_{jt}(\theta) - \ln \sigma_{0t}(\theta) = \alpha p_{jt} + \lambda_j - \lambda_t + \xi_{jt}$$

Taking a time and cross-sectional difference yields $\tilde{\xi}_{jt} \equiv (\xi_{jt} - \bar{\xi}_j) - \frac{1}{N}\sum_j \left(\xi_{jt} - \bar{\xi}_j\right)$, with $\bar{\xi}_j \equiv \frac{1}{T}\sum_t \xi_{jt}$. $\tilde{\xi}_{jt}$ is assumed orthogonal to observed price instruments z_{jt}.

The instrument z_{jt} needs to vary with j and t—an instrument that varies only with time, or cross-sectionally, will be differenced out in the first stage due to the presence of time and good fixed effects. We use a combination of instruments: first, lagged wholesale price is a valid instrument under the assumption that ξ_{jst} is iid but that determinants of costs are persistent; second, different wholesalers are based in different countries, so good-specific exchange rates vary across goods so that these countries' exchange rates might be relevant for marginal costs and hence, wholesale costs.

3.6.4 Moment-generating model results

3.3 summarizes the results of the demand estimation. The inclusion of brand and time dummies goes far in solving the endogeneity problem between prices and unobserved (to the econometrician) quality; the average estimated price elasticity becomes larger than one, but not by much. Since we invert the demand system to recover marginal costs, it is crucial that the elasticities be reasonable, otherwise the marginal costs will be estimated too low—and perhaps even negative. The inclusion of instruments pushes the elasticity well above two, validating their efficacy.

In 3.7 we plot normalized marginal costs against normalized qualities (a random sampling of points is omitted). There is a very small, but nonetheless positive, relationship between cost and quality (coefficient of 0.01, $p < 0.05$). Although the positive correlation seems marginal, it may well be large enough to induce the kind of selection our intuition suggests. Moreover, it is clear that there is a huge degree of quality variation among goods with the same costs—variation that will certainly be important for welfare calculations.

3.7 Results

After fitting the model to the pre-shock data, we estimate the fixed cost of entry as 173880 rubles, while the mean of the exponential distribution of scrap values is estimated as 43470 rubles. Converted into dollars using the pre-shock exchange rate, these are roughly $5800USD$ and $1450USD$ respectively.

We are interested in comparing the performance of the model with quality heterogeneity to a more standard model with only cost heterogeneity. To that end, we fit a model with only cost heterogeneity to the data, and perform the same counterfactual on both models. Firms in the model with cost heterogeneity will have a fixed quality, set equal to the mean of qualities in the data. Our counterfactual is to double the brands' marginal ruble costs to capture the effects of the exchange rate shock.

3.7.1 Model Fit

First, we compute a measure of model fit by looking at the change in the price index. To construct a quantity weighted price index equivalent to the one in 3.3.3, we weight by the quantities in the pre-shock period and use the old and new prices to compute price indices for each period; We then take the ratio of these indexes. Denote with t objects pre-shock, and $t+1$ objects post-shock:

$$\Delta P = \frac{\sum_{x \in A} p_{t+1} \sigma_t(x) \tilde{s}_t(x)}{\sum_{x \in A} p_t \sigma_t(x) \tilde{s}_t(x)}$$

With only cost heterogeneity, the price index exhibits incomplete pass-through, with $\Delta P = 1.099$—the price index rises roughly 10%. As expected, with quality heterogeneity the price index increase is dampened: we find that $\Delta P = 1.043$, an increase of roughly 4%. This number is surprisingly close to the data despite not attempting to match this moment.

We also compute the average within-state change in prices for both models by taking the ratio of pre- and post-shock prices at each state, and then taking the simple average over those prices. The quality heterogeneity model exhibits within-state average price increases of 17.9% while the model without quality heterogeneity exhibits average price increases of

17.4%. Since we do not observe the marginal costs of the wholesalers themselves it is difficult to determine whether this is reasonable; we present it for completeness and to highlight the fact that in our logit model, markups do indeed adjust.

3.7.2 Counterfactual

To get a sense of whether or not our model can deliver reallocation towards low cost, low quality goods, we show how the distribution of firms in each cost state changes post shock. For the model with no quality heterogeneity, the cost quartile completely describes a firm's position in the distribution; for the model with heterogeneity, the cost quartile is the marginal distribution integrating over the distribution of qualities in that quartile.

Results are presented in 3.4. In the leftmost column are the discretized states used in computation, at the pre-shock levels. The first two columns present (1) the initial distribution of firms across states and (2) the ratio of the fraction of firms in that state post cost shock to the fraction of firms in that state pre cost shock. The last two columns present analogues for the model with quality heterogeneity, where the distribution over costs is now the marginal distribution (integrating over the quality dimension.) Notice that for the lowest cost level (marginal cost of 146.09 rubles) there is a 9 percent increase in the number of firms in the model with quality heterogeneity, and only a 6 percent increase in the number of firms in the model without quality heterogeneity.

There are two conclusions to be drawn from this analysis: first, since cost and quality are positively correlated in the data, the model can indeed generate a redistribution towards low cost, low quality goods. The introduction of a flexible second dimension of good heterogeneity offers a straightforward and intuitive solution to the question of why lower quality goods seem to benefit during a depreciation with no income shock. Second, the model with quality heterogeneity can actually induce a greater reallocation towards lower cost quartiles. As previously mentioned, a mechanism in the model that can rationalize this fact is the second-order impact of quality heterogeneity on the competitive environment.

Finally, to compute the change in welfare, we use the standard logit formula for the

consumer surplus to market participation:

$$W = \log\left(\sum_j \exp(\alpha p_j + \lambda_j - \lambda_0)\right)$$

Taking ΔW, the ratio of post to pre shock welfare, we find that, as expected, the model without quality heterogeneity predicts lower welfare losses after the exchange rate shock. The model with only cost heterogeneity has $\Delta W = 0.857$, a roughly 14% decrease in welfare, while the model with both cost and quality heterogeneity has $\Delta W = 0.791$, a roughly 21% decrease in welfare. Interestingly, the model with both cost and quality heterogeneity exhibits a smaller increase in the price index, but a fall in welfare that is 50% larger than the model without such heterogeneity. Quality heterogeneity evidently dampens the effect of the cost shock on the price index while simultaneously increasing the welfare cost to consumers through a lower quality selection of goods.

3.8 Conclusion

We use a novel online retail dataset that spans Russia's enormous currency depreciation in late 2014 to analyze the role of compositional shifts in product offerings for consumer welfare and price dynamics. We document in the data that there is a reallocation towards relatively low cost *and* low quality goods after the shock. We take care to show that this reallocation is not driven by an income shock or "flight from quality". For models with a single dimension of product heterogeneity this represents a puzzle: the highest quality firms and lowest cost foreign firms should remain after a cost shock, while low quality and high cost firms should exit.

To make progress on resolving this question, we construct a model of demand and firm dynamics where consumers value low prices and high qualities, and firms draw product cost and quality from a joint distribution on entry. A positive correlation between marginal cost and quality can explain quality downgrading as a byproduct of high cost, high quality firms exiting. We verify that the correlation holds in the data and calibrate the model to pre-shock moments. A counterfactual cost shock shows that high cost firms disproportionately exit

while the share of low cost, low quality firms increases, which has sizable repercussions for consumer welfare.

The paper shows how a richer model of product cost and quality heterogeneity can explain the quality downgrading observed in the data. It refrains from proving general statements about the conditions on the joint distribution of cost and quality that would be necessary to ensure quality downgrading in a crisis. Moreover, the model generates rich predictions on how the distribution of markups change in response to a cost shock; indeed, we rely on a consequence of those predictions when arguing for why the model with cost and quality heterogeneity induces a greater reallocation towards lower cost goods than the model with cost heterogeneity alone. A more thorough understanding of the patterns in markups and pass-through that can be generated by introducing meaningful quality heterogeneity is a fruitful avenue for future research.

3.9 Appendices

3.9.1 Appendix A: Tables

Table 3.1: Within-SKU Pass-through

	Dependent Variable: $\Delta\bar{p}_{i,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta_{\tau_1} usdrub_t$	0.911***	0.474***			-0.448	-0.011		
	[0.172]	[0.142]			[0.881]	[1.484]		
$\Delta_{\tau_2} usdrub_{t-\tau_1}$	-0.125	-0.016			-0.552	0.034		
	[0.192]	[0.322]			[0.726]	[1.081]		
$\Delta_{\tau_1} eurrub_t$			1.083***	0.624***			-0.111	2.004*
			[0.160]	[0.218]			[0.807]	[1.156]
$\Delta_{\tau_2} eurrub_{t-\tau_1}$			-0.262	-0.255			-0.462	-1.216
			[0.193]	[0.351]			[0.635]	[1.016]
SKU FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,942	1,303	14,942	1,303	1,106	97	1,106	97
R^2	0.471	0.355	0.479	0.363	0.280	0.353	0.280	0.442
Imported	Yes	Yes	Yes	Yes	No	No	No	No
Target Group	All	Shirts	All	Shirts	All	Shirts	All	Shirts

Table 3.2: Parameters and moments

Parameter	Interpretation	Identifying empirical moment
α	Coefficient on price in utility	Average elasticity
γ	Monopolist's markup	Avg. markup over obs. wholesale cost
λ_0	Outside good utility	Normalize minimum quality to 0
(μ, Σ)	*Ex ante* joint dist. of cost and quality	Discretized EDF
κ	Entry cost	Number of firms entering
K	Mean scrap value	Rate of firm exit

Table 3.3: Structural demand parameter estimates

	OLS	OLS	IV1	IV2
α	-0.00012***	-0.00032***	-0.00049***	-0.00052***
	(0.000)	(0.000)	(0.000)	(0.000)
Fixed effects		j,t	j,t	j,t
Instrument			lagged price	lagged price, exchange rate
$\alpha \cdot \gamma \cdot p(1-s)$ (Mean price elasticity)	0.416	1.25	2.08	2.26
R^2	0.045	0.554	0.758	0.7515
N	89058	69676	11857	9120

*** = .1%, ** = 1%, * = 5% significance. Fixed effect j,t denotes fixed effects at the product SKU, season -month level.

Table 3.4: Reallocation towards low cost goods

	No quality het.		Quality het.	
Marginal cost	Initial dist.	New/Initial	Initial dist.	New/Initial
146.09	0.37	1.06	0.40	1.09
305.44	0.23	0.82	0.31	0.83
590.89	0.18	0.67	0.17	0.57
2884.25	0.21	0.77	0.12	0.57

3.9.2 Appendix B: Figures

Figure 3.1: Overlapping Generations of Goods

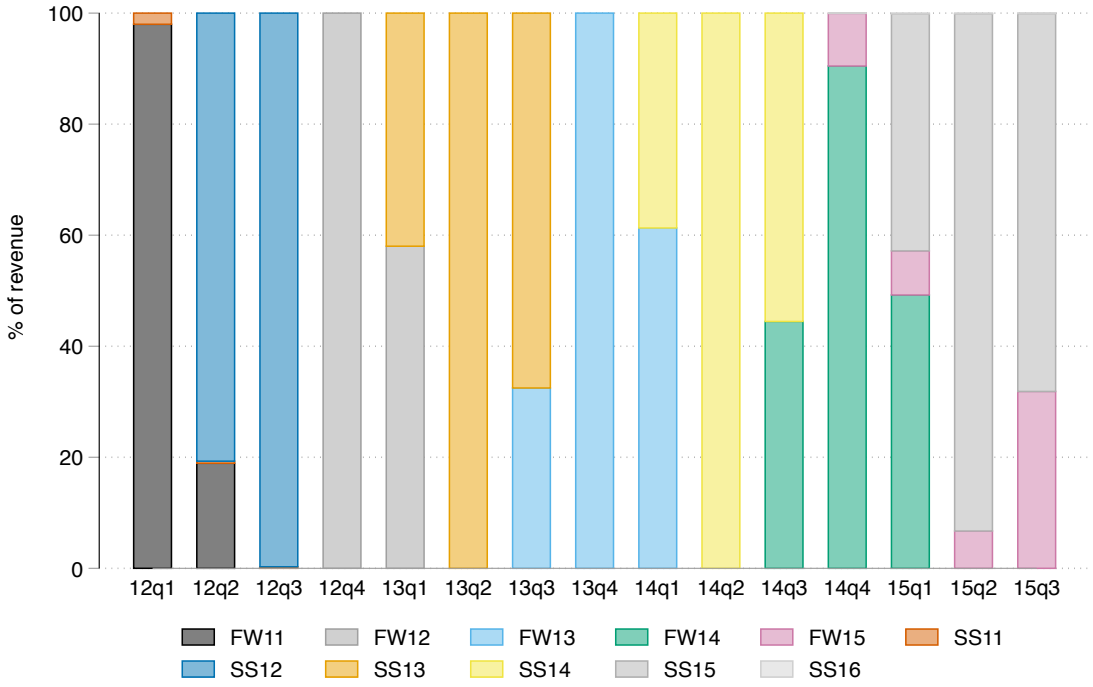


Figure 3.2: Cost of Goods Sold

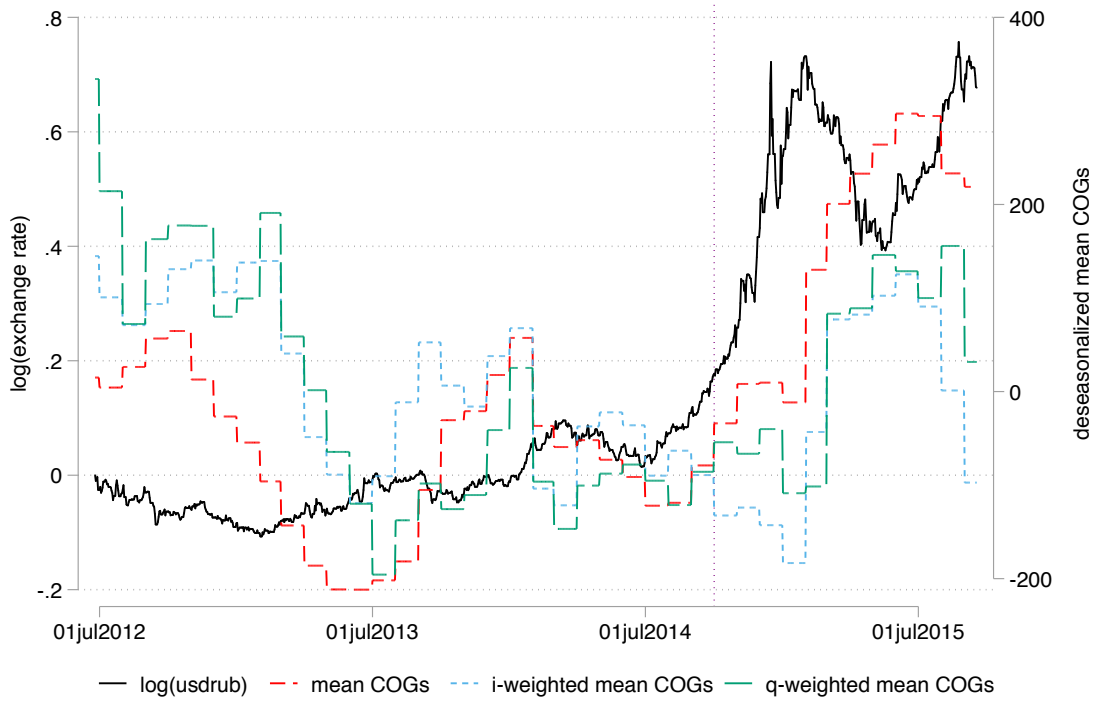


Figure 3.3: Price Indexes

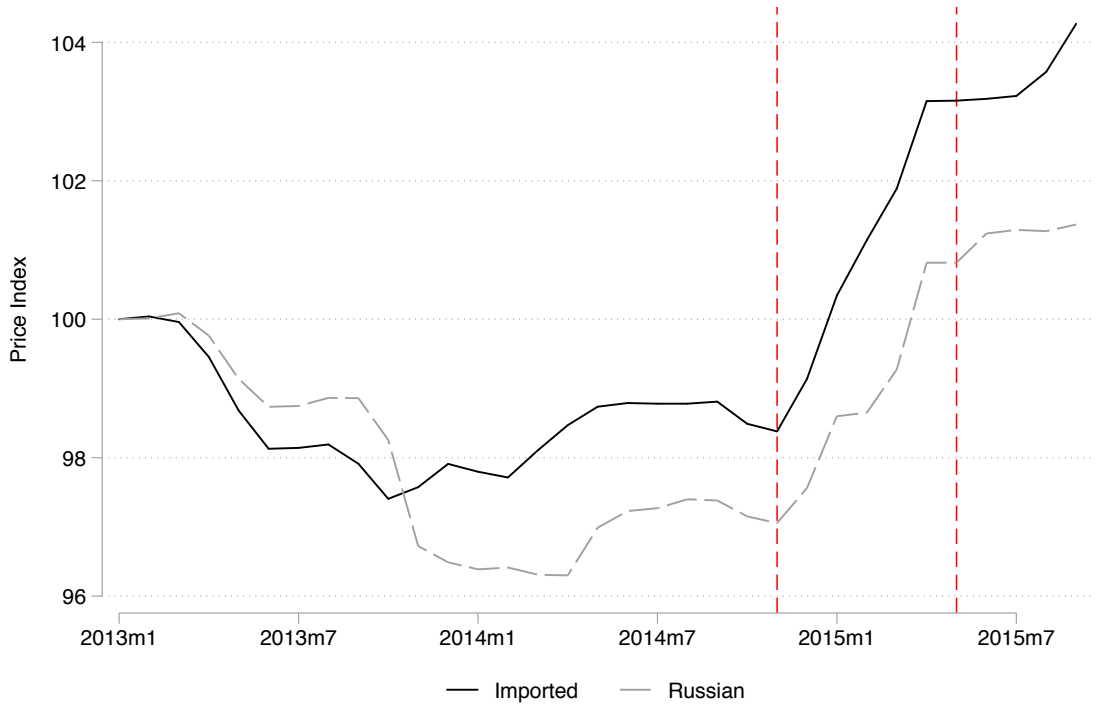
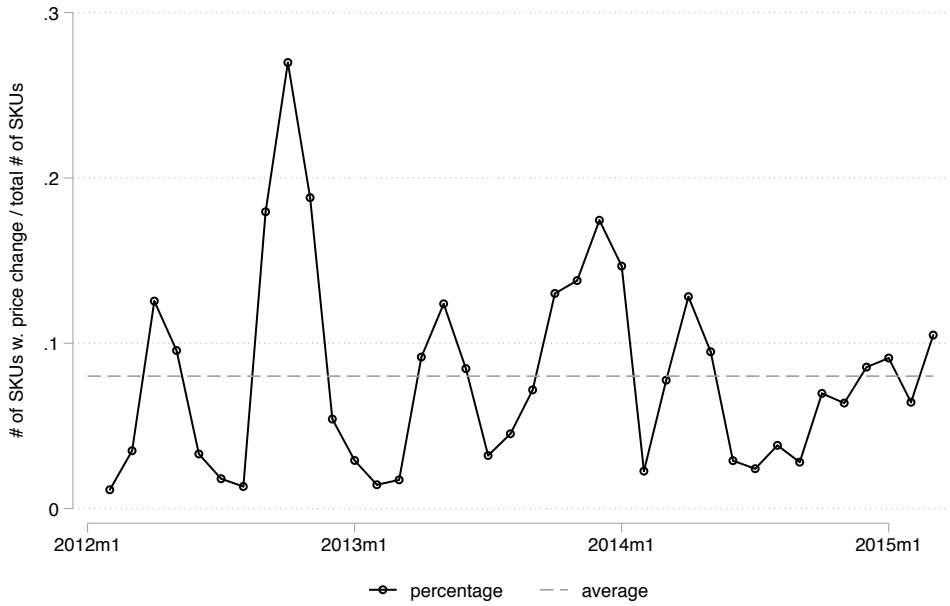


Figure 3.4: Price Adjustments

Frequency of Price Changes



Weighted Frequency of Prices Changes

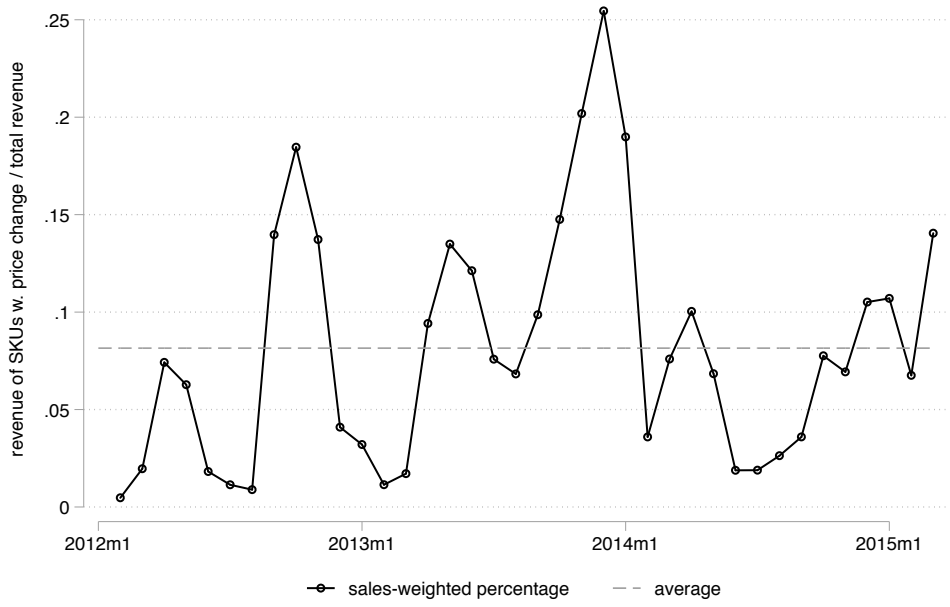
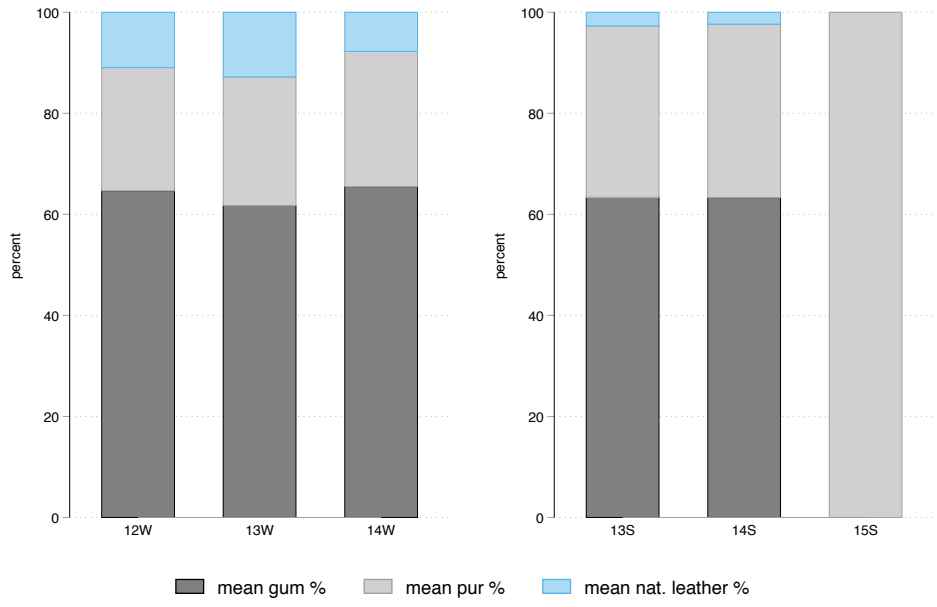


Figure 3.5: Fabrics & Quality Index

Shirt Fabric Types



Quality Index for Shirts

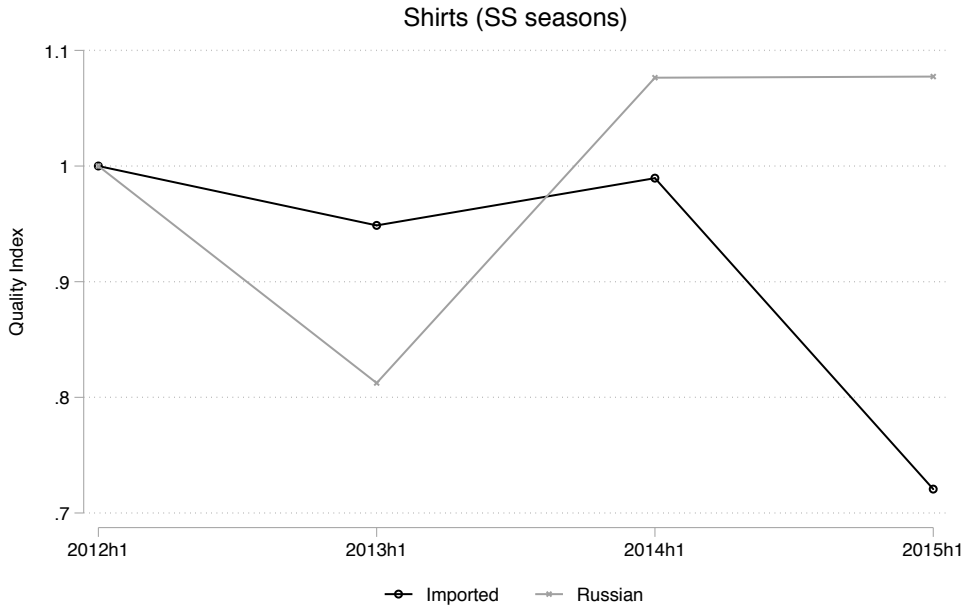
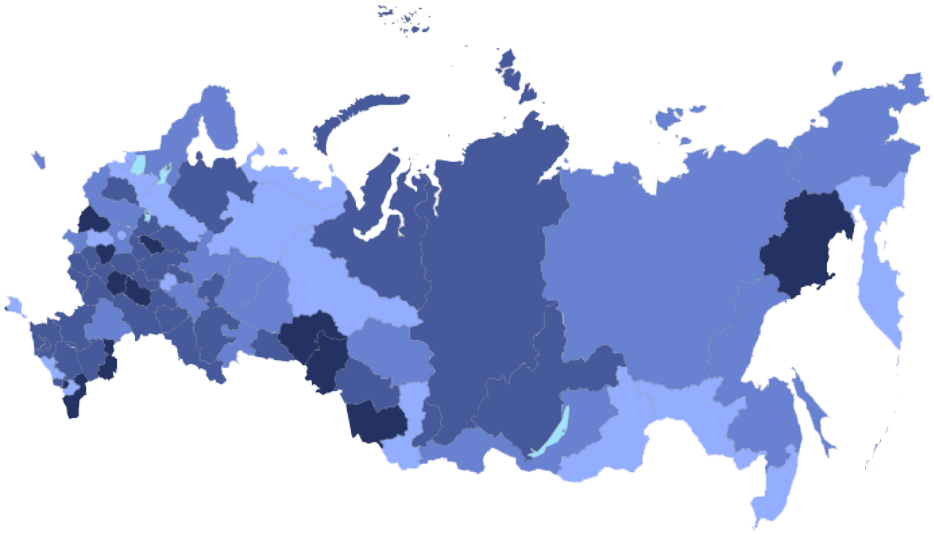


Figure 3.6: Demand Channel

Regional Growth (2015)



Income Effect

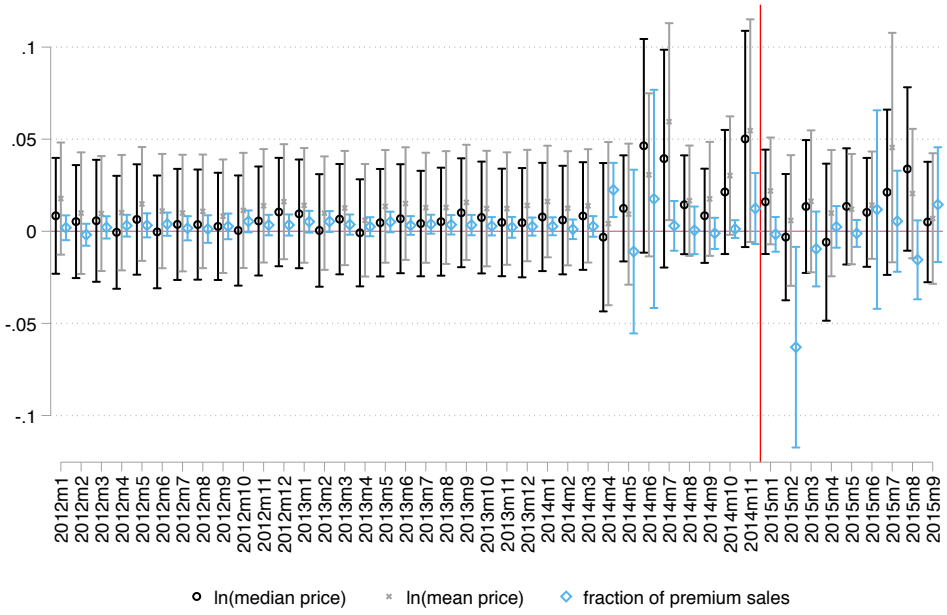
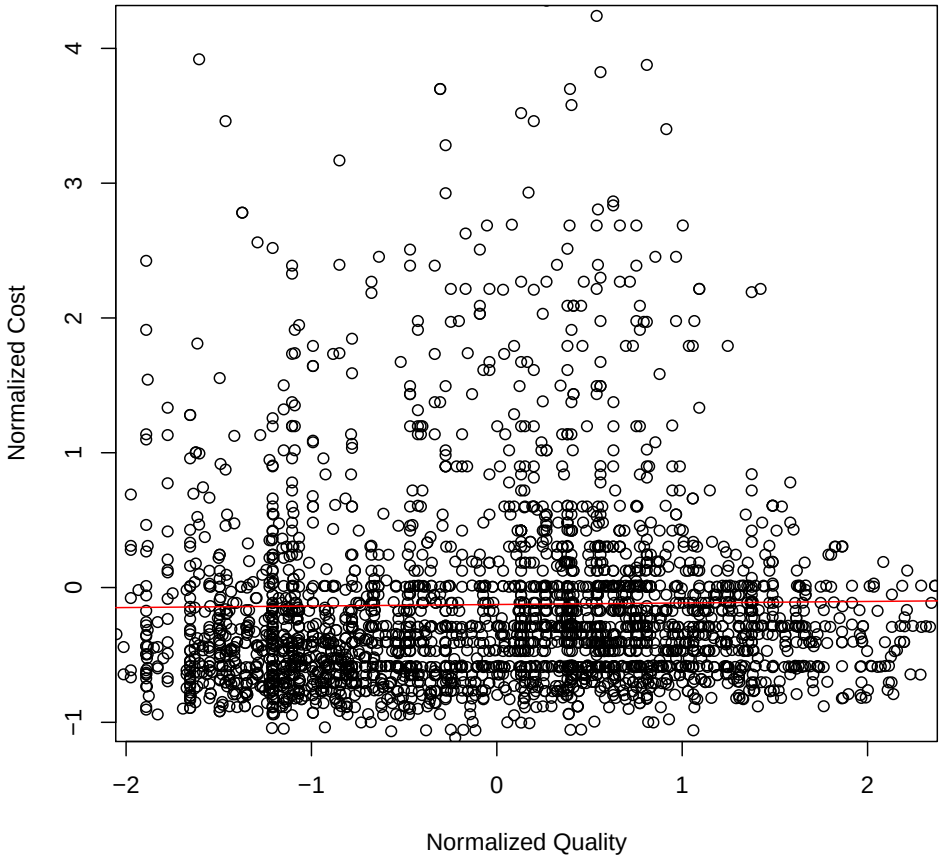


Figure 3.7: Small, positive relationship between cost and quality



Bibliography

Ambrus, Attila, Eric Chaney, and Igor Salitsky, "Pirates of the Mediterranean: An Empirical Investigation of Bargaining with Transaction Costs," Technical Report, Economic Research Initiatves at Duke 2016.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "Importers, Exporters, and Exchange Rate Disconnect," *The American Economic Review*, July 2014, *104* (7), 1942–78.

_ , _ , and _ , "International Shocks and Domestic Prices: How Large Are Strategic Complementarities?," *Working Paper*, 2016.

Argentesi, Elena and Lapo Filistrucchi, "Estimating Market Power in a Two-Sided Market: The Case of Newspapers," *Journal of Applied Econometrics*, 2007, *22* (7), 1247–1266.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-market, trade costs, and international relative prices," *The American Economic Review*, 2008, *98* (5), 1998–2031.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin, "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 2007, *75* (5), 1331–1370.

Becker, Gary S., Dennis W. Carlton, and Hal S. Sider, "Net Neutrality and Consumer Welfare," *Journal of Competition Law and Economics*, 2010, *6* (3), 497–519.

Bems, Rudolfs and Julian Di Giovanni, "Income-induced expenditure switching," *Working Paper*, 2014.

Berry, Steve and Elie Tamer, "Identification in Models of Oligopoly Entry," *Advances in Economics and Econometrics*, 2006, *2*, 46–85.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica*, 1995, pp. 841–890.

_ , _ , and _ , "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Berry, Steven T, "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 1994, pp. 242–262.

Bils, Mark and Peter J. Klenow, "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 2004, *112* (5), 947–985.

Björkegren, Daniel, "The Adoption of Network Goods: Evidence from the Spread of Mobile Phones in Rwanda," Technical Report, Brown University 2015.

Board, Oliver, "Competition and Disclosure," *The Journal of Industrial Economics*, 2009, *57* (1), 197–213.

Boik, Andre, "The Economics of Universal Serivce: An Analysis of Entry Subsidies for High Speed Broadband," *Information Economics and Policy*, 2017, *Forthcoming.*

Broda, Christian and David E Weinstein, "Understanding international price differences using barcode data," Technical Report, National Bureau of Economic Research 2008.

Brown, Jennifer, Tanjim Hossain, and John Morgan, "Shrouded Attributes and Information Suppression," *Quarterly Journal of Economics*, 2010, *125*, 859–876.

Burstein, Ariel and Nir Jaimovich, "Understanding movements in aggregate and product-level real exchange rates," *Working Paper*, 2012.

_ , Martin Eichenbaum, and Sergio Rebelo, "Large Devaluations and the Real Exchange Rate," *Journal of Political Economy*, 2005, *113* (4), 742–784.

Burstein, Ariel T, Joao C Neves, and Sergio Rebelo, "Distribution costs and real exchange rate dynamics during exchange-rate-based stabilizations," *Journal of Monetary Economics*, 2003, *50* (6), 1189–1214.

Cai, Hongbin and Ichiro Obara, "Firm Reputation and Horizontal Integration," *The RAND Journal of Economics*, 2009, *40* (2), 340–363.

Chandra, Ambarish and Allan Collard-Wexler, "Mergers in Two-Sided Markets: An Application to the Canadian Newspaper Industry," *Journal of Economics and Management Strategy*, 2009, *18*, 1045–1070.

Chen, Natalie and Luciana Juvenal, "Quality and the great trade collapse," Technical Report, University of Warwick, Department of Economics 2015.

_ and _ , "Quality, trade, and exchange rate pass-through," *Journal of International Economics*, 2016, *100*, 61–80.

Coibion, Olivier, Yuriy Gorodnichenko, and Gee Hee Hong, "The Cyclicality of Sales, Regular and Effective Prices: Business Cycle and Policy Implications," *The American Economic Review*, 2015, *105* (3), 993–1029.

Cravino, Javier and Andrei A. Levchenko, "The distributional consequences of large devaluations," Technical Report, National Bureau of Economic Research 2016.

Crawford, Gregory and Ali Yurukoglu, "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 2012, *102* (2), 643–685.

Dafny, Leemore, Kate Ho, and Robin Lee, "The Price Effects of Cross-Market Hospital Mergers," Technical Report, National Bureau of Economic Research 2016.

Devereux, Michael B and Charles Engel, "Exchange rate pass-through, exchange rate volatility, and exchange rate disconnect," *Journal of Monetary Economics*, 2002, *49* (5), 913–940.

Dranove, David and Ginger Zhe Jin, "Quality Disclosure and Certification: Theory and Practice," *Journal of Economic Literature*, 2010, *49* (4), 935–963.

Dutz, Mark, Jonathan Orszag, and Robert Willig, "The Liftoff of Consumer Benefits from the Broadband Revolution," *Review of Network Economics*, 2012, *11* (4).

Economides, Nicholas and Benjamin E. Hermalin, "The Economics of Network Neutrality," *RAND Journal of Economics*, 2012, *43* (4), 602–629.

Eichenbaum, Martin, Nir Jaimovich, Sergio Rebelo et al., "Reference Prices, Costs, and Nominal Rigidities," *The American Economic Review*, 2011, *101* (1), 234–62.

Ellison, Glenn and Alexander Wolitzky, "A Search Cost Model of Obfuscation," *The RAND Journal of Economics*, 2012, *43*, 417–441.

_ and Sara Fisher Ellison, "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 2009, *77*, 427–452.

Ericson, Richard and Ariel Pakes, "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 1995, *62* (1), 53–82.

Evans, David S., "The Antitrust Economics of Multi-Sided Platform Markets," *Yale Journal on Regulation*, 2003, *20* (2), 325–381.

_ , "The Web Economy, Two-Sided Markets, and Competition Policy," *Working Paper, University of Chicago*, 2010.

_ and Richard Schmalensee, "The Antitrust Analysis of Multi-sided Platform Bussiness," Technical Report, National Bureau of Economic Research 2013.

Feenstra, Robert C, Joseph E Gagnon, and Michael M Knetter, "Market share and exchange rate pass-through in world automobile trade," *Journal of International Economics*, 1996, *40* (1), 187–207.

Fitzgerald, Doireann and Stefanie Haller, "Pricing-to-market: evidence from plant-level prices," *The Review of Economic Studies*, 2014, *81* (2), 761–786.

Forbes, Silke J., Mara Lederman, and Michael Wither, "Quality Disclosure when Firms Set Their Own Quality Targets," Technical Report, Rotman School of Management 2017.

_ , _ , and Trevor Tombe, "Do Firms Game Quality Ratings? Evidence from Mandatory Disclosure of Airline On-Time Performance," Technical Report, Rotman School of Management 2011.

Fudenberg, Drew and Jean Tirole, "Sequential Bargaining with Incomplete Information," *Review of Economic Studies*, 1983, *50*, 221–247.

Gabaix, Xaivier and David Laibson, "Shrouded Attributes, Consumer Mypoia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 2006, *121*, 505–540.

Gans, Joshua, "Weak Versus Strong Network Neutrality," *Journal of Regulatory Economics*, 2015, *47* (2), 183–200.

Gardete, Pedro M., "Cheap-Talk Advertising and Misrepresentation in Vertically Differentiated Markets," *Marketing Science*, 2013, *32* (4), 609–621.

Goeree, Michelle Sovinsky, "Limited Information and Advertising in the U.S. Personal Computer Industry," *Econometrica*, 2008, *76* (5), 1017–1074.

Goldberg, Linda S and José Manuel Campa, "The sensitivity of the CPI to exchange rates: Distribution margins, imported inputs, and trade exposure," *The Review of Economics and Statistics*, 2010, *92* (2), 392–407.

Goldberg, Pinelopi Koujianou and Rebecca Hellerstein, "A structural approach to identifying the sources of local currency price stability," *The Review of Economic Studies*, 2012, p. rds015.

Gopinath, Gita and Oleg Itskhoki, "Frequency of Price Adjustment and Pass-Through," *The Quarterly Journal of Economics*, 2010, *125* (2), 675–727.

_ and Roberto Rigobon, "Sticky Borders," *The Quarterly Journal of Economics*, 2008, *123* (2), 531–575.

_ , Pierre-Olivier Gourinchas, Chang-Tai Hsieh, and Nicholas Li, "International prices, costs, and markup differences," *The American Economic Review*, 2011, *101* (6), 2450–2486.

Gowrinsankaran, Gautam, Aviv Nevo, and Robert Town, "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry," *American Economic Review*, 2015, *105* (1), 172–203.

Grennan, Matthew, "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices," *American Economic Review*, 2013, *103* (1), 145–177.

Grossman, Sanford J., "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 1981, *24*, 461–489.

Hansen, Lars Peter, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50* (3), 1029–54.

Honoré, Bo and Aureo de Paula, "Interdependent Durations," *Review of Economic Studies*, 2010, *77* (3).

Jia, Panle, "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Jin, Ginger Zhe, "The Effect of Information on Product Quality," *The Quarterly Journal of Economics*, 2003, *118*, 409–451.

_ , "Competition and Disclosure Incentives: an Empirical Study of HMOs," *RAND Journal of Economics*, 2005, *36* (1), 93–112.

Jovanovic, Boyan, "Truthful Disclosure of Information," *Bell Journal of Economics*, 1982, *13*, 36–44.

Kehoe, Patrick and Virgiliu Midrigan, "Prices are sticky after all," *Journal of Monetary Economics*, 2015, *75*, 35–53.

Khandelwal, Amit, "The long and short (of) quality ladders," *The Review of Economic Studies*, 2010, *77* (4), 1450–1476.

Lee, Robin S. and Ariel Pakes, "Multiple Equilibria and Selection by Learning in an Applied Setting," *Economics Letters*, 2009, *104* (1), 13–16.

_ and Tim Wu, "Subsidizing Creativity through Network Design: Zero-Pricing and Net Neutrality," *Journal of Economic Perspectives*, 2009, *23* (3), 61–76.

Mathios, Alan D., "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market," *The Journal of Law & Economics*, 2000, *43* (2), 651–678.

Medina, Pamela, "Within-firm responses to import competition: quality upgrading and exporting in the Peruvian apparel industry," *Job Market Paper*, 2017.

Merlo, Antonio and Charles Wilson, "A Stochastic Model of Sequential Barganing with Complete Information," *Econometrica*, 1995, *63*, 317–399.

_ and Xun Tang, "Identification and Estimation of Stochastic Bargaining Models," *Econometrica*, 2012, *80* (4), 1563–1604.

Milgrom, Paul R., "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 1981, *18*, 380–391.

Nakamura, Emi and Dawit Zerom, "Accounting for Incomplete Pass-Through," *The Review of Economic Studies*, 2010, *77* (3), 1192–1230.

Nevo, Aviv, John L. Turner, and Jonathan W. Williams, "Usage-Based Pricing and Demand for Residential Broadband," *Econometrica*, 2016, *84*, 411–443.

Peitz, Martin and Florian Schuett, "Net Neutrality and Inflation of Traffic," *International Journal of Industrial Organization*, 2016, *46*, 16–62.

Scherbakov, Oleksander, "Measuring Consumer Switching Costs in the Television Industry," Technical Report 2015.

Stiglitz, Joseph E., "The Theory of 'Screening,' Education, and the Distribution of Income," *American Economic Review*, 1975, *65*, 66–74.

Weintraub, Gabriel Y., C. Lanier Benkard, and Benjamin Van Roy, "Markov Perfect Industry Dynamics With Many Firms," *Econometrica*, 2008, *76* (6), 1375–1411.

Wollman, Thomas, "Trucks Without Bailouts: Equilibrium Product Characteristics for Commerical Vehicles," Technical Report, University of Chicago Booth School of Business 2016.

Yildiz, Muhamet, "Waiting to Persuade," *Quarterly Journal of Economics*, 2004, *119* (1), 223–248.