10.14754/CEU.2018.11

Three Essays in Industrial Organization

by

Péter Zsohár

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy at
Central European University

Supervisor: Miklós Koren

Budapest, Hungary

10.14754/CEU.2018.11

10.14754/CEU.2018.11

CENTRAL EUROPEAN UNIVERSITY
DEPARTMENT OF ECONOMICS AND BUSINESS

The undersigned hereby certify that they have read and recommend to the Department of Economics and Business for acceptance a thesis entitled **"Three Essays in Industrial Organization"** by **Peter Zsohar**.
Dated: November 29, 2018

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Chair of the Thesis Committee:

Laszlo Matyas

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Advisor:

Miklos Koren

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Internal Examiner:

Sergey Lychagin

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
External Examiner:

Robert Somogyi

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
External Member:

Balazs Murakozy

10.14754/CEU.2018.11

Central European University
Department of Economics and Business

Author:	Péter Zsohár
Title:	Three Essays in Industrial Organization
Degree:	Ph.D.
Dated:	November 29, 2018

Hereby I testify that this thesis contains no material accepted for any other degree at any other institution and that it contains no material previously written and/or published by another person except where appropriate acknowledgement is made.

Signature of the author: ....

10.14754/CEU.2018.11

Abstract

The thesis consists of three chapters. The first two papers are related to the economic policy question of network neutrality, while the third paper studies competition on the retail gasoline market. The three papers build on a diverse set of tools. The first chapter is a computational theory paper where the main emphasis is on a new theoretical modeling framework. The second chapter involves empirical data analysis and numerical methods to conduct counterfactual analysis related to an actual economic policy matter. The third chapter is a methodological paper where I develop and demonstrate the working of a new empirical framework.

Chapter 1
Net Neutrality in a Dynamic Platform Market Environment

The contribution of this paper is a dynamic industry model that enables researchers and policy makers to directly analyze the anecdotal positive feedback loop (the so called "virtuous cycle") that lies at the heart of the Federal Communication Commission's (FCC) Open Internet Order that aims to establish the highly debated net neutrality principle as an effective rule on the Internet. The discrete time multi-agent dynamic stochastic game is an extended version of the framework of Ericson and Pakes (1995) which is played by two types of players (i) two Content Providers (CPs) differentiated by their product quality and (ii) a single Internet Service Provider (ISP) characterized by an actual network capacity. Both types have costly dynamic controls over their states. Agents' per-period payoffs represent profits from a platform market game. The novelty in my approach is that it allows for (i) both the no termination fee and the no paid priority forms of net neutrality, (ii) endogenous price setting and side-payments among agents in the stage game, (iii) dynamic restructuring of the industry as a response to the change in regulation, and (iv) direct analysis of the effects of net neutrality on the virtuous cycle. Although these features allow for more realistic interactions among agents they come at the price of analytic tractability. Hence, the model is solved numerically for a range of parameters

by standard Newton-Raphson and Gaussian methods. Then I compare equilibrium strategies under four different regimes (i) net neutrality, (ii) termination fee, (iii) paid priority, and (iv) simultaneous termination and priority fees. In the comparisons I focus on three key outcomes: investment to network capacity, innovation on the content provider side and consumer welfare.

Keywords: net neutrality, dynamic game, dynamic programming

Chapter 2

Counterfactual Analysis of Net Neutrality in a Calibrated Model

The Federal Communication Commission's (FCC) Open Internet Order seeks to regulate the Internet relying on an anecdotal feedback loop, the Virtuous Circle. Despite the scale of the issue there is very little or no empirical evidence that supports the claims of proponents or opponents of the proposed regulation. This paper makes an attempt to produce comparable ballpark figures for the potential economic effects of net neutrality in the United States. I use the model developed in my first chapter to study the outcomes of the counterfactual regulatory regimes in consideration. In order to reflect the actual market environment and to accommodate a greater number of firms it is necessary to increase the state space. However, this inflates the computational burden dramatically and deterministic solution methods are no longer applicable. For the dynamic game I propose a reinforcement learning algorithm closely following the stochastic algorithm of Pakes and McGuire (2001) but modified to handle the two types of agents in my model. To solve the stage game I use a parallelized homotopy algorithm to increase efficiency and robustness to different sets of parameter values. Then I calibrate the fundamental parameters of the model using various public data sources to match key observed moments. Using the computed equilibrium policies from all regulatory regimes I compare the alternative outcomes.

Keywords: net neutrality, counterfactual analysis, reinforcement learning, homotopy method

Chapter 3
Delineation of Market Areas Using Sparse Learning and Spatial Regularization

Market definition is a key part in industrial economic analyses both in antitrust and business cases. However, current state-of-the-art methods require either too strong assumptions about the market environment or too many proprietary data sources. These features may make them unattractive or even inapplicable in court cases or in the everyday business setting. This paper applies modern statistical methods to analyze the mutual influence among prices of competing products. The key idea is to apply spatial regularization via the fused lasso to filter the noisy price data and to get rid of spurious associations. Then the procedure described in the paper can be used to identify market boundaries and to analyze the existence of pricing pressure in arbitrary product or geographic spaces. The main advantages of the method are (i) it is simple, (ii) it requires only publicly available data, (iii) it doesn't rely on any specific theoretical model, and (iv) it extends the current bivariate time-series analysis based methods to high dimensional settings with many products. To demonstrate the potential usefulness the method is applied to weekly consumer prices of gasoline stations in Hungary to highlight markets that could be potentially harmed by a hypothetical merger between to competitors.
Keywords: market definition, spatial pricing pressure, fused lasso, gasoline

10.14754/CEU.2018.11

Acknowledgements

I am indebted to my advisor, Miklós Koren who offered excellent professional and personal guidance throughout the years. Working closely with him on several research projects played a key role in my professional development. His integrity and attitude towards honest science will serve as a model for my work in the rest of my life.

I wish to thank Ariel Pakes for recognizing my passion for rigorous empirical work and offering his help in deepening my knowledge in the field. The year I spent at Harvard University under his supervision was a great learning experience and an exceptionally inspiring year of my life.

I am very grateful to the examiners of my thesis, Sergey Lychagin and Róbert Somogyi who gave lots of constructive comments that helped me improve the quality of my papers.

The Department of Economics of Central European University provided an intellectually stimulating academic environment for my studies. I benefited a lot from conversations with professors and my peers in the Ph.D. program.

I greatly acknowledge financial support of the Rosztoczy Foundation and the Pallas Athéné Domus Scientiae Foundation.

I am grateful to my parents who always encouraged me to pursue my own goals and career path. In the recent years of my life nothing could have been as refreshing and joyful as spending time with my children, Abigél and Olivér. Last but not least, I would like to thank my wife, Rita, for her never ending selfless support. Life would have been much less fun without you.

10.14754/CEU.2018.11

Contents

List of Tables viii

List of Figures ix

1 Net Neutrality in a Dynamic Platform Market Environment 1

1.1 Introduction 1

1.2 Net neutrality 3

1.2.1 Motivating example 3

1.2.2 The FCC and the virtuous cycle 4

1.2.3 Concerns and harmful incentives 5

1.2.4 Dominance of video streaming 7

1.3 Related literature 8

1.4 Model 9

1.4.1 Overview 9

1.4.2 Agents 12

1.4.3 Congestion 15

1.4.4 Termination and paid priority 17

1.4.5 Static equilibrium 20

1.4.6 Dynamic framework 21

1.5 Lessons From the Analysis 24

1.5.1 Static equilibrium 25

10.14754/CEU.2018.11

1.5.2 Dynamic equilibrium 25
1.6 Extensions 31
1.7 Alternative equilibrium concepts 32
1.8 Conclusion 33

2 Counterfactual Analysis of Net Neutrality in a Calibrated Model 35
2.1 Introduction 35
2.2 Data sources 36
2.3 Calibration 37
2.3.1 Annual timing assumption 37
2.3.2 Static part 38
2.3.3 Calibrated values and sanity checks 42
2.3.4 Dynamic game 44
2.4 Computational algorithm 46
2.4.1 Homotopy method 46
2.4.2 Reinforcement learning 48
2.5 Results 50
2.5.1 Static counterfactuals 50
2.5.2 Dynamic equilibria 52
2.6 Conclusion 52

3 Delineation of Market Areas Using Sparse Learning and Spatial Regularization 54
3.1 Introduction 54
3.2 Existing Empirical Approaches 55
3.2.1 Price tests 55
3.2.2 Structural approach 56
3.3 The Proposed Method 57

3.3.1 Intuition behind the methodology 57
3.3.2 Elements of the statistical model 59
3.3.3 Full model 63
3.3.4 Tuning Parameters 63
3.3.5 Translating coefficients to competitors 64
3.3.6 Discussion 67
3.4 Simulations 68
3.4.1 Results 69
3.5 Application 70
3.5.1 Spatial structure 71
3.5.2 Prices 72
3.5.3 Estimation 75
3.5.4 Results 76
3.5.5 Hypothetical merger 77
3.6 Conclusion 77

Bibliography **79**

A Appendix for Chapter 1 **84**
A.1 Simple Model with Net Neutrality 84
A.1.1 Static equilibrium with constant bitrates 85
A.1.2 Static equilibrium with increasing bitrates 87
A.1.3 Dynamic equilibrium with CP adjustment 87
A.1.4 Dynamic equilibrium with both CP and ISP adjustment 91
A.2 Effects of Termination Fees 91
A.3 Effect of Paid Prioritization 96

10.14754/CEU.2018.11

List of Tables

1.1 Summary of static equilibria 26
1.2 Summary of dynamic equilibria 29

2.1 Summary of static counterfactuals, monthly values 51
2.2 Summary of dynamic equilibria, annual values 53

3.1 Simulation results – Estimated market boundaries 70
3.2 Summary statistics of estimation results 76
3.3 Summary statistics of estimation results – without extreme values 77

10.14754/CEU.2018.11

List of Figures

1.1 The Virtuous Cycle of investment, innovation and consumer demand 4
1.2 Stage game flowchart. Solid lines indicate monetary flows while dashed lines indicate network traffic. 20
1.3 CP prices in static equilibrium 26
1.4 ISP prices in static equilibrium 27
1.5 CP profits in static equilibrium 27
1.6 ISP profit in static equilibrium 27
1.7 ISP profits in static equilibrium as a function of termination and priority fees 28
1.8 CP value function in dynamic equilibrium 28
1.9 ISP value function in dynamic equilibrium 29
1.10 CP investment in dynamic equilibrium 30
1.11 ISP investment in dynamic equilibrium 30
1.12 Stationary distribution of the system in the dynamic equilibrium 30

2.1 Evolution of daily average relative delays in video streaming 39
2.2 The evolution of download traffic for the average user over the day 43
2.3 Daily total peak hour download traffic 44
2.4 Daily total peak hour download traffic against daily average relative delays 46

10.14754/CEU.2018.11

3.1 Illustration of how estimated model coefficients are translated to competitors . . . 66

3.2 Simulated data set . . . 69

3.3 An example of a typical $\hat{\beta}$ with $R^2 = 0.99$. . . . 70

3.4 Spatial Structure of Gasoline Stations . . . 72

3.5 Trend and correlation of prices . . . 73

3.6 Distribution of available observations . . . 75

3.7 Network of Competition . . . 78

3.8 Markets with at least 15% change in HHI . . . 78

A.1 Prices and profits in the net neutrality model with constant bitrates . . . 85

A.2 Market shares, consumer surplus and network traffic in the net neutrality model with constant bitrates . . . 86

A.3 Decreasing marginal utility of content resolution . . . 87

A.4 Prices and profits in the net neutrality model with increasing bitrates . . . 88

A.5 Market shares, consumer surplus and network traffic in the net neutrality model with increasing bitrates . . . 89

A.6 Content provider states and corresponding product quality levels and bitrates . . . 89

A.7 Value functions and investment in the net neutrality model with CP adjustment. Note that for better visibility the horizontal axis shows states (ω) rather than the corresponding qualities (δ). . . . 90

A.8 Value functions and investment in the model with net neutrality . . . 92

A.9 Stationary distributions in the model with net neutrality . . . 92

A.10 Prices and profits in the model with termination fees . . . 93

A.11 Optimal termination fee schedule in the model with termination fees . . . 94

A.12 Market shares, consumer surplus and network traffic in the model with termination fees 94
A.13 Value functions and investment in the model with termination fees 95
A.14 Stationary distributions in the model with termination fees 95
A.15 Prices and profits in the model with paid prioritization . . 97
A.16 Optimal priority fees in the model with paid prioritization 97
A.17 Market shares, consumer surplus and network traffic in the model with paid prioritization 98
A.18 Value functions and investment in the model with paid prioritization 99
A.19 Stationary distributions in the model with paid prioritization 99

10.14754/CEU.2018.11

Chapter 1

Net Neutrality in a Dynamic Platform Market Environment

1.1 Introduction

Net neutrality is best understood as a network design principle. It requires all bits traveling over the internet to be treated equally. The internet, however, is not a single network. It is a complex structure of several interconnected networks of different sizes and scales. These specific interconnection points where networks exchange traffic connect hundreds of millions of users at the edges of the networks day-by-day. It is not unrealistic to consider the internet as the most important platform for economic growth, innovation, competition, and free expression in the 21st century. Net neutrality, however, restricts the way how interconnecting agents in the internet ecosystem are allowed to interact which may change their incentives for investments or innovation. Potentially this may have significant effects on the evolution of the network through the functioning of the interconnection relationships, and as a consequence, on our society as a whole.

There is a long standing public debate over the potential effects of the regulation and, not surprisingly, net neutrality became a frequent guest in US news headlines. Nothing illustrates better the importance of net neutrality than the fact that in 2014 the Federal Communications Commission's proposed Open Internet Order – that was meant to fix net neutrality as a rule – generated nearly 4 million public comments on its website.

Given the scale of the potential long-run effects of net neutrality one would expect an exceptionally thorough economic analysis by the regulator. However, the FCC received substantial critiques for its lack of proper economic analyses both from policymakers and academic economists alike.[1] Even Tim Brennan, the former chief economist of the FCC admitted that in the Open Internet Order *"a fair amount of the economics was wrong, unsupported, or irrelevant"*.[2] Therefore, it is of great importance to better understand what kind of interactions and incentives (the lack of) net neutrality regulations might induce and to what market outcomes they might lead. Indeed, the bulk of the economics literature focuses on the effects of the two rules proposed by the FCC. The great variety of models that each focus on certain aspects of the problem in isolation lead to better understanding of agents' incentives. However, the range of controversial findings signals how important it is to consider the effects of the proposed interventions in a single model where players' incentives and possible interactions are accurately represented.

My contribution to the literature is a model that enables researchers and policymakers to study either the individual or the combined effects of the two proposed rules on a variety of market outcomes in a single framework while allowing for a rich set of interactions among agents. An additional distinctive feature of my model is that is able to generate the kind of firm behavior that the FCC uses in its argumentation for its net neutrality rules. This makes it possible to directly assess the arguments in the Open Internet Order. With that in mind, I use the model to study the effects of the ban on termination fees and paid priority services in a baseline calibration where I take all values from the industry model of Pakes and McGuire (1994).

I find that that ISPs underinvest in the regimes with paid prioritization to earn profits from the priority fees and this will decrease CPs' returns to investment. As the equilibrium can be computed only by numerical methods, I perform robustness checks with alternative parameter settings. The main results are qualitatively the same after the perturbations.

The rest of the paper is organized in the following way. Section 1.2 gives

[1]See Katz (2017) for specific net neutrality related issues or Faulhaber et al. (2017) for a general critique of the FCC's economics related practices.

[2]Source: Brennan (2016).

10.14754/CEU.2018.11

a more detailed introduction to net neutrality through discussing the related questions and concerns while section 1.3 provides a brief overview of the related literature. In section 1.4 I describe the details of the model. In section 1.5 I specify a baseline set of parameters and present the results of my analysis. Sections 1.6 and 1.7 discuss potential extensions and alternative equilibrium concepts.

1.2 Net neutrality

1.2.1 Motivating example

It is instructive to illustrate the typical questions that arise regarding the consequences of the lack of net neutrality through an example. Following Greenstein et al. (2016) I use the typical example of Comcast, the largest cable and broadband provider in the U.S., and Netflix, the giant television and movie-streaming service. Suppose that on a Friday night you want to watch a movie on Netflix. Unfortunately, however, you are not alone with your plans for the evening. Millions of fellow Comcast subscribers intend to do the same, the majority of whom will have to experience serious performance issues because Comcast's network capacity has reached its limits and the network is congested. This immediately leads us to a seemingly endless set of question: What would happen if Comcast asked Netflix to pay for faster and more reliable access to its subscribers? Would Netflix be likely to agree to this request? Would Netflix charge you more for the movie? Would Comcast raise its broadband subscription fee for this improved service? How would the deal affect Comcast's future investments into network capacity? What would happen to the rest of the movie streaming services, like e.g. Amazon Instant Video or Youtube? Would that deter or encourage new firms to enter the market? If such a deal was struck, in what ways would consumer or producer surplus change? The example above is about paid prioritization, one of the two main network management practices that the FCC has banned entirely. The other one being the ban on blocking, or more generally a ban on termination fees. In the context of the previous example imagine that Comcast would demand a price (a termination fee) from Netflix for each of its subscribers or otherwise it would block

10.14754/CEU.2018.11

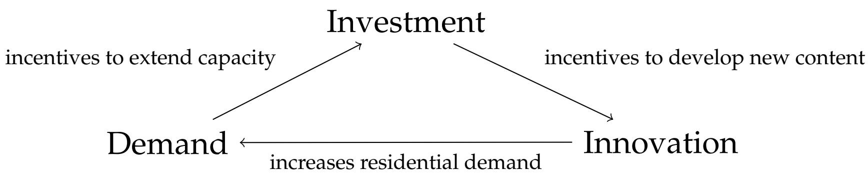


Figure 1.1: The Virtuous Cycle of investment, innovation and consumer demand

its contents. All the questions above apply here as well.

It turns out that such deals are not unprecedented. E.g. in 2003 Cox and Comcast blocked VPNs, in 2005 Madison River Communications blocked VoIP services, in 2007 Comcast throttled peer-to-peer programs, in 2012 AT&T restricted the use of Apple's FaceTime to certain customers, and in 2014 Comcast gave Netflix a special treatment for a cost – just to name a few significant cases.

1.2.2 The FCC and the virtuous cycle

To address the issue the FCC released its Open Internet Order to promote and preserve an open internet. They argue that net neutrality is essential in achieving these goals and a ban on termination fees and paid prioritization would suffice to ensure neutrality. However, as I pointed out in the introduction, the FCC did not perform a proper rigorous economic analysis and their whole argumentation was based on an anecdotal positive feedback loop of innovation, demand and investment, called the *virtuous cycle* and presented visually on Figure 1.1. The FCC informally argues that given abundant network capacity the innovative uses of the network by content providers (e.g. Netflix, Youtube, Facebook, Google etc.) lead to increased end-user demand for the broadband services of telco and cable companies (e.g. Comcast, Verizon, Time Warner Cable, AT&T, etc.), which strengthens incentives for network infrastructure developments, which in turn lead to further innovative uses of the network, and so on. As one can see, the main trade-off to consider when trying to assess net neutrality is related to the *balance of innovation and investment*.

Proponents of net neutrality emphasize that any price raise or termination fee charged by internet service providers would create a barrier to entry for content providers and also negatively impact the already active

10.14754/CEU.2018.11

smaller content provider firms – where a significant part of innovation comes from (e.g. start-ups). On the other hand, opponents of net neutrality argue that internet service providers wouldn't be able to recoup the investments into network capacity without imposing additional fees and net neutrality would just hinder future developments. Both stories are plausible and nothing prevents the two channels to exist simultaneously. It is just a question of which effect is stronger and whether the modeling framework can accommodate these behaviors.

1.2.3 Concerns and harmful incentives

The FCC further argues that broadband providers not only have the ability to engage in the banned practices but also have significant incentives to do so. Here I briefly reiterate on the main questions and concerns raised previously and their potential consequences in a more general way.

Excessive pricing

In the case of termination fees the concern is that broadband providers would act as gate-keepers over their subscriber base. As local competition is very limited residential users don't have alternatives for broadband access and providers could act as monopolies setting excessively high prices for content providers.

Content Creation and Innovation

It is reasonable to associate content innovation with successful market entry to the digital content provision market. This is based on the remark that creation of a new (differentiated) variety is usually associated with a small startup (think of Facebook, Twitter, etc.) who enters the market with this single product. The price charged to this side of the market generally makes entry less profitable and thus may lead to less innovation on the content provider side. This is important because we think that broader variety and better content brings higher welfare to consumers. However, there is another mechanism that may be beneficial

10.14754/CEU.2018.11

for total welfare. By making entry more costly entrants will be selected by the expected profitability of their products. We may expect to see an increase in average product quality of entrants and a higher probability of setting foot in the industry. This means that sunk costs of product development for potentially less successful entrants can be saved which is important for total welfare.

Capacity and Investment

Internet service providers argue that the zero-pricing rule prohibits the appropriation of returns to their investments. From public policy point of view we might expect that infrastructure development leads to better user experience, increased demand and thus increased incentives to supply quality content. So a non-zero pricing might have a dual effect on content creation. First, it could reduce innovation through increasing barriers to entry (see the subsection above). Second, it increases incentives to invest in capacity through which it increases consumer demand and thus the incentives to entry the content provision market. Depending on the actual trade-off between these effects it may be interesting to see the outcome of e.g. a capacity investment subsidy together with the zero-pricing rule.

Moreover if paid prioritization is allowed then internet service providers may have an incentive to keep their network capacity low, creating artificial scarcity that allows them to extract more money from content providers.

Foreclosure and Fragmentation

Net neutrality raises a number of antitrust related issues. First, if a uniform pricing is allowed then broadband providers have an incentive to raise prices to capture the high willingness to pay of large established content providers and thus excluding the smaller ones who cannot keep up with the price. Second, some content providers might be able to foreclose rivals from local markets by paying for exclusive agreements. Both of these lead to internet fragmentation, which means that consumers at different internet service providers are not able to access the same set of

10.14754/CEU.2018.11

content providers. However, the presence of cross-side network externalities should weaken this effect. Furthermore, if small companies are able to use scalable web services, such as Amazon's, to deliver content then the question is rather to what extent the prices will be transmitted by Amazon.

Competition in related markets

A very important feature of the industry is that traditionally the dominant broadband access providers are established cable companies that have a primary business in a field that is directly competing with the digital content offered by online content providers. Indeed, in the video space alone, CBS and HBO have recently announced new plans for streaming their content free of cable subscriptions. Being forced to allow free usage of their networks access provider firms are hurt by the increased competition from rivals who have a sunk cost advantage and are able to price more aggressively. The concerns with this is that they have an incentive to favor their own service and to degrade the quality of their online competitors.

1.2.4 Dominance of video streaming

Internet content producing companies are very diverse. However, there is small number of large dominant firms and a lot of small firms (e.g. startups, who might eventually get large). The annual Sandvine Global Internet Phenomena Report from 2016 provides some interesting statistics. Their measurements show that in North America around 75% of prime-time downstream traffic comes from the ten largest firms and more than 90% of this traffic is related to video streaming. Not surprisingly, the five largest firms are all video streaming sites like Netflix, YouTube and Amazon Video.[3] Based on this observation and the fact that congestion is caused by the traffic-intensive applications I will focus primarily on the online streaming video industry in my model.

[3] The exact market shares are Netflix 35.15%, YouTube 17.53%, Amazon Video 4.26%, iTunes 2.91%, Hulu 2.68%

1.3 Related literature

The (almost single) agreement in the literature seems to be that it is best to think of net neutrality in a two-sided market environment where the internet service provider acts as a platform and lets content providers interact with subscribers.

However, the main differentiating dimension in the literature is whether the authors focus on the termination fee or paid prioritization part of the net neutrality rules as set out by the FCC.

The simplest type of analysis in the literature concerns the effects of the changes in prices due to the introduction of the termination fee, like in Economides and Tag (2012). Similarly, Njoroge et al. (2013) also study the effect of the termination fee but they also take into account the platform's investment incentives.

The stream of papers that study paid priority and network capacity investments with a fixed set of content providers include Choi and Kim (2010),Cheng et al. (2011) and Kramer and Wiewiorra (2012). Although their results are not unambiguous, they all show that prioritization distributes more profits to the platform.

A notable exception is Peitz and Schuett (2016) who study both the termination fee and paid priority rule in a single but simplified industry model. In their model net neutrality leads socially suboptimal amounts of traffic.

Economides and Tag (2012) show that competition among internet service providers is not likely to change that positive effects of net neutrality while Bourreau et al. (2015) show that paid priority increases both capacity investment and content innovation.

Research papers also vary in the set of interactions they allow across or within sides of the market. Most papers don't allow monetary transaction between the two sides or don't acknowledge within side competition among content providers. An exception for the former Gans and Katz (2016), who introduce side payments between the two sides.

Review articles such as Lee and Wu (2009), Schuett (2010), Kramer et al. (2013) and Greenstein et al. (2016) are the most useful sources in listing

10.14754/CEU.2018.11

the potential question and concerns, however, they do not offer a model or data to conduct further analysis.

My work is also related to the work on investment incentives in the context of general platform markets. Alexandrov and Deb (2012) study the interaction of price discrimination and investment incentives of the platform and Belleflamme and Peitz (2010) study platform competition and seller investment incentives. Their results show that discrimination increases platform investment and that easy access to the platform could lead to underinvestment.

Regarding the dynamic modeling framework my work builds heavily on the dynamic industry model of Ericson and Pakes (1995). Out of the large number of papers that use some variation of this dynamic framework, my work is closest to Markovich (2008), Markovich and Moenius (2009) and Chen et al. (2009) who study dynamic platform market environments. The computational algorithm I use to compute the equilibrium is a slightly modified version of Pakes and McGuire (1994) which is also the source for the models parameters.

1.4 Model

In this section I describe all building blocks of the modeling framework that I use to study the potential effects of switching from a neutral regime to non-neutral ones. First I describe the model in general and then I gradually introduce and discuss all features.

1.4.1 Overview

The model itself is an infinite horizon discrete time multi-agent dynamic stochastic game with per-period profits determined by the actual regulation in place. There are two main building blocks of the model. The dynamic game builds on the dynamic industry model of Ericson and Pakes (1995) (EP) and extends the framework to allow for two kinds of firms with separate value and policy functions. The static game that determines per-period payoffs builds on an oligopolistic price setting game which is adapted to a platform market environment.

10.14754/CEU.2018.11

There are three different types of agents in the model:

- A mass of residential consumers (subscribers) who desire to purchase and consume digital content
- A discrete set of competing vertically differentiated streaming video content providers (CPs) who distribute their products over the internet
- A single Internet Service Provider (ISP) who is intermediating between the two sides and lets subscribers and CPs physically connect over its network.

In any time period the state of the dynamic game is fully characterized by a collection of payoff relevant state variables that contains all information that is relevant for firms to play the stage game and to choose their dynamic controls. The three kinds of state variables include (i) a list of CPs' product characteristics, (ii) the ISP's network capacity and (iii) a scalar congestion parameter. All state variables evolve endogenously and are controlled by firms' policies. In particular, state variables (i) and (ii) evolve according to firms' investments into their product quality and network capacity, respectively. The congestion parameter, however, depends on lagged values of the other two state variables and it plays a role in subscribers' expectations of future network performance.

Other important features of the stage game include (i) the presence of cross-side network externalities, (ii) within-side competition effects on the CP side, (iii) side payments between the two sides of the market and (iv) both per-user transaction and fixed priority access fees charged by the platform.

Why use a dynamic model?

Switching to the dynamic modeling framework from a static game has three main advantages. First, it allows for an equilibrium that is not only a single state but a probability distribution over the industry states. In the dynamic framework this means that all states that have a positive weight in the equilibrium distribution are visited infinitely often with frequencies proportional to their probability weights. The equilibrium

10.14754/CEU.2018.11

distribution is determined by the dynamic responses of agents to each others' actions. In many cases this is much closer to the way we actually think about industries and it makes stylized behaviors as described by the virtuous cycle directly feasible and observable. Second, separating the problem into dynamic and static parts makes it feasible to solve the platform market problem in the stage game that allows for a rich set of interactions among agents both within and across sides. A traditional platform market approach would become far too complex.[4] Third, note that the dynamic game is completely agnostic to how per-period profits are determined. This makes the dynamic framework suitable for analysis of various equilibria resulting from different payoff functions without having to rewrite the model. The approach is common in the industrial organization literature and is well fit for the type of counterfactual analysis I am interested in.

The price, of course, for these advantages is the elegance and completeness of analytic tractability. As a consequence the equilibrium can be computed only by using numerical dynamic programming methods. Although existence or unicity are not guaranteed in general, Doraszelski and Satterthwaite (2010) give formal conditions for the existence of a stationary Markov-perfect equilibrium in the EP-style industry models.[5]

I believe that the inherently dynamic nature of the virtuous cycle justifies the use of dynamics. In particular, a simple sequential decision timing assumption (e.g. investments first, prices second) would allow for any two consecutive links in the cycle, but would leave the loop open without allowing for explicit feedback – as it is the case in general with two period models. E.g. without the internet subscription channel it might be an attractive choice as it is common in the literature. On the other hand, assuming all decisions are made simultaneously I might exclude behavior that is implied by the virtuous cycle. E.g. the dynamic model would allow an ISP to invest into capacity without observing several high-quality CPs but only anticipating such behavior (i.e. increased CP investment) in a future industry state with the resulting higher network capacity.

[4]The multi-sided market approach of Weyl (2010) and Veiga et al. (2017) might be an elegant analytic alternative to the dynamic framework.

[5]I don't provide a formal equilibrium characterization in this paper but I refer to their work when it is applicable.

10.14754/CEU.2018.11

In the following section I introduce the details of the model's primitives through the model absent congestion and featuring the net neutrality regulation in place. Building on this baseline model I gradually introduce the features that give rise to the three additional alternative regulatory regimes I consider for the static game. Finally I provide the details of the dynamic game.

1.4.2 Agents

Subscribers

In every period of the game a mass of M residential users face a two-stage decision. First, they decide whether or not to subscribe for the ISP's broadband service based on the available content offered by CPs. Second, those who decided to subscribe can choose one from the available set of CPs or may opt for not choosing any. I let subscribers to have an outside option in both choice situations for two reasons. One is that assuming a fixed broadband share would exclude a link from the virtuous cycle making it impossible for the dynamic feedback loop to emerge, decrease the CPs' investment incentives by shutting down the market expansion effect, make the ISP's revenues independent of network content quality and network capacity/congestion by making demand perfectly inelastic. The other reason is that share of broadband subscribers is an important market outcome that is of policy interest. In the second choice I leave users an option of 'casual internet browsing' instead of choosing a particular content provider, that is the outside option can be thought of as using all other services on the internet.

Subscribers are identical in their preferences regarding how much they value the digital goods offered by the CPs and internet access up to an idiosyncratic taste shock. However, they don't fully observe their tastes on the CPs' products until they subscribe to the ISP. Thus, when deciding on broadband subscription they base their decision on the expected utility from the CP choice situation. I introduced this feature so that only consumers with internet access can have a precise understanding of what utility they may have from a digital good.

Formally, consumer i's utility from content provider j is $u_{ij} = v_j + \varepsilon_{ij}$, where v_j denotes the *observed* part of utility. I consider several specifications for v_j but for now I will use simply $v_j = \delta_j - p_j$, where δ_j denotes the mean utility of CP j who charges price p_j for its content. For the ε errors I assume type I extreme value distribution and I normalize the observed utility of casual browsing (v_0) to zero. None of these come without a cost. First, the logit assumption introduces the well-known unlimited love of variety property, while the second assumption basically restricts the set of outcomes by fixing a free parameter.

The individual choice probability for each consumer to choose CP j, or in other words, the market share of CP j among consumers who have internet access is

$$s_j = \frac{\exp\{v_j\}}{1 + \sum_k \exp\{v_k\}}.$$

Given the distribution of ε, before subscribing to the ISP the expected utility of the CP choice situation for a given consumer is

$$V = \ln\left\{\sum_j \exp\{v_j\} + 1\right\},$$

and the utility of internet access for consumer i is $U_{i1} = V - P + \epsilon_{i1}$, where P is the net price charged by the ISP. For the ϵ terms I assume type I extreme value distribution and I normalize the observed utility of no internet access (V_0) to zero. Thus, the probability for a consumer to subscribe to the ISP is

$$S = \frac{\exp\{V - P\}}{1 + \exp\{V - P\}},$$

which also equals the share of consumers with broadband access.

It is important to emphasize that all these values are specific to a given period. If the available set of CPs or their product offering changes it has two effects on CPs' market shares. A direct (or competitive) effect on the consumers with internet access and an indirect (or market expansion) effect on the market share of the ISP. Note, that this latter effect is one crucial part of the virtuous cycle.

Before I turn to CPs, it is instructive to clarify that although consumers have rational expectations regarding their utilities from CPs' products, they are not forward looking in a sense that they don't form expectations for future periods on the sets of available CPs. Also they don't have a dynamic role in a sense that their preferences are independent from history, a notion that I will relax later when I introduce congestion.

Content providers

In every period there are J CPs with the corresponding set of mean utilities (product qualities) $\{\delta_j\}_{j=1}^{J}$ selling their products to the ISP's subscribers. Note, that the specification of consumer utility implies a systematic vertical differentiation among content providers based on the mean utilities and an idiosyncratic horizontal differentiation due to consumer taste heterogeneities. In the online video streaming industry one way to interpret the vertical characteristic is to think of it as video resolution. I assume that CPs have video libraries of similar sizes but with different titles and consumers don't prefer systematically any of the libraries. Vertical differentiation comes from the different video resolutions of content availability in . Ideally, the higher the resolution the better people perceive the product on average. However, higher resolution means not only more desired products but higher bitrates which lead to increased marginal content distribution costs and will have an important interaction with network capacity and, as I show in turn, might lead eventually to network congestion.

CPs face a marginal cost (mc^{cp}) per user proportional to their product quality (resolution) that covers the costs of content distribution in the network. This marginal cost consist of two multiplicative components. The first component (b_j, the bitrate) specifies how many bits are needed to transmit the content of CP j, and the second component (c^{cp}) is the price of content distribution per bit that is common for all CPs in the industry.

Formally, CP j sets its price to maximize profits as given in

$$\Pi_j^{cp}(p_j, p_{-j}, P) = \underbrace{MS(P,p)s_j(p_j,p_{-j})}_{\text{total demand for CP } j}\left(p_j - c^{cp}b_j\right),$$

where for clarity I directly indicated that market shares depend on prices.[6]

[6]For the ease of notation I will only denote the dependence on prices where it serves a well-defined purpose.

Internet service provider

The role of the ISP is to intermediate between subscribers and CPs. It operates a network with a certain capacity (κ) that carries bits traveling from CPs to subscribers. The ISP incurs costs of network management depending on the volume of total network traffic. Traffic is determined by the demand for CPs' products and their bitrates.

The ISP then sets its price to maximize static profits which are given as

$$\Pi^{isp}(P,p) = \underbrace{MS(P,p)}_{\text{\# subscribers}} P - c^{isp} \underbrace{MS(P,p) \sum_j s_j(p_j, p_{-j}) b_j}_{\text{total traffic}} .$$

Notice that higher quality product offerings by CPs have three effects on the ISP's profits. First, they increase revenues by increasing the subscribers' expected utility from the CP choice situation therefore increasing demand for the ISP's broadband service. This is exactly what I called previously as the market expansion effect and is directly related to the virtuous cycle. Second, they increase the ISP's costs on the extensive margin (i.e. attracting more subscribers) Third, they also increase costs on the intensive margin as higher quality products with higher bitrates consume more bandwidth. This latter channel is often overlooked in pro net neutrality arguments.

1.4.3 Congestion

I introduced congestion as a state variable that contains information about other the lagged state variables. In this section I start from a much more intuitive interpretation.

Let us first forget for a second about the source of congestion and let us focus only on the consequences. In a congested network bits arrive at their destinations with a lag. This has a negative effect on subscribers' experience as they can't enjoy the content in its full glory. To model this decrease in subscribers' utility due to a congested network it would be natural to simply discount the utility derived from internet access by a factor $\mu \in (0,1]$, e.g. like in $U_{i1} = \mu V - P + \epsilon_{i1}$. This approach sounds reasonable as congestion affects all agents connected to the network.

However, it misses the important notion that because of the differences in required bandwidths across different products congestion may have a differential effect on user experience – a point that is often emphasized by opponents of net neutrality. To reflect this I choose to model the observed part of subscribers' utility as $v_j = \mu\delta_j - p_j$, where $\mu \in (0, 1]$ is a scalar factor that discounts the mean value derived from product j and both the form of utility and the value of the parameter is known to all subscribers and firms. Note that modeling congestion in this way the utility from 'casual browsing' that is not bandwidth intensive is not affected and higher quality products will contribute more to the decrease in V. Also, this approach will turn out to be useful later when I introduce paid priority.

Regarding the source of congestion, it occurs when network traffic exceeds network capacity. It would be convenient to define the congestion parameter as a fraction of capacity and traffic, like e.g. $\mu = \min\left\{\frac{\kappa}{T}, 1\right\}$, where total traffic in the given period is given by $T = MS\sum_j s_j b_j$. However, there is a logical flaw in this reasoning because firms also consider the effects of potential congestion on subscribers' utility and set their prices accordingly. This means that the proper definition would be

$$\mu = \min\left\{\frac{\kappa}{MS(P, p, \mu)\sum_j s_j(p_j, p_{-j}, \mu)b_j}, 1\right\}$$

and one would have to solve for μ together with prices in the static equilibrium for each possible set of products marketed and network capacities (i.e. all industry states).

There are several reasons why I choose to follow an alternative approach where the congestion parameter μ is *inherited* from the previous period and is defined recursively across periods as

$$\mu = \min\left\{\frac{\kappa^{-1}}{T^{-1}}, 1\right\},$$

where κ^{-1} and T^{-1} are network capacity and total traffic values from the preceding period. First, there is an obvious computational reason. The computational burden on the dynamic part of adding congestion realized in the previous period as a new state variable with a relatively low number of potential values is way less than that of losing the analytic first- and second-order conditions used in the computational algorithm for the stage game.Second, it relaxes the implicit assumption that consumers anticipate congestion rates correctly. In other words I assume that subscribers are *not network engineers* and they don't track each other's consumption and how bits travel along a complicated network. More formally, I assume that consumers form adaptive expectations with perfect adjustment regarding the evolution of congestion.[7] Third, I find this modeling choice more intuitive and realistic as it is natural that consumers base their decisions on their past experiences.[8]

1.4.4 Termination and paid priority

In the preceding sections I introduced the model with the net neutrality regulations in place. That is the ISP was bound to set a zero price to the CP side of the market and was not allowed to offer CPs a paid differentiated service without congestion. In this section I relax the ban on both termination and paid priority fees and show how the full model nest all regulatory regimes in consideration.

Termination

I define the termination fee (t) as a per transaction fee that CPs have to pay to the ISP after each of their subscribers. In effect the termination fee essentially increases the marginal cost of each CP by the same t amount and increases the profit margin of the ISP. Profits are now defined by the set of equations below.

$$\Pi^{isp}(P,p,t) = MS(P,p,t)\left(P + \sum_j s_j(p_j,p_{-j})(t - c^{isp}b_j)\right) \tag{1.1}$$

$$\Pi_j^{cp}(p_j,p_{-j},P,t) = MS(P,p,t)s_j(p_j,p_{-j})\left(p_j - (t + c^{cp}b_j)\right) \tag{1.2}$$

Unfortunately, as it was pointed out in Greenstein et al. (2016) this approach suffers from the indeterminacy of prices. This is because consumers care only about the total price they pay (e.g. $P + p_j$) but not

[7] I would like to thank Sergey Lychagin for pointing this out to me.

[8] E.g. when renewing a contract with, say, Netflix, naturally one would consider their own past experience or would check the Netflix Speed Index that Netflix publishes for each ISP based on their previous network performance.

about the split of this total price between the ISP and the CPs. To alleviate this problem I assume that the ISP first commits to a termination fee and then all firms (including the ISP) set prices. I also assume that the ISP is able to compute its profits resulting from all potential termination fees and selects the one with the highest payoffs.

Paid priority

I define the paid priority service offered by the ISP to be a temporary capacity extension technology that gives CPs exempt from congestion. Formally, this means the consumers are able to get the full mean utility from a product even though μ is less than one. I modify the observed utility of subscribers to

$$v_j = \left[(1-\mu)G_j + \mu\right]\delta_j - p_j,$$

where $G_j \in 0, 1$ and $G_j = 1$ if and only if CP j pays for priority. Note that for $G_j = 0$ subscribers perceive the mean utility from product j still as $\mu\delta_j$ while for $G_j = 1$ the term in square brackets reduces to one and the product can be fully enjoyed by subscribers.

The ISP makes the priority service available to all CPs for the same price F, which I call the priority fee. I assume that there are no direct costs associated with the provision of the priority service and the ISP finances the temporary capacity extension from the collected priority fees. Although it is not a too far fetched assumption that ISPs can "easily plug in a new cable" at any internet exchange point between their servers and their interconnecting peers, there are two natural requirements that are completely missing from the current model. First, there is an unlimited availability of the temporary capacity which ideally should be not more than a certain (small) fraction of actual capacity. Second, by providing the prioritized service the available capacity for regular (non-priority) traffic remains the same which is considered to be an important trade-off as it can widen the gap between the quality of regular and priority traffic. To address these issues I limit the number of CPs that are allowed to buy priority traffic to one. This will reduce the traffic flowing through the priority channel, introduce a rudimentary rationing mechanism (the one who pays the most gets the priority service), and limit the gains from priority traffic.

A related issue is that I assume that providing the priority channel is more "costly" for the ISP when the network is congested. To reflect this I discount the ISP's revenues from prioritization by μ.[9]

After including the priority service option profits become

$$\Pi^{isp}(P, p, F, G) = MS(P, p) \left(P - \sum_j s_j(p_j, p_{-j}) c^{isp} b_j \right) + \mu F \sum_j G_j \tag{1.3}$$

$$\Pi_j^{cp}(p_j, p_{-j}, P, F, G) = MS(P, p) s_j(p_j, p_{-j}) \left(p_j - c^{cp} b_j \right) - F G_j. \tag{1.4}$$

Unfortunately by introducing the binary decisions the problem becomes a little bit more difficult computationally. I use the same approach as with the termination fee and assume a sequential process. First the ISP commits to a priority fee, then the CP that is best off by buying the priority service gets it and all other CPs remain in the congested network. I assume that the ISP is able to compute CPs' profits and to pick a priority fee that leads to the highest level of its own profits.

Full model and static profits

In the full model I allow the ISP to set both a termination and a priority fee. Figure 1.2 illustrates the structure of the model and depicts money flows along with network traffic. The figure shows that ISP charges a termination fee to both CPs for every user they interact with but it only charges the one-time priority fee to the first CP. The two CPs have product qualities (δ_1, δ_2) but the ISP's network is congested so the second CP's product is only perceived as $\mu\delta_2$. Residential subscribers are the source of all revenues in the model. The ISP collects P from everyone who subscribes and CPs collect their prices (p_1, p_2) from those who decide to buy their products on top of the internet subscription.

[9] A more appropriate way of introducing a gap between the fee that is paid by a CP and the net amount that an ISP gets would be using $G_j F$ in the CP profit function and $F \sum_j G_j - C(\mu) \sum_j G_j$ for the ISP, where $C(\mu)$ is a decreasing function of μ representing costs related to providing the priority channel for one CP given congestion level μ. In the current setup I just use the special case $C(\mu) = (1 - \mu)F$.

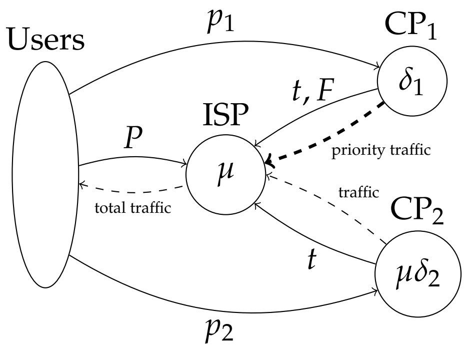


Figure 1.2: Stage game flowchart. Solid lines indicate monetary flows while dashed lines indicate network traffic.

Thus the general profit functions are

$$\Pi^{isp} = MS\left(P + \sum_j s_j(t - c^{isp}b_j)\right) + \mu F \sum_j G_j \tag{1.5}$$

$$\Pi_j^{cp} = MSs_j\left(p_j - (t + c^{cp}b_j)\right) - FG_j. \tag{1.6}$$

I consider four profit function specifications for the static game. In particular, for the net neutrality model I set $t = 0$ and $F = \infty$, for the termination fee model I don't constrain t but set $F = \infty$, for the paid priority model I only apply the zero-pricing rule as $t = 0$ but allow for $F < \infty$ and in the full model I allow the ISP to freely set both t and F. The profits from the four alternative static games will be used as per-period payoff functions in the dynamic game.

1.4.5 Static equilibrium

For any given pair of values for t and F the first-order conditions of the ISP's and the CPs' profit maximization problem lead to the following system of $J + 1$ nonlinear equations:

$$(1 - S)\left(P + \sum_j s_j(t - c^{isp}b_j)\right) - 1 = 0 \tag{1.7}$$

$$\left(1 - Ss_j\right)\left(p_j - (t + c^{cp}b_j)\right) - 1 = 0, \qquad j = 1 \dots J \tag{1.8}$$

I assume symmetry of content providers within a given type j. This applies to both the price setting and the service choice decision. To solve for optimal fees I compute prices for all (t, F) pairs over a fine grid and pick the one that leads to the highest ISP profit.

1.4.6 Dynamic framework

As described in the model overview I adopt an extended version of the dynamic industry model of EP. The equilibrium concept is that of a stationary Markov Perfect Equilibrium (MPE). Basically there are two main conditions for a set of dynamic policies to constitute an MPE:

- Each agent chooses optimal policies given her perceptions of likely future industry structures
- These perceptions are consistent with the behavior of other agents

The difficulty introduced by my extension is that agents perceptions have to be formed over two kinds of agents' behaviors and their best responses have to be consistent with both perceptions.[10] These primitives are very important because the iterative solution algorithms typically rely heavily on them. Below I will briefly show how to derive the values and policies as functions of firms' perceptions of their competitors likely future states. Then I use the derived expressions in a slightly modified version of the Gaussian numerical algorithm of Pakes and McGuire (1994).

The rest of this section builds heavily on the handbook chapter by Doraszelski and Pakes (2007). As EP-style industry models have become standard in the literature here I just describe the parts of the dynamic model that are either unique to the setting or especially important for understanding. For additional details I refer the reader to their paper and to the literature cited by them.

State space and transitions

In each period the state of the industry is fully described by the triplet (ω, κ, μ), where

- ω is list of all CPs' states taking values from the set of integers $\Omega = \{1, \ldots, \bar{\omega}\}$. Higher states correspond to higher mean utilities and

[10]These kinds of *double fixed point* problems are similar to the difficulties encountered in dynamic demand models. See Aguirregabiria and Nevo (2013) for a recent review.

higher bitrates.[11]

- κ denotes the capacity state of the ISP's network on the set of integers $K = \{1, \ldots, \bar{\kappa}\}$. Similarly to CP qualities the corresponding capacity values V_κ increase with the states.

- μ is the congestion parameter as perceived by residential users that depends on historical realizations of traffic and network capacity. μ is discretized to a grid on $(0, 1]$.

In each period content providers can invest in product development that may result in a new, better product. The outcome of the investment is stochastic with the following density function:

$$p(\nu|x) = \begin{cases} \frac{\alpha x}{1+\alpha x}, & \text{if} \quad \nu = 1 \\ \frac{1}{1+\alpha x}, & \text{if} \quad \nu = 0 \end{cases} \tag{1.9}$$

Note that both the amount spent (x) and the investment technology parameter (α) increase the probability of a successful investment outcome ($\nu = 1$). The ISP has access to a similar investment technology but with different success rate (τ)

$$p(\psi|X) = \begin{cases} \frac{\tau X}{1+\tau X}, & \text{if} \quad \psi = 1 \\ \frac{1}{1+\tau X}, & \text{if} \quad \psi = 0 \end{cases} , \tag{1.10}$$

where X denotes the amount spent on investment by the ISP.

In each period the industry is hit by an industry-wide shock (η) that is common across all firms in the industry. This shock can take values 0 or 1 with probabilities λ and $(1 - \lambda)$, respectively. The next period state of a firm with quality level ω_i and investment outcome ν is given by $\omega_i' = \omega_i + \nu - \eta$, while the evolution of the ISP's state is given by $\kappa' = \kappa + \psi - \eta$. Note that the common industry shock has a dual role. First, it generates correlation among firms' profits which is a general observation when looking at industrial data. Second, it creates incentives for continuous development by modeling decreasing quality of products. This is an alternative way to model that over time consumers get used to a certain quality level and might not value it as much as initially.

[11] There is a one-to-one mapping among the elements of Ω the mean utilities ($\{\delta_j\}_{j=1}^J$) and the bitrates ($\{b_j\}_{j=1}^J$)

In particular I implicitly assume that the value derived from watching an HD movie in 2007 is comparable to watching a 4k movie in 2017 rather than an HD movie in 2017. In this way the model allows for continuous development and does not constrain the industry in the steady state.

As discussed previously the evolution of μ conditional on the current state (ω, κ, μ) is deterministic and is given by

$$\mu' = \min\left\{\frac{V_\kappa}{T_{\omega,\mu}}, 1\right\} \tag{1.11}$$

Value functions and investments

In this section I briefly show how to derive the values and policies as functions of firms' perceptions of their competitors likely future states.

In each period incumbent firms first enjoy payoffs from the static game corresponding to the current industry state and regulation. Then they simultaneously decide about their investments in order to maximize their continuation values. At the very end of the period random shocks and investments realize and the industry transitions to the next period with the new state.

The value of an incumbent CP i with individual state ω_i in an industry state (ω, κ, μ) is defined as

$$V_{cp}(\omega_i, \omega_{-i}, \kappa, \mu) = \Pi_{cp}(\omega_i, \omega_{-i}, \mu) + \max_{x_i}\left\{-x_i + \beta \mathrm{E}\left[V_{cp}(\omega_i', \omega_{-i}', \kappa', \mu', \phi')|x_i, \omega_i, \omega_{-i}, \kappa, \mu\right]\right\}. \tag{1.12}$$

To derive analytic formulas for optimal investment strategies let us denote the incumbent CPs' perceived probability of the next period value of its competitors' states (ω_{-i}') and the ISP's state (κ') conditional on the current state and common shock by $q(\omega_{-i}'|\omega_i, \omega_{-i}, \kappa, \mu, \eta)$ and $z(\kappa'|\omega_i, \omega_{-i}, \kappa, \mu, \eta)$. Then the continuation value can be written as

$$\mathrm{E}\left[V_{cp}(\omega_i', \omega_{-i}', \kappa', \mu')|x_i, \omega_i, \omega_{-i}, \kappa, \mu\right] = \sum_{\nu} W_{cp}(\nu|\omega_i, \omega_{-i}, \kappa, \mu) p(\nu|x_i) \tag{1.13}$$

where

$$W_{cp}(\nu|\omega_i, \omega_{-i}, \kappa, \mu) = \sum_{\eta, \kappa', \omega_{-i}'} V_{cp}\left(\omega_i + \nu - \eta, \omega_{-i}', \kappa', \min\left\{1, \frac{V_\kappa}{T_{\omega,\mu}}\right\}\right) \times q(\omega_{-i}'|\omega_i, \omega_{-i}, \kappa, \mu, \eta) z(\kappa'|\omega_i, \omega_{-i}, \kappa, \mu, \eta) p(\eta) \tag{1.14}$$

10.14754/CEU.2018.11

With this formulation optimal investment policy is given as

$$x(\omega_i, \omega_{-i}, \kappa, \mu) = \max \left\{ 0, \frac{-1 + \sqrt{\beta\alpha(W_{cp}(1|\omega_i, \omega_{-i}, \kappa, \mu) - W_{cp}(0|\omega_i, \omega_{-i}, \kappa, \mu))}}{\alpha} \right\} \tag{1.15}$$

if $W_{cp}(1|\omega_i, \omega_{-i}, \kappa, \mu) \geq W_{cp}(0|\omega_i, \omega_{-i}, \kappa, \mu)$, and $x(\omega_i, \omega_{-i}, \kappa, \mu) = 0$ otherwise. We can now write the value function as

$$V_{cp}(\omega_i, \omega_{-i}, \kappa, \mu) = \Pi_{cp}(\omega_i, \omega_{-i}, \mu) - x_i(\omega_i, \omega_{-i}, \kappa, \mu) + \beta \sum_{\nu} W_{cp}(\nu|\omega_i, \omega_{-i}, \kappa, \mu) p(\nu|x_i(\omega_i, \omega_{-i}, \kappa, \mu)). \tag{1.16}$$

The ISP's value function solves the problem

$$V_{isp}(\omega, \kappa, \mu) = \Pi_{isp}(\omega, \mu) + \max_{X} \left\{ -X + \beta \mathrm{E} \left[V_{isp}(\omega', \kappa', \mu') | X, \omega, \kappa, \mu \right] \right\}. \tag{1.17}$$

As there is only one single ISP the continuation value corresponding to a specific investment outcome depends only on the ISP's perceived probabilities of the next period states of CPs (Q).

$$\mathrm{E} \left[V_{isp}(\omega', \kappa', \mu') | X, \omega, \kappa, \mu \right] = \sum_{\psi} W_{isp}(\psi|\omega, \kappa, \mu) p(\psi|X) \tag{1.18}$$

where

$$W_{isp}(\psi|\omega, \kappa, \mu) = \sum_{\eta, \omega'} V_{isp}\left(\omega', \kappa + \psi - \eta, \min\left\{ 1, \frac{V_\kappa}{T_{\omega,\mu}} \right\} \right) Q(\omega'|\omega, \kappa, \mu) p(\eta). \tag{1.19}$$

Optimal ISP investment is derived analogously to incumbent CPs and it allows to express the integrated value function of the ISP in a similar fashion as

$$V_{isp}(\omega, \kappa, \mu) = \Pi_{isp}(\omega, \mu) - X(\omega, \kappa, \mu) + \beta \sum_{\psi} W_{isp}(\psi|\omega, \kappa, \mu) p(\psi|X(\omega, \kappa, \mu)). \tag{1.20}$$

1.5 Lessons From the Analysis

Static profits provide the investment incentives for firms and as such they are the foundation of the dynamic equilibrium. Therefore, I find it instructive to have a look at static profits before the computed dynamic policies.[12]

[12] For a step-by-step detailed analysis of certain channels please see the Appendix.

1.5.1 Static equilibrium

To gain insights to what is going on in the static game in the four alternative regimes Table 1 provides information on profits, prices and fees. I pick four different congestion rates ranging from 1 to 0.9 (no, low, mid and high, respectively) and take averages across all industry states. Note, that this average is not weighted by the stationary distribution.

In the neutral regime both profits and prices decrease with the severity of congestion (top two panels).

This decreasing pattern is similar for the termination regime but with two notable differences: (i) the introduction of the termination fee opens a drastic gap between ISP and CP prices and (ii) CP profits fall significantly, while ISP profits increase only marginally compared to the neutral regime.

In contrast, under the priority regime I find that the ISP can increase its profit with the severity of the congestion while CPs can maintain their profits at the level of the neutral regime. This is because once CPs pay for prioritization subscribers don't notice the effect of congestion and the ISP can extract the gains from the CPs.

Lastly, when I allow for both termination fees and paid priority I find an interesting mix of the two regimes. First, ISP profits are the highest and they follow the pattern similar to priority profits while CP profits are lowest and follow the pattern of the termination regime. Second, both ISP and CP prices follow the level of termination regime prices but they are only slightly decreasing like in the priority regime.

Figure 1.7 illustrates how the different termination fees affect the profits of the ISP who faces a moderately congested network and two CPs (1 high quality and 1 middle quality). Note how profits drop beyond a certain priority fee. This is the fee where CPs stop buying the priority access and rather face congestion.

I would like to emphasize once again that these averages were taken across states that may or may not be in the recurrent class and thus they are just here to show that the static profits seem to make sense under the alternative regimes. To see the variation across industry structures Figures 1.3-1.6 depict prices and profits for all possible industry structures with two CPs.

1.5.2 Dynamic equilibrium

In Table 2 I report four sets of equilibrium outcomes for the four alternative regimes. All outcomes are weighted averages using the stationary distribution of states.

10.14754/CEU.2018.11

Table 1.1: Summary of static equilibria

outcome	ISP profit				CP profit			
congestion	no	low	mid	high	no	low	mid	high
model								
neutral	2.96	2.90	2.73	2.52	0.37	0.35	0.29	0.17
termination	3.10	3.06	2.94	2.78	0.18	0.16	0.11	0.04
priority	2.96	2.97	2.99	3.01	0.37	0.35	0.28	0.15
both	3.10	3.11	3.12	3.13	0.18	0.16	0.13	0.06

(a) Average profits across industry states

outcome	ISP price				CP price			
congestion	no	low	mid	high	no	low	mid	high
model								
neutral	6.09	5.99	5.70	5.10	7.86	7.84	7.78	7.67
termination	4.82	4.66	4.48	4.24	9.42	9.51	9.57	9.70
priority	6.09	6.06	5.97	5.83	7.86	7.86	7.85	7.84
both	4.82	4.74	4.76	4.81	9.42	9.46	9.31	9.13

(b) Average prices across industry states

outcome	termination fee				priority fee			
congestion	no	low	mid	high	no	low	mid	high
model								
neutral	-	-	-	-	-	-	-	-
termination	1.74	1.86	1.97	2.17	-	-	-	-
priority	-	-	-	-	-	0.04	0.16	0.40
both	1.74	1.79	1.64	1.45	-	0.03	0.11	0.25

(c) Average fees across industry states

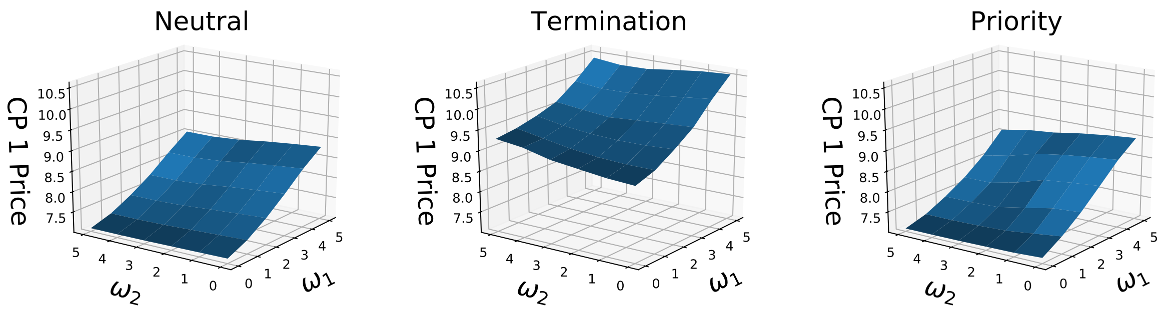


Figure 1.3: CP prices in static equilibrium

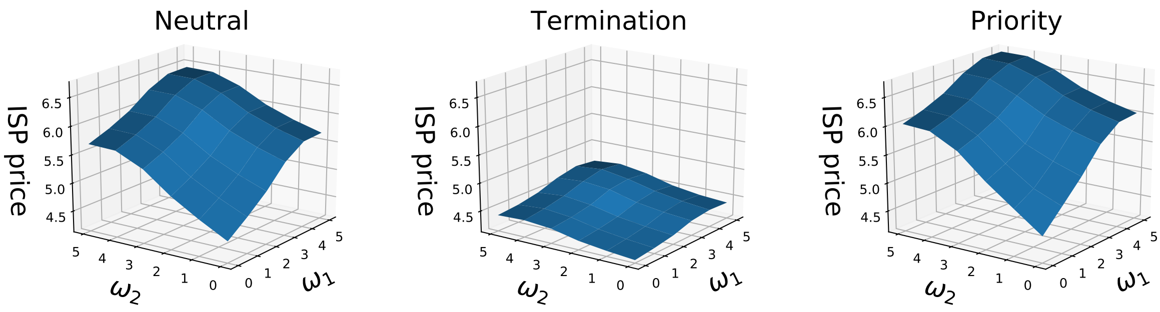


Figure 1.4: ISP prices in static equilibrium

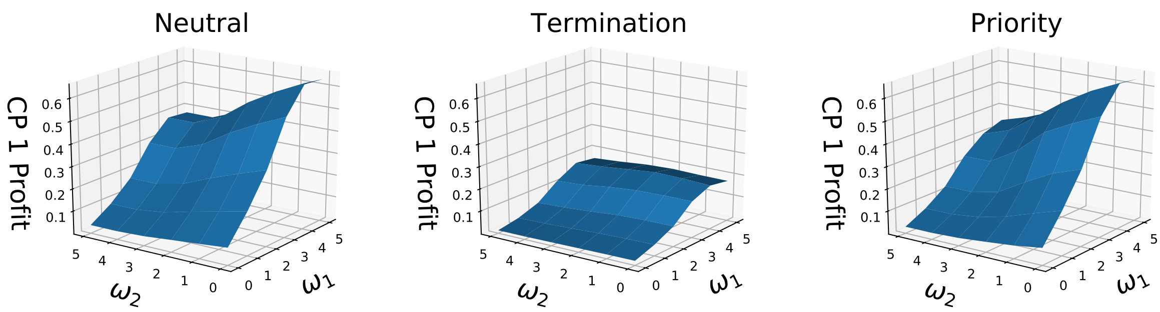


Figure 1.5: CP profits in static equilibrium

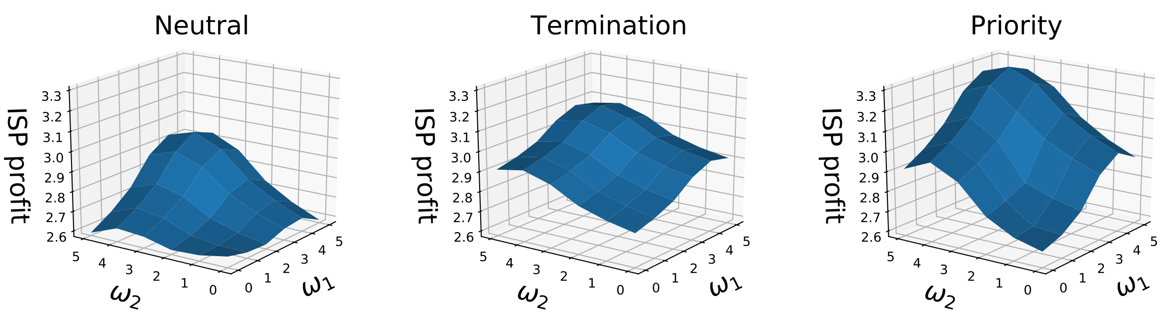


Figure 1.6: ISP profit in static equilibrium

10.14754/CEU.2018.11

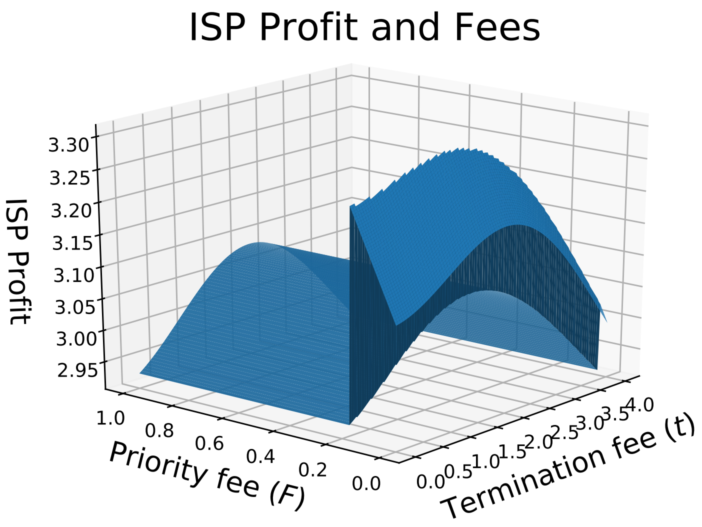


Figure 1.7: ISP profits in static equilibrium as a function of termination and priority fees

The top panel shows that both termination fees and paid prioritization increase the gap between ISP and CP values. The gap is largest when both termination and priority fees are allowed.

ISP investments are lower for non-neutral regimes and the effects are reflected in lower CP quality, lower network capacity and more severe congestion. Interestingly the recurrent class is significantly larger in the priority regimes that may signal that the industry is less stable with paid prioritization. One can verify this on Figure 1.12 which depicts the stationary distribution across all industry structures.

The effects on consumer surplus are not significant for the current parameterization but total welfare is highest in the neutral regime and it is less sensitive to the introduction of paid prioritization than to termination fees.

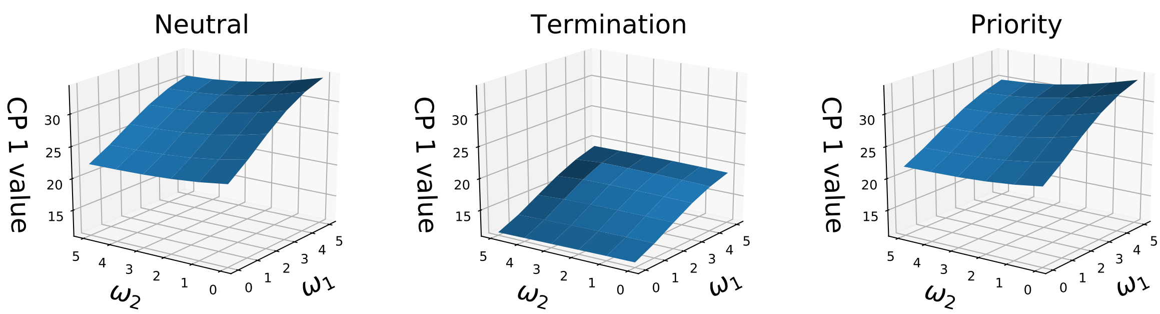


Figure 1.8: CP value function in dynamic equilibrium

Table 1.2: Summary of dynamic equilibria

	ISP value	ISP investment	CP value	CP investment
neutral	324.252	0.055	30.012	0.144
termination	325.542	0.050	17.924	0.156
priority	324.610	0.051	29.637	0.145
both	325.206	0.047	14.900	0.144

(a) Expected values and investments in the stationary distribution

	CPs	level	capacity	congestion	recurrent class
neutral	2.000	4.251	2.366	1.000	2.65%
termination	2.000	3.560	2.291	0.999	2.03%
priority	2.000	4.230	2.313	0.998	4.27%
both	2.000	3.806	2.008	0.991	3.51%

(b) Mean industry characteristics in the stationary distribution

	consumer surplus	industry profit	total welfare
neutral	6.00	4.19	10.19
termination	6.00	3.98	9.98
priority	6.00	4.18	10.18
both	6.00	3.88	9.88

(c) Expected consumer surplus and welfare in the stationary distribution

	broadband share	network traffic	priority share
neutral	69.94%	2.12	0.00%
termination	69.99%	1.65	0.00%
priority	69.88%	2.12	2.92%
both	69.83%	1.54	5.74%

(d) Other important outcomes in the stationary distribution

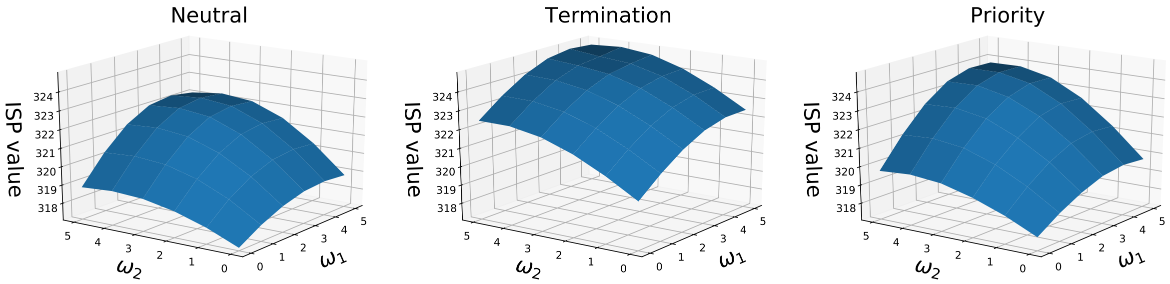


Figure 1.9: ISP value function in dynamic equilibrium

10.14754/CEU.2018.11

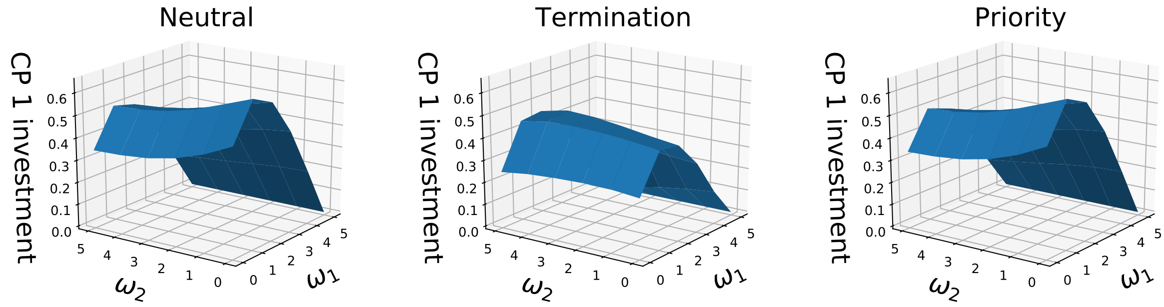


Figure 1.10: CP investment in dynamic equilibrium

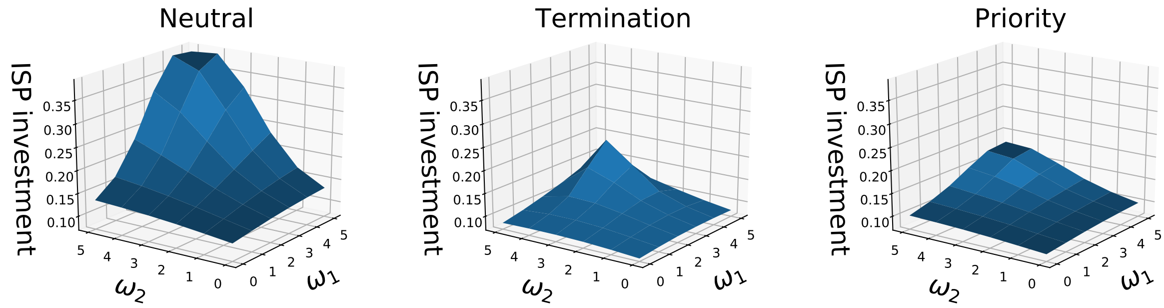


Figure 1.11: ISP investment in dynamic equilibrium

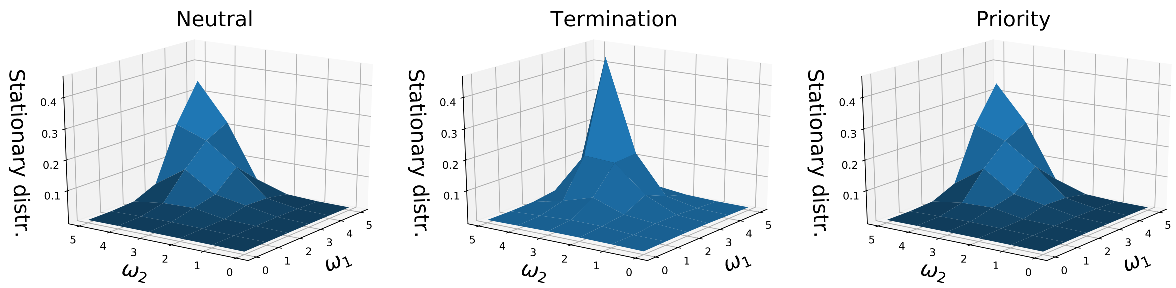


Figure 1.12: Stationary distribution of the system in the dynamic equilibrium

1.6 Extensions

Below I describe and discuss several extensions that could make the model more realistic.

Multiple speed tiers Throughout the paper I assume that consumers are offered a single speed tier. In reality, of course, there are multiple speed tiers available but it is important to note that the current definition of broadband speed (at least 25 Mbps) is enough to seamlessly enjoy an HD movie, while non-broadband tiers typically offer significantly slower connections that are not designed for video streaming. For this reason I only focus on broadband subscription and assume one single broadband speed tier. Modeling higher speed tiers should be necessary for subscribers that use several data-intensive applications simultaneously (e.g. large households might stream multiple movies at the same time on multiple screens) but for the main purpose of this paper it would excessively complicate the model.

Usage based pricing In the model consumers are only offered an unlimited data internet access plan. However, in reality ISPs are - to a limited extent - allowed to rely on usage-based pricing (UBP) practices to manage increasing demand for internet content (i.e. manage network congestion). My understanding is that UBP is essentially the consumer counterpart of termination fees. It was shown by Nevo et al. (2016) that it redistributes surplus from consumers to the ISP and eliminates inefficient (least valued) traffic – paralleling my findings with the termination fee. I have chosen not to model UBP because (i) unlike in other countries it is not widely applied in the U.S., (ii) even the lowest data caps would allow streaming an HD movie every single day in a month and (iii) it would likely not solve the prime-time congestion issue as it is hard to shift the movie experience to another time or location (daily commute, workplace etc.) because of its nature (e.g. need for big screen, darkness, family etc.).

Surge pricing Surge pricing is a technique used by an increasing number of companies for matching demand to supply. To this date I am not aware of any applications to managing network traffic but it is possible that in case of permanent severe congestion it may help to ease the load on the network. However, I find it unlikely to see surge pricing implemented for such purposes in residential networks because (i) current experience shows that consumers simply dislike when prices change frequently over time and (ii) it would make billing much more complex than a flat fee, which is of key concern to avoid the "bill shock" at the end of the month.[13]

[13]Consumers' aversion towards frequent price changes is not only related to surge pricing (like e.g. in

10.14754/CEU.2018.11

Subscriber heterogeneity Individual specific price sensitivity or mean utility coefficients would be a natural way to introduce heterogeneity to subscribers' preferences. Also, one interesting exercise would be to introduce correlation among the two to reflect that patterns of content consumption are related to wealth and demographics.

Subscriber lock-in To introduce history dependence to the ISP's market share I could exogenously set a fraction of consumers who stay with the ISP. This would need an additional state variable of lagged ISP share, though. As an alternative it is possible to use imperfect adjustment in subscribers' adaptive expectations. This could be done without increasing the state space by computing the new congestion as a (possibly random) convex combination of the actual and previous values.

Vertical integration It is possible to modify the static model in a way that the ISP and a selected CP maximize their joint profits. A simpler approach would be to let the ISP exogenously gain profits from the portion of its subscribers who choose the outside option in the CP choice problem and call it 'watch TV' instead of 'casual browsing'.

Multi-homing Multi-homing of consumers on the CP side might be important for both CP profits and the amount of network traffic that in turn affect investment incentives. It would be possible to extend the current setup to allow for a choice of multiple CPs but it would complicate the computation of the stage-game equilibrium.

Costs of congestion Currently I assume that congestion does not increase the costs of network management. To relax this assumption it would be possible to inflate the ISP's marginal cost by the congestion parameter.

1.7 Alternative equilibrium concepts

The current Markov Perfect Equilibrium approach has two serious drawbacks: It is hard to compute the equilibrium policies with a large number of firms and it requires firms to have full and symmetric information on the states of all other firms. If I would want to generalize the model to include more firms, e.g. by either simply increasing the number of firms or by allowing for different content types (i.e. allow for horizontal differentiation) then in the current framework the equilibrium would be almost impossible to compute and the informational assumptions would be questionable at best.

case of Uber) but also to dynamic pricing practices (like e.g. in the very recent case of Amazon) where firms systematically change prices for various reasons including experimentation and obfuscation etc.

However, in reality the market for online content is characterized by a few dominant firms and many small firms (i.e. a competitive fringe). As a result in a model that allows for such market structures, it may be reasonable to assume that firms with many competitors are more sensitive to changes in the dominant firms' states and it may be unrealistic to expect that firms have resources to keep track of the evolution of all rivals.

There are several approaches in the literature to address these issues. To some extent kinds of generalizations of the idea in Krusell and Smith (1998) to the framework of Ericson and Pakes (1995).

One solution to the problems with many firms and full information is offered by Weintraub et al. (2008). They define a new equilibrium concept that they call oblivious equilibrium, in which each firm is assumed to make decisions based only on its own state and knowledge of the long-run average industry state, but where firms ignore current information about competitors' states. This approach is appealing but it is not suitable for concentrated industries and informational asymmetries. In a follow-up paper Benkard et al. (2015) introduce the partially oblivious equilibrium that allows for a set of strategically important firms whose states are always monitored by every other firm in the market. This alleviates the problem of symmetric information but works only with relatively 'light-tail' distribution of firms. In a recent work Ifrach and Weintraub (2017) define a new equilibrium concept that they call moment-based Markov equilibrium, in which firms keep track of their own state, the detailed state of dominant firms, and few moments of the distribution of fringe firms' states.

An alternative approach to introduce asymmetric information is offered by Fershtman and Pakes (2012) where firms may not observe some part of their competitors' states only some publicly observed actions. They define the restricted experience based equilibrium where agents form expectations about the likely future outcomes of their actions' based on their past play using a simple learning algorithm. Unfortunately, their approach is not directly applicable for an industry with a large number of firms. However, with an appropriate modification to firms' information sets it may be an appealing option for my setting.

1.8 Conclusion

In the preceding sections I introduced a model that allows for a rich set of interactions among subscribers, internet service providers and content providers to address the questions and concerns of the net neutrality debate. I believe that this approach leads to more realistic incentives and, therefore, leads to better approximation of agents' likely behavior under different regulatory environments in this complex market.

10.14754/CEU.2018.11

I think the main task for the future is to perform thorough testing of the findings in a large variety of parameter settings and to find an appropriate set of parameters that would generate quantitatively meaningful results. Also, future work on a proper equilibrium concept seems inevitable to accommodate a large number of firms and different content types.

10.14754/CEU.2018.11

Chapter 2

Counterfactual Analysis of Net Neutrality in a Calibrated Model

2.1 Introduction

Residential broadband internet provision and online video streaming are very large industries. In 2015 there were 89 million broadband subscribers (households) in the U.S. alone. The aggregate revenues of internet service providers from residential high-speed internet access approached $40.6 billion and total network investments exceeded $6.6 billion. The top four online video streaming firms accounted for more than 60% of prime-time download traffic.[1]

Surprisingly, neither did the Federal Communications Commission conduct rigorous analyses to investigate the potential effects of their network neutrality laws, nor has been published any peer-reviewed empirical work that is suitable for counterfactual analysis. This paper attempts to produce ballpark counterfactual estimates based on the following strategy. I take the structural model of chapter 1 and assume that currently observed market outcomes constitute an equilibrium in the static game. This allows me to calibrate the model parameters to match observed moments derived from public data sources. Then I solve the model with state of the art numerical algorithms to assess the effects of the introduction of termination fees and paid prioritization.

My key findings are that the lack of net neutrality favors the ISP and paid prioritization might be worth to investigate more as it makes it possible to restore the value lost due to congestion.

[1] I have chosen to exclude mobile ISPs from the analysis mainly as they are less relevant to the prime-time entertainment market and the mobile network's capacity is a small fraction of the fixed network's capacity. However, both might change as a response to a widespread adoption of zero-rating for high definition video content.

10.14754/CEU.2018.11

2.2 Data sources

Here I give a brief overview of the several different sources of publicly available data that I use.

Measuring Broadband America
MBA is a program administered by the FCC in the United States. They distributed 5000 whiteboxes to households scattered all over the United States which collect various internet performance related measurements in every hour. The distribution of whiteboxes matches the distribution of the number of households with broadband subscription across states and also the number of subscribers in various speed tiers which makes it a small but representative sample. I use two main measurements from the dataset: total traffic (in bytes) for each hour and a video performance related test.

SEC filings
Form 10-K and Form 8-K are invaluable sources for company revenues, profits, membership numbers and capital expenditures. I processed the public filings of major streaming video content providers (Amazon, Hulu, Netflix, and Youtube) and major publicly traded internet service providers offering a wireline service and above the yearly 1 billion $ revenue threshold (15 companies including AT&T, CenturyLink, Comcast, Time Warner Cable, and Verizon).

Internet companies' reports
Sandvine is a networking equipment company who publishes various internet traffic and performance measurements in its widely cited Global Internet Phenomena Report series. I use their content and firm specific traffic measurements in peak-hour periods. The other piece of industry source I use is Cloudflare, an internet company that provides a content delivery service on a global scale. I use their estimated average price for carrying network traffic.

Market research companies
I use the number of monthly unique viewers for streaming video content providers and their average monthly time spent on watching videos from the Advertising & Audiences report of Nielsen. On the other hand Comscore measures and reports ad-based online video audiences and the number of videos watched in their monthly Online Video Rankings.

American Community Survey
To gauge the size of the market for internet subscription services and to compute broadband penetration in the United States I use the total number of households and the number of households with a broadband internet subscription from the annual ACS 1-year estimates.

10.14754/CEU.2018.11

Academic research

In the lack of directly connected published empirical work I use estimates of consumer preferences from related empirical studies. I use the estimates for consumers surplus derived from and willingness to pay for internet subscription from Nevo et al. (2016) and Liu et al. (2017). I use consumers' willingness to pay for not seeing advertisements from Train (2016) who studies demand for online video streaming.

Internet search

I collect content providers' prices using the Internet Archive Wayback Machine available at *www.archive.org*.

2.3 Calibration

The goal of the calibration is to find values for the parameters of the model that generate behavior which is close to observed outcomes in the data. If incentives are accurately represented then the model outcomes should be in the ballpark of the actual effects of the changes in regulation.

To calibrate the parameters I will use observed market outcomes from 2015 in the online video streaming industry and assume that the actual values constitute an equilibrium. Then I will set parameters such that in equilibrium those outcomes are matched. The dynamic modeling framework has the advantage that the static and dynamic games can be calibrates separately. First I describe the parameters needed for the static game and I go through the moments I use to set their values. Then I present the calibration of the dynamic part.

2.3.1 Annual timing assumption

For both parts of the model it is important to set the time frequency of agents' decisions because observed market outcomes should be interpreted accordingly. To consolidate the timing of subscription and pricing decisions across the two sides of the market I use an annual frequency. This is more consistent with both ISP and CP price changes which are rare and are much closer to annual rather than monthly decisions.[2] The annual frequency is also better for modeling investments which are typically "lumpy" in a sense that it takes time to deploy a better broadband infrastructure, or to implement a new content standard or to shoot a new season of a show in higher resolution. The only part where the new timing assumption is a bit less plausible compared to a monthly model is the formation of consumers' expectations of future network congestion.

[2] In fact, several major ISPs and Amazon actually do offer annual contracts.

10.14754/CEU.2018.11

Since several pieces of data are reported on the monthly level I calibrate the static game to match those monthly outcomes and then scale up everything that goes into the dynamic model as input to the annual level.

2.3.2 Static part

There are seven types of parameters in the static game:

- market size (M)
- congestion (μ)
- CP product mean utilities (δ)
- utilities of outside options (V_0, v_0)
- price sensitivity of consumers in the internet subscription decision (a)
- product bitrates (b)
- marginal costs (c^{cp}, c^{isp})

Market size and congestion are calibrated entirely from data without relying on the model. However, for mean utilities, outside options and price sensitivity I will make heavy use the double-logit specification of the consumers' two-stage decision and data on broadband penetration, market shares of Netflix, Amazon, Hulu and Youtube, prices for content and internet subscription and consumers' willingness to pay for internet access. For the cost related parameters and bitrates I will rely on the first order conditions of optimality in the stage game and standard video bitrates.

Market size

I set total market size to $M = 118$ million which is the number of US households in 2015 according to the American Community Survey. As both broadband subscription and online streaming video are tied to physical homes I find the unit choice of a household appropriate.

Congestion

To gauge congestion I use data from the Measuring Broadband America (MBA) program that were collected from participating households' network routers. The whiteboxes attached to the routers perform several network quality measurements in every

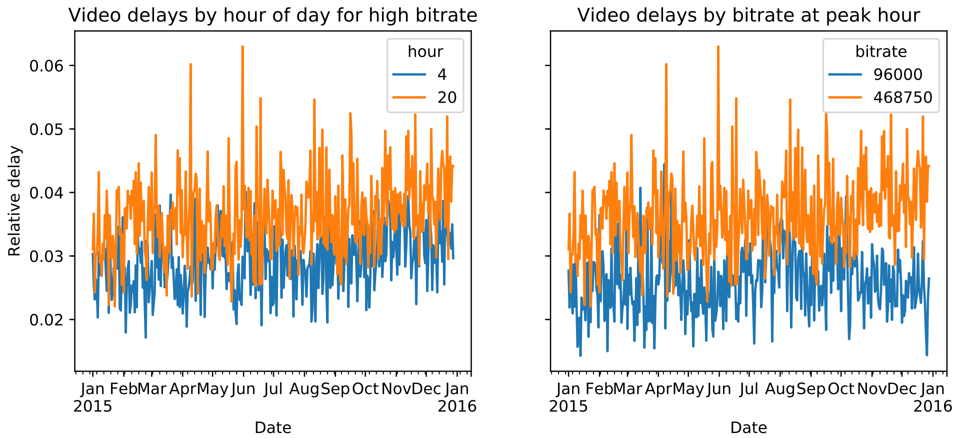


Figure 2.1: Evolution of daily average relative delays in video streaming

hour. One of the tests conducted is a video streaming test. In every hour the device tries to fetch a high and a low bitrate video file from a server and it records the time the video was lagging during the test. I take the measured delay for each household from the peak hour (8-9 pm) and normalize it by the length of the video which was fixed for all tests to get a relative delay. Figure 2.1 presents the average daily relative delay across all households for the sample period and confirms that delays are more severe for the peak hour and higher quality videos are more sensitive to network congestion. One can see that the high bitrate video was interrupted for ca. 4% of its duration on average while this value was just ca. 2.5% for the low bitrate video. I will use the measurement for the high bitrate video as a proxy for the extent of quality degradation and attribute it to a congested network with parameter $\mu = 1 - 0.04 = 0.96$.[3]

CP mean utilities, outside options and price sensitivity

Recall, that consumers have the following two decisions to make:

1. Subscribe to the ISP to have internet connection or not
2. Conditional on having internet connection whether or not to subscribe to an available CP

I will work backwards from the CP choice. To calibrate mean utilities that consumers get from enjoying the CPs' products I use market shares determined by the model as

$$s_j = \frac{\exp\{\mu\delta_j - p_j\}}{\exp\{v_0\} + \sum_k \exp\{\mu\delta_k - p_k\}}, \forall j = 1 \ldots 4.$$

[3] I pick the congestion from the high quality video measurements because the low quality video's bitrate is not sufficient for any sensible movie experience with current technology.

and match them to market shares reported by market research companies. A recent survey conducted by Parks Associates, a company specializing in emerging consumer technology products and services, reveals that in 2015 52%, 24% and 14% of broadband households had subscriptions to Netflix, Amazon and Hulu, respectively. To get Youtube's share among broadband households I use the data reported by ComScore, a large media research company that publishes its monthly Video Update, where it lists the unique viewers for ad-supported video sites. In 2015, Youtube had ca. 162 million monthly unique American viewers which I divide by 2.6 (the average household size according to the American Community Survey) and combine it with the fact that 76% of U.S. households had broadband subscription in 2015 to conclude that 69% of U.S. broadband households watched Youtube in an average month.

There is a complicating factor, however, as because of multi-homing of consumers market shares add up to more than one . In fact, half of the households subscribing to a monthly paid streaming video service subscribe to more than one and possibly most of them watch Youtube as well.[4] As the model is not able to accommodate multi-homing customers I normalize the market shares of Netflix, Amazon, Hulu and Youtube, so that they add up to 87%, the share of all U.S. broadband households that watches at least one streaming video in a month.[5] Finally I end up with shares that add up to 87% as $s = (52\%, 24\%, 14\%, 69\%) \times 0.547 = (28.4\%, 13.1\%, 7.7\%, 37.8\%)$ for Netflix, Amazon, Hulu and Youtube, respectively. To correct for this market "shrinkage" I will scale up the profits by the factor of $1/0.547$ when comparing profits from the model to observed quantities.

To calibrate the mean utilities, on top of market shares I use historical monthly prices for the video subscription services and I use the median household income to determine the forgone wages spent on watching advertisements on Youtube.[6] In this way I get monthly prices of $p = (\$10, \$8.25, \$8, \$6)$ for Netflix, Amazon, Hulu and Youtube, respectively.

Now that I have s, p and μ I can perform a contraction mapping similar to Berry et al. (1995) conditional on any value of v_0 which I set to $v_0 = 4.35$.[7] In this way I end up with mean utilities $\delta = (15.76, 13.13, 12.31, 11.89)$ for Netflix, Amazon, Hulu and Youtube, respectively.

[4]See Sean Buckley, "About 50% of OTT video subs have multiple subscriptions, Parks says", Fierce Cable (Jan. 4, 2017).

[5]See Nielsen, Advertising and Audiences 2015

[6]Viewership minutes come from the ComScore Video Update report.

[7]v_0 is a free parameter in the model. However, I set it to a value that leads to equal marginal costs between the ISP and CPs.

Note, that the expected utility from the CP choice situation (V) can be determined based on the now available quantities as

$$V = \ln\left(\sum_k \exp\{\mu\delta_k - p_k\} + \exp\{v_0\}\right).$$

The calibrated quantities δ, μ, v_0 and observed prices p lead to $V = 6.39$. Next I use the estimates of Liu et al. (2017) for U.S. households' willingness to pay (wtp) for internet connection. Using their estimates and the median upload and download characteristics of plans in the Measuring Broadband America program I get a WTP for a monthly connection of about \$60.[8] Together with V and using the specification for observed utility in the ISP subscription problem ($V - aP$) this pins down the price sensitivity parameter as $a = V/WTP = 0.105$. The two final pieces needed to calibrate the utility of the outside option (V_0) are the share of broadband households (S) and the ISP price (P). Broadband share is set to $S = 76\%$ based on the American Community Survey, as noted eariler. For the internet access price I use the median price (\$45) paid by households participating in the Measuring Broadband America survey.[9] As a result I set $V_0 = 0.44$ to satisfy

$$S = \frac{\exp\{V - aP\}}{\exp\{V_0\} + \exp\{V - aP\}}.$$

Costs and bitrates

Marginal costs of content distribution and bitrates are pinned down by the first-order conditions of optimality.

$$\left(1 - Ss_j\right)\left(p_j - c^{cp}b_j\right) - 1 = 0, \qquad j = 1\ldots4 \tag{CP}$$

$$a(1-S)\left(P - c^{isp}\sum_j s_j b_j\right) - 1 = 0 \tag{ISP}$$

Note that the three missing pieces are b, c^{cp} and c^{isp} everything else is either data or has been calibrated already.

I start with the first-order conditions of CPs. Initially I set bitrates to $b_{init} = (3.5, 3.5, 3.5, 2.5)$ Mbps for Netflix, Amazon, Hulu and Youtube, respectively. These values are standard minimum bitrates required for HD streaming movies on these sites.[10] Then I use the first oder conditions (CP) to compute the corresponding marginal costs for all four

[8]The median broadband upload speed in the sample is 9 Mbps while the median download speed is 41 Mbps.

[9]Based on cable companies SEC filings this matches almost perfectly the average revenue per high-speed internet subscriber, which is \$44.32.

[10]See the collection at https://www.lifewire.com/internet-speed-requirements-for-movie-viewing-1847401

content providers individually. Specifically I get $c^{cp}_{init} = (2.49, 2.04, 1.98, 1.84)$ measured in \$/Mbps/user. However, I would like to have a common unique value for the marginal cost of content distribution across CPs, so I set $c^{cp} = 2.09$ \$/Mbps/user, the mean of the individual marginal costs. Then I re-set bitrates to CP specific values that satisfy the first-order conditions with the common marginal cost. The new values are $b = (4.18, 3.42, 3.32, 2.2)$ Mbps.

To calibrate the ISP's marginal costs I use the first-order condition of the ISP. All terms except for c^{isp} are set by now and the only value that satisfies the equality is $c^{isp} = 2.09$ \$/Mbps/user.[11]

2.3.3 Calibrated values and sanity checks

The final set of parameters that I adopt for the static game:

- $\delta = (15.76, 13.13, 12.31, 11.89)$ for Netflix, Amazon, Hulu and Youtube
- $b = (4.18, 3.42, 3.32, 2.2)$ in the same order, measured in Mbps
- $(v_0, V_0) = (4.35, 0.44)$
- $a = 0.105$
- $c^{cp} = c^{isp} = 2.09$ in \$/Mbps/user
- $\mu = 0.96$
- $M = 118$, millions of U.S. households.

Note, how all parameters seem to make sense. The price parameter (a) leads to a price elasticity of -1.14 in the internet access decision. Bitrates are standard bitrates for HD-streaming movies. The marginal content-carrying related costs are in an acceptable range. For such high-volume players in the internet eco-system, a \$2-8 unit price said to be acceptable.[12] Moreover, if I compute the optimal prices for this set of parameters, I match prices and market share perfectly.

Next I compare revenues and profits predicted by the model to actual data reported in company SEC filings. Netflix's monthly revenue in 2015 was \$550 million and my model predicts \$255 million. However, if I account for the lack of multi-homing in the

[11] Note that c^{isp} equals the marginal cost for the CPs. This is fully intentional by selecting a v_0 parameter that led to an a that scaled the ISP's first-order condition so that the solution for c^{isp} is 2.09.

[12] See Cloudflare, Internet transit costs report.

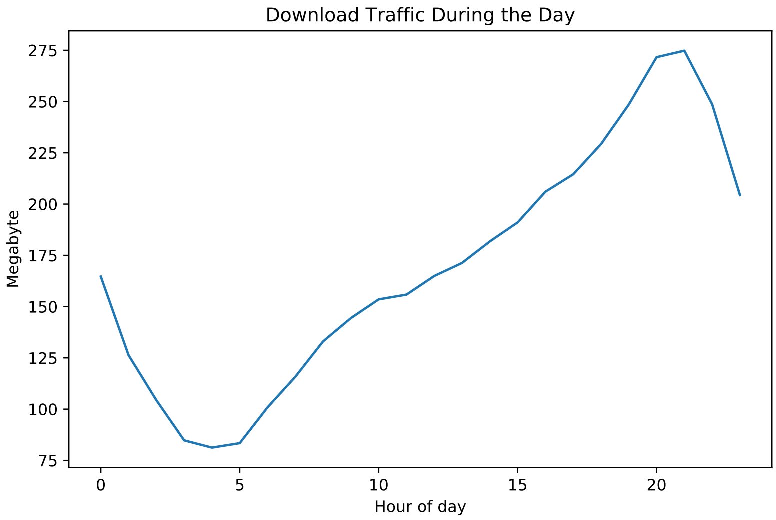


Figure 2.2: The evolution of download traffic for the average user over the day

model and upscale profits by the factor of (1/0.547) I get $466 million. As an alternative, now I compare the profit/revenue ratios which should be independent of the scaling/shrinking effect. Netflix reports a 20% ratio, while my model predicts 12.76%.

For a reasonable ISP comparison, I add up the revenues of all large publicly traded ISP based on their SEC filings. My model predicts $4042 million monthly revenues and a 56% profit/revenue ratio for the single ISP, while aggregate company revenues add up to $3383 million in a month with a 41% profit/revenue ratio.

Traffic check

To compare model traffic to actual volumes I first I have to approximate the actual values. In order to do that I use an MBA measurement to aggregate individual household level browsing data to the country level. In each hour whiteboxes record the amount of downloaded data at every household participating in the program. Figure 2.2 shows the evolution of downstream traffic over a day at an average household. One can see that the peak is around 8-9 pm and traffic volume is ca. 4 times higher than in the early morning hour. It is important to distinguish the peak hour, because I only have traffic share estimates of video streaming for that period. To approximate peak hour total download traffic first I compute the mean download traffic only for the peak hour across all households in the MBA sample for every day of the sample period. Then I scale it up by the number of households with broadband connection. I get the number of such households for Jan 1 of years 2014, 2015 and 2016 from the American Community Survey and then linearly interpolate between the three dates to get a daily value for the entire sample period. Figure 2.3 shows the evolution of country level peak hour download traffic over the sample period. The mean peak hour total downstream

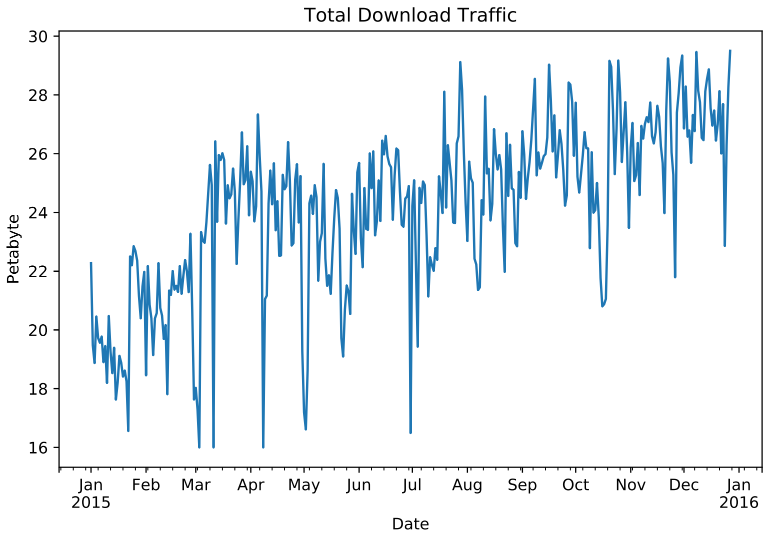


Figure 2.3: Daily total peak hour download traffic

traffic is 25.8 Petabytes per hour. To approximate the share of Netflix, Amazon, Hulu and Youtube I use the annual Sandvine Global Internet Phenomena Report from 2016. Their measurements show that in North America the joint downstream traffic share of Netflix, Amazon, Hulu and Youtube in the peak hour was 59.6%. Combined with the upscaled MBA measurement data this leads to an estimated 15.4 Petabytes downstream traffic in the peak hour that my model has to match.

The raw output of my model is 244 Terabit per sec bitrate network traffic of video streaming. In order to make the two numbers comparable I adjust for the following factors:

- Upscale by (1/0.547) to account for multi-homing
- Convert to Petabytes per hour (8×3.6)
- Adjust for time spent with daily online video streaming (5.48/60)

For the last adjustment I use the Nielsen Total Audience Report 2015, that reports the average daily minutes spent on streaming video to be 5.48 minutes. Taken the three together I get peak hour downstream video traffic of 18.6 Petabytes per hour as an output of my model. This is 20% more than my approximate number for the actual value but it is indeed very similar in magnitude.

2.3.4 Dynamic game

The calibration of the dynamic game consists of three main parts:

- The state space grid for all three kinds of states (δ, κ, μ)
- Discounting and depreciation (β, η)
- Investment efficiency (α, τ)

State space grid

Note that the exact calibrated mean utilities are only used directly for the sanity checks and the "as if" counterfactuals in section 2.5.1. For the dynamic game I create a grid of mean utilities and corresponding video bitrates so that firms can invest in order to move across the quality ladder. To do so I extrapolate the ratio of the calibrated mean utilities and video bitrates to extend it in both directions with a lower valued/less bandwidth intensive and a higher valued/more bandwidth intensive product. For theoretical and numerical considerations I alter the values to ensure that the resulting profit functions are bounded and satisfy the sufficient conditions for existence of a Markov-Perfect Equilibrium in a very similar dynamic industry model of Ericson and Pakes (1995). The final values that I use in the algorithm:[13]

- $\delta = (11.56, 12.81, 13.78, 14.74, 15.44, 15.77)$
- $b = (2.25, 2.45, 2.8, 3.17, 3.77, 4.49)$

To set the capacity grid I take the approximated total downstream traffic and create a grid that goes beyond that limit so that the ISP can always reach a state where its capacity exceeds the maximum potential network traffic. There are two reasons I believe that there is a constantly binding capacity constraint in the peak hour and as such total peak hour traffic can be interpreted as network capacity. First, every day in the sample period there is always congestion in the peak hour. Second, Figure 2.4 confirms that there is a positive relationship between the traffic volumes and delays and that there might be a binding capacity constraint in the background.

For the congestion parameter I take the most severe congestion from the sample period ($\mu_{min} = 0.938$) and extend the grid so that the industry could reach states with twice as severe congestion values.

Discounting and depreciation

I set $\beta = 0.8$ to match the average annual treasury yield curve rate in 2015. To set the probability of the onset of a deprecation shock (η) I use the industry stylized fact that in

[13]The upper and lower boundaries were selected as when a monopolist incumbent CP stops to invest.

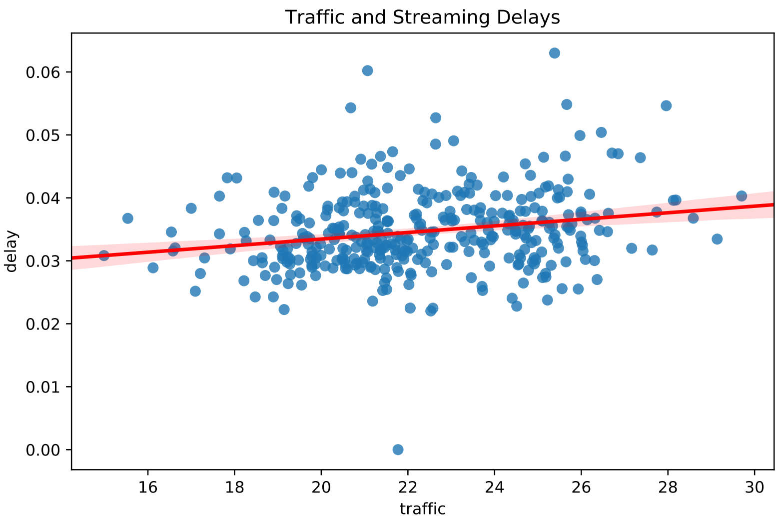


Figure 2.4: Daily total peak hour download traffic against daily average relative delays

the last almost 10-12 years there were no two consecutive years without a major video resolution upgrade that became a new standard afterwards.[14] To match that I set the the probability of such a shock to 0.775, which implies a probability of the industry not seeing a hit within two years not more than 5%.

Investment

To set the ISP's investment parameter (τ) I try to match the ratio of aggregate cable ISP network investments (capital expenditures) and aggregate revenues attributed to high-speed internet services, both as reported in SEC filings. I set the parameter of CPs' investment (α) similarly, where I take the ratio of investments in technology and development to revenues. I re-run the dynamic game until both ratios should are in the reported 5-10% bounds for the neutral regime.

2.4 Computational algorithm

Just like the model the computational algorithm also consists of two separate parts corresponding to the static and dynamic parts.

2.4.1 Homotopy method

The first-order conditions of optimality in the static game lead to a system of nonlinear equations. The system is relatively small, all functions are smooth and differentiable,

[14]With the exception of 3D television, which has almost disappeared from the market by 2017.

10.14754/CEU.2018.11

the analytical Jacobian of the system is available and is relatively cheap to evaluate. Basically I have everything which makes standard Newton-like methods an attractive choice. However, this system has to be solved many times under constantly changing conditions. Indeed, the large number of potential industry states and the varying parameters require the algorithm to be robust to all such settings.

Unfortunately the performance of Newton's method depends heavily on the initial values. Practically this means that after any change to the parameters I would need to come up with new values that are in the neighborhood of the final solution (i.e. prices). To alleviate the problem I've implemented a homotopy algorithm that establishes a connection between two different sets of industry states or parameter settings one of which already has a computed solution.

The idea behind the homotopy method is the following. Imagine that you have to solve a system, that is difficult to solve. Then instead of solving the hard problem directly, start with an other or similar problem that is easy to solve. Then gradually change the easy system to be more "similar" to the difficult problem and use your initial values from the preceding computations. More formally, suppose you want to solve the system $F(x) = 0$.

1. Find a system, say $F_0(x) = 0$, that is easy to solve.
2. Define the homotopy map as $H(x, \lambda) = \lambda F(x) + (1 - \lambda)F_0(x)$.
3. Start with $\lambda = 0$ and solve $H(x, \lambda) = 0$. Label the solution as x_λ.
4. Iterate over a fine grid of $\lambda \in [0, 1]$ values and solve $H(x, \lambda)$ using initial values from the preceding iterations.
5. The final solution corresponding to $\lambda = 1$ is the solution to $F(x) = 0$.

Although this procedure describes the general idea quite well there are many additional details to the homotopy method. It is possible that the solution path, i.e. the set of consecutive (x_λ, λ) pairs, does not represent a proper functional relationship between x_λ and λ. There is actually a reason that step 4 does not specify the relationship of successive λ values. Although in simple problems it is sufficient to gradually increase λ sometimes it may be necessary to decrease it or simply restart the whole procedure from a range of new initial values. Depending on the properties of $H(x, \lambda)$ solution paths can be non-smooth or even discontinuous. Algorithms that can deal with such complications are called path following algorithms and they utilize additional information on $F(x)$ and in general $H(x, \lambda)$ at the actual locations when selecting the consecutive λ and x values for the next iteration. In my implementation I allow for simple turning points, where λ has to be increased but not for branching and, therefore it can't trace out multiple equilibria. However, I've not encountered a single case where the solution path was non-monotonic in λ.

First I used the algorithm to compute a solution for a simple monopoly industry using $F_0(x) \equiv x - a$, where a is a constant, and thus $F_0(x) = 0$ is indeed easy to solve. Then when advancing to more complicated industry structures I set a to the monopoly price and use λ to gradually change the state vector. In this way I can utilize the good local convergence properties of Newton's method and only worry about the initial values of the very first monopoly problem as the homotopy ensures the appropriate values for the successive problems.

2.4.2 Reinforcement learning

In mathematical terms the equilibrium of the dynamic game is a solution of a functional equitation. There is a very extensive literature of numerical dynamic programming methods that potentially could be used. Doraszelski and Pakes (2007) give a good survey of the methods and their properties that have been applied to dynamic industry models.

The main computational challenge in multi-agent dynamic programming problems with a large state space lies in the computation of agents' continuation values. The continuation values are based on the agents' expectations of their competitors' likely actions. This is a complicated object and different algorithms construct it differently. To address the computational burden of simply iterating over all possible outcomes of competitors' actions to assess the probability of state-to-state Pakes and McGuire (2001) (PM) suggest a stochastic approximation. Their key observation is that $W(\text{investment outcome}|\text{state})$ – the expected value of future payoffs given a certain investment outcome conditional on being in the current state – summarizes the agents' expected future returns and it is fully sufficient for making investment decisions. As such W fully determines agents' behavior and it is sufficient to iterate over W instead of value or policy functions and stop when the actual W generates policies and values that constitute an equilibrium. Their algorithm approximates W iteratively based on simulating agents' behavior instead of numerically integrating over all possible actions.

The attractive feature of the stochastic PM algorithm is that it can be given a behavioral interpretation of constantly learning agents. This basically means that agents are no longer expected to perform complex calculations such as numerical integration to compute optimal policies just to follow their best responses dictated by their W values and to update these expectations after the realizations of their competitors' actions. This constant experimentation by following some greedy policy that is based on the actual expected values resulting from possible actions accompanied by a learning rule to update expectations is essentially the same general idea that is behind temporal difference learning and reinforcement learning algorithms which are well-known in operations research and artificial intelligence.

The procedure that I describe below follows PM but it was modified to allow for both kind of agents in the model. First set an initial value for W. In practice it is useful to set it to a value that is expected to approximate the relative differences among states, like $\pi/(1-\beta)$. Also my experience shows that with larger state spaces it is further useful to "inflate" the part of W that corresponds to the successful investment outcome to encourage more active experimentation. Then select an initial industry state ω and iterate over the following steps at each iteration k:

1. Given the actual value of $W(\cdot|\omega^k)$ compute firms' optimal investment for the current industry state ω^k.
2. Using these policies simulate the investment and determine ω^{k+1}, the new state of the industry.
3. To update the values in $W(\cdot|\omega^k)$:
   (a) Evaluate the value function for every firm i with state ω_i^k in the new industry state ω^{k+1} using the values in $W(\cdot|\omega^{k+1})$.
   (b) Set the value in $W(\cdot|\omega^k)$ to a weighted average of the evaluated value function at the new industry state and the actual value of $W(\cdot|\omega^k)$.
4. Continue with the iteration at state ω^{k+1}.

There are several important things to note. First, notice that the algorithm is asynchronous as it updates the values in W only corresponding to the actual state that is visited. Second, the evolution of the industry states is determined by simulated random draws. There is no guarantee that the algorithm will visit each state enough times so that the averaged values across visits to a state give a good approximation to W. However, it is only required that the algorithm gives us good approximation of W at states that are in the recurrent class.[15] Third, to check whether the algorithm has converged it would be necessary to stop the algorithm after several iterations and perform a test that involves the same burdensome numerical integration for each state that is done in deterministic algorithms. However, I follow the method of Fershtman and

[15]There is an additional problem with states that are on the boundary of the recurrent class. That is, from boundary states it is possible to get to states that are not in the recurrent class only firms' equilibrium policies are such that it never happens in equilibrium. Thus firms' in boundary states are required to evaluate options that belong to states that are not expected to be visited infinitely often, and thus the algorithm won't produce a good approximation to their corresponding values in W. See the recent work of Asker et al. (2016) on how one might try to alleviate the problem.

10.14754/CEU.2018.11

Pakes (2012) who propose an algorithm that avoids the need for a such a computationally intensive test. First I simulate the industry without updating W for a large number of iterations so that I can be confident that it has wandered to the recurrent class. Then I simulate the model for R periods. If R is sufficiently large, the visited states constitute the recurrent class where I have to check the consistency of W with the outcomes from the policies generated by W. Fershtman and Pakes (2012) show that if W and the average of realized firm values over the simulated sample paths are close to each other then W generates equilibrium policies on the recurrent class.

2.5 Results

2.5.1 Static counterfactuals

To see what would happen to the industry with the current market players (Netflix, Amazon, Hulu, Youtube and an aggregate ISP) with either termination fees or paid prioritization I use the calibrated values and compute the static equilibrium without allowing for dynamic restructuring. Table 1 summarizes the results. As I have discussed earlier, the neutral regime captures the current market outcomes fairly well.

After the introduction of the termination fee one can see that prices adjust according to expectations. The ISP charges a per-transaction termination fee to the CPs who transmit it to the customers. As a result while the ISP can lower its price all CPs have to raise their prices. This in turn raises ISP profits but decreases CP profits. It seems that the ISP's lower price can't compensate consumers for the higher content prices and consumer surplus drops. A interesting outcome is that traffic drops proportionally much more than broadband share.

In the case of paid prioritization one can see that the highest quality CP (Netflix) buys the priority access and it is the only CP that can maintain its neutral level profits. Moreover, it can further raise its price since the competitors' products became relatively worse. The ISP seems to be successful in extracting the value from the restored content quality and is able to significantly raise its profits. For other CPs this is the worst of the scenarios with even lower profits than with termination fees. Consumer surplus, broadband share and traffic all increase.

In general I conclude that results are in line with the pattern from the baseline theoretical model from chapter 1.

10.14754/CEU.2018.11

Table 2.1: Summary of static counterfactuals, monthly values

outcome	ISP price ($/m)	CP price ($/m)			
firm	aggr.	YT	HU	AM	NF
model					
neutral	45.1	5.99	7.99	8.24	9.99
termination	42.3	6.42	8.46	8.70	10.43
priority	46.6	5.93	7.98	8.23	10.14

(a) Prices from static counterfactuals

outcome	ISP profit ($m)	CP profit ($m)			
firm	aggr.	YT	HU	AM	NF
model					
neutral	2263	86.8	13.3	23.9	59.5
termination	2284	78.1	11.8	21.3	53.3
priority	2354	73.7	11.1	20.0	59.6

(b) Profits from static counterfactuals

outcome	consumer surplus ($m)	broadband share (%)	traffic (Tb/s)
model			
neutral	2069	76	446
termination	2010	74.6	408
priority	2097	76.6	484

(c) Consumer surplus, broadband share and traffic volume from static counterfactuals

10.14754/CEU.2018.11

2.5.2 Dynamic equilibria

In Table 2 I report the dynamic equilibrium outcomes weighted by the stationary distribution. There are no dominant or clearly dominated alternative regulatory regimes. However, it seems that the lack of net neutrality in general favors the ISP and makes the CPs worse off. It is not surprising if we think about how the ISP gets more tools to extract value from CPs.

It is interesting to see some arguments supported by the outcomes and complemented by the full picture, parts of which are often overlooked. E.g. allowing paid prioritization indeed leads to less ISP investment and a more congested network – as net neutrality proponents emphasize – but on the other hand it leads to higher CP investment and restoring the value lost from congestion is valued by consumers. Or termination fees are indeed used by the ISP for network developments – as net neutrality opponents claim – but they are transmitted to the customers and squeeze CP margins and discourage content development.

2.6 Conclusion

In this paper I showed a potential way to calibrate and solve numerically a complex dynamic industry model that tries to capture realistic incentives of players in the prime-time online streaming video market. Then I conduct counterfactual analysis of different types of net neutrality regulation. My conclusion is that (i) regulation should be very careful as there are serious distributional consequences in the non-neutral regimes in favor of the ISP and that (ii) introducing paid prioritization may restore value by saving sensitive content from congestion.

Table 2.2: Summary of dynamic equilibria, annual values

	ISP value ($tr)	ISP investment ($b)	CP value ($b)	CP investment ($m)
neutral	135	3.84	2.75	161
termination	137	4.01	2.46	149
priority	141	3.19	2.47	172
both	143	3.17	2.43	168

(a) Expected values and investments in the stationary distribution

	CPs	level	capacity (Tb/s)	congestion	recurrent class
neutral	4	3.51	475	0.965	20.20%
termination	4	3.39	448	0.987	14.47%
priority	4	3.78	502	0.941	27.03%
both	4	3.77	456	0.958	25.81%

(b) Mean industry characteristics in the stationary distribution

	consumer surplus ($b)	industry profit ($b)	total welfare ($b)
neutral	2.26	2.54	4.80
termination	2.20	2.55	4.75
priority	2.29	2.62	4.91
both	2.27	2.57	4.84

(c) Expected consumer surplus and welfare in the stationary distribution

	broadband share	network traffic (Tb/s)	priority share
neutral	76.7%	474	-%
termination	72.5%	439	-%
priority	77.2%	504	14.15%
both	76.2%	472	13.80%

(d) Other important outcomes in the stationary distribution

10.14754/CEU.2018.11

Chapter 3

Delineation of Market Areas Using Sparse Learning and Spatial Regularization

3.1 Introduction

The main objective of any empirical market delineation exercise is to help organizing the way we think about the economic activity we observe and provide sound foundations for further economic or antitrust analyses. E.g. firms' market shares can only be defined with respect to a certain geographical or product market and therefore both the EU and US jurisdictions explicitly require antitrust authorities to undertake a market definition exercise as the first step in an investigation before progressing to evaluate competitive effects. As a consequence, market delineation has to be performed quickly, frequently and in a variety of different settings using only some basic routinely collected data. The fact that the task is quite challenging and is considered to be a potentially decisive factor in court decisions explains the increasing number of available methods. This paper contributes to the literature by offering a practical method for identifying markets that compromises between simplicity and the amount of structural assumptions to combine the attractive features of the different existing empirical approaches while alleviating some of their problems.

The idea behind almost all methods is that pricing constraints that restrict a firm's ability to increase prices arise directly from firms who compete in the same market. In principle one could infer competitive relationships among firms if pricing constraints could be estimated from the data and this is indeed what most methods are designed to accomplish. The two main differentiating dimension of approaches in the literature are (i) whether they are based on the simplest statistical assumptions or on a specific model derived from economic theory and (ii) whether they require proprietary

10.14754/CEU.2018.11

or costly to collect data on top or routinely observed data like prices and locations. Analysis of pairwise correlations and structural demand modeling represent the two opposite ends of the spectrum. In Section 3.2 I argue that notwithstanding their attractive features both approaches have significant drawbacks which might make them unattractive or even completely inapplicable in certain cases.

In Section 3.3 I describe an empirical procedure that has its origins in the statistical learning and signal denoising literature. There are two key assumptions I maintain.

1. Pricing constraints among competing firms result in co-movements in prices.
2. The First Law of Geography holds.[1]

The role of the first assumption is to justify the use of regression analysis for detecting statistical relationships among firms' prices. The second assumption motivates the use of spatial regularization in the regression analysis. By applying spatial regularization I constrain the relationship between a given firm's prices and any pair of two other firms' which are close together in some underlying space.[2]

In Section 3.4 I demonstrate the working of the method on simulated data and show that in a simplified example it is possible to recover market boundaries even in a high-dimensional settings with significant multicollinearity. In Section 3.5 I apply the procedure to a hypothetical merger case in the Hungarian retail gasoline market and show how one can use the method to highlight areas of potential harm that should be rigorously investigated by the antitrust authority.

3.2 Existing Empirical Approaches

To explain what motivated the new approach for identifying competitors this section gives a brief overview of empirical market delineation practices.

3.2.1 Price tests

The aim of statistical tests is to identify the set of products whose prices move together "sufficiently" so that it can be inferred that they belong to a single market. Following the pioneering work of Slade (1986) the most common statistical price-based approaches today are (i) price correlations, (ii) Granger causality tests, (iii) stationarity tests, and (iv) cointegration tests.

[1] The law states that "everything is related to everything else, but near things are more related than distant things" and is credited to Waldo Tobler. It is also known as the First Law of Spatial Econometrics.

[2] E.g. if firms A and B are close to each other in location or have very similar products then the competitive pressure they impose on the prices of any firm C should be similar.

These methods have a number of advantages. First, they are atheoretic (or model-free) in a sense that they don't require a specific economic model, thus they are less prone to model misspecification. Second, they rely almost exclusively on data that is publicly available such as prices, thus there is no need to collect proprietary information on e.g. firm specific costs. And third, up to a varying degree they are relatively simple to explain in front of court and are easy to implement.

On the other hand one immediate disadvantage of these methods is that because of the pairwise nature of the analysis they can't reliably quantify the magnitude of pricing constraints that competing products simultaneously impose on each other. Furthermore, as Coe and Krause (2008) show in their simulation studies the relatively small sample sizes that are available in actual empirical analyses seriously limit the use of the sophisticated econometric techniques. For these latter reasons the advanced techniques proposed in the existing literature are very difficult to implement successfully and as a consequence simple correlation analysis remains the most widely used and accepted price-based (statistical) method in antitrust cases.

3.2.2 Structural approach

As an alternative to price-based tests in 1982 the U.S. Department of Justice introduced the "hypothetical monopolist" or "SSNIP" test as a method for delineating markets. The test begins by defining a narrow market and asking whether a hypothetical monopolist could profitably implement a small but significant and non-transitory increase in price (SSNIP). If sufficient numbers of consumers are likely to switch to alternative products so that the price increase is unprofitable, then the firm or cartel lacks the power to raise price and the relevant market needs to be expanded. The next closest substitute is added and the process is repeated until the point is reached where a hypothetical cartel or monopolist could profitably impose a 5% price increase. The set of products/locations so defined constitutes the relevant market.

The differentiated products oligopoly model of Berry et al. (1995) became the cornerstone of structural approaches in economics and it allows for the exact implementation of the thought experiment prescribed by the merger guidelines' SSNIP test. In the past decades the use of such models has been promoted by academic economists as a theoretically superior approach to merger analysis in differentiated product industries (e.g., Nevo (2000) or Geroski and Griffith (2003)).

However, as Gaynor et al. (2013) point out, while the conceptual exercise prescribed by the hypothetical monopolist test is straightforward, in practical antitrust cases the implementation is not. This is partly because of data limitations, and partly due to the reluctance of courts to rely on complex structural econometric models. Indeed, structural methods need data on important socio-economic demand factors and costs that are either proprietary or very costly to obtain, especially given the short time frame. This is why in practice a much simpler version of the hypothetical monopolist test has been adopted by several competition authorities worldwide and has been the flagship of model based tests ever since. The simplified version uses a specific economic model to evaluate consumer switching behavior but the primitives of the model like demand and cost functions are typically not estimated rigorously.

Of course, if one has reliable estimates of demand at hand, the test can be implemented in a direct way that is consistent with the conceptual exercise. However, to alleviate the burden of increased data requirements there are several informal approaches that try to approximate the demand system with back-of-the-envelope calculations. Unfortunately, it turns out that even if using the correct economic model the demand estimates arising from these simple calculations are too imprecise to generate correct predictions for substitution patterns. Gaynor et al. (2013) show that these informal approaches imply elasticities ranging from 2.4-3.4 times as large as those calculated from a proper structural model. Furthermore, their analysis shows that different structural methods generate results that are largely consistent with each other but inconsistent with such ad hoc or plug-in methods.

Based on all of the considerations above it is clear that in general an ideal method would work from a minimal set of assumptions on the underlying data generating process (like price tests), use only publicly observed routinely collected data (like price tests) yet it is able to assess competitive relationships among stations simultaneously (like structural models).

3.3 The Proposed Method

In this section first I argue that the proposed statistical model can be motivated by a simple and fairly general economic model. Then I gradually introduce the building blocks of the statistical model.

3.3.1 Intuition behind the methodology

I assume that in each period all J (potentially very large) number of operating firms set prices to maximize their profits. Firms are myopic and they don't engage in multi-period strategies. The profit function of firm i in period t depends on its own price and its competitors' prices and period-specific demand conditions. Note that the set of i's competitors, $\mathcal{M}_i$, is only a subset of all operating firms and does not necessarily include all of them. Let us denote this profit function by $\Pi_t(p_i, p_{-i})$, where p_{-i} includes prices of all operating firms except for i. The dependence on t expresses variation over time in common demand and supply conditions. However, it is important to note that this variation is assumed to have no effect on the set of competitors, $\mathcal{M}_i$, which is assumed to be constant over time. Thus, for firms that are not competitors of firm i the partial derivative of the profit function is zero by definition, that is

$$\frac{\partial \Pi_t(p_i, p_{-i})}{\partial p_j} = 0 \Leftrightarrow j \notin \mathcal{M}_i.$$

Throughout the paper I assume that in each period the market is in equilibrium. Practically this means that all prices set by operating firms satisfy a system of first-order conditions. Let us denote the condition for firm i with the scalar valued function F_i, then the relationship among equilibrium prices can be described implicitly as

$$F_i(p_i, p_{-i}) = 0.$$

If we assume for a second that p_i could be expressed explicitly as a function of p_{-i} then we could write this relationship as

$$p_i = \mathcal{F}_i(p_{-i}).$$

For all firms in $\mathcal{M}_i$ the function $\mathcal{F}_i$ has a non-zero derivative and for all firms not in $\mathcal{M}_i$ the derivative is zero. As we are interested only in identifying the set $\mathcal{M}_i$ the task is equivalent to identifying non-zero derivatives of $\mathcal{F}_i$. Unfortunately we don't generally know $\mathcal{F}_i$ without assuming a particular structural form. To alleviate the problem I propose a linear approximation $\tilde{\mathcal{F}}_i$ to $\mathcal{F}_i$, or more formally

$$p_i = \mathcal{F}_i(p_{-i}) \approx \tilde{\mathcal{F}}_i(p_{-i}) = \beta' p_{-i}$$

If prices are strategic complements then the sign of derivatives of $\mathcal{F}_i$ will be approximated well by the signs of the β coefficients of the linear function $\tilde{\mathcal{F}}_i$. However, $\mathcal{F}_i$ does not generally exist and there are cases when there is only an implicit relationship among prices through F_i. The logic of the approximation approach is very similar.

There are several advantages of this approach. First, it is very easy to estimate $\tilde{\mathcal{F}}_i(p_{-i})$ as this is just a linear function of observed prices. Second, by estimating the parameters of $\tilde{\mathcal{F}}_i(p_{-i})$ the derivatives are immediately available, thus one can argue that firms with zero coefficients are not direct competitors to firm i. Third, sparse learning techniques make it easy to estimate $\tilde{\mathcal{F}}_i(p_{-i})$ with a very large number of operating firms, and will identify non-zero coefficients *automatically* without the need to specify a minimum threshold for a non-zero coefficient. This means that the practitioner does not have to worry about what coefficients are "too low" to consider the two products to be on the same market because the algorithm will set those coefficients to zero.[3]

[3]The root of very small but positive elasticities is the unbounded support of taste shocks in the widely used discrete choice models that lead to non-zero predicted market shares for the most inferior (in the characteristics) product.

The drawbacks, of course, include the lack of dynamics like history dependence or forward looking behavior allowed by the assumptions. More sophisticated behaviors (e.g. with non-monotonic relationships among prices) may be ruled out.

3.3.2 Elements of the statistical model

In this section I describe a procedure that is capable of identifying competitors of a chosen firm based on their price co-movements. As set out in the previous section, the formal goal is to identify non-zero coefficients in a linear approximation to the equilibrium price relationships $\mathcal{F}$, given as

$$p_i = \mathcal{F}_i(p_{-i}) \approx p_{-i}\beta_i$$

where $p_i \in \mathbb{R}^T$, $p_{-i} \in \mathbb{R}^{T\times(J-1)}$ and $\beta_i \in \mathbb{R}^{(J-1)}$. For notational simplicity I will drop the i superscript and denote the dependent variable and the matrix of independent variables with y and X, respectively. Then the baseline specification becomes the simple linear least squares model.

$$\min_{\beta} ||y - X\beta||_2^2 \tag{3.1}$$

In the forthcoming sections I gradually introduce the tools developed in the statistical learning literature and show how to apply them to the problem.

Sparse learning

Although appealing at first, the huge problem with the simple OLS model is that it can not be estimated by usual techniques as the number of parameters (firms) is usually much larger than the number of observations (periods). Such settings with a very large number of *potentially* important predictors are called high-dimensional problems. They arise frequently in cases when practitioners cannot generally exclude covariates based on qualitative reasoning. However, having a large number of *potential* predictors does not automatically mean having a large number of *actual* effects. Therefore, in such settings it pays off to *bet on sparsity* and use statistical techniques that lead to sparse solutions.[4] In the statistical learning literature *sparsity* refers to the vector of estimated parameters. There are a number of potential advantages like sparse models can be faster to compute, easier to understand, yield more stable inferences, and are feasible to compute with more regressors then observations.[5]

[4]Hastie et al. (2001) coined the informal "Bet on Sparsity" principle. Sparse learning methods assume that the truth is sparse in some basis. If the assumption holds true then the parameters can be efficiently estimated. If the assumption does not hold then no method will be able to recover the underlying model without a large amount of data per parameter.

[5]See Hastie et al. (2016) for a thorough and formal review of sparse learning methods.

10.14754/CEU.2018.11

To relate this to the market delineation context recall that the task is to identify the set of competitors to a given firm. It is reasonable to expect the algorithm to select a narrower subset of firms that sell a similar product. The difficult part is to avoid imposing a priori such restrictions on the methodology and to let the data "speak" for themselves.[6]

l_1 and l_2 regularization

Sparsity of the estimated parameter vectors are achieved by regularization. The cornerstone of sparse learning is l_1-norm regularization introduced in Tibshirani (1996), which became popular under the name Least Absolute Shrinkage and Selection Operator, or just simply *lasso*. In the lasso model the usual least squares objective function is augmented with an additional penalty term to get the objective function

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda_1 ||\beta||_1 \tag{3.2}$$

where the penalty parameter λ_1 governs the weight of the l_1 penalty in the objective function.It turns out that this problem has very appealing properties. Most importantly it yields sparse solutions for β setting many of it's entries explicitly to zero and shrinking the non-zero elements to some extent towards zero when compared to the OLS estimate.[7] This is important in discovering true signals in high dimensional settings and makes out of sample predictions more robust to noise in the data. In the market delineation context we expect the lasso to select a small subset of most influential firms (i.e. assign non-zero coefficients to them) that affect the prices of our chosen firm.

An important drawback of the lasso solution is that it will typically include only one from a set of equally important but highly correlated regressors. This is a serious problem if we are interested in groups of covariates that are thought to have similar predictive powers. In the market delineation context this may occur with many close substitutes whose prices are highly correlated. Thus, the simple lasso has a risk of throwing out a close competitor because it's price is very similar to the price of another close competitor. One solution to that would be to set up a *group lasso* estimation, that allows groups of coefficients to be together zero or non-zero. (Yuan and Lin (2006)) However, this approach would require the analyst to explicitly define the groups (e.g. the set of close competitors), which contradicts the purpose of the whole exercise. Therefore I use l_2-norm regularization (also called the *ridge* penalty, Hoerl and Kennard (1970)) in addition to the lasso penalty.

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda_1 ||\beta||_1 + \lambda_2 ||\beta||_2^2 \tag{3.3}$$

[6] E.g. spatial weighting schemes impose declining structure on the coefficients.

[7] When the OLS estimate is applicable, of course.

This combined specification is called the *elastic net*, following Zou and Hastie (2005), which retains most of the lasso sparsity without the risk of throwing out one of two highly correlated (even identical) but important regressors. This is a consequence of the ridge penalty which is capable of keeping even identical regressors in the model and spreading their effect evenly among them.

Thus, we can can expect the elastic net to select a small subset of firms including all the influential ones potentially with highly correlated prices among themselves.

Spatial regularization

The elastic net is already a big step towards an applicable model but unfortunately it also introduces a new problem. In markets characterized by sellers with multiple products the prices of all products of a given seller might be highly correlated due to common cost or other supply related shocks. As a consequence, in the market delineation context we might expect the elastic net to select products that are not competitors to our product of interest but only produced by the same firms that produce the truly competing products.

To illustrate this, consider the retail gasoline market with typically a few number of large chains with many stations each. Suppose we are interested in identifying the competitors of a given station. It is reasonable to expect that the competitors are stations in some neighborhood of our chosen station. If we run the elastic net on historical prices of all stations we might end up with a set of identified competitors and many additional stations that belong to the chains of competitors.

In this section I will argue that utilizing the spatial relationships among regressors in some underlying space might help to identify the true effects we are interested in.

Fusion of parameters

Consider an estimation problem where we are interested in the effects of a potentially large set of regressors $\{x_j \in \mathbb{R}^T\}_{j=1}^J$ on an outcome $y \in \mathbb{R}^T$. Suppose that the regressors can be ordered according to some rule, e.g. time. From domain expertise we happen to know that regressors close to each other (e.g. from two consecutive periods) tend to have similar effects. This knowledge could be built into the loss function by adding a term that penalizes differences in estimated effects corresponding to neighboring regressors as in (3.4).

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda \sum_{j=2}^{J} |\beta_j - \beta_{j-1}| \tag{3.4}$$

It is important to note two things:

- The ordering is completely independent of the actual *values* of the regressors.
- The ordering is defined *exclusively* on the regressors and it does not specify the regressors' relationship to the outcome variable.

As a result the fusion approach yields an estimated vector of coefficients that is sparse in the *consecutive* differences. In practice that means that by adding a fusion penalty it restricts the set of values that the elements of the coefficient vector may take and the estimated coefficients will be a piecewise constant function in the order. In some sense fusing the coefficients is a spatial homogenization method and it was originally developed in the signal denoising literature for trend filtering (Land and Friedman (1996)).

Fortunately, it is possible to combine the fusion penalty with other methods to get an estimated coefficient vector that retains sparsity on top of being fused. This is exactly the basis of the *fused lasso* proposed by Tibshirani et al. (2005) as in (3.5).

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda_1 ||\beta||_1 + \lambda_2 \sum_{j=2}^{J} |\beta_j - \beta_{j-1}| \tag{3.5}$$

Spatial structure

Note, that the fused lasso is not applicable when regressors can not be ordered. However, it should be recognized that it is not explicitly the ordering that is useful for us but the difference in *neighboring* coefficients. In this sense if it is possible to define neighbors of a coefficient in some arbitrary underlying (possibly multi-dimensional) space then the neighborship structure can be represented by an undirected graph. Indeed, if $G = (V, E)$ is a graph with set of nodes V end set of edges E then $\sum_{(i,j)\in E} |\beta_i - \beta_j|$ is the generalized definition of the fusion penalty and the model

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda_1 ||\beta||_1 + \lambda_2 \sum_{(i,j)\in E} |\beta_i - \beta_j| \tag{3.6}$$

is called the *generalized fused lasso*. The generalized fusion penalty has it's roots in image denoising where the method is called total variation denoising (Rudin et al. (1992)) and there is an increasing number of applications to spatial trend filtering in the field of neuroscience (Watanabe et al. (2014)) or more recently in the social sciences (Wang et al. (2014)).

In the market delineation context the neighborship structure becomes interpretable if the marketed products can be represented in some space with an appropriate distance metric. This could either be actual geographic space (e.g. in the case of gas stations) or the product characteristics' space (e.g. in the case of cars or cellphones etc.).

However, it seems that it is up to the researcher what products are considered neighbors. This may be unwanted when automation is essential due to time limitations. To overcome this problem and, again, to let the data speak for themselves I propose to perform a spatial Dealunay triangulation procedure among the products' locations to get an undirected neighborhood graph. For the workings of the triangulation procedure I refer the reader to Section 3.5 where it is demonstrated how to form a neighborhood graph of actual gas stations' locations.

3.3.3 Full model

It is useful to take a step back and remind ourselves what do we expect from our statistical model. There are three key points I require from the method:

- Sparse coefficient vector with non-zero entries for competitors
- Accommodate highly correlated regressors
- Spatial homogeneity in the coefficient vector

The last point simply means that if two products are close to each other in geographical or product characteristics' space then their competitive relationship to any other product in question should be similar.

After adding all three types of penalties to the least-squares loss to achieve the desired outcomes the baseline model to be estimated becomes

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda_1||\beta||_1 + \lambda_2||\beta||_2^2 + \lambda_3||M\beta||_1, \tag{3.7}$$

where $M \in \mathbb{R}^{|E| \times |V|}$ is the incidence matrix corresponding to graph G.

It is important to emphasize that the spatial structure that is put on the coefficients (i) comes entirely from the data (i.e. firm locations or product characteristics) and (ii) does not specify the spatial relationship between the outcome variable and any of the regressors. When setting up the estimation problem in this way nothing requires the model to select influential regressors *from the neighborhood* of the outcome variable therefore I can avoid imposing a priori the desired structure on the estimated coefficients.

3.3.4 Tuning Parameters

It is possible to rewrite the model in (3.7) as

$$\min_{\beta} ||y - X\beta||_2^2 + \alpha \left\{ \delta \left(\gamma||\beta||_1 + (1-\gamma)||\beta||_2^2 \right) + (1-\delta)||M\beta||_1 \right\}, \tag{3.8}$$

to get a more interpretable form. It is now clear that α alone controls the weight of the whole penalty term, δ specifies the proportions between the elastic net and the fusion term and γ splits the weight on the elastic net term between the lasso and ridge penalties.

Normally, such hyper-parameters are tuned by (nested) model selection techniques based on information criteria or out-of-sample prediction accuracy, such as k-fold cross-validation. However, with the lack of abundant observations I choose not to follow this approach and interpret my results as a lower bound to the optimal performance of the method. I set $\gamma = 0.95$ as this is an amount of l_2 regularization that is usually considered a good choice when model selection is not available.[8] Unfortunately there is no consensus in the literature on how to select the best amount of spatial regularization, so I opted for a conservative method and set $\delta = 0.9$ for a moderate amount of fusion in the penalty term.

To select α I use the minimal amount of regularization that is required to achieve a fixed degrees of freedom. Tibshirani and Taylor (2011) showed that the degrees of freedom in fused lasso-type problems is best approximated by the number of connected components in graph G (the graph corresponding to incidence matrix M) with edges among nodes with differing coefficients deleted. For easy interpretation I set the required degrees of freedom to 10, and pick the smallest α that matches the criterion.[9]

3.3.5 Translating coefficients to competitors

To translate the estimated coefficients to a set of competitors I propose the following procedure. First, as noted earlier, I am not interested in the size of coefficients per se. To identify competitors of a given product I take the vector of estimated coefficients ($\hat{\beta}$) from the full model in (3.8) and select all nodes that have a non-zero corresponding coefficient. Then I look for connected components in the subgraph of G corresponding to the selected nodes. The whole procedure can be summarized in a series of Figures 3.1a – 3.1f.

Step 1. Perform the triangulation procedure to arrive at a neighborship graph of *all* locations. Pick a central product whose set of competitor products are going to be estimated. Figure 3.1a shows one such graph with the red dot indicating the location of the central product and the black dots indicating the potential competitors.

[8] For more on this please see Hastie et al. (2016).

[9] This may seem too simplistic but this is actually the approach used in state-of-the-art applications as in Wang et al. (2014) or Watanabe et al. (2014). The problem is that fusion is a very restrictive regularization technique and it typically does not increase prediction accuracy. It is best used to constrain the coefficients to reflect some principle.

Step 2. Remove the central product and its adjacent edges from the graph. Assign a coefficient to each node. Figure 3.1b shows such a graph with three example nodes with their coefficients. Run the penalized regression in (3.8) with the central node's prices on the left hand side and using the incidence matrix of the graph in the penalty term. Note how the absolute differences of connected nodes' coefficients enter the loss function. The procedure is completely blind regarding the location of the central product and the coefficients.

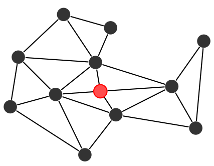
Step 3a. Figure 3.1c shows a typical result of the penalized regression. The black nodes indicate products with non-zero estimated coefficients. Coefficients of gray nodes where set to zero by the l_1 penalty term in the loss function. Drop all edges between the zero and non-zero components of the graph and count the connected components with non-zero nodes. In this case there is only one such connected component and it has common nodes with the central products neighborhood. This set of nodes is the identified set of competing products in this example.

Step 3b. Figure 3.1d shows an alternative potential outcome when the non-zero connected component is not unique. In such cases I apply the first law of geography as a weak guiding principle and I pick the component that has a common node with the neighborhood of the central product. If there are several such components then I pick one based on the actual distance from the central product.

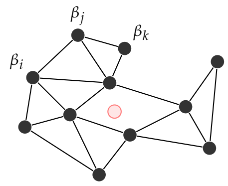
Step 3c. Figure 3.1e shows the situation when there is no such non-zero component that intersects with the central product's neighborhood. In such cases I follow the spatial econometrics literature and I impose an additional penalty term to the objective function in (3.8) to achieve a set of coefficients that is non-increasing with the distance to the central product. E.g. Figure 3.1f shows to model re-estimated with entries such as $|\beta_j - \beta_i|$ and $|\beta_j - \beta_k|$ replaced with $\theta|\beta_j - \beta_i| + (1-\theta)(\beta_i - \beta_j)$ and $\theta|\beta_j - \beta_k| + (1-\theta)(\beta_k - \beta_j)$ because the node with β_j is closer to the central product than nodes with β_k and β_i. Essentially, I replace $||M\beta||_1$ in the penalty term with a convex combination of $||M\beta||_1$ and the sum of $M\beta$ and arrange the values in each row of M so that the value 1 corresponds to the node of the edge that is farther from the outcome.[10]

[10]It is important to note that unlike using a standard spatial decay function I am not imposing a shape on the coefficients as a hard constraint. My method fully allows for two equidistant coefficients to have different contributions to the penalty term.

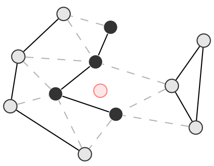
10.14754/CEU.2018.11



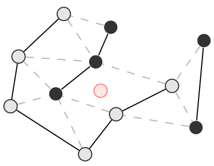
(a) Original graph of locations after the triangulation procedure



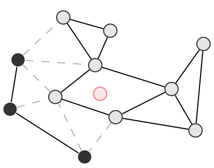
(b) The graph of locations after central product was removed



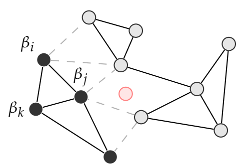
(c) Estimation results with one non-zero connected component



(d) Estimation results with two non-zero connected components



(e) The estimated non-zero connected component doesn't contain any products from the central product's neighborhood



(f) Penalizing spatial non-monotonicity yields a non-zero component with an element from the central product's neighborhood

Figure 3.1: Illustration of how estimated model coefficients are translated to competitors

3.3.6 Discussion

Asymmetric relationships

Up to this point the description of the procedure focused mainly on identifying competitors of one given product.[11] However, one might be interested in re-running the procedure for multiple or all products to learn about the pattern of competing relationships among them. In this case it is important to distinguish between the cases when product i was identified as a competitor of j or the other way around. Note that the relationship is not restricted to be symmetric.[12] Sometimes this can be a very useful property e.g. in a case where one product is a market leader while the other is a fringe product it is reasonable if the market leader's price heavily affects the fringe product's price but not the other way around.

Market areas

By allowing for directional relationships the resulting network of competition can be described by a directed graph. It may happen that this network consists of several disconnected or loosely connected components which might be a signal of independent market areas. This can be important information for competition economists or market analysts alike.

The reflection problem

Common shocks to supply and demand is the central challenge for price-correlation-based market definition. This is related to the classic reflection problem: do prices of neighboring firms co-move because firms respond to each others pricing strategies or because firms react to the some spatially correlated shocks in cost or demand?[13] To remain minimalistic in data requirements the current model relies only on price data and as such has only very limited power in separating the effects of common shocks on co-movements. This is a weakness of the approach that one has to bear in mind when interpreting the results. Using additional spatially varying controls could alleviate the problem (at least partly) but it would contradict the point of the exercise and could be entirely infeasible in other settings.

In the next sections I will demonstrate the workings of the model on both simulated and actual data.

[11] Recall that a competitor is defined as another firm whose prices put a constraint on the prices of the actual firm in question.

[12] In my experience the relationships are symmetric in the majority of the cases but it should depend on the exact application.

[13] I would like to thank Sergey Lychagin for pointing this out.

10.14754/CEU.2018.11

3.4 Simulations

Here I demonstrate the working of the method by applying it to a simulated data set. The focus will be on support recovery, that is how well can the method recover the set of regressors with non-zero effects.

The simulated data mimics some important moments of the gasoline data set that will be used in the next section. As in the previous section $X \in \mathbb{R}^{T\times J}$ will denote the matrix of covariates which are drawn from a J-dimensional multivariate normal distribution and $y \in \mathbb{R}^T$ will represent the T-vector of outcomes computed as

$$y = X\beta + \varepsilon,$$

where β represents the vector of J true effects assigned to the covariates and ϵ is the vector of T independent normally distributed disturbances.

The key features of the simulated data set:

- Large number of covariates ($J = 1000$) compared to the number of observations ($T = 100$).
- Highly correlated covariates with pairwise correlations ranging from 0.85 to 0.95.[14]
- Sparse effect vector, meaning that the fraction of non-zero elements in β is small (10%).
- Each covariate and corresponding effect are assigned a location on the $[0, 1]$ interval. Locations are increasing from the first to the last column of X.
- The location of the outcome variable is 0.
- The values in β are arranged in decreasing order which represents that effects are decreasing with distance from the outcome variable.

The two panels of Figure 3.2 help to understand the details.

The left panel plots the β coefficients as a function of their location and shows that only the covariates located in $[0, 0.1]$ (also separated by a dashed vertical line) have non-zero effects and the effects are linearly decreasing with the distance from 0. In the market delineation context these covariates represent the set of competitors. Essentially, the rest of the regressors with zero effects will introduce the noise in the estimation procedure because of the strong multicollinearity. As a consequence, the outcome variable will also be highly correlated with the rest of the covariates through the 100 covariates

[14] The covariance matrix for the multivariate normal draws was simulated using the so-called random factors method.

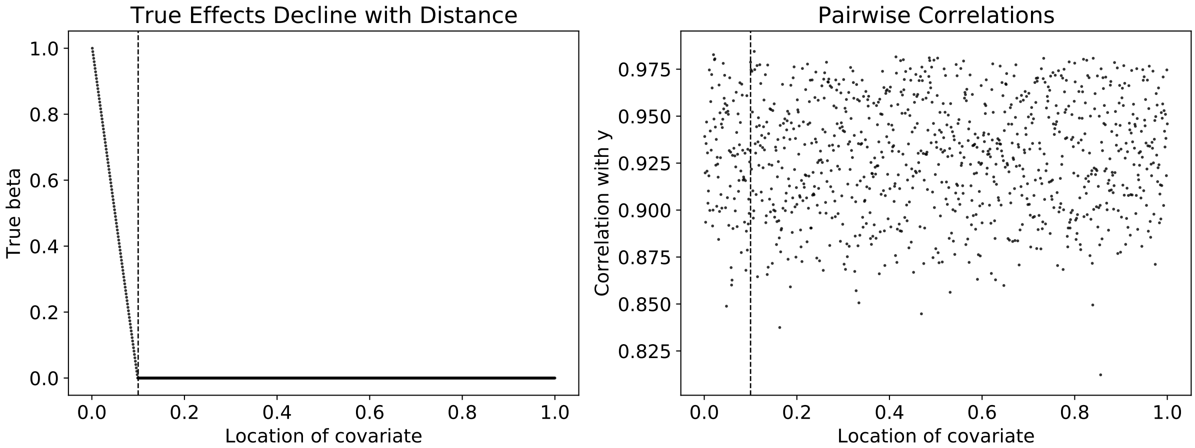


Figure 3.2: Simulated data set

with positive effects. The right panel of Figure 3.2 shows these pairwise correlations between the outcome variable and the covariates.

It can be seen immediately that there are no clear clustering patterns in the pairwise correlations so there is a need for a multivariate method that also exploits the underlying spatial structure. As in the previous section the spatial structure will be represented by the neighborship graph of the covariates.

I set the standard deviation of the disturbances to be 10% of the standard deviation of $X\beta$.[15]

3.4.1 Results

In addition to the baseline specification described in the previous section I performed experiments with several setups for the data generating process. All specifications were estimated 50 times with the default values for the tuning parameters. From each run I take the location of the first zero covariate as my estimate for the market boundary.

Figure 3.3 shows a typical result from one single run. There are three things to note on the figure:

- Because of spatial regularization estimates form a piecewise constant function.
- The lasso penalty pushes most of the coefficients explicitly to zero.
- Coefficients drop to zero around the market boundary.

[15]Increasing the variance of the disturbances would lower the correlation among the outcome variable and the regressors. I want to avoid that to keep the highly correlated spurious regressors.

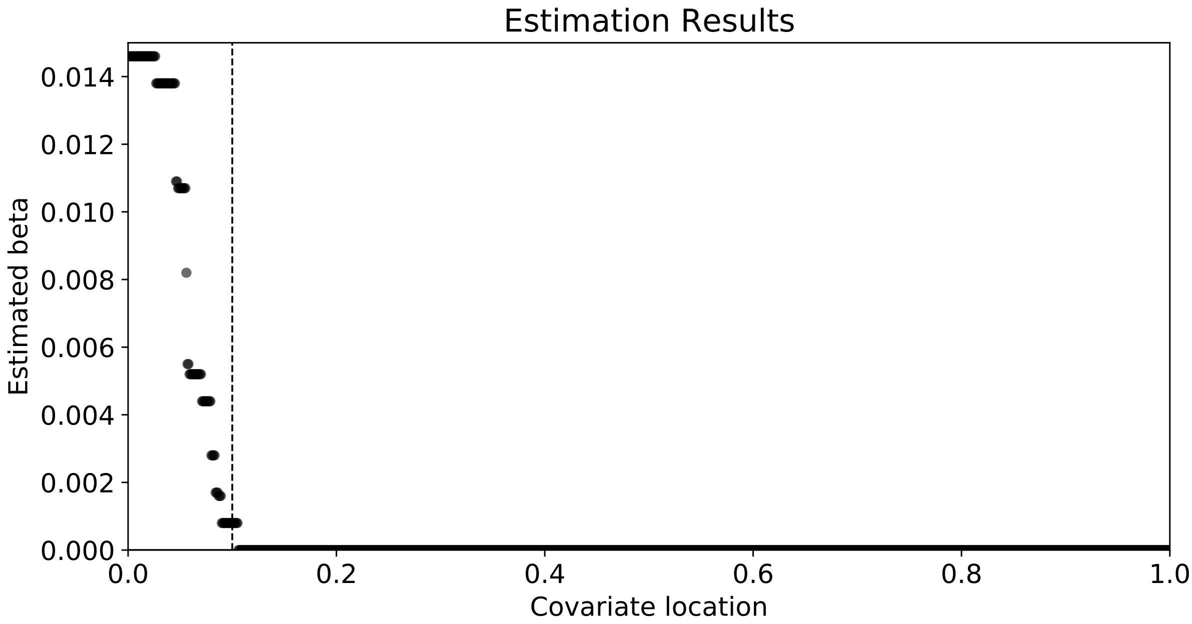


Figure 3.3: An example of a typical $\hat{\beta}$ with $R^2 = 0.99$.

Table 3.1: Simulation results – Estimated market boundaries

	mean			median		
boundary	0.10	0.25	0.50	0.10	0.25	0.50
T						
100	0.0982	0.2072	0.2362	0.0960	0.2150	0.2745
250	0.1064	0.2418	0.4034	0.0990	0.2390	0.4225
500	0.1031	0.2474	0.4437	0.0990	0.2410	0.4490

Table 3.1 shows both averages and medians across replications for each of the nine separate scenarios. In general it can be concluded, that increasing the sample size improves precision and it is possible to get reasonable estimates for the market boundary even with the relatively low sample size of 100 observations for 1000 variables.

3.5 Application

To illustrate how the method could be used in an actual analysis I will apply it to a data set on weekly retail gasoline prices in Hungary. I will identify the set of competitors of stations involved in a hypothetical merger case using the method described in the previous sections.

The source of the data set is an online price comparison site *www.holtankoljak.hu* and it contains prices for 103 weeks of about 1200 stations from October 2006 to December

10.14754/CEU.2018.11

2008. The market is characterized by a few number of chains with several stations each and a number of independent vendors.[16] The three largest chains are MOL, SHELL and OMV with market shares in terms of the number of stations of 26.7%, 14.1% and 12.9%, respectively.

In the application below I will identify the set of competitors of all SHELL and OMV stations where a hypothetical merger between the two chains would raise significant antitrust concern because of the change of concentration in their local markets. It is important to note that the analysis is only meant to serve as a screening stage before further rigorous merger scrutiny and it's main contribution is to highlight cases of potential harm.

3.5.1 Spatial structure

I use the geographical location of stations to form a neighborhood graph for spatial regularization. Station addresses were geocoded on a city level to the center of the city. To assign a unique location to all stations I added a random shock to all locations. Shocks were weighted by the areas of corresponding cities to get realistic distance patterns. Unique locations are needed because putting multiple stations to one location would imply very unrealistic neighborship relations. There is an error introduced to the assumed underlying spatial structure by this random coordinate assignment but this won't be an issue in an actual application where locations are fully observed. One possibility would be to replicate the full procedure for a large number of randomly drawn locations or to manually geocode the stations on the rooftop level.

To get a neighborhood network I performed a spatial Delaunay triangulation on the unique station locations. In very broad terms this procedure minimizes the number of sharp angles in the triangles. The resulting network includes some disproportionately long edges around the graph boundary which would tie the coefficients of very distant stations so I drop the longest 2%. Figure 3.4 shows the network of all stations to be used in the estimation. As previously, the graph is represented by the incidence matrix M, where each row corresponds to an edge in the graph with 1 and -1 at the corresponding nodes and zeros elsewhere.

The triangulation is needed to come up with an underlying graph structure, which in turn can be used to penalize differences in coefficients that share an edge in the graph. In principle I could use a fully connected network of stations with edges weighted by some inverse distance metric to penalize the differences in coefficients but that would result in a huge increase in computational time and also would restrict the set of possible patterns that can arise.

[16] As this data set has been thoroughly analyzed in Farkas et al. (2009) and in Bekes et al. (2011) I refer the interested reader to their papers for a deeper understanding of the Hungarian retail gasoline market.

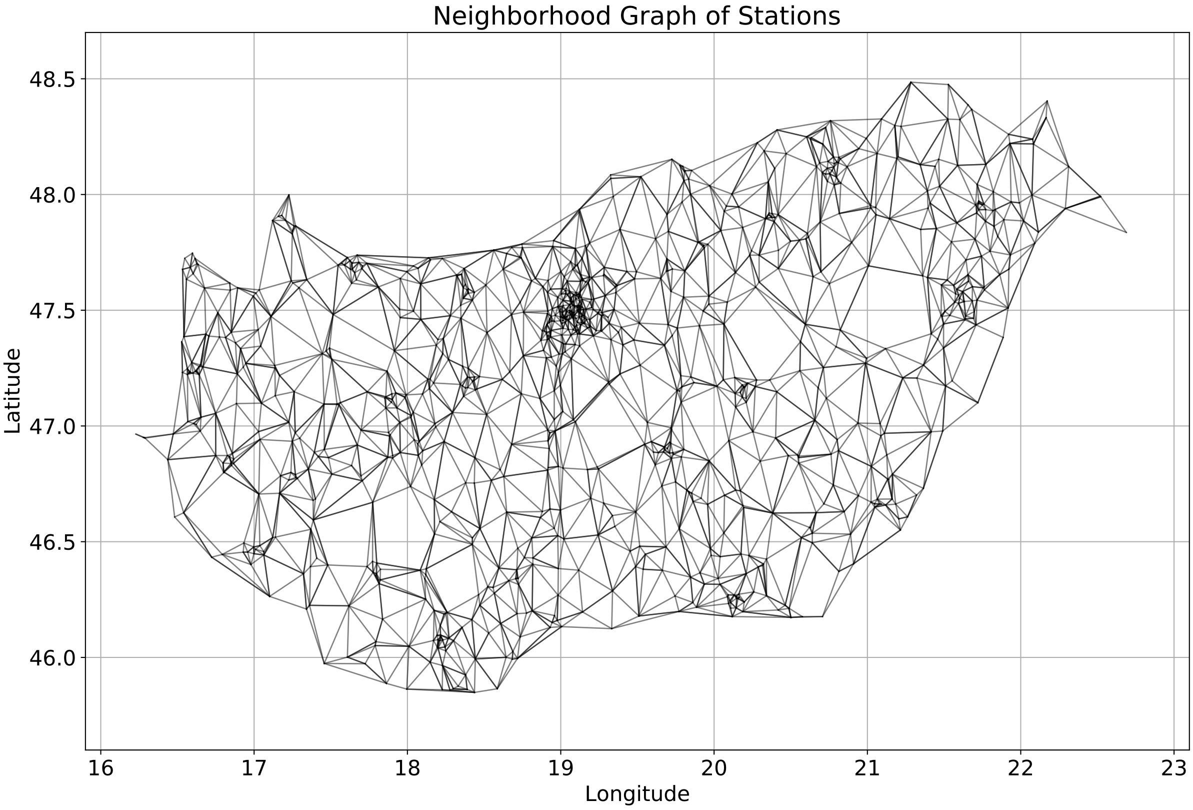


Figure 3.4: Spatial Structure of Gasoline Stations

3.5.2 Prices

Prices varied a lot over the analyzed period. Most of the trend can be attributed to common costs which lead to high pairwise correlations among prices. The left panel of Figure 3.5 depicts the evolution of prices over time while the right panel shows the histogram of all elements in the pairwise correlation matrix of prices. One can see, that in such a high-dimensional setting it would be very hard to rely on simple bilateral correlation analysis to identify competitors. First differencing might be a good solution to reduce correlations, especially in the case of daily prices. Other detrending methods could work also but it is the very purpose of this exercise to show how powerful spatial regularization is even when applied to simple price levels without any pre-processing.

The only pre-processing I did was imputation of missing values which process is described in the next section below. The reader not interested in the details may skip this section.

Imputation of missing prices with spectral regularization

Except for very fortunate cases missing observations are prevalent in empirical studies. Under some conditions it is possible to exclude observations that are only partially

10.14754/CEU.2018.11

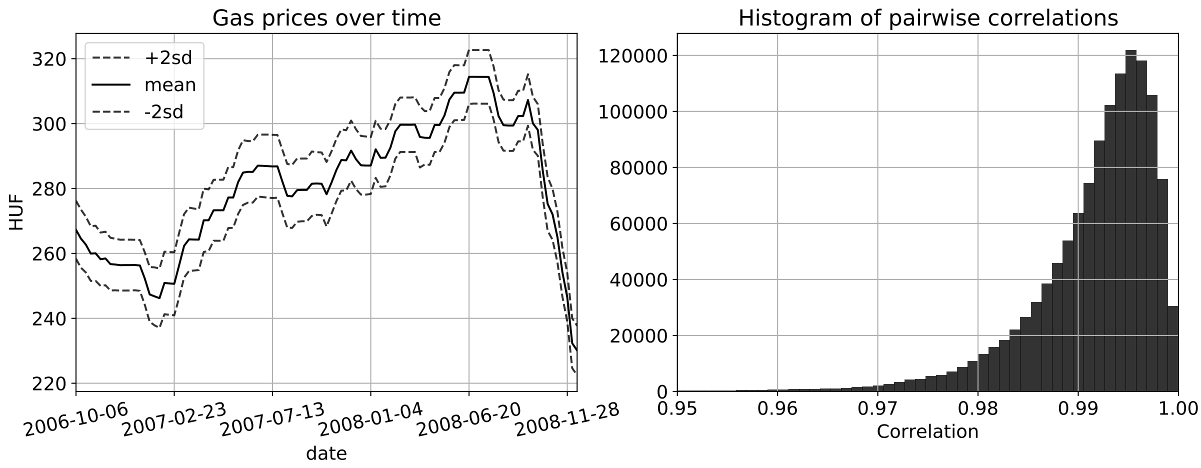


Figure 3.5: Trend and correlation of prices

-

observed without sacrificing unbiasedness and statistical power.[17] However, in the high-dimensional setting it is common that the cardinality of the set of observations that have valid values for all variables of interest is low or even zero. In general terms the usual practice to get around this problem is to either exclude some problematic variables (e.g. those with too many missing values) or to use imputation where it is believed not to introduce excessive bias. One should be very cautious when dropping variables entirely to avoid accidental exclusion of potentially important effects. Certainly, in market delineation exercises there are some cases when a set of regressors corresponding to products or locations could be excluded without loss of generality purely based on qualitative reasoning. E.g. in a densely populated country one can be confident in the independence of two distant locations. In the empirical application below I will demonstrate the working of the method using prices of gas stations where it would be natural to follow this logic to exclude distant competitors and impute only the closest stations but for the sake of illustration I decided to impute all missing values. This will certainly lessen the signal to noise ratio so I will interpret the results accordingly as a worst case.

There are several methods that can be used to impute missing values to a data set. The econometrics literature prefers model-based imputation techniques to ad hoc approaches like mean imputation or hot deck imputation because of their superior performance. However, a sophisticated method with good theoretical properties like a Markov Chain Monte Carlo-type multiple imputation procedure may pose a significant computational challenge. For both computational and prediction accuracy reasons I opted for an approach based on low-rank matrix factorization that is non-

[17]Here I focus only on the case where the mechanism which determines which values are missing is ignorable. In cases where the selection mechanism matters one should consider estimating the appropriate model.

10.14754/CEU.2018.11

traditional in economics but is a well-developed and highly adopted technique in the statistics and machine learning literature with superior performance.

Because of its rare occurrence in economics research here I provide a brief overview of Rahul Mazumder and Tibshirani (2010) who introduced the state-of-the-art SOFT-IMPUTE algorithm for learning missing values in large-scale problems. Without any restrictions on the degrees of freedom in the completed matrix the matrix completion problem is underdetermined since the missing entries could be assigned arbitrary values. Thus matrix completion often seeks to find the lowest rank matrix. Suppose we want to learn the missing elements of a matrix $X_{m\times n}$ where $\Omega \subset \{1,\ldots,m\} \times \{1,\ldots,n\}$ denotes the indices of the observed elements. Then we could consider to solve for an approximation $Z \in \mathbb{R}^{m\times n}$ in the following problem:

$$\begin{aligned} \text{minimize} \quad & \text{rank}(Z) \\ \text{subject to} \quad & \sum_{(i,j)\in\Omega} \left(X_{ij} - Z_{ij}\right)^2 \leq \delta, \end{aligned} \tag{3.9}$$

where $\delta \geq 0$ is a regularization parameter controlling the tolerance in the training error. Unfortunately this problem is in general NP-hard, but there are tractable algorithms that achieve exact reconstruction with high probability. The basis of SOFT-IMPUTE is the idea to solve the convex relaxation to (3.9) by using the nuclear norm of Z instead of the rank constraint.[18] Rewriting the problem in Lagrange form yields the following semidefinite programming problem

$$\underset{Z}{\text{minimize}} \quad \frac{1}{2} \sum_{(i,j)\in\Omega} \left(X_{ij} - Z_{ij}\right)^2 + \lambda ||Z||_* \tag{3.10}$$

which can be solved efficiently by SOFT-IMPUTE.

The algorithm performs low-rank singular value decompositions in every iteration. It is advised to remove row and/or column means from a matrix before performing a low-rank SVD or running matrix completion algorithms. Likewise we may also wish to standardize the rows and or columns to have unit variance. This makes the algorithm more stable and improves prediction accuracy. However, this double normalization is not straightforward, especially with missing matrix entries. In this paper I use the method developed by Hastie et al. (2014) that is fast, memory efficient and suitable for very large and sparse matrices while allows for missing data.

Figure 3.6 shows the histogram of sample sizes that are available for each gas station. The overall missing rate is around 20%, and around 80% of stations have a missing rate of less than 20%. Prices can be missing for several non-random data-recording reasons. E.g. if a station started or closed operation in the middle of the period or it was simply

[18]The nuclear norm of a matrix ($||\cdot||_*$) is the sum of its singular values.

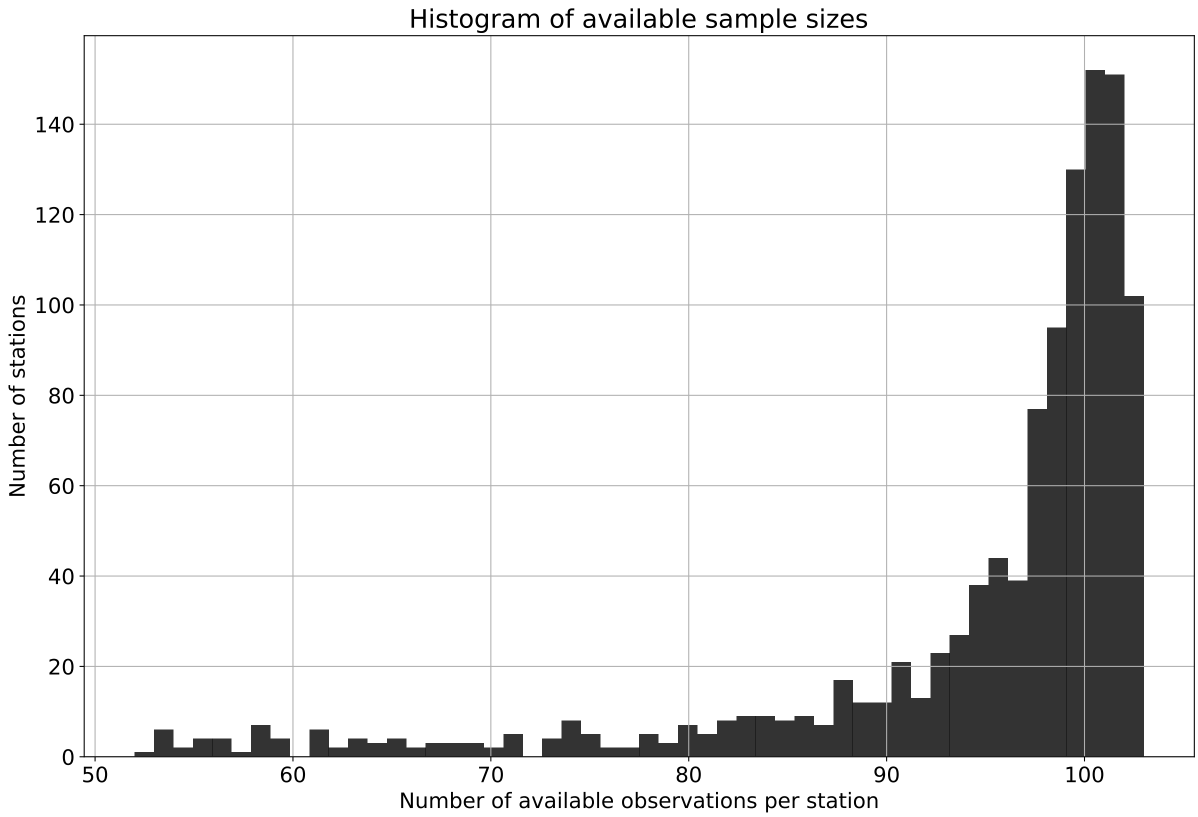


Figure 3.6: Distribution of available observations

not sampled by the price comparison site. I choose to impute prices of stations that have at least 50% of the observations which leaves me 1110 stations.

To illustrate the method I take a balanced subsample of the data (107 stations) and randomly delete 20% of the observations. Using SOFT-IMPUTE I was able to impute observations with a root mean squared error of 1.008 which is less than 0.35% of the average price. For comparison, the best performing k-nearest neighbors method with 7 neighbors resulted in a RMSE of 1.1959 which is about 20% larger.

3.5.3 Estimation

I estimate the full model in (3.8) for all stations separately to identify the sets of competitors for each station. That means that each of the 1110 stations' time series of length 103 will serve once as an outcome (y) and 1109 times as a regressor (as a column of X). For each estimation run I remove the node corresponding to the outcome station from the graph. In this way the estimation procedure remains "blind" to the spatial relation between the outcome and the regressors and will require only that neighboring stations have similar coefficients. I use the default parameters (δ, γ) for the weights inside the penalty term and tune α for each estimation run to have 12 degrees of freedom.

Table 3.2: Summary statistics of estimation results

statistic percentile	number of competitors	D_{min} (km)	D_{median} (km)	D_{mean} (km)	D_{max} (km)
10	1	0.48	1.35	1.39	1.74
25	2	1.12	2.85	3.08	4.18
50	6	2.75	6.28	6.74	10.10
75	20	5.51	18.75	19.61	35.15
90	84	10.41	50.45	51.84	105.89

3.5.4 Results

First it is useful to have a look at the identified sets of competitors. For each station I have counted the number of competitors, the distance to the closest, to the median, to the mean and to the most distant competitor. Table 3.2 summarizes the distributions of the five statistics in the population. There are two important things to note. First of all, the estimates seem to make sense. It is well known that the gasoline market exhibits strong local competition (see e.g. Pinkse et al. (2002)). This fact is supported by the result that half of the stations have at most 6 competitors and all competitors within 10 kilometers (6.25 miles). Also, by looking at Figure 3.7 one can confirm that stations in areas with higher station density (e.g. stations within certain cities) have more competitors and are closer to their competitors. The second immediate observation is that, unfortunately, the limits of the method and the data manifest themselves in a number of extreme cases with an unreasonably large number of identified competitors for the top ca. 10-15% of stations (in terms of the number of competitors). There are a number of potential reasons for these performance issues like e.g. insufficient frequency of observations, simulated locations or affected stations just work differently and the statistical model is just not able to capture their behavior.[19] Table 3.3 repeats the statistics from Table 3.2 with the top 10% of stations (in terms of the number of competitors) excluded. In this restricted sample the ratio of symmetric competing relationships is 84.2%.[20]

[19]One explanation could be that many of the problematic stations are very close to the border and in Bekes et al. (2011) it was shown that pricing at stations along the border works differently than at other locations.

[20]For visualization purposes on Figure 3.7 I only show symmetric relationships of stations after excluding the problematic top 10%.

Table 3.3: Summary statistics of estimation results – without extreme values

statistic percentile	number of competitors	D_{min} (km)	D_{median} (km)	D_{mean} (km)	D_{max} (km)
10	1	0.45	1.11	1.16	1.40
25	1	1.05	2.43	2.56	3.37
50	4	2.47	4.58	4.84	7.15
75	9	4.58	9.35	10.34	16.14
90	17	7.78	16.88	16.56	31.52

3.5.5 Hypothetical merger

In this section I will analyze the case of a hypothetical merger between the second and third largest competitors on the market. The question is whether the merged parties will have sufficient market power to raise prices. To address that I use my estimates and compute the change in market concentration at all stations other than the merging parties. Figure 3.8 shows stations with an increase in HHI of at least 15%. The resulting graph has several independent connected components that correspond to markets that could be potentially monopolized by the merger. After this preliminary screening step the competition authority could go and analyze thoroughly the competitive effects specifically at those markets even independently.

3.6 Conclusion

In this paper I present an approach for identifying competing firms using modern statistical learning techniques. This method makes it possible for researchers and practitioners to identify sets of competing firms based solely on publicly observed price data even in high-dimensional settings. As the application shows, the results then can be readily used by antitrust authorities to detect areas of potential harm.

In the future it would be interesting to see more applications to real data. One potentially interesting exercise would be to connect the estimates to local socio-demographic variables and see what explains competitive relationships. Although the simulation studies show unbiasdness it would be important to focus also on the theoretical properties of the estimation method.

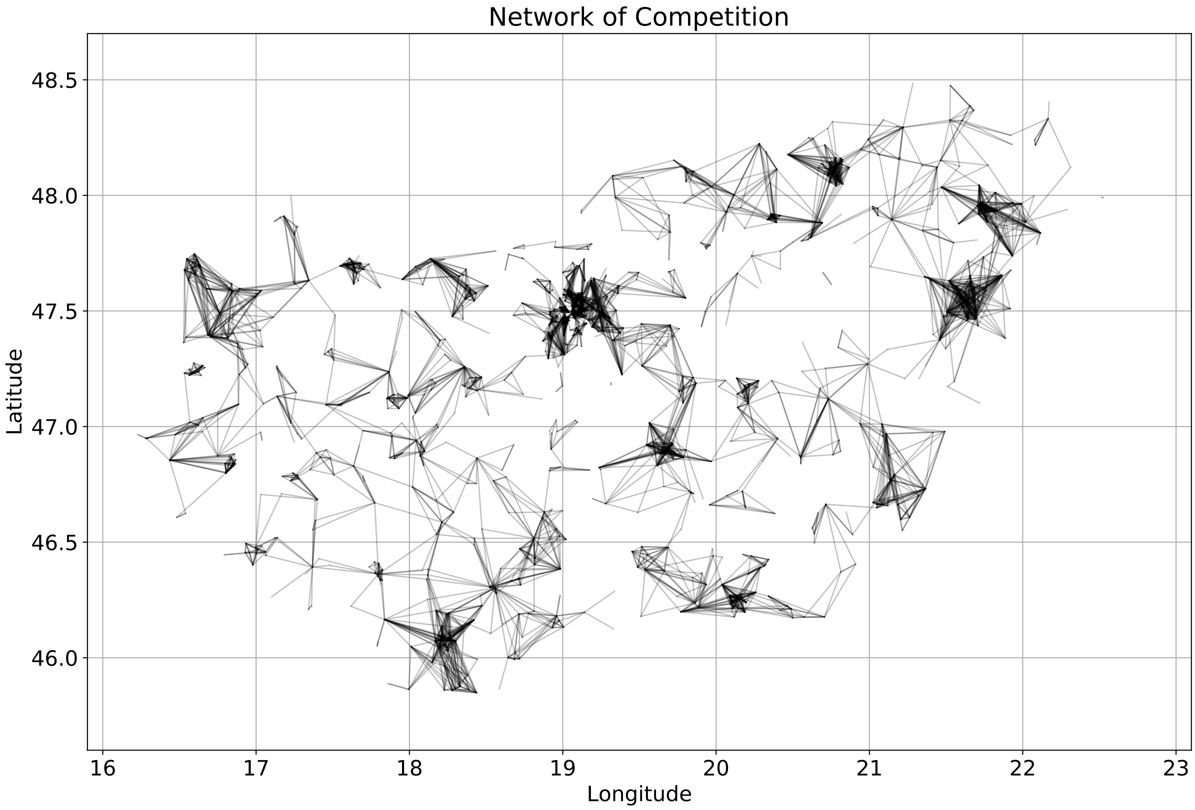


Figure 3.7: Network of Competition

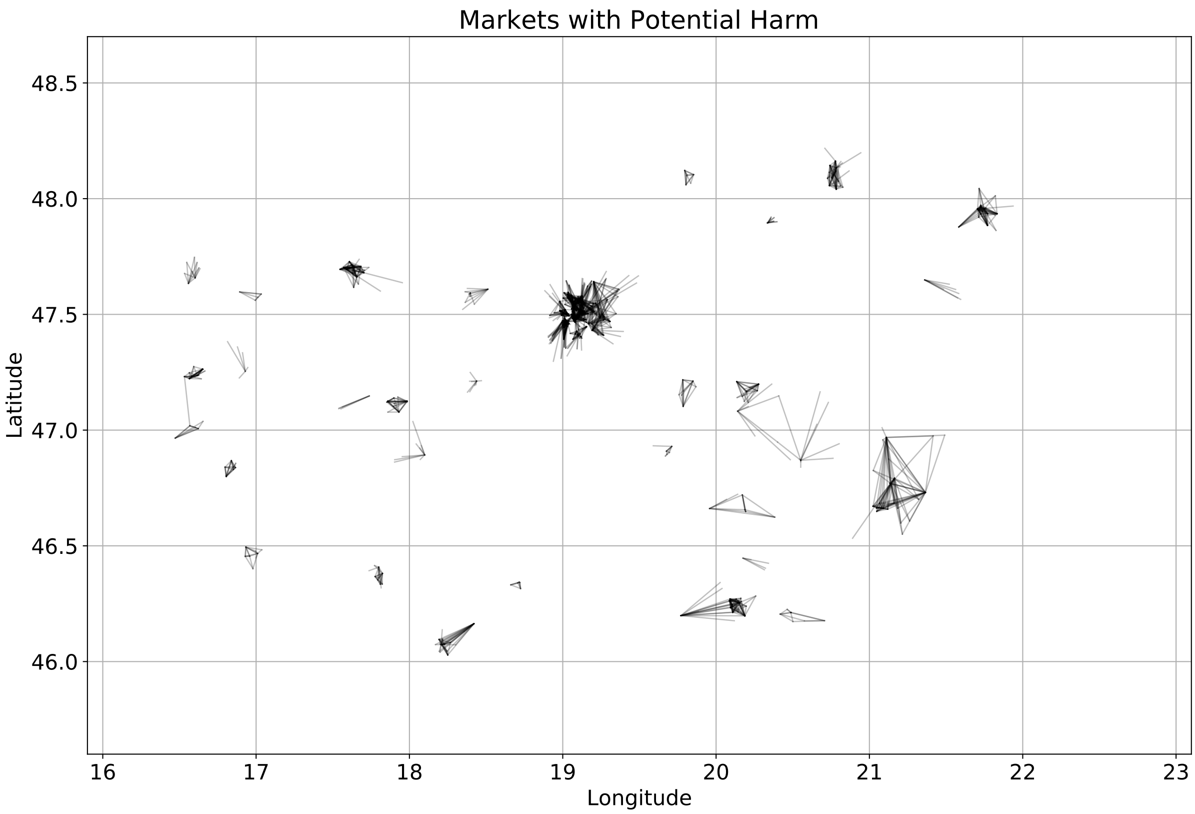


Figure 3.8: Markets with at least 15% change in HHI

Bibliography

Aguirregabiria, Victor and Aviv Nevo, *Recent Developments in Empirical IO: Dynamic Demand and Dynamic Games*, Vol. 3 of *Econometric Society Monographs*, Cambridge University Press,

Alexandrov, Alexei and Joyee Deb, "Price discrimination and investment incentives," *International Journal of Industrial Organization*, 2012, *30* (6), 615 – 623.

Asker, John, Chaim Fershtman, Jihye Jeon, and Ariel Pakes, "The Competitive Effects of Information Sharing," Working Paper 22836, National Bureau of Economic Research November 2016.

Bekes, Gabor, Miklos Koren, and Peter Zsohar, "Geographical analysis of Hungarian Retail gasoline prices," 2011.

Belleflamme, Paul and Martin Peitz, "Platform competition and seller investment incentives," *European Economic Review*, 2010, *54* (8), 1059 – 1076.

Benkard, C. Lanier, Przemyslaw Jeziorski, and Gabriel Y. Weintraub, "Oblivious equilibrium for concentrated industries," *The RAND Journal of Economics*, 2015, *46* (4), 671–708.

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

Bourreau, Marc, Frago Kourandi, and Tommaso Valletti, "Net Neutrality with Competing Internet Platforms," *The Journal of Industrial Economics*, 2015, *63* (1), 30–73.

Brennan, Tim, "Is the Open Internet Order an "Economics-Free Zone"?," *Perspectives from Free State Foundation Scholars*, 2016, *11.*

Chen, Jiawei, Ulrich Doraszelski, and Joseph E. Harrington Jr., "Avoiding market dominance: product compatibility in markets with network effects," *The RAND Journal of Economics*, 2009, *40* (3), 455–485.

Cheng, H. K., S. Bandyopadhyay, and H. Guo, "The debate on net neutrality: A policy perspective," *Information Systems Research*, 2011, *22* (1), 60–82.

Choi, Jay Pil and Byung-Cheol Kim, "Net neutrality and investment incentives," *The RAND Journal of Economics*, 2010, *41* (3), 446–471.

Coe, Patrick J. and David Krause, "An Analysis of Price-Based Tests of Market Antitrust Market Delineation," *Journal of Competition Law and Economics*, 2008, *4* (4).

Doraszelski, Ulrich and Ariel Pakes, *A Framework for Applied Dynamic Analysis in IO*, Vol. 3, New York: Elsevier,

— **and Mark Satterthwaite**, "Computable Markov-perfect industry dynamics," *The RAND Journal of Economics*, 2010, *41* (2), 215–243.

Economides, Nicholas and Joacim Tag, "Network neutrality on the Internet: A two-sided market analysis," *Information Economics and Policy*, 2012, *24* (2), 91 – 104.

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

Farkas, David, Gergely Csorba, and Gabor Koltay, "Prices and concentration on the Hungarian retail gasoline market," *Hungarian Economic Review*, 2009.

Faulhaber, Gerald, Hal Singer, and Augustus Urschel, "The Curious Absence of Economic Analysis at the Federal Communications Commission: An Agency in Search of a Mission," *International Journal of Communication*, 2017, *11* (0).

Fershtman, Chaim and Ariel Pakes, "Dynamic Games with Asymmetric Information: A Framework for Empirical Work*," *The Quarterly Journal of Economics*, 2012, *127* (4), 1611–1661.

Gans, Joshua S. and Michael L. Katz, "Net Neutrality, Pricing Instruments and Incentives," Working Paper 22040, National Bureau of Economic Research February 2016.

Gaynor, Martin S., Samuel A. Kleiner, and William B. Vogt, "A Structural Approach to Market Definition With an Application to the Hospital Industry," *The Journal of Industrial Economics*, 2013, *61* (2), 243–289.

Geroski, Paul and Rachel Griffith, "Identifying anti-trust markets," 2003.

Greenstein, Shane, Martin Peitz, and Tommaso Valletti, "Net Neutrality: A Fast Lane to Understanding the Trade-Offs," *Journal of Economic Perspectives*, May 2016, *30* (2), 127–50.

Hastie, Trevor, Rahul Mazumder, Jason Lee, and Reza Zadeh, "Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares," 2014.

— , **Robert Tibshirani, and Jerome Friedman**, *Elements of Statistical Learning*, Springer, 2001.

— , — , **and Martin Wainwright**, *Statistical Learning with Sparsity*, CRC, 2016.

Hoerl, Arthur E. and Robert W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 1970, *12* (1), 55–67.

Ifrach, Bar and Gabriel Y. Weintraub, "A Framework for Dynamic Oligopoly in Concentrated Industries," *The Review of Economic Studies*, 2017, *84* (3), 1106–1150.

Katz, Michael L., "Wither U.S. Net Neutrality Regulation?," *Review of Industrial Organization*, Jun 2017, *50* (4), 441–468.

Kramer, J., L. Wiewiorra, and C. Weinhardt, "Net neutrality: A progress report," *Telecommunications Policy*, 2013, *37* (9), 794–813. Cited By :48.

Kramer, Jan and Lukas Wiewiorra, "Network Neutrality and Congestion Sensitive Content Providers: Implications for Content Variety, Broadband Investment, and Regulation," *Information Systems Research*, 2012, *23* (4), 1303–1321.

Krusell, Per and Anthony A. Smith, "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 1998, *106* (5), 867–896.

Land, Stephanie R. and Jerome H. Friedman, "VARIABLE FUSION: A NEW ADAPTIVE SIGNAL REGRESSION METHOD," 1996.

Lee, Robin and Tim Wu, "Subsidizing Creativity through Network Design: Zero-Pricing and Net Neutrality," *Journal of Economic Perspectives*, 2009.

Liu, Yu-Hsin, Jeffrey Prince, and Scott Wallsten, "Distinguishing Bandwidth and Latency in Households Willingness-to-Pay for Broadband Internet Speed," Working Paper March 2017.

Markovich, Sarit, "Snowball: A dynamic oligopoly model with indirect network effects," *Journal of Economic Dynamics and Control*, 2008, *32* (3), 909 – 938.

— **and Johannes Moenius**, "Winning while losing: Competition dynamics in the presence of indirect network effects," *International Journal of Industrial Organization*, 2009, *27* (3), 346 – 357.

Mazumder, Trevor Hastie Rahul and Robert Tibshirani, "Spectral Regularization Algorithms for Learning Large Incomplete Matrices," 2010.

Nevo, Aviv, "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *The RAND Journal of Economics*, 2000, *31* (3), 395–421.

__ , **John L. Turner, and Jonathan W. Williams**, "Usage-Based Pricing and Demand for Residential Broadband," *Econometrica*, 2016, *84* (2), 411–443.

Njoroge, P., A. Ozdaglar, N. E. Stier-Moses, and G. Y. Weintraub, "Investment in two-sided markets and the net neutrality debate," *Review of Network Economics*, 2013, *12* (4), 355–402.

Pakes, Ariel and Paul McGuire, "Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model," *The RAND Journal of Economics*, 1994, *25* (4), 555–589.

__ **and** __ , "Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the curse of Dimensionality," *Econometrica*, 2001.

Peitz, M. and F. Schuett, "Net neutrality and inflation of traffic," *International Journal of Industrial Organization*, 2016, *46*, 16–62.

Pinkse, Joris, Margaret E. Slade, and Craig Brett, "Spatial Price Competition: A Semi-parametric Approach," *Econometrica*, 2002, *70* (3), 1111–1153.

Rudin, Leonid I., Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, 1992, *60* (1), 259 – 268.

Schuett, Florian, "Network Neutrality: A Survey of the Economic Literature," *Review of Network Economics*, 2010, *9* (2), 1–15.

Slade, Margaret E., "Exogeneity Tests of Market Boundaries Applied to Petroleum Products," *The Journal of Industrial Economics*, 1986, *34* (3), 291–303.

Tibshirani, Robert, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B*, 1996, *58*, 267–288.

__ , **Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight**, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, *67* (1), 91–108.

Train, Kenneth, "Mixed logit with a flexible mixing distribution," *Journal of Choice Modelling*, 2016, *19* (Supplement C), 40 – 53.

Veiga, Andre, E. Glen Weyl, and Alexander White, "Multidimensional Platform Design," *American Economic Review*, May 2017, *107* (5), 191–95.

Wang, Yu-Xiang, James Sharpnack, Alex Smola, and Ryan J. Tibshirani, "Trend Filtering on Graphs," 2014.

Watanabe, Takanori, Daniel Kessler, Clayton Scott, Michael Angstadt, and Chandra Sripada, "Disease prediction based on functional connectomes using a scalable and spatially-informed support vector machine," *NeuroImage*, 2014, *96* (Supplement C), 183 – 202.

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

Weyl, E. Glen, "A Price Theory of Multi-sided Platforms," *American Economic Review*, September 2010, *100* (4), 1642–72.

Yuan, Ming and Yi Lin, "Model selection and estimation in regression with grouped variables," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 2006, *68*, 49–67.

Zou, Hui and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, *67* (2), 301–320.

10.14754/CEU.2018.11

Appendix A

Appendix for Chapter 1

The goal of this section is to analyze the effects of certain model features one-by-one in a simplified version of the model presented in the main text. First I focus on a model with one ISP and one single CP where there is no dynamic adjustment possible. Then I add some additional layers of complexity and gradually open up the dynamic adjustment possibilities for CP and ISP states.

When introducing termination fees or paid prioritization before examining what happens to investments I will take a closer look at the static equilibrium as profits from the stage game are the main drivers of the evolution of industry states. Firms invest in order to achieve states with higher payoffs. Analyzing outcomes from the stage game is therefore useful in understanding the arising investment patterns.

A.1 Simple Model with Net Neutrality

I start from the simplest model specification featuring one ISP and one single CP and fix CP product quality (δ) and congestion (μ) to a constant. Then I vary δ and μ over a range of values and compute the static equilibrium. To keep the model as simple as possible I make the following simplifications:

- Set the market size (M) to 1
- Set marginal costs of content distribution (c^{isp}, c^{cp}) to 0
- Normalize the utility from no internet subscription (V_0) and casual browsing (v_0) to 0
- Set the price coefficient (z) to 1
- Set bitrate of CP product to a constant b regardless of the value of δ

I re-computed the static equilibrium for a grid of $\delta \in [5, 10]$ and $\mu \in [0.8, 1]$.

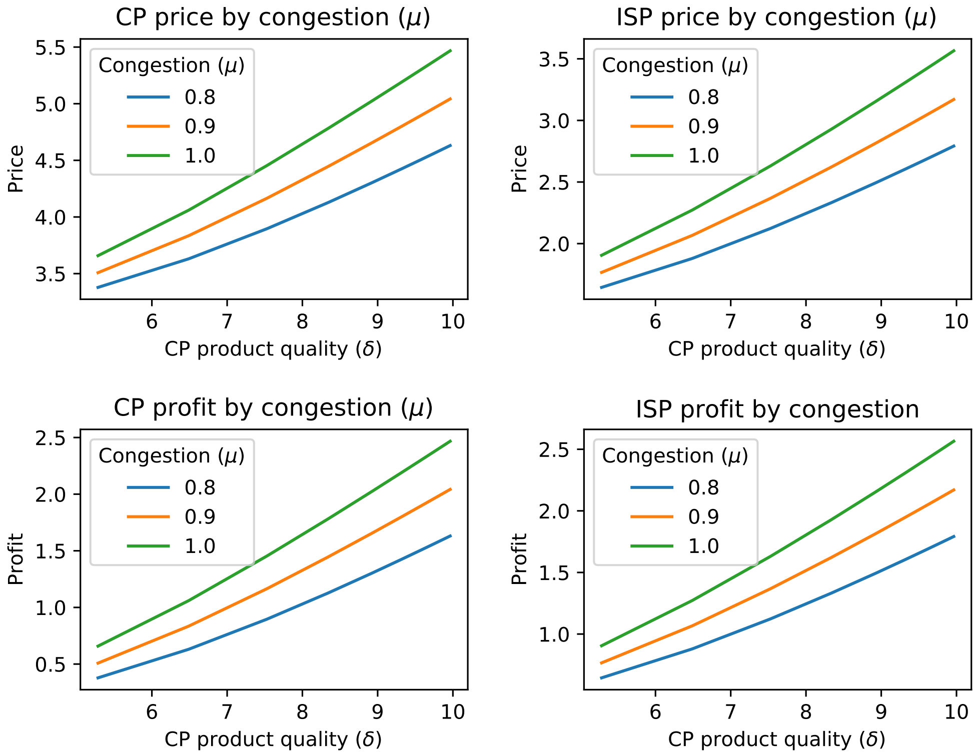


Figure A.1: Prices and profits in the net neutrality model with constant bitrates

A.1.1 Static equilibrium with constant bitrates

Figure A.1 and Figure A.2 summarize the outcomes of the plain vanilla net neutrality model. Price patterns are increasing with CP product quality and decreasing with the severity of congestion. This makes sense as the higher the product quality the more attractive it is to consumers so the CP can raise its prices. Also, a higher quality CP increases the value of internet subscription so the ISP can also raise its prices. The widening gap between the prices under different congestion severities shows that the attractiveness of higher quality products is more sensitive to congestion.

However, there is one fundamental flaw in this model that makes it impossible to use it in the dynamic framework. The growth in profits is unbounded in δ so the CP would wish to invest forever to get into higher states making the problem not stationary. To account for that in the next section I introduce bitrates that increase with product quality and set marginal costs of content distribution (c^{isp}, c^{cp}) to 1. This is one of the most important pieces of the model that connects ISP capacity to CP product quality.

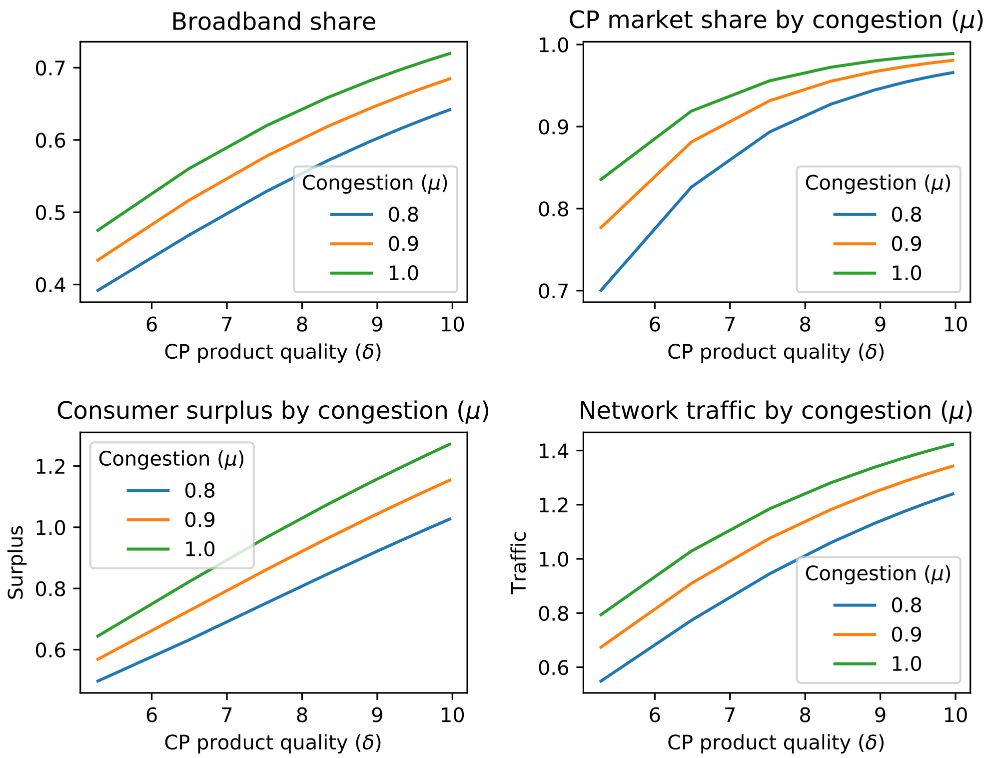


Figure A.2: Market shares, consumer surplus and network traffic in the net neutrality model with constant bitrates

10.14754/CEU.2018.11

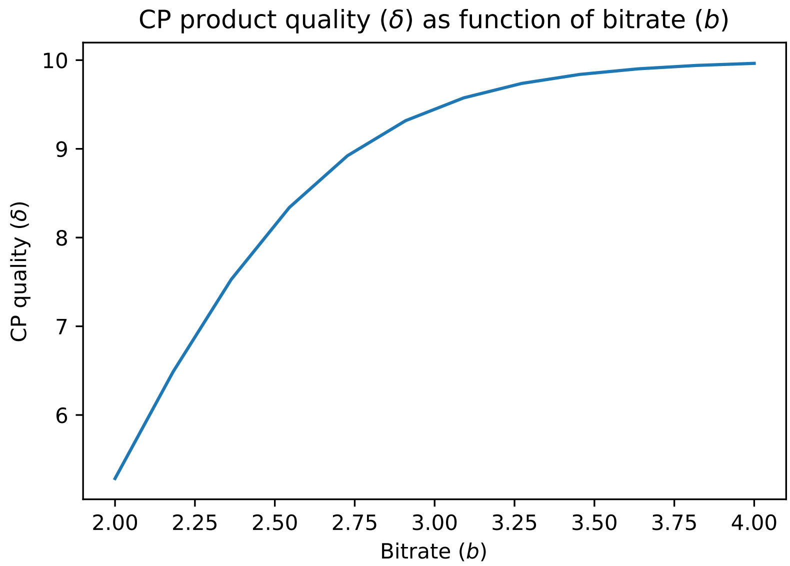


Figure A.3: Decreasing marginal utility of content resolution

A.1.2 Static equilibrium with increasing bitrates

Figure A.3 shows how product quality increases with bitrate. Note the decreasing marginal utility of increase bitrate. I have chosen this on purpose to reflect that the higher the video resolution the harder it is to distinguish quality but bitrates still rise proportionally. One common reason for that is e.g. the lack of suitable screens. Figure A.4 and Figure A.5 summarize the outcomes of the net neutrality model with increasing bitrates. From the profit figures it becomes immediate that profits are bounded as very high quality products are too costly to produce. Even the steep price increase can't justify the high costs. Except for that nothing else has changed qualitatively.

A.1.3 Dynamic equilibrium with CP adjustment

To see what kind of investment behavior is triggered by the profits from the model I first allow the CP to reposition its product by investing into product quality. First I set up a grid of 12 states using increasing bitrates and product qualities from Figure A.3. The states (ω) and the corresponding product qualities (δ) and bitrates (b) are shown on Figure A.6.

For this exercise I shut down the ISP capacity investment channel and assume that there is no congestion. Figure A.7 plots the both value functions, CP investment and the stationary distribution of the Markovian system. One can see that CP investment is in sync with the value function and that investment vanishes beyond a certain state. A typical pattern that could arise is of a CP who starts investing heavily from the low states until it reaches a good state and then keeps investing small amounts to stay on

10.14754/CEU.2018.11

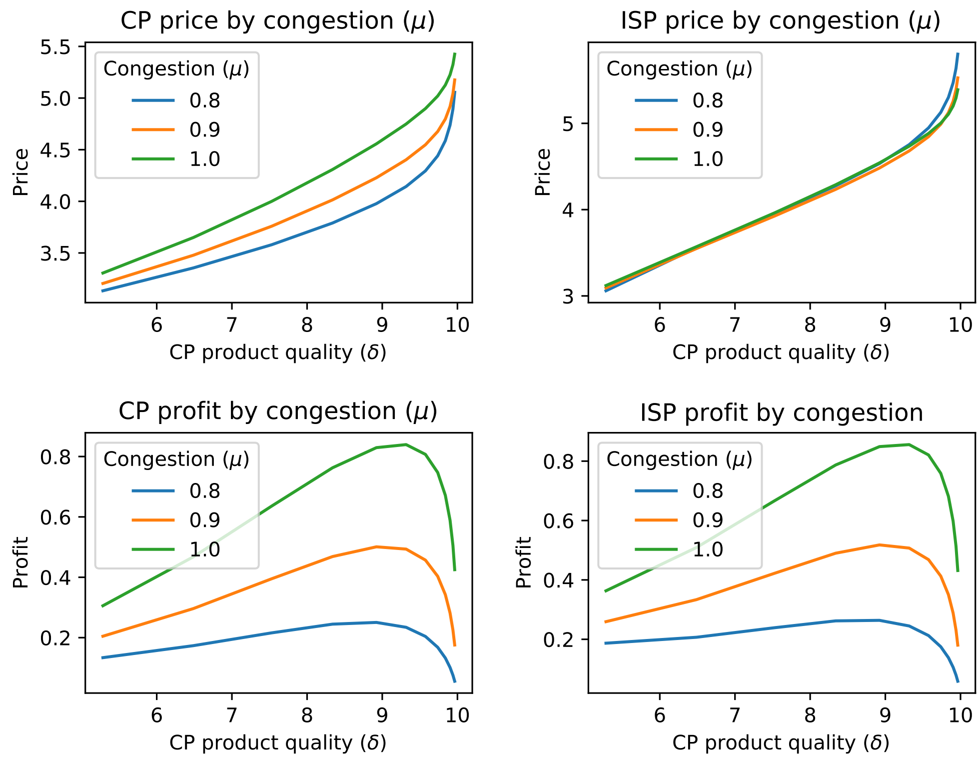


Figure A.4: Prices and profits in the net neutrality model with increasing bitrates

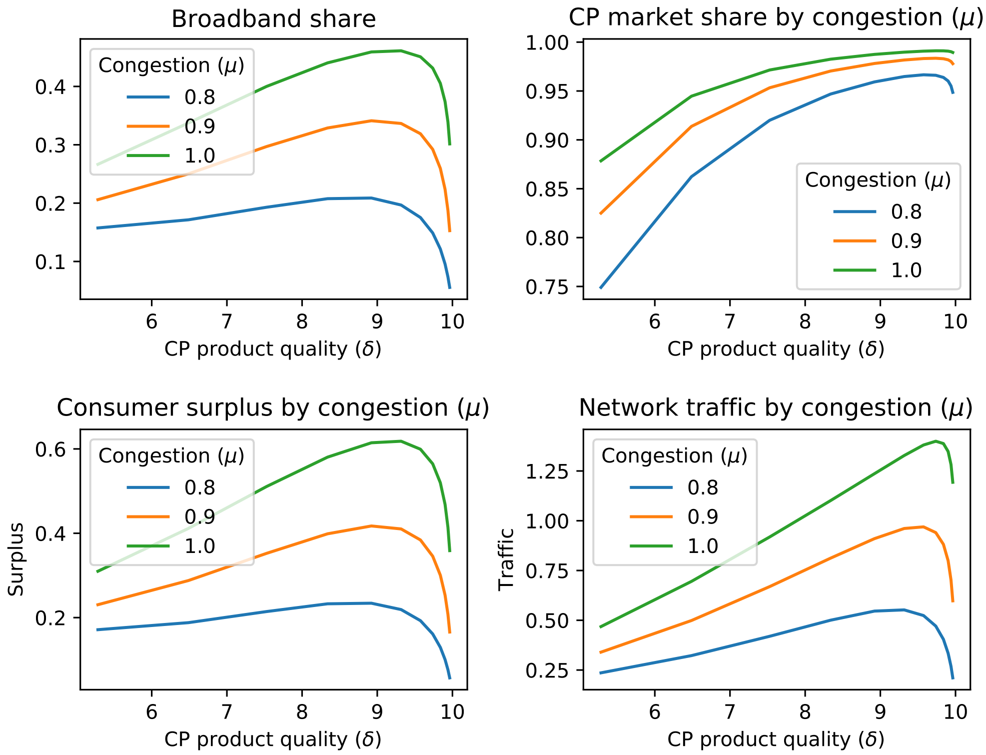


Figure A.5: Market shares, consumer surplus and network traffic in the net neutrality model with increasing bitrates

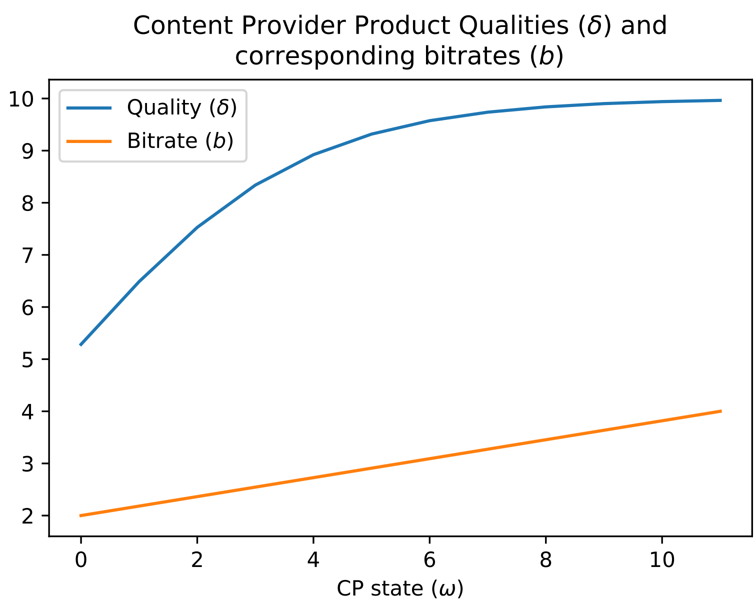


Figure A.6: Content provider states and corresponding product quality levels and bitrates

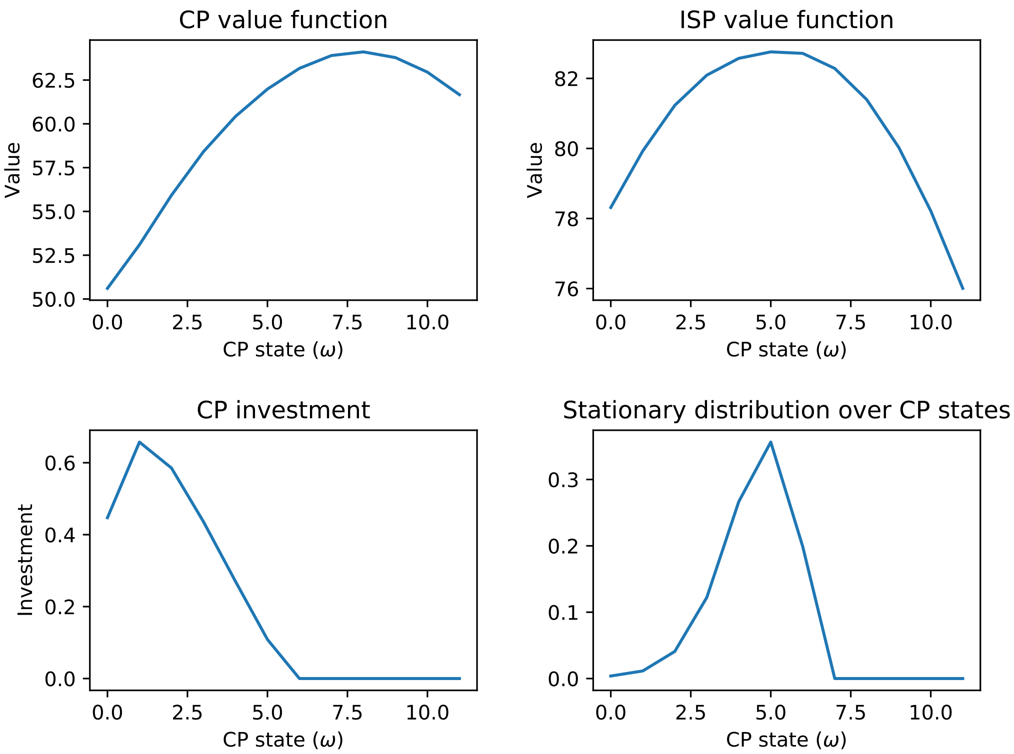


Figure A.7: Value functions and investment in the net neutrality model with CP adjustment. Note that for better visibility the horizontal axis shows states (ω) rather than the corresponding qualities (δ).

top. The stationary distribution confirms this as it is spread over the middle and lower CP states and is mostly concentrated around the middle states.

Interestingly, however, the two value functions have maxima at different states. In the general model this can be a driver of strategic behavior from both sides.

A.1.4 Dynamic equilibrium with both CP and ISP adjustment

In this section I present results of the dynamic equilibrium when both the ISP and the CP can invest in order to achieve higher states. I specify potential ISP capacity levels (κ) to an equidistant grid of 10 states over $[0.25, 2.5]$. The highest capacity level is set such that it exceeds the highest potential traffic volume.

Figure A.8 and Figure A.9 summarize the outcomes. First, it is important to note that the value functions are similar in shape to ones seen in the previous section. However, in this case the level is somewhat lower, because by adding more lower capacity levels to the potential ISP states firms realize that it can get worse than the current state.

Second, the value functions for the same CP state are higher for higher capacity levels. In general both value functions and CP investment follow the order of ISP capacity levels which confirms that higher network capacity means better perspectives for firms. The one exception is ISP investment where the logic is the opposite. This is because with high capacity there is abundance of free capacity and there is no need to invest. On the other hand, when capacity is low the ISP wants to invest more to get to a state with a higher value.

A.2 Effects of Termination Fees

In this section I will continue with the same baseline model but allow the ISP to charge a per-transaction termination fee to the CP. Figure A.10 highlights the differences in prices and profits for a moderate amount of congestion.[1] First, note that CP prices increase substantially but this is accompanied by a severe drop in profits. On the other hand the ISP is able to raise profits by not only keeping its price low but also keeping it at a constant low level. To explain this puzzle one has to take into account the transfers between the ISP and the CP. Figure A.11 helps to see what is going on. On the left plot one can see that for each δ the ISP's profit is maximized at a different termination fee. From the center plot it can be seen that the optimal fee schedule is increasing with CP product quality and it is very similar across different congestion levels. Finally, the right plot shows that the final (or total) price that a residential user pays for internet

[1] For other congestion values the results are qualitatively the same.

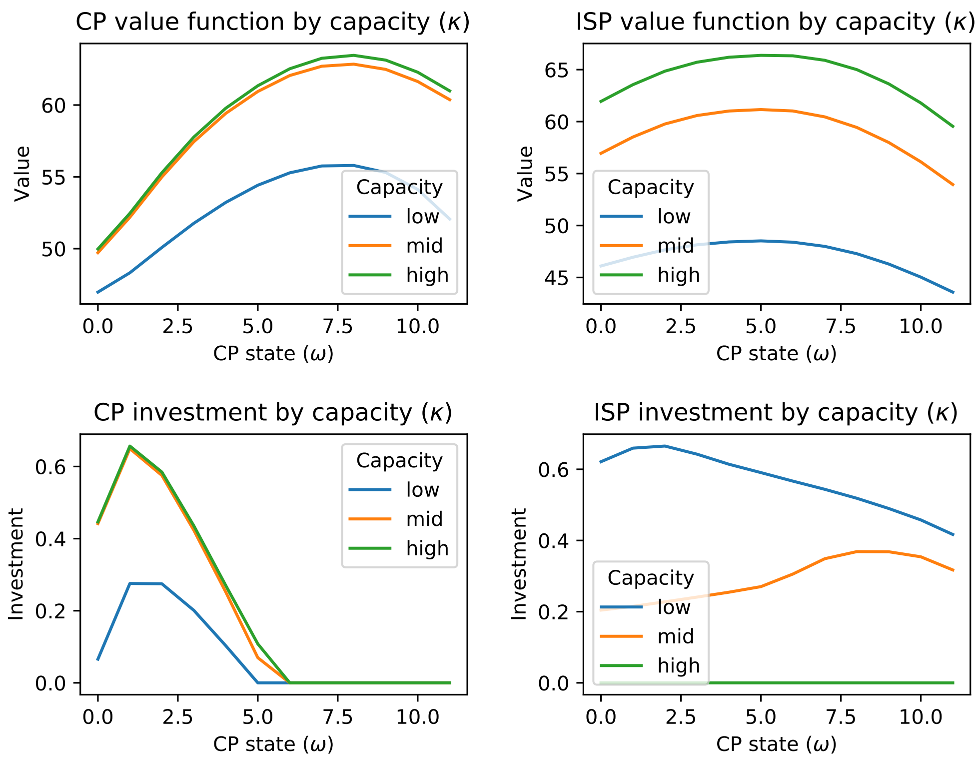


Figure A.8: Value functions and investment in the model with net neutrality

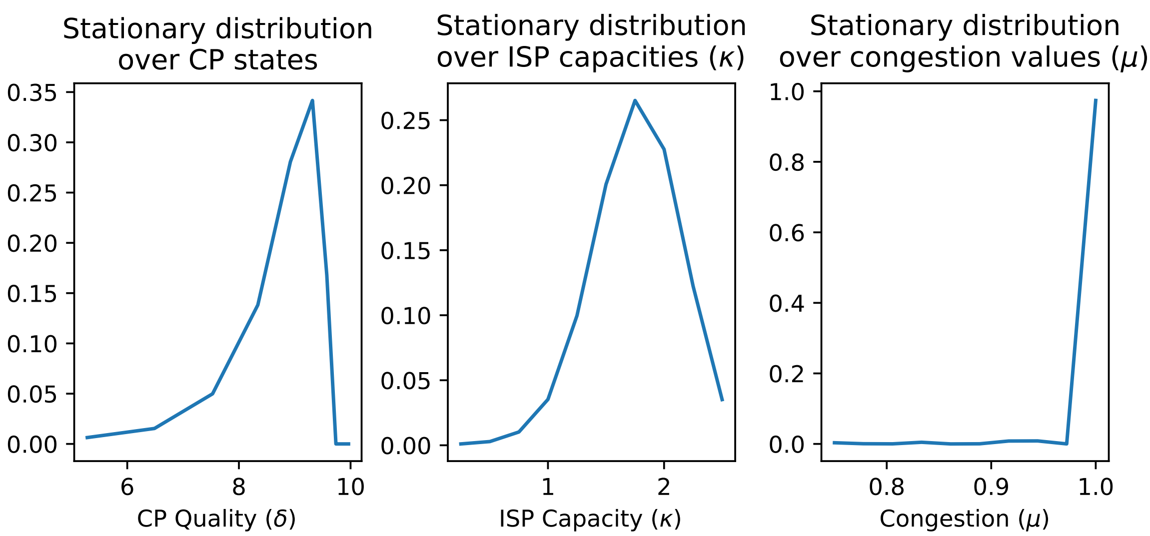


Figure A.9: Stationary distributions in the model with net neutrality

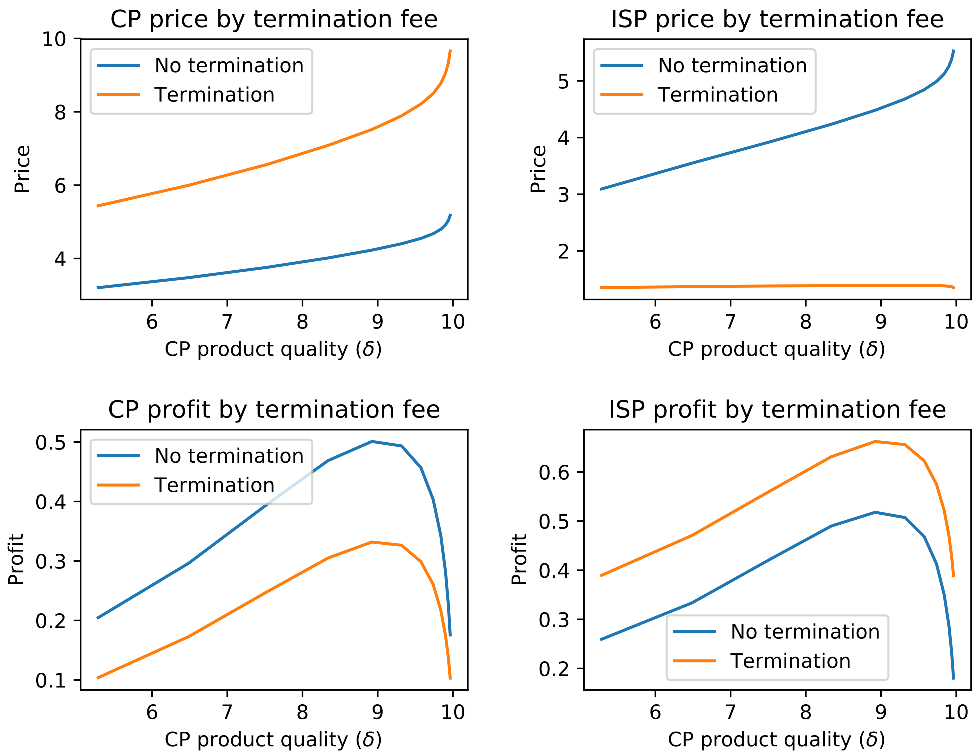


Figure A.10: Prices and profits in the model with termination fees

access and streaming content is roughly the same. This means that the critique of the termination fee that it is going to be transmitted to the user is only half of the truth, because on the other hand the ISP has incentives to lower the subscription price to counter-balance.

Figure A.12 reveals that the high CP prices lead to a loss in market shares. Given that overall internet subscription has increased this means that the streaming video sector has shrunken and users have turned to less bandwidth-intensive casual browsing instead. Interestingly under the current parameterization this has a positive effect on consumer surplus.

One conclusion of this section is that introducing the termination fee leads to an unequal split of revenues from the user between the ISP and the CP. The drop in CP profits might have serious consequences for investment decisions. To examine these possibilities I computed the dynamic equilibrium. Figure A.13 shows the expected drop in CP value and investment across all states and the increase in ISP value. As a consequence of smaller traffic ISP investment has dropped a little, too. Figure A.14 confirms that because of the lower investment levels the stationary distribution has shifted to the left for both CP product qualities and ISP capacities.

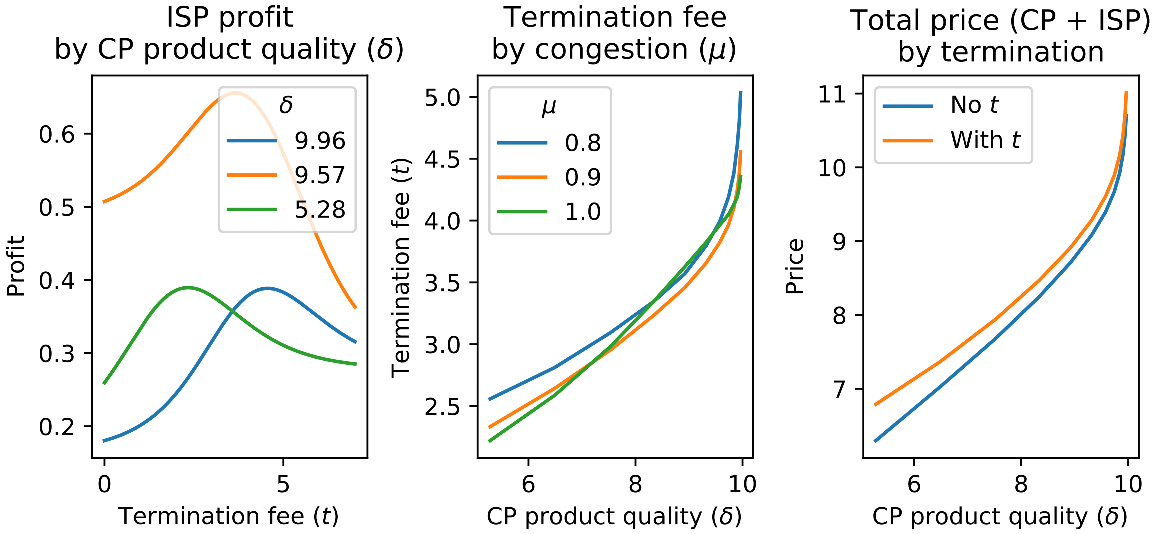


Figure A.11: Optimal termination fee schedule in the model with termination fees

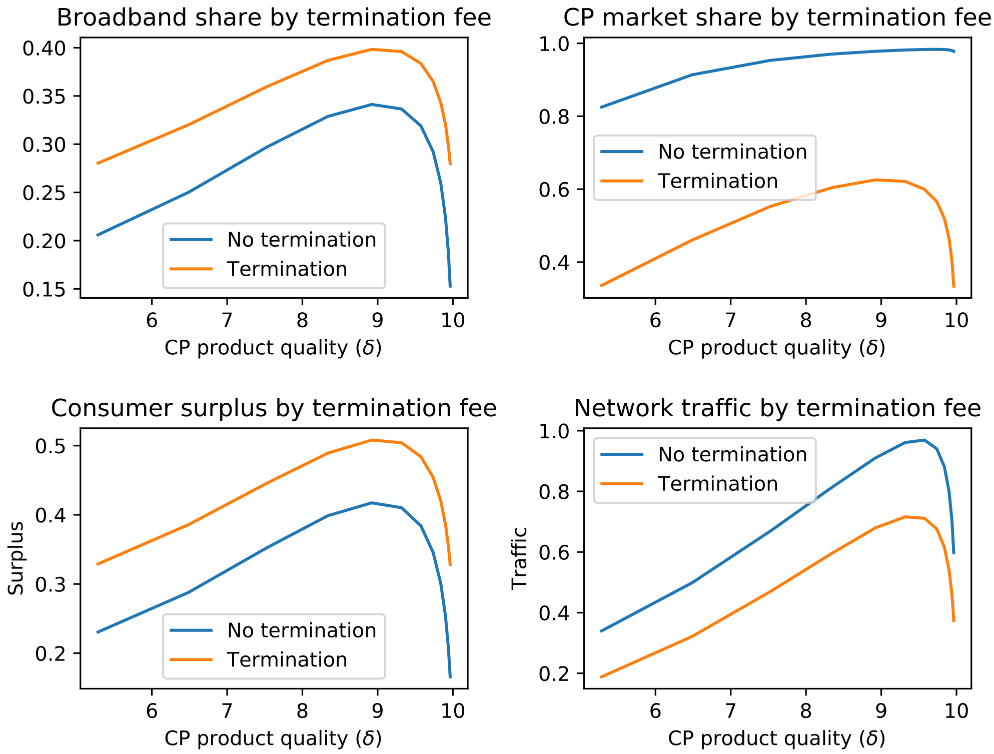


Figure A.12: Market shares, consumer surplus and network traffic in the model with termination fees

10.14754/CEU.2018.11

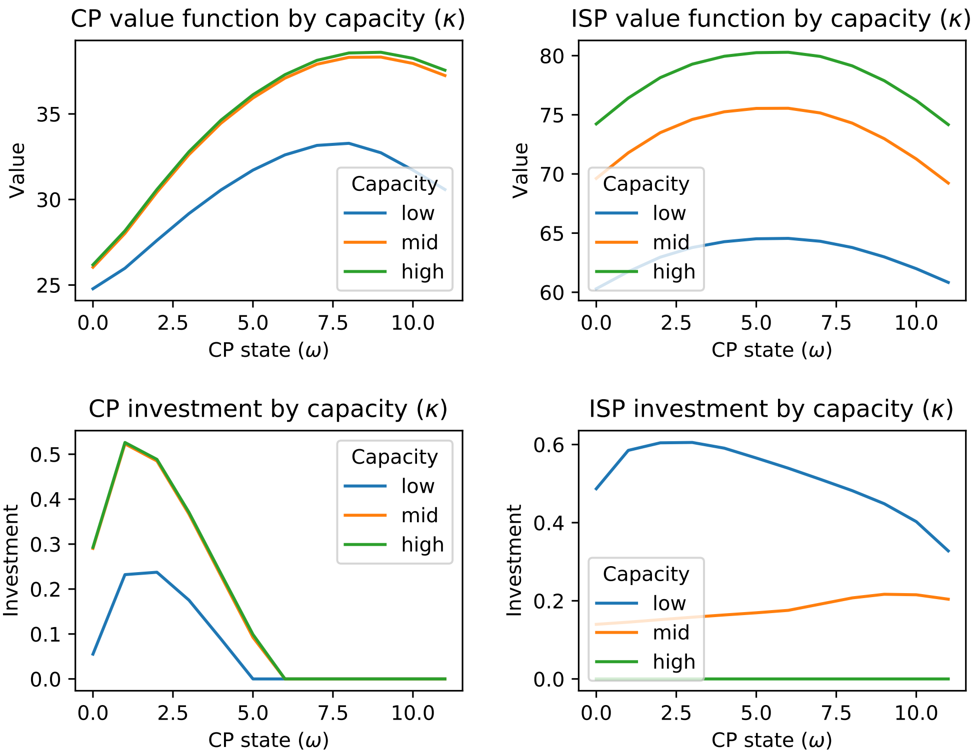


Figure A.13: Value functions and investment in the model with termination fees

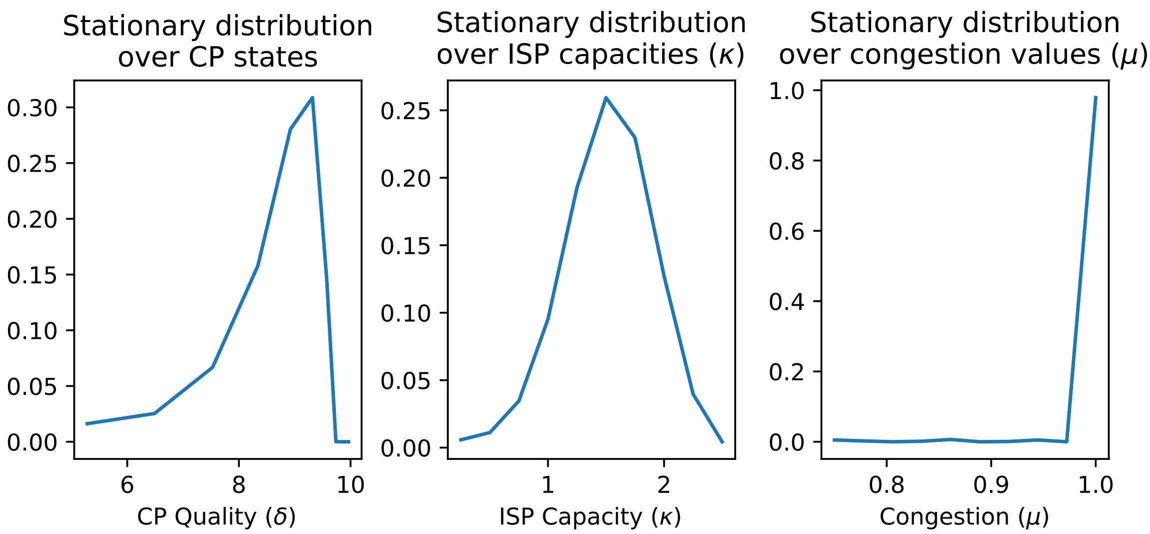


Figure A.14: Stationary distributions in the model with termination fees

A.3 Effect of Paid Prioritization

In this section I allow the ISP to offer a paid priority service to the CP to avoid the negative consumption effects of congestion. Figure A.15 shows that in the case of moderate congestion both CP and ISP prices increase but only ISP profits change as a result. This is because the priority service is no longer mandatory for the CP and it is incentive compatible. The ISP will offer only contracts that leave the CP with at least as large profit as in the neutral regime.

To better understand where the extra profit of the ISP is coming from it is worth to take a look at the optimal fee structure. The left plot on Figure A.16 shows how the ISP determines what fee to charge in a given period depending on the product quality of the CP. It will raise the fee until the CP would be indifferent between buying or not. Of course, this only works in case there is congestion that affects the perceived quality of the CP. In this way the ISP can enjoy all gains from restoring the quality of the CP from $\mu\delta$ to δ. The right plot confirms that the more severe the congestion the more is to gain from the priority service.

According to Figure A.17 the changes in other outcomes like broadband share, CP market share and consumer surplus suggest that there is a lot of value lost if there is congestion and it sounds like a good idea to design a market mechanism that is able to restore this value. However, since the restored value is distributed unevenly to the ISP it may create incentives for the ISP to sustain socially suboptimal congestion.

Under the current parameterization it is not that striking but Figure A.18 confirms this. First, note that the gap between low capacity and high capacity ISP values is much smaller and overall levels are also a little higher compared to net neutrality. This in turn naturally leads to less investment to capacity, too. Maybe the most interesting is the clearly visible under-investment in high CP states where the ISP can extract the most value from the CP.

Otherwise the rest is very similar to the net neutrality results with a slightly lower CP value and investment. According to Figure A.19 the stationary distributions are also similar with a little more mass on some congested states.

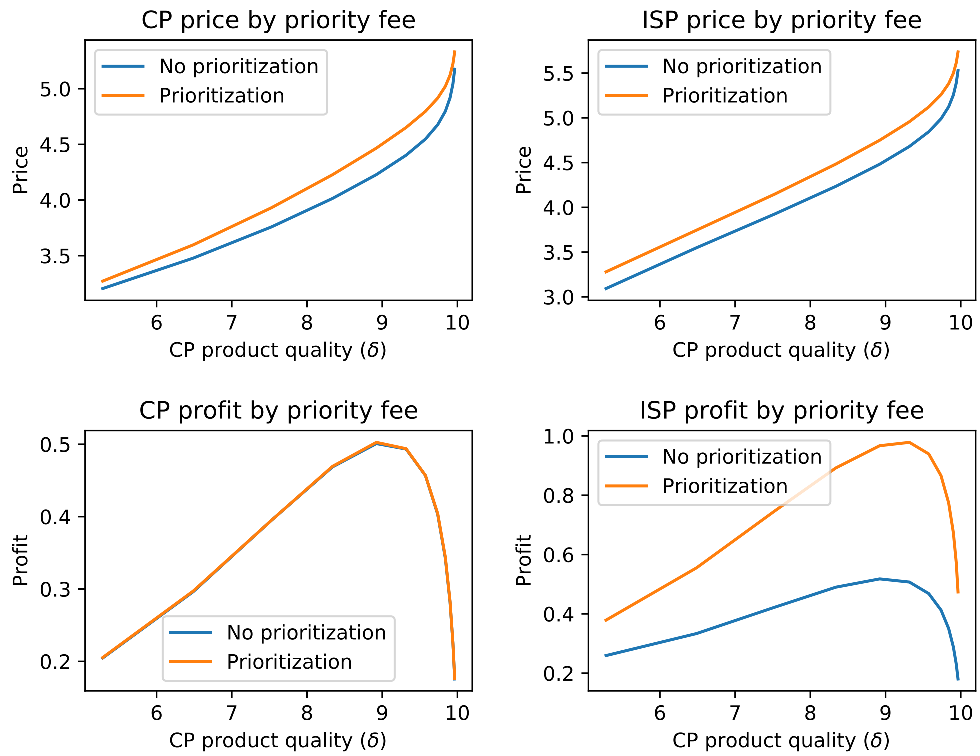


Figure A.15: Prices and profits in the model with paid prioritization

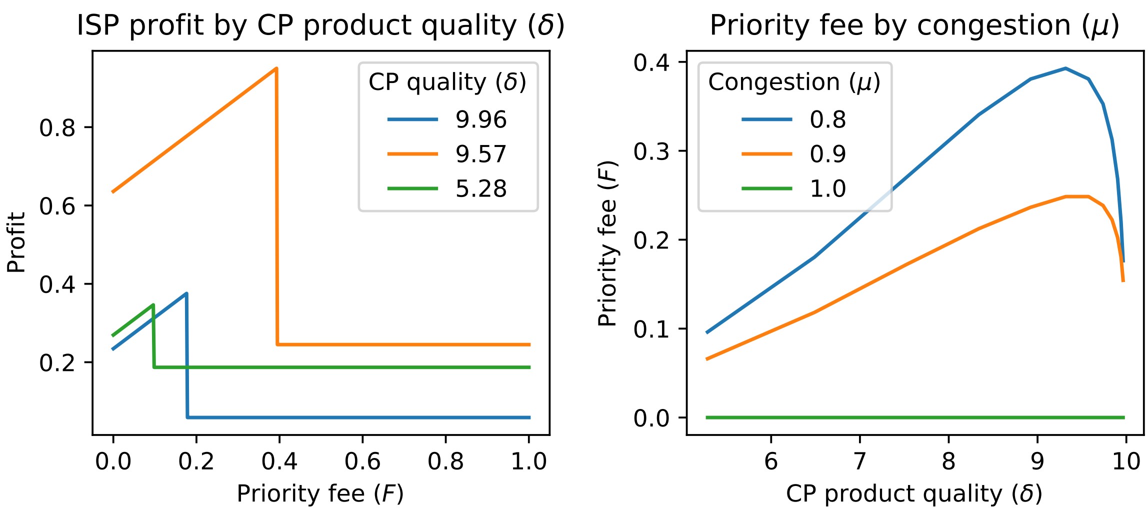


Figure A.16: Optimal priority fees in the model with paid prioritization

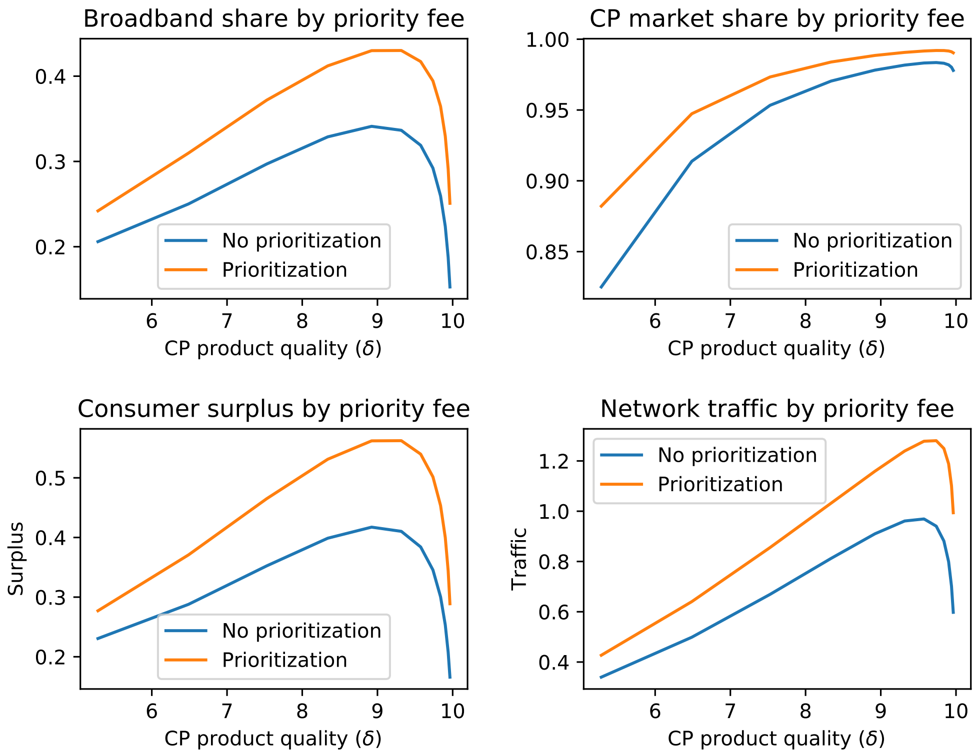


Figure A.17: Market shares, consumer surplus and network traffic in the model with paid prioritization

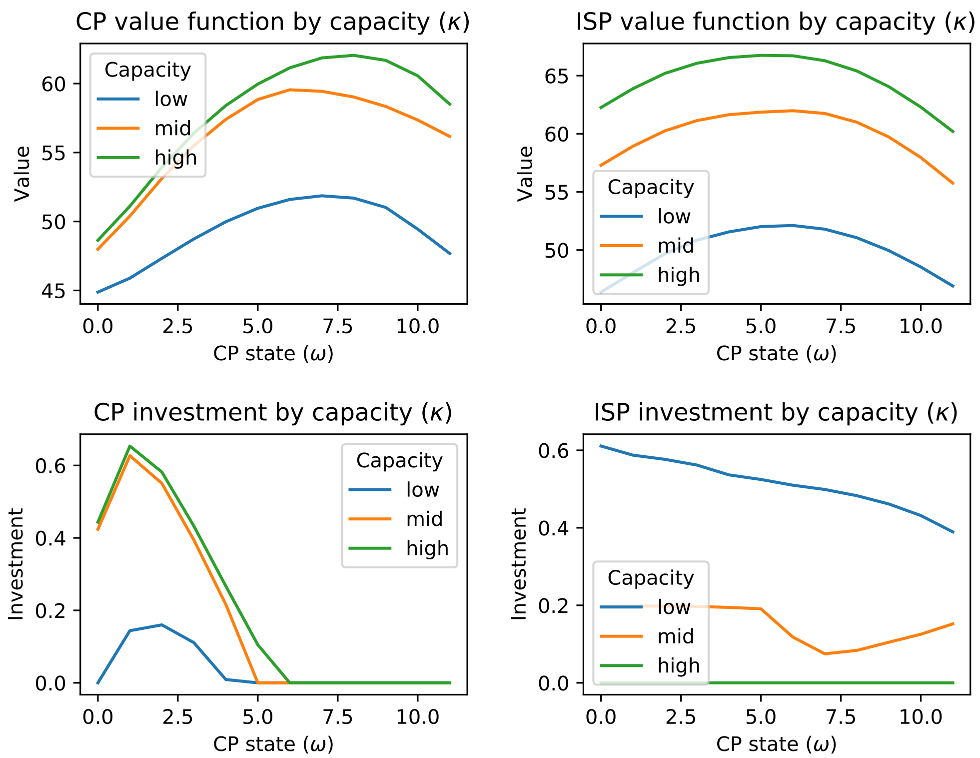


Figure A.18: Value functions and investment in the model with paid prioritization

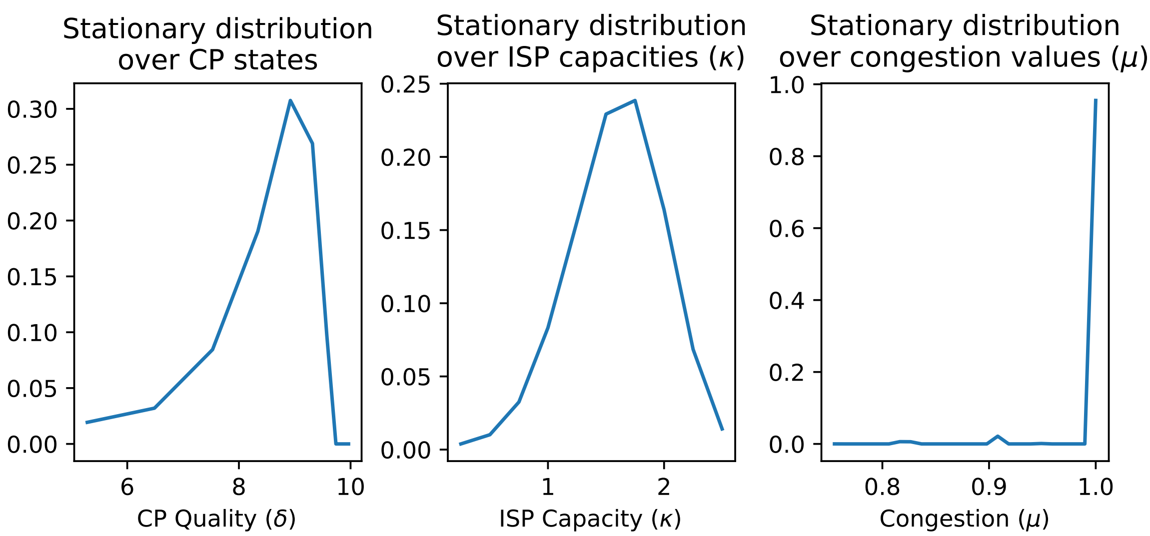


Figure A.19: Stationary distributions in the model with paid prioritization