Mergers and Investments in the Wireless Industry *

Evgeni Drynkin †

Stanford GSB

Abstract

This paper studies the outcomes of hypothetical T-Mobile/Sprint and AT&T/T-Mobile mergers in the U.S. wireless telecommunications industry. I propose a model in which consumers trade off price and network coverage, so firms have to compete on both price and investment. The key finding is that had T-Mobile and Sprint merged in 2009, consumers would have benefited from expanded network coverage. The two firms would have increased profits due to less duplication on the investment side. An acquisition of T-Mobile by AT&T, on the other hand, would have harmed consumers because it would not have resulted in better coverage. Additionally, the outcomes of the T-Mobile/Sprint merger vary across geographic areas. Markets with high population density or flat terrain typically have a strong initial Sprint or T-Mobile presence, and would therefore experience lower, often negative, changes in consumer surplus as a result of the merger. Conversely, markets where the merging parties struggle to enter separately, mainly those with lower population density and harder to cover terrain, benefit more because the merger would diversify carrier choices.

Keywords: Merger analysis, wireless telecommunications, dynamic investment.
JEL Classification: L13, O33, L96, L51.

Latest draft is available at:
https://github.com/edrynkin/research_papers/raw/master/JMP.pdf

*I am grateful to my main advisor, Ali Yurukoglu, for continuous support and many hours of in-depth discussions during every stage of the project. I am also thankful to my advisors, Lanier Benkard, Peter Reiss, and Paulo Somaini, for bringing up important points and their vision to the project. I appreciate multiple discussions of the paper with Wes Hartmann and Brad Larsen. Of course, the paper would not be complete without dozens of questions and comments from participants of IO seminar, IO workshop, and Quantitative Marketing workshop at Stanford. All remaining mistakes and typos are my own.

†edrynkin@stanford.edu

1 Introduction

This paper studies mergers in the wireless telecommunications industry. Its main contribution is the estimation of changes in price, 4G technology investment, and welfare that would have followed two hypothetical mergers – the AT&T/T-Mobile merger and the T-Mobile/Sprint merger – had they happened in 2009. Both of these mergers were proposed to the FCC in the past decade, with the commissioners resolving to prohibit the former and clear the latter. The explanations provided to justify these decisions can be boiled down to a single question – does the positive effect of network quality improvements outweigh the negative effects of increased market concentration?[1]

Theory on how mergers with endogenous investments affect consumer surplus, profits, and total welfare is ambiguous. The tension between market power and appropriability of returns goes back at least as far as Schumpeter (1942) and Arrow (1972).[2] As wireless telecommunications is a highly concentrated industry with just four major nationwide firms in the U.S., authorities would likely view a merger with suspicion fearing higher prices and lower consumer surplus. On the other hand, higher appropriability from larger market shares after a merger could incentivize a combined firm to invest more. This tradeoff between price and quality makes the impact of the merger on the final consumer ambiguous.

I propose and estimate an empirical model of interaction between firms in the market that can generate different effects depending on parameter values, similar to Goettler and Gordon (2011). The model here is tailored to several factors specific to the wireless telecommunications industry – nationwide pricing in combination with localized investment, spatial variation in service quality, large heterogeneity in consumer preferences, and endogenous quality decisions.

My model has a flexible demand system that allows me to generate rich substitution patterns. Because I have individual consumer data, I allow each individual to have their own preferences that vary by demographics and location. The demand system also accounts for the fact that consumers value coverage quality in the area surrounding the zip code in which they live. This paper implements a novel technique to feasibly estimate a demand system that generates flexible substitution patterns between goods using individual level choice data. This technique is a generalization of the often used group-wise (bucketing) estimation.

Second, I specify and estimate a dynamic model of investment. Having a dynamic model for investment is crucial, as firms time their investments strategically in complex ways that will have an effect on the final market structure and competition. In addition, the data clearly suggests that the firms prefer to consistently postpone deployment in

[1]In 2019 FCC and DOJ cleared a proposed merger between T-Mobile and Sprint as "[t]wo of the FCC's top priorities are *closing the digital divide in rural America* and *advancing United States leadership in 5G*, the next generation of wireless connectivity. The commitments made today by T-Mobile and Sprint would substantially advance each of these critical objectives." In 2011, FCC's response on proposed AT&T and T-Mobile merger was: "... [FCC] concludes that significant *harms to competition* are likely to result, primarily in the form of *increased prices for consumers, reduced incentives for innovation, and decreased consumer choice.*"

[2]See Shapiro (2011) for a wonderful review of the subject and Aghion et al. (2005) for one of a multitude of possible theoretical frameworks that generate inverted U-shape curve linking competition and innovation.

some areas, waiting for a better time to invest. The dynamic model explains these delays by linking them to declines in costs over time.

I combine the demand and investment models into a single equilibrium model which operates at multiple aggregation levels. Firms invest at the zip code level and set prices at the national level. At the same time, demand is determined at the county level and serves as an input for both nationwide pricing and zip code investments.

I estimate the proposed model using a unique dataset created specifically for my analysis. I combine several different sources on historic coverage with 10-K reports for four major U.S. wireless carriers, adding elevation data from the USGS as a cost shifter. To estimate the model, I find a set of parameters that match predicted market performance to observed performance and employ instrumental variables to account for endogenous quality choices in demand. These underlying, or structural, parameters allow me to compute the equilibrium in a counterfactual market structure that has never been observed in actual data.

The main finding of this paper is that a hypothetical merger between AT&T and T-Mobile in 2009 would harm consumers by almost $30bn in net present value, while a merger between T-Mobile and Sprint in 2009 would result in consumer welfare gains of around $15bn. The results of each merger also dramatically differ in their effects on the investments and profits of merging and non-merging parties. The effect of the AT&T/T-Mobile merger is almost entirely confined to price increases, resulting in higher profits for every firm in the market. In contrast, the T-Mobile/Sprint merger incentivizes the merging parties to invest, while total industry profits increase slightly with a substantial increase in profits for merging parties and a decrease for non-merging parties.

The growth in consumer welfare under the T-Mobile/Sprint merger comes from various sources. First, the merged firm will have more incentives to invest than either of the two original parties. This results in better average service quality. Second, many areas would experience an increase in the number of alternatives after the merger, and a richer choice set would lead to some welfare gains. Finally, though the prices would go up slightly immediately after the merger, they are estimated to be lower in the medium-term. The effects of the merger also differ across geographic areas. Markets with a strong initial presence of one of the merging parties, typically urban areas, experience low or negative changes in consumer welfare.

Total profits under the T-Mobile/Sprint merger go up by as much as $5.7bn. However, total profits for non-merging parties go down by $4bn, as competition becomes tougher in the high quality segment of the market. Since total investments become more efficient due to broader coverage and reductions in duplicate investments, the total welfare effect of the merger is positive. The AT&T/T-Mobile merger, in contrast, barely affects AT&T's incentives to deploy new facilities, instead softening price competition without adding enough competitive pressure in the deployment game. As a result, total profits of merging and non-merging parties grow by $12bn and $27bn respectively, while consumer surplus goes down by almost $30bn.

The structural approach of this paper is complementary to the methods that are currently used to assess of the effects of wireless telecommunications mergers. Rather than evaluating the outcomes of mergers directly, I model the interaction between the firms based on a set of assumptions and underlying set of structural parameters. this is in contrast to the analysis of wireless telecommunications using difference-in-difference (DiD) or synthetic control (SC) techniques. The method in this paper benefits from being able to make welfare statements and not relying on data from other countries that can be a poor benchmark due to institutional and geographic differences.

The rest of the paper is organized as follows. Section 2 describes the industry and data used in the estimation. Section 3 presents descriptive statistics that help to summarize the key characteristics and trends in the wireless industry. Section 4 provides a high level overview of the model and describes how its parts are connected. Section 5 presents the demand model and pricing and Section 6 presents the investment model. Section 7 outlines the results of the estimation. Section 8 summarizes counterfactual simulations. Section 9 is the conclusion.

2 Industry and data

2.1 Industry overview

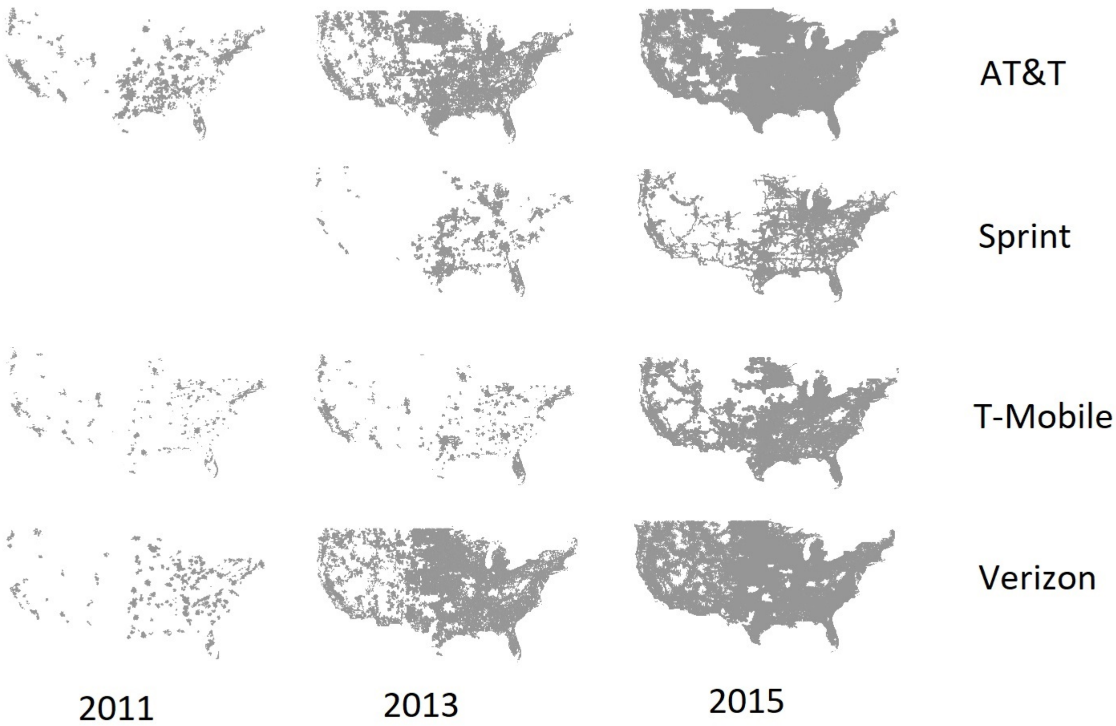
The wireless telecommunications industry is a large and growing part of the U.S. economy. As of 2016, facility-based operators accounted for 416 million connections with total revenue of around $200bn. In recent years, the connections base has been growing by 5-7% annually. An important market characteristic is that large investments need to be made to provide high quality services. Between 2010 and 2016, major carriers invested around $200bn in the wireless telecommunication industry, accounting for approximately one full year of revenue. These investments have led to a tremendous expansion of the coverage network – the number of cell sites grew by 60% over the last ten years.

The industry currently consists of four major players: AT&T, Sprint, T-Mobile, and Verizon. AT&T and Verizon are higher quality service providers, accounting for approximately a third of the market each. Sprint and T-Mobile split the rest of the market. Over time, the U.S. wireless industry is gradually becoming more concentrated. A large industry merger happens on average every six years, with a recently proposed $27bn merger between T-Mobile and Sprint.

Aside from facility based providers, there are also mobile virtual network operators (MVNOs) operating from the facilities of the facility based carriers. I left this market completely outside the scope of analysis for the current paper for two reasons. First and foremost, I am focusing on investment decisions, which are made by facility based operators. Therefore, modeling their decisions is a priority. Second, the share of MVNOs is smaller than the share of facility based operators by orders of magnitude.

The wireless telecommunications industry has several important characteristics. First, prices are set at the national level. However, although pricing is uniform across geographic areas, pricing itself demonstrates a lot of non-linearity. In particular, monthly payments depend on voice minutes, text messages, data included in the contract, as well as the number of lines under a given contract with typical nonlinearities for all of these pricing components. In later years, variability between contracts has significantly de-

Figure 1: 4G coverage over time.



Notes: Heterogeneity in timing of investment and available carriers over different geographic locations. The coverage maps are generated for 4G technology coverage for mainland U.S. Sprint had 4G first rolled out in 2012.

clined. Most contracts now include unlimited voice and text communication, restricting the choice to two variables – data and the number of lines.

Second, carriers compete on service quality. Since I am considering facility based carriers, quality is a variable the firms set themselves. The market is characterized by substantial heterogeneity in coverage across geographical areas, time, and carriers. Some general investment patterns can be observed for all carriers in Figure 1. At the beginning, the carriers target cheap and populated markets to implement a technology. Over time, they start to invest in other areas to gradually improve coverage all over the country. Incentives to invest vary between carriers, with those that have larger market shares tending to invest more. This difference between the two largest and the third and fourth companies is especially striking in less populated areas. Conversely, the largest cities usually get near perfect coverage almost immediately after the introduction of new generation technologies.

Third, different demographics have different usage patterns for wireless devices. For example, younger and wealthier people are more likely to own a smartphone rather than a simple cell phone (FCC (2017)), making them more likely to use enhanced wireless functionality such as streaming, IoT, or browsing. These differences in usage could lead to different positioning for different customers.

Finally, the market has three different levels of aggregation. As I mentioned previously, prices are determined at the national level. Investment decisions, on the other hand, are highly localized and can only affect a single zip code. However, consumers care about coverage at the semi-local level. Since the main advantage of wireless is mobility, consumers care about having coverage in a larger local area, where they are likely to commute, work, or relax, rather than in just their residential zip code. Though there might be some benefits from a carrier covering areas far from a person's location, they are a secondary concern for the vast majority of the population.

2.2 Data

The main piece of my model is a flexible demand system. To estimate how preferences change over individuals and locations, I use individual-level GfK MRI survey data from 2009-2016.[3] This survey is a repeated random sample representative of the population for that year. The size of the survey varies slightly from year to year, but is around 25,000 households annually. The GfK MRI surveys provide the data on the household location, their choice of wireless carrier, and their demographic characteristics. The characteristics that I use in the analysis are the annual income of the household and the age of the head of the household.

To allow for more variation in preferences due to individuals' locations, I make preferences dependent on the geographic area where the individual lives. The idea behind this source of variation comes from the fact that more businesses or people owning mobile devices could lead to higher utility from wireless telecommunications. I thus need to control for some aggregate demographics of the local areas. In my case, I use median area income and age from the Census data.

The main trade-off consumers face in their choice is quality versus price. As discussed previously, a quality measure should reflect quality in the areas close to the individual's location. I assume that the area that consumers care about is the county of the their zip code. During the selected time period, there were a total of three technologies that the carriers were still using – 2G, 3G, and 4G. These technologies are nested, with 2G being the most basic technology and 4G being the most advanced one. Hence, if a firm has 4G coverage in some particular zip code, it supports 2G and 3G in that area as well. I denote q^l_{jct} the share of zip codes of county c where firm j in year t had l as the best available technology. A quality that firm j provides in year t within county c is then defined as a vector $q_{jct} = (q^{2G}_{jct}, q^{3G}_{jct}, q^{4G}_{jct})$.

The mobile deployment data comes from two different sources that are then combined into a single dataset. The first source of deployment data is a proprietary dataset from a driving test conducted by Mosaik Solutions.[4] This is a historic dataset that covers 2009-2013. The results of the driving test are reported as a set of binary indicators by zip code for every carrier and technology. The decision on whether a particular zip code is assigned zero or one is based on the centroid methodology – one is assigned if and only if the geometric center of a given zip code is covered. The second source is the FCC 477 Mobile Deployment Form. It is a self-reported set of coverage maps that every facility based carrier must annually present to the FCC. The first filing year was 2014, so I use the data from the 2014-2016 reports. Each set includes a shapefile for every technology that the carrier has implemented. To make this dataset consistent with the one provided by Mosaik Solutions, I apply the centroid methodology by intersecting coverage shapefiles with zip codes shapefiles obtained from the US Census web site. I will refer to combined dataset as “coverage data” hereafter.

As mentioned above, the actual price that a given household pays is non-linear and almost impossible to observe. I thus use a single nationwide number for a given carrier-year pair as a proxy for price. A reasonable candidate for this role is the average service revenue per user (service ARPU, or ARPU). I use FTC wireless competition reports to obtain measures of service revenues and total connections for every firm on the market. I treat this number as a price of service, and will refer to it as “price” for the rest of the paper.

A potential problem with this definition of the price is its low variation over time and between carriers, as can be seen in Figure 2. Though there is some price decline after 2013, with T-Mobile dropping the price a little earlier than its competitors, the overall differences in prices between carriers are close to constant over time. To overcome this issue, I use another demand shifter. I argue that ad expenditures can serve as such a shifter to control for time changes in relative “branding”. I use 10-K reported advertising expenditures by the carriers as a proxy for actual ad expenditures.

Another valid concern is that deployments are correlated with unobserved qualities in the demand system. To address this problem, I use USGS data on terrain as an instrument to estimate demand. I calculate the terrain ruggedness index from a US elevation map. The terrain ruggedness index (TRI) is a commonly accepted measure that quantifies the topographic heterogeneity of underlying landscapes. TRI is defined as the root of a mean squared local elevation.[5] I then calculate the terrain ruggedness index for a given geographic area, denoted TRI_c, as the average terrain ruggedness index of county c. The TRI has proven itself useful as an exogenous shifter for costs of

[5]Formally, for a given point on the plane (x, y), TRI is defined as:

$$TRI(x,y) = \sqrt{\frac{\sum_{(i,j)\in\{-1,0,1\}^2} \Big(elev(x+i, y+j) - elev(x,y)\Big)^2}{8}}. \tag{1}$$

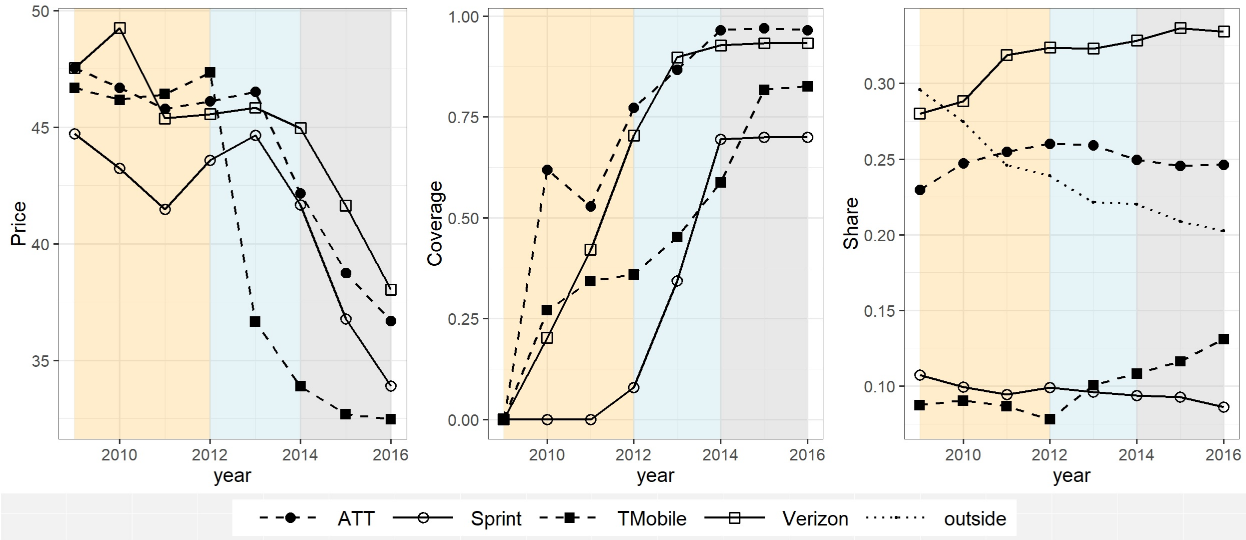
For example, if underlying terrain has the following elevation map with point (x, y) being at the center:

1	2	3
4	5	6
7	8	9

then the corresponding TRI is:

$$TRI(x,y) = \sqrt{\frac{\sum_{i=1}^{9}(5-i)^2}{8}} = \sqrt{7.5} = 2.739. \tag{2}$$

Figure 2: Nationwide summary statistics for main market variables.



Notes: Statistics for mainland U.S. Price is in $/month. Coverage is the share of zip codes covered by 4G technology in a given year.

economic activity in different settings (see, for example, Nunn and Puga (2012)).

3 Descriptive statistics

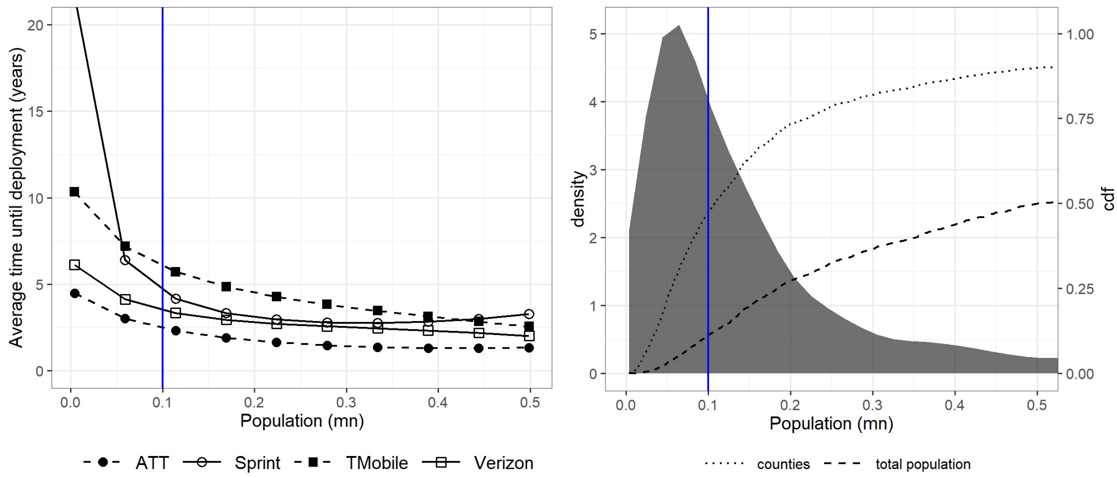
This section presents some major trends and patterns observed in the data that I will later incorporate into the structural model. I start by looking at time trends at the nationwide level of aggregation shown in Figure 2.

Over the analysis period, the prices declined for each of the four firms. The fall in prices combined with expanded coverage and roll-out of 4G led to nearly 10 p.p. higher penetration by the end of the period. However, the relative acquisition of these new wireless customers was heterogeneous among firms as well as over time. I split 2009-2016 into three sub-periods demonstrating the possible interplays of the competing effects.

In the early period, highlighted in yellow in Figure 2 and spanning 2009-2012, the extensive margin comes almost fully from expansion of AT&T and Verizon. Over that period the prices of all firms remained nearly constant. Deployment of 4G technology, on the other hand, improved quality of the provided service. These effects resulted in fewer consumers choosing an outside option. As AT&T and Verizon have faster roll-out of 4G, their quality increased relative to T-Mobile and Sprint, explaining the small decline in shares of two smaller carriers and an increase in shares of the market leaders.

The period highlighted in blue and spanning period 2012-2014 is characterized by a substantial price drop by T-Mobile and gradual 4G roll-out from both T-Mobile and Sprint. As qualities and prices of AT&T and Verizon stay relatively constant over that period, so are the market shares. However, an increase in quality in combination with significant price reduction attracted many new customers to T-Mobile. Most of those

Figure 3: County level summary statistics for population.



Notes: Average time is calculated as inverse probability. The blue line corresponds to rural regions. County cdf plots the share of counties with at most given population. Total population cdf plots the share of population that lives in counties with at most given population.

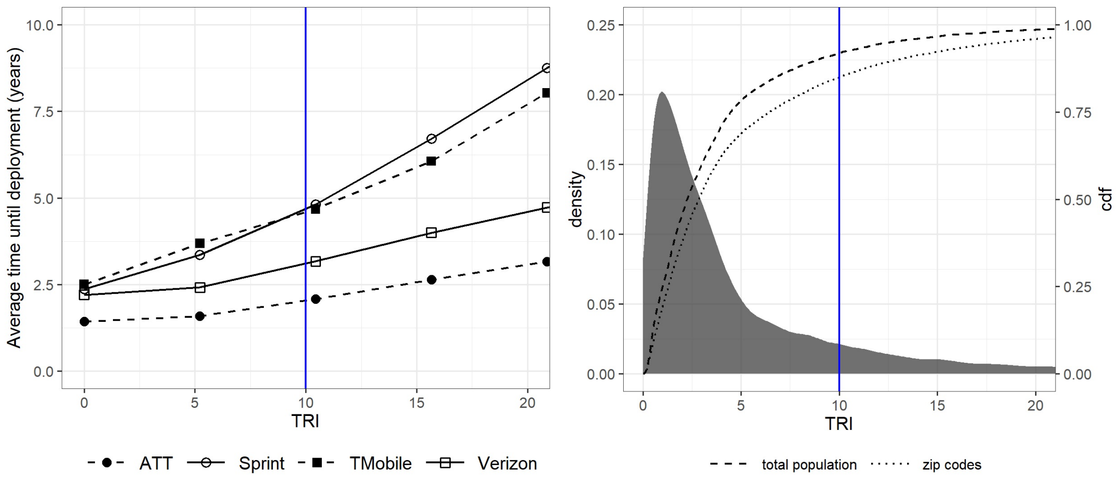
customers are either new consumers or consumers who switched from Sprint. There is little or no cross substitution from market leaders. This fact aligns well with the idea that consumers who previously stayed outside of the market or subscribed to Sprint are the ones with higher sensitivity to price and/or lower value for quality.

The final sub-period starts in 2014 and ends in 2016 is highlighted in grey in Figure 2. It is a period of substantial quality improvement by T-Mobile and price decrease by its rivals. As T-Mobile became a closer substitute for AT&T and Verizon, consumers who have stronger preferences for quality than previously subscribed to T-Mobile started to consider it a viable option. Lower prices set by AT&T and Verizon, however, prevented mass switching of their customer base to T-Mobile. Yet, better coverage of T-Mobile attracted many customers from struggling Sprint or an outside option.

Next, I provide some stylized facts on investment incentives in the market. Figure 3 and Figure 4 show the effects population and TRI have on how fast an average zip code gets covered. Figure 3 demonstrates what is often called "digital divide" between urban and rural America. The blue line separates low populated counties from highly populated ones. In counties with population above 100,000, all four carriers deploy within five years after 4G technology becomes available, and for those above 500,000 – within three years. In contrast, counties with population below 100,000 people have a very limited number of carriers providing high quality most of the time, with T-Mobile and Sprint taking more than 10 years on average to deploy in some of the counties. Those low populated counties account for almost half of all counties in the country but only 12% of U.S. population. To disentangle the cost and benefit part of this "rural divide", I specify cost and revenue functions that depend on population.

Another dimension "dividing America" that is usually left out of public debates is the landscape of surrounding area as shown in Figure 4. T-Mobile and Sprint struggle

Figure 4: Zip code level summary statistics for TRI.



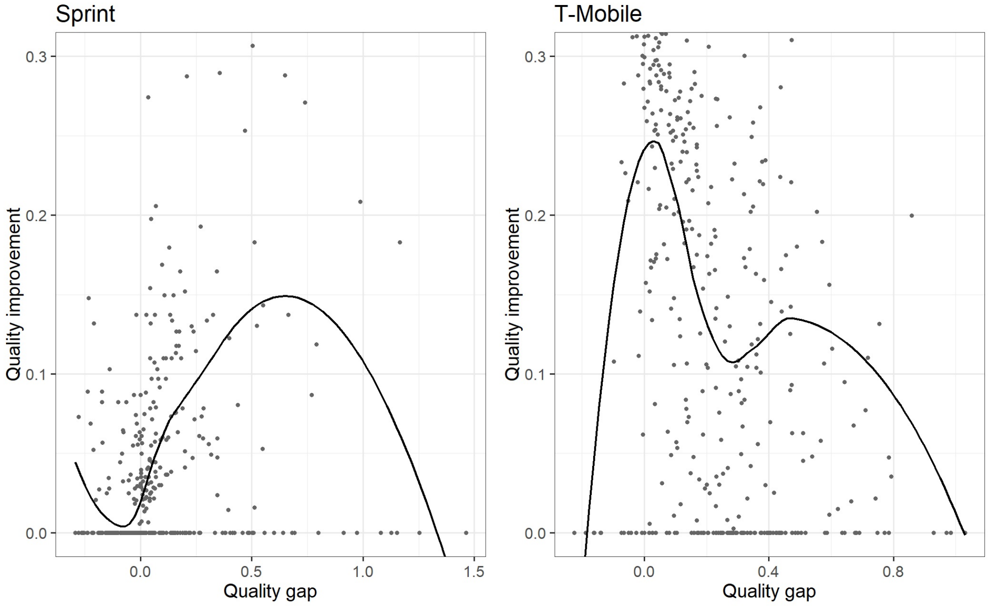
Notes: Average time is calculated as inverse probability. The blue line corresponds to high ruggedness areas. County cdf plots the share of zip codes with at most given TRI. Total population cdf plots the share of population that lives in zip codes with at most given TRI.

in a similar fashion for extremely rugged terrains. Those areas account for around 10% of population and 15% of zip codes. I interpret this stylized fact as evidence for costs depending on ruggedness.

Appropriability and dynamics are important forces shaping the structure of the wireless market. If it is indeed the case, one should be able to see some patterns in entrance decisions across various geographic locations. Presumably, this effect is especially important for the carriers with smaller shares, T-Mobile and Sprint. Figure 5 provides some evidence that this is indeed the case. This graph shows how the size of investment, defined as the difference between next and current period quality indices, for the named carriers depends on the quality gap. The quality gap is a measure representing how far a given carrier falls behind its competitors. There are many other equilibrium effects besides those relative presences that affect the size of deployments. In order to highlight the effect of interest as much as possible, I restrict the sample to a single year of 2009. I also restrict the analysis to the counties where the gaps are positive. The results confirm the idea of equilibrium entry concerns as a result of lower returns when stronger competitors are present. With extremely high gaps, the qualities demonstrate strategic substitute pattern with optimal action going down with more investment from competitors. In contrast, with tight competition, the firm has stronger incentives to deploy to preempt its competitors, making the decisions strategic compliments. As a result, the size of investment demonstrates the inverse U-shape as a function of quality gap. The different types of the responses over the different parts of the domain is another industry specific detail that can be generated within my model.

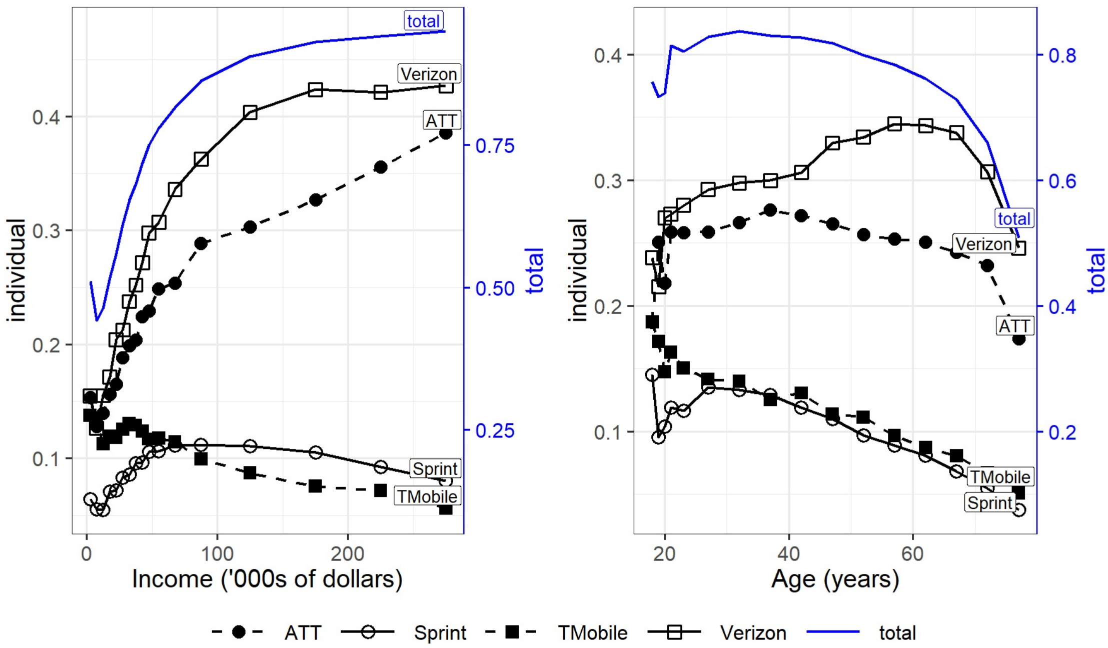
Finally I document the variation in consumer preferences. Overall wealthier and younger people have stronger preferences for wireless. The poorer population is more price sensitive, substituting quality for price by choosing Sprint or T-Mobile more often.

Figure 5: Inverse U-shape in investment decisions for Sprint and T-Mobile.



Notes: Quality is defined as estimated average utility from deployment levels for a given county and carrier. Quality gap is defined as the difference between average quality deployed by the competitors and own quality. Subsample for year 2009 is shown.

Figure 6: Individual choices variation.



Notes: Based on GfK MRI survey data. Left y-axis and black lines show shares of individual carriers. Right y-axis and blue line show total penetration rate (sum of four shares). Income stands for income of a household, Age is the age of the head of household.

Despite younger people having stronger tastes for quality and a larger mobile penetration rate, they do not have such a strong brand preferences. In contrast, older cohorts tend to disregard Sprint and T-Mobile choosing Verizon or AT&T as their primary carrier. To account for all of these effects, I propose a very flexible demand system.

I conclude this section with a simplified version of the analysis I will later exploit in a full scale structural model. The goal is to demonstrate the main identification strategy within a more familiar econometric approach. At a high level, the estimation uses TRI as an instrument for quality. Abstracting away from the endogeneity of 2G and 3G deployments, I estimate regressions that only treat 4G deployment as the endogenous explanatory variable. Table 6 summarizes results obtained by OLS and IV regressions. The first stage of the IV regression demonstrates the expected negative sign on TRI. More importantly, TRI is strongly significant justifying its usage as an instrument. Comparing the second stage of the IV to OLS regression reveals that not accounting for endogeneity is likely to cause biases in the coefficients. Despite this simplified estimation procedure being instructive, it is overly simplistic, and does not account for endogeneity of 2G and 3G deployments, heterogeneity in preferences, and supply side of the model. I now move to a full scale structural model to incorporate these factors into the analysis.

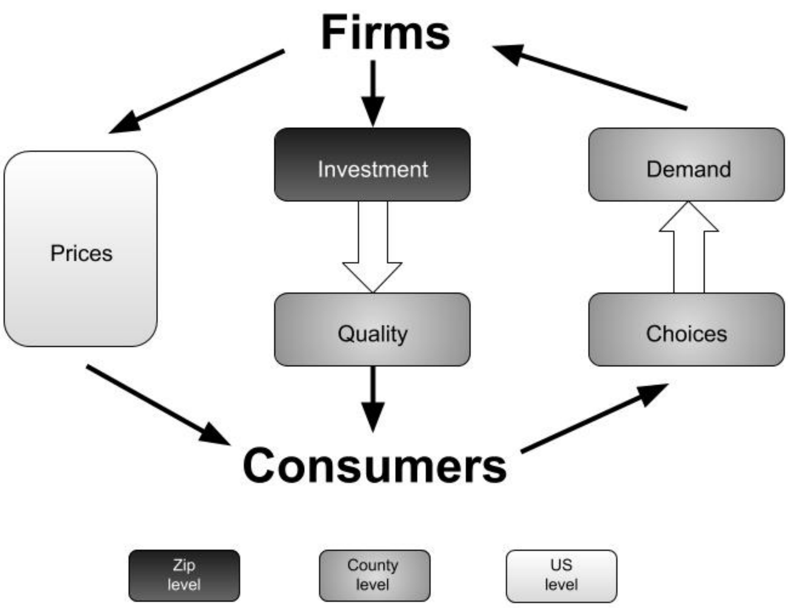
4 Model overview

Before going into details on the methods used later, I give an overview of the timing of the decisions and the geographic levels at which decisions are made. The actions include: (i) investment and pricing for firms; and (ii) carrier choice for consumers. As a result, the model has to reflect the fact that those actions affect market performance differently. We specify three different aggregation levels:

1. **Nationwide level:**
    - Pricing game.
    - Investment strategies.
2. **County level:**
    - Observed and unobserved qualities.
    - Demands.
3. **Zip code level:**
    - Investment decisions.

These levels have strong connections. For example, even though *investment decisions* are made at the zip code level, the overall investment strategy, i.e., the description of how the firm is *expected to behave* is a nationwide strategy. Of course, in equilibrium those must be consistent. More details on interactions between decision layers are presented later in the paper. Another important aspect of the model is the horizon of the decision's effect. The decisions can be divided into two groups: (i) static problems – pricing and demands; and (ii) dynamic problem – investments. The aforementioned structure is summarized in Figure 7.

Figure 7: Summary of the model structure.



Demand is modeled as the demand system with heterogeneous agents. The specification of demand employs a standard empirical industrial organization approach following Berry (1994) and Berry et al. (2004). In our demand system every consumer is endowed with some underlying preferences that uniquely define a good this consumer chooses. Aggregation over individual choices generates aggregate demand.

Given individual demands, the firms aggregate it to a national level an then solve a pricing game. I assume static Bertrand with heterogeneous products equilibrium prices. The static prices and corresponding profits then serve as inputs for investment problem.

In the investment game, firms trade off current deployment costs, extra demand coming from higher quantity, and benefits from waiting for a cheaper investment opportunity. The firms have a choice of deployment in any zip code at any time. Deployment in a given zip code will affect all future demands in the whole county containing the zip code. The firms thus evaluate the extra benefits from higher demands given current deployments by competitors, characteristics of the market, and assumed future behavior of every participant in the market and compare it with cost of investment in a given market, accounting for cost shock realization and the fact that waiting can also lead to lower costs.

The investment game is similar to Björkegren (2018) and Goettler and Gordon (2011). However, there are some important differences due to institutional details and available data. In particular, Goettler and Gordon (2011) evaluate investment decisions in the personal computer microprocessor industry, which is characterized by long run but homogeneous demand. In our case demand is short run and heterogeneous. Björkegren (2018) analyzes the effect an early entrant would have on the roll-out of the system and adoption of wireless telecommunication over time in the context of Rwanda. Björkegren (2018) has detailed data on tower location in Rwanda. I build a similar model with additional data on individual choice. As a result, the present paper can construct a more flexible demand system at the price of simpler investment model allowing only for discrete investment decisions rather than the analysis of the evolution of the whole network. Another modeling difference is that this paper includes dynamic investment game and allows for dynamic adjustment in prices. In Björkegren (2018) the firms commit to price and investment path upfront, which is institutionally more relevant for the Rwandan context.

The investment model also introduces some innovations, as the firms take the multiple decisions on different levels of demand aggregation. The way we address multiplicity of decisions is by assuming that the firm is a collection of independent entities, where every entity tries to maximize profits by own actions but does not coordinate those actions in order to maximize profits by joint actions. Different aggregations and entities put limitations on the applicability of Bajari et al. (2007). We combine the idea of Bajari et al. (2007) with a mean-field approximation introduced in Weintraub et al. (2006). As investment leads to changes in quality of the good purchased by consumer, the specification of this paper fits directly into demand estimation with endogenous product characteristics. Goettler and Gordon (2011) proposed a dynamic framework to analyze quality improvement over time. However, their formulation was continuous and product improvement did not have any logical cap. In this paper, investment is a discrete decision that is irreversible and can only take on two values, 0 or 1. Aguiar and Waldfogel (2018) generate endogeneity in quality through dynamics in entry and exit of the firms in the market. The wireless telecommunications industry, in contrast, has extremely persistent structure over time. Wollmann (2018) is similar to Aguiar and Waldfogel (2018) with dynamics on the products rather than firms generating the effect. Petrin and Seo (2016) take somewhat different approach and used firm optimality conditions to correct on potential endogeneity of unobserved quality. The present paper treats optimal investment conditions as a dynamic quality decision and unobserved products characteristics as orthogonal to other product characteristics.

5 Demand and Pricing

5.1 Model

Detailed data on individual consumers characteristics and choices makes it possible to have a flexible demand specification. In contrast, limited data on price variation requires additional functional form assumptions to obtain both identification and higher precision of the estimates.

I assume that every consumer has a random utility of the following form:

$$u_{ijct} = \sum_{l \in \{2G,3G,4G\}} \beta_i^l q_{jct}^l + \alpha_i p_{jt} + FE_{ij} + \gamma_{ij}^{FE} t + \gamma_i^{ad} \log AD_{jt} + X_c \gamma_i^X + \xi_{ijct} + \epsilon_{ijct} \quad (3)$$

where:

- $j \in \{0, 1, \ldots, 4\}$ is the identity of the firm. We normalize the characteristics of outside option, $j = 0$, to 0;
- l is the type of technology, which in the period of consideration varies between either of 3 types: 2G, 3G, and 4G;
- c is the market identity. We associate a market with a county;
- $t \in \{2009, \ldots, 2016\}$ is year;
- q^l_{jct} is the share of zip codes in county c where the best technology deployed by firm j at year t is l;
- p_{jt} is the price of firm j at year t;
- FE_j is the set of firm fixed effects;
- γ^{FE}_{ij} is the firm specific time-trend in quality level;
- AD_{jt} is the advertising spendings;
- X_c is the county level control variables;
- ξ_{ijct} is heterogeneity that is unobserved by econometrician, but is observed by firms and consumers;
- ϵ_{ijct} is idiosyncratic shock. We assume that those shocks are iid across all dimensions, and follow type-I extreme value distribution.

Most demand models of this type impose some functional form, typically linear in individual level characteristics Z_i, on β_i, α_i, and γ_i, while the fixed effects are assumed to be constant over i.[6] The novelty of my specification comes from the fact that I allow all individual specific coefficients to vary in arbitrary way with their demographics.[7] The restriction is that the coefficients are functions of underlying individual characteristics Z_i, that is $\beta_i = \beta(Z_i)$, $\alpha_i = \alpha(Z_i)$, $FE_{ij} = FE_j(Z_i)$, $\gamma^{FE}_{ij} = \gamma^{FE}_j(Z_i)$, $\gamma^{ad}_i = \gamma^{ad}(Z_i)$, $\xi_{ijct} = \xi_{jct}(Z_i)$.

Price is nationwide and hence it varies only over jt combinations. The model is thus not identified with arbitrary jt fixed effects. Instead, I propose the following functional form on FE_{jt}:

$$FE_{jt}(Z_i) = FE_j(Z_i) + \gamma^{FE}_j(Z_i)t + \gamma^{ad}(Z_i)\log AD_{jt}. \tag{4}$$

Despite the fact that for any fixed Z_i choice satisfies IIA, aggregate demand does not, allowing for flexible substitution patterns. The idea is essentially the same as in BLP

[6]The effects of consumer specific tastes have been found to be essential for a variety of industries including among others: cable television (Crawford and Yurukoglu (2012)), new cars (Berry et al. (2004)), alcohol beverages (Griffith et al. (2019)), and soda (Dubois et al. (2017)).

[7]The functional form that is used in present paper is most similar to Griffith et al. (2019) and Dubois et al. (2017).

with random coefficients. As different groups of consumers tend to choose goods with different characteristics, the distribution of those tastes matters for determination of own and cross elasticities. The fact that it fully comes from observed distribution of tastes is irrelevant for this matter.

Combining all assumptions, I end up with the following equation defining the utility of the consumer:

$$\begin{aligned}u_{ijct} = \sum_{l\in\{2G,3G,4G\}} \beta^l(Z_i)q^l_{jct} + \alpha(Z_i)p_{jt} + \\ FE_j(Z_i) + \gamma^{FE}_j(Z_i)t + \gamma^{ad}(Z_i)\log AD_{jt} + \\ \gamma^X(Z_i)X_c + \xi_{jct}(Z_i) + \epsilon_{ijct}\end{aligned} \quad (5)$$

Another limitation of the price data is that the relative prices of different alternatives do not vary much. Even though T-Mobile dropped its prices earlier than other carriers, which is one of the major identification variations for the price sensitivity, the fact that I use only a proxy for the prices makes estimates noisy. To extract more information from the limited relative price variation, I impose functional form assumption on the cost side. The model assumes that the firms compete à la Bertrand with differentiated products. I also assume that there is no private information at the pricing stage. Hence, as the demand in every period is fully determined by variables observed in that period, the pricing game is essentially static. Under Bertrand competition the following first order condition (FOC) must hold:

$$\sum_i \frac{\partial s_{ijt}(p_t)}{\partial p_{jt}}(p_{jt} - MC_{jt}) + \sum_i s_{ijt}(p_t) = 0 \quad (6)$$

Due to logit specification of the individual level demand function, the partial derivative has a form:

$$\frac{\partial s_{ijt}(p_t)}{\partial p_{jt}} = \alpha_i s_{ijt}(p_t)\Big(1 - s_{ijt}(p_t)\Big) \quad (7)$$

This function can be inverted for every function of parameters $\theta(\cdot)$. $MC(\theta(\cdot))$is then a vector of marginal costs rationalizing observed prices. In closed form:

$$MC_{jt} = p_{jt} + \frac{\sum_i s_{ijt}(p_t)}{\sum_i \frac{\partial s_{ijt}(p_t)}{\partial p_{jt}}} \quad (8)$$

The cost side assumption states how marginal costs evolve over time:

$$\log MC_{jt} = \kappa^{MC}_{\tilde{j}(t)0} + \kappa^{MC}_1 t + \kappa^{MC}_2 q^{4G}_{jt} + \omega^{MC}_{jt}, \quad (9)$$

where $\tilde{j}(t)$ is j for AT&T, Sprint, and Verizon, while for T-Mobile $\tilde{j}(t)$ is split into two separate indicators: before 2014 and from 2014 on. This is done to control for cost changes due to acquisition of MetroPCS in 2013 with the merger being finalized by 2014.

5.2 Identification and estimation

There are three main sources of variation that make it possible to estimate the demand model. Those sources operate on different levels of aggregation: individual characteristics, county level controls and quality, and nationwide prices, fixed effects, and advertising. I start with the highest aggregation level.

The first source of variation that identifies price sensitivity comes from relative changes in shares and prices over time. The basic assumption here is that after properly controlling time and firm fixed effects, levels of ad expenditures, and observed quality the relative change in prices is exogenous. This idea stated formally can be summarized as:

$$\mathbb{E}_c[\xi_{jct}(Z_i)] = 0, \quad \forall j, t, Z_i. \tag{10}$$

As was mentioned above, relative price variation is small compared to noise in the price data. I extract additional information on price sensitivity by adding an extra moment on cost evolution. Formally, I impose:

$$\mathbb{E}[\log AD_{jt}\omega_{jt}^{MC}] = 0. \tag{11}$$

This assumption holds, for example, if ad spendings are predetermined, i.e., 2014 expenditures are set in 2013, while the shocks in marginal costs are contemporaneous. This assumption is weaker than imposing $\mathbb{E}[\log AD_{jt} \log MC_{jt}] = 0$, which does not hold in the data. The way this moment helps to improve precision for price sensitivity estimate is through price itself. Both $\log AD_{jt}$ and ω_{jt}^{MC} are positively correlated with price. The difference between them is that ω_{jt}^{MC} depends on price sensitivity through the inversion formula. This would lead to different decompositions of prices into $\log AD_{jt}$ and ω_{jt}^{MC} for different values of α. The moment allows to prioritize values of parameters that violate the moment less.

A second source of variation comes from variation over counties. Consider two otherwise identical counties. If they have different market shares, it must be because they have different ξ's. Since there is no way to observe "otherwise equal" counties in the data, I control for other variables in Equation 5. The only variables that vary at the county level are X_c and q_{jct}^l. The assumption I impose is:

$$\mathbb{E}_c[\xi_{jct}(Z_i)X_c] = 0, \quad \forall j, t, Z_i. \tag{12}$$

This assumption states that controlling on prices, deployments, county variables and nationwide variables, unobserved taste shocks are orthogonal to county controls.

Conversely, assuming $\mathbb{E}[\xi_{jct}(Z_i)q_{jct}^l] = 0$ would be problematic. The strategic choice of areas to invest comes from returns on investment, which depends on relative tastes for a carrier in a given county, $\xi_{jct}(\cdot)$. As a result, there is a natural correlation between q_{jct} and $\xi_{jct}(\cdot)$. To overcome this problem, I use an instrumental variable.

As an instrument for deployments I use the terrain ruggedness index (TRI). Covering areas with variable elevations is more costly.[8] I impose that taste shocks are uncorrelated

[8]For example, FCC lists physical obstacles among some other factors that can negatively affect coverage

with local landscape once all variables from Equation 5 are included. Formally, the identifying moment is:

$$\mathbb{E}_c[\xi_{jct}(Z_i)TRI_c] = 0, \quad \forall j, t, Z_i. \tag{13}$$

TRI only varies over c, so all the variation coming from j and t is irrelevant for this set of moments as long as specification for $FE_{jt}(Z_i)$ is correct. The control is then equivalent to control for tastes being systematically different in areas with different terrains. The combined set of moments I use for identification of demand and pricing is then:

i. For every j, t, and Z_i:

$$\mathbb{E}[\xi_{jct}(Z_i)] = 0 \tag{14}$$
$$\mathbb{E}[\xi_{jct}(Z_i)TRI_c] = 0 \tag{15}$$
$$\mathbb{E}[\xi_{jct}(Z_i)income_c] = 0 \tag{16}$$
$$\mathbb{E}[\xi_{jct}(Z_i)age_c] = 0 \tag{17}$$

ii.

$$\mathbb{E}[\log AD_{jt}\omega_{jt}^{MC}] = 0 \tag{18}$$

The last piece of variation, the consumer level choice differences, makes it possible to trace back parameters as functions of Z_i. To better understand how, I now go into details on the estimation.

5.2.1 Estimation technique

The model outlined above fits into the general framework of Doudchenko and Drynkin (2019).[9] More details, as well as theoretical properties of the procedure can be found there. Here I briefly outline the main steps of a two-step estimation algorithm.

Estimation is done in two steps. In the first step, I estimate a set of functions $\{\widehat{s}_{jct}(Z)\}_{jct}$ for all jct combinations. I fit a logit regression for every jct. Formally, at the first stage I solve the sequence of the problems:

$$\max_{\theta_{jct}} \prod_{i \in \mathcal{I}_{jct}} sig(Z_i\theta_{jct})^{a_i}(1 - sig(Z_i\theta_{jct}))^{1-a_i} \tag{19}$$

The idea for the second step comes from the fact that once the estimates $\{\widehat{s}_{jct}(Z)\}_{jct}$ are obtained, the optimization can be easily done, as the dependence on Z drops out. Indeed, once a particular value Z_0 is fixed, the problem fits directly into the framework of Berry (1994) which is an IV-estimation of transformed shares, in this case $\{\widehat{s}_{jct}(Z_0)\}_{jct}$. The assumptions above insure that there are enough moments to identify all parameters in the demand system.

in its "Understanding Wireless Telephone Coverage Consumer Guide".

[9]There are some recent developments in structural econometrics designed to incorporate modern machine learning techniques into structural demand estimation. Approaches alternative to theirs can be found in Compiani (2018), Gillen et al. (2015), and Gillen et al. (2014).

The only difference with the procedure described previously is that the moment condition (ii) is aggregate rather than Z_i-wise. Appendix A describes in detail, how this complication can be handled. I use a grid of 36 points over income-age pairs, and use cubic splines to interpolate to other values Z_i.

6 Investment

At the lowest level of aggregation, a zip code, a firm makes a discrete investment decision. As period of analysis is starts when 4G roll-out started, I restrict the action set of the firms to a single decision on whether to upgrade a particular zip code to 4G technology. If the firm decides to invest at time t, it moves to an absorbing state, that is the firm obtains 4G in that zip code forever starting $t+1$ and paying investment costs upfront, i.e., in period t.

I now move to the description of a strategy of a firm in period t. I assume that the firm acts on limited information. That information includes its own and competitors' current average unobservables in the county, $\bar{\xi}_c = \frac{1}{I_c}\sum_i \xi_c(Z_i)$, current average quality in the county, $\bar{q}_c = \frac{1}{I_c}\sum_i q_c(Z_i)$, the TRI and population of the county, (TRI_c, pop_c), and TRI of a particular zip code, TRI_z. These assumptions have different rationales behind them. That strategy is only a function of $(\bar{\xi}_c, \bar{q}_c)$ is made to simplify the state space. In fact, the true state is the whole distribution of $\xi(\cdot)$ and $q(\cdot)$. However, this full state space would not be tractable. The simplification, that only ties strategy to means rather than the full distribution, is common in macroeconomics papers with heterogeneous agents, following Krusell and Smith (1998). Similar ideas hold for the dependence of the strategy on (TRI_c, TRI_z), rather than on the whole distribution of $\{TRI_{z'}\}$ for $z' \in c$. The assumption that the strategy only depends on pop_c, however, does not impose any restrictions. Indeed, it does not matter what is the particular distribution of population over the zip codes within the county, as the demand only depends on the county level quality.

Denote now $\lambda_{zt} = (\bar{\xi}_{ct}, \bar{q}_{ct}, TRI_{ct}, pop_{ct}, TRI_{zt})$. Let $T_t(\lambda'; \lambda)$ be the density of the transition process, for transition from λ at period t to λ' at $t+1$. The only two evolving states are $\bar{\xi}_{ct}$ and $\bar{q}_{ct}$. The evolution of $\bar{q}_{ct}$ is described in detail below, as it requires some extra assumptions. For $\bar{\xi}_{ct}$ I assume an exogenous $AR(1)$ process. That is, for some ρ and σ^{ξ}, the following holds:

$$\bar{\xi}_{jct} = \rho\bar{\xi}_{jct-1} + \sigma^{\xi}\epsilon^{\xi}_{jct} \tag{20}$$

with ϵ^{ξ}_{jct} being $\mathcal{N}(0,1)$ independent of all other variables in the model.

To fully describe optimal investment problem of a firm, one needs to go one layer above, to the county-level, as the returns on investments depend on the demand in the whole county. The key assumption here is that each county comprises a large number of zip codes drawn at random from some distribution. Under this assumption, the mean field approximation technique from Weintraub et al. (2006) can be used. The evolution of $\bar{q}_{ct}$ is then defined by equation:

$$\bar{q}_{jct} = \frac{1}{|zip_c|} \sum_{z=1}^{|zip_c|} \bar{q}_{jzt} \tag{21}$$

where:

$$\bar{q}_{jzt} = \frac{1}{I_c} \sum_{i=1}^{I_c} \Big(\bar{q}_{jzt-1}(Z_i) + \sigma_{jt-1}(\lambda_{zt-1}) \big(\beta^{4G}(Z_i) - \bar{q}_{jzt-1}(Z_i) \big) \Big) \tag{22}$$

with the initial condition:

$$\bar{q}_{jz,2009} = \frac{1}{I_c} \sum_{i=1}^{I_c} \sum_{l \in \{2G,3G,4G\}} \beta^l(Z_i) \mathbb{I}[d_{jz,2009} = l] \tag{23}$$

As a result, any profile of strategies σ has associated transition densities $T(\lambda'; \lambda)$.

Another simplification comes from the fact that one can linearize non-linear county level profits around realized paths to evaluate the effect of 4G deployment in a given zipcode z:

$$\Delta\pi_{jzt'} = pop_c \sum_{t>t'} \delta^{t-t'} (p_{jt} - MC_{jt})[s^1_{jct} - s^0_{jct}] \approx$$
$$pop_c \sum_{t>t'} \delta^{t-t'} (p_{jt} - MC_{jt}) \left[\frac{\partial s_{jct}}{\partial q_{jct}} \Delta q_{jzt,t'} \right] \tag{24}$$

Using some algebra,[10] it can be shown that:

$$\Delta q_{jzt,t'} = \frac{1}{I_c} \sum_{i=1}^{I_c} \frac{1}{|zip_c|} \Big(\beta^{4G}(Z_i) - \sum_{l \in \{2G,3G,4G\}} \beta^l(Z_i) \mathbb{I}[d_{jz,t} = l] \Big) \prod_{t''=t'+1}^{t-1} (1 - \sigma_{jt''}(\lambda_{zt''})) \tag{25}$$

Averaging over many forward simulations, I end up with $E\Delta\Pi_{jzt}$.

To calculate the flow of profits in Equation 24, the firms rely on future path of p_{jt} and MC_{jt} at national level. The process for prices corresponds to equilibrium pricing decisions given observed realizations of all other variables in the demand system. As parameters $\alpha_i = \alpha(Z_i)$ are known, so is every term in Equation 6 except for p_t. I thus can solve the system for p^*_t, for a Bertrand-Nash equilibrium. This completes the description of evolution process for p_t.

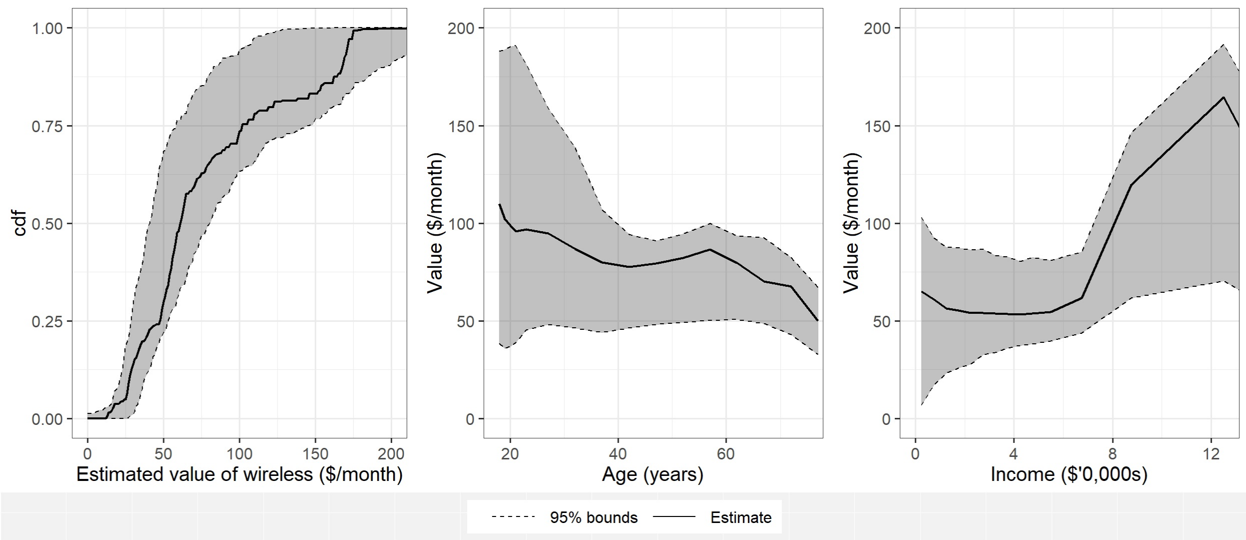
Next step is to estimate costs. To do this I evaluate the NPV for the future deployment costs in zip code z, if today decision is to wait. I assume that the cost of deployment have the following form:

$$C^l_{jt}(pop_c, TRI_z, u_{jzt}) = c^l_0 (c_\gamma)^{t-2009} pop_c^{c^l_{pop}} TRI_z^{c^l_{TRI}} u_{jzt} \tag{26}$$

where parameters c^l to be estimated. $u_{jzt} \overset{iid}{\sim} \log \mathcal{N}(-\sigma^2/2, \sigma^2)$. Upon non-investment,

[10] For more details see Appendix D.

Figure 8: Estimated value of wireless.



Notes: Total value is calculated as $-(\beta_i^{2G} + \beta_i^{3G} + \beta_i^{4G})/\alpha_i$. Distribution is taken with respect to GfK MRI survey. Shaded areas show 95% confidence intervals.

the expected continuation costs equal (omitting l):

$$C_{jzt'} = c_0 c_\gamma^{t'-2009} pop_c^{c_{pop}} TRI_z^{c_{TRI}} \sum_{t>t'} (\delta c_\gamma)^{t-t'} \mathbb{E}\left[u_{jzt}|\sigma_{jt}(\lambda_{zt})\right] \sigma_{jt}(\lambda_{zt}) \prod_{t''=t'+1}^{t-1} (1 - \sigma_{jt''}(\lambda_{zt''})) \tag{27}$$

Averaging over simulations I end up with EC_{jzt}.

Finally, as all the necessary quantities are simulated, the problem is within the Bajari et al. (2007) approach. Indeed, the best response in this problem can be described as a bang-bang strategy with switching threshold defined by equation:

$$E\Delta\Pi_{jzt} + C_{jt}(pop_c, TRI_z, u_{jzt}^*) = EC_{jzt} \tag{28}$$

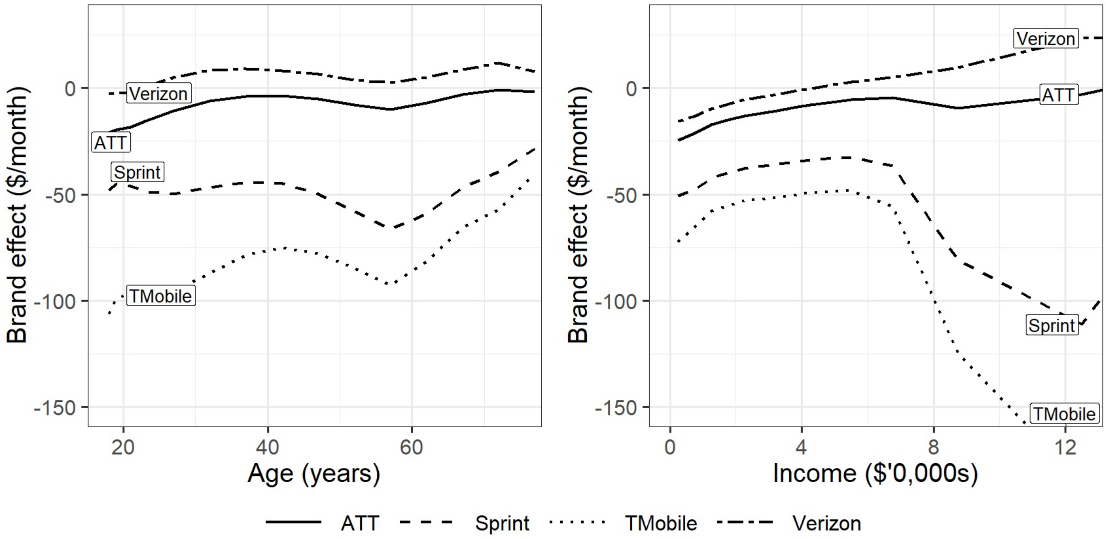
As u_{jzt}^* can be uniquely mapped to $BR_{jzt} = F_u(u_{jzt}^*)$, the rest of the problem is a direct optimization over parameters to match the best responses to the observed behavior as closely as possible.

7 Results

7.1 Demand and marginal cost estimation

Recall from Section 5 that this paper jointly estimates demand and marginal costs. As I allow for rich heterogeneity on the demand side, the demand is not easily summarizable. I thus compute various descriptive statistics based on the estimates, rather than try to fully describe the underlying functions. An example of full estimate function and corresponding t-statistics is shown in Figure 16 in Appendix B.

Figure 9: Estimated brand effects.



Notes: Fixed effects for different types of consumers by carrier. Plotted values are averages of $-FE_{iij}/\alpha_i$ over individuals with corresponding characteristics.

Figure 8 demonstrates the preference variability that the model generates when applied to the value of wireless technology when 4G is available. The left panel depicts the CDF of values in the population. The predicted value distribution is dispersed from \$20 to \$180 per month. The distribution has two modes at around \$70 and \$170 per month. The middle and right panels demonstrate how preferences vary across age and income dimensions. The effect of age is mild with two little dips between twenty to forty and sixty to 80 years. Income has a much stronger effect on the value of the technology for consumers. For lower incomes the value is around \$50 per month on average. The curve becomes steep at around \$70,000 per year income mark. Every additional \$10,000 in annual income increases the valuation by \$10-15 per month going as high as \$150 per month for the richest population.

Another major part of the demand system is the significant importance of "branding", defined as the value $FE_j(Z_i)$. As a particular firm targets certain groups of population, those groups have higher FE_j relative to some benchmark. This definition only reflect *relative* branding with respect to that benchmark. In the present case, a natural candidate for the benchmark is the outside good with fixed effects normalized to 0 for all consumers. Figure 9 shows how FE_j changes over age and income for different carriers. There is no consistent difference in branding over age cohorts, however, there is a clear segmentation over income dimension. As with the value of service, a significant change can be observed around \$70,000 per year with consumers above that threshold having much lower unobserved quality value for Sprint and T-Mobile as compared to AT&T and Verizon.

The demand model generates heterogeneous substitution patterns across individuals with different characteristics. To verify the degree of heterogeneity in demands, I calculate average implied elasticities for groups of population with low and high incomes.

Table 1: Implied elasticities for different types of consumers.

∂p_i (%)	∂s_j (%) ≤\$35,000/year				∂s_j (%) ≥\$85,000/year				
AT&T	-1.40	0.36	0.38	0.38	-0.40	0.14	0.15	0.14	AT&T
Sprint	0.12	-1.42	0.12	0.12	0.04	-0.47	0.04	0.04	Sprint
T-Mobile	0.10	0.10	-1.49	0.10	0.02	0.02	-0.48	0.02	T-Mobile
Verizon	0.57	0.54	0.57	-1.25	0.24	0.24	0.24	-0.32	Verizon

Notes: The comparison of the implied elasticities for the two groups of the U.S. population. Average implied elasticities for households with annual income of \$35,000 or less are shown on the left panel, and for those with annual income ob \$85,000 or above are shown on the right. The order of carriers in columns and rows is alphabetical, i.e., row/columns 1 through 4 correspond to AT&T, Sprint, T-Mobile, and Verizon respectively.

Table 1 summarizes the results. As those are the groups with reasonably homogeneous preferences within themselves, each of substitution matrices demonstrates logit-like elasticities. However, there is more than three-fold difference in price sensitivity between the groups in both own- and cross-price elasticities. As expected, richer people are less likely to switch the carriers in response to the same price change. The effect comes from two sources. First, richer people have higher value of quality to value of money ratio as we discussed above. Second, their optimal choices are different from the choices of less wealthy cohorts. Figure 14 and Figure 15 in Appendix B show side-to-side the predicted and observed shares of different carriers for income and age. Though the fit is not perfect due to smoothness restrictions imposed on the coefficients of the model as functions of individual characteristics, the overall trends and patterns of the data are mirrored by my demand model.

The marginal costs estimates come in two parts: the actual realizations of MC_{jt} obtained by inverting the price using optimal pricing condition, and values of the MC_{jt} equation parameters. Actual realizations of MC_{jt} are used in the counterfactual analysis to have a closer match to actually realized data. As there is only one path realization of MC_{jt} the use of realized shocks is a more suitable approach. However, to calculate equilibrium, one need to account for expectations that the firms form. Those expectations depend on structural parameters parameterizing MC_{jt} rather than actual realizations. Below I present results on both, parameters as well as imputed realized values.

I start with actual realizations as they serve as the departing point for determining underlying structural parameters. The underlying parameters are obtained through a corresponding regression of logs of realized MC_{jt} on the respective controls. Actual, or, more precisely, imputed, MC_{jt} are obtained as follows. I make use of a static Bertrand-Nash condition given by Equation 6 but now treat MC_{jt} as unknown, while prices and shares are plugged in at the actual values, while $\alpha(Z_i)$ is provided as an output of demand estimation. Rearranging terms leads to the following expressions on MC_{jt}:

$$\widehat{MC}_{jt} = \frac{\sum_i \frac{\partial \widehat{s}_{ijt}}{\partial p_{jt}} p_{jt} + \sum_i \widehat{s}_{ijt}}{\sum_i \frac{\partial \widehat{s}_{ijt}}{\partial p_{jt}}} = \left(\frac{1}{\widehat{\epsilon}_{jt}} + 1 \right) p_{jt} \tag{29}$$

where $\widehat{\epsilon}_{jt}$ is the estimate of the aggregate demand elasticity of good j in year t. The results obtained in this way are shown in Figure 17 in the Appendix B. Even absent of the merger the markups went up over time for all four carriers. In contrast, the marginal costs of providing wireless service declined consistently over time. Overall, from the static competition perspective the industry can be viewed as concentrated with markups being as high as 75 to almost 100 percent in most recent years. For the largest carriers the markups have been consistently above 70 percent over the whole analyzed period.

Moving to the expected costs part, the results are summarized in Table 2. T-Mobile's base level marginal costs remained the same after the acquisition of MetroPCS. This finding casts doubt on a story of operational costs efficiencies resulted from a merger. In contrast, it favors a story of the acquiring firm keeping its business practices and resulting firm inheriting marginal costs of that firm. Second, marginal costs go down quickly – annual decrease rate is around 30% demonstrating higher efficiencies companies developed over time. This can be attributed to improvement in many business practices over that period, such as automation and digitalization of support services, and some technology improvements, such as denser network within the same coverage. As an anecdotal evidence of close to zero marginal costs in later years, in June 2017 and in March 2019 Sprint offered a free year of service to any customer switching from their competitors. At the same time, the operational maintenance of 4G customers leads to higher costs. On average, switching the whole country to 4G leads to more than a tripling of marginal costs. Finally, leading firms, AT&T and Verizon, have substantially lower marginal costs, which is particularly important during the earlier time periods. These difference in costs reinforce their strong position in the industry. In later periods the absolute difference in costs between carriers declines, as overall marginal costs go down, eliminating this effect over time.

7.2 Investment costs

I now turn to summarizing costs of deployment. Estimates obtained from the data are shown in Table 3. The immediate interpretation for c_γ can be given as the rate of technological improvement. As the data suggests, the period of my analysis was characterized by a rapid technological improvements with average 30% annual decrease in the costs of deployment.

The parameters c_{pop} and c_{TRI} are the elasticities of costs with respect to population and TRI respectively. First, the assumption that TRI is a cost shifter, hence a relevant instrument, is confirmed by the data through my findings. Indeed, TRI shifts costs of deployment substantially for upgrade from either no deployment or 3G initial deploy-

Table 2: Estimation results for marginal costs evolution process.

Coefficient	Value	s.e.
κ_1^{MC}	-0.35^{***}	0.08
κ_2^{MC}	1.22^{**}	0.58
AT&T	-1.19^{***}	0.34
Sprint	-0.33	0.26
T-Mobile (2009-2013)	-0.79^{**}	0.32
T-Mobile (2014-2016)	-0.80^{*}	0.44
Verizon	-1.73^{***}	0.31
$\text{sd}(\omega_{jt}^{MC})$	0.51	

p-values: *=0.1, **=0.05, ***=0.01.

Table 3: Estimation results for investment problem.

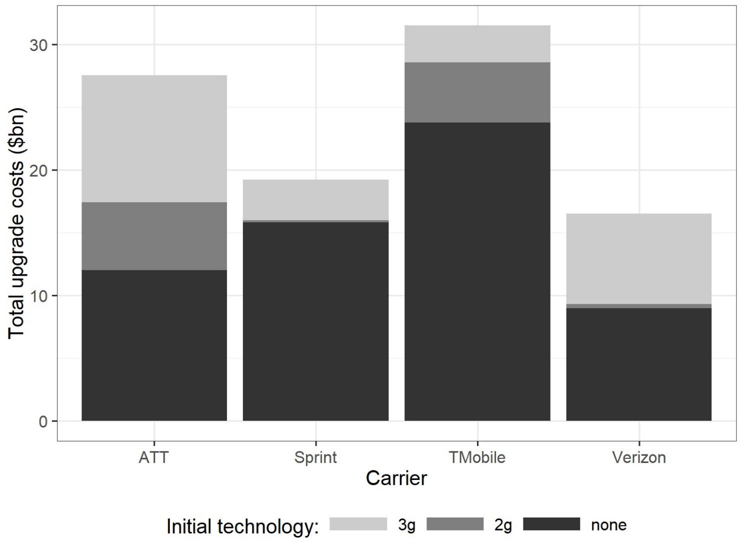
Coefficient	none	2G	3G
c_0	1.036	0.106	0.005
c_{pop}	0.955	0.430	0.166
c_{TRI}	0.417	0.001	0.205
c_γ		0.69	
σ		1.33	

ment. It does not have much effect on upgrade from 2G initial technology, however, this is a rare case in the data.

Second, it can be seen that elasticity of costs with respect to county population is decreasing as one moves from no to 3G initial deployment. The elasticity of upgrade from no initial deployment is close to 1. The conclusion is that a large part of investment that requires a particular area to be covered is essentially linear in population. The rural areas thus do not face *cost disadvantages* to get at least the initial 2G coverage. However, they often do not get covered by smaller carriers because of *return disadvantages* that carriers face by investing in those areas. Since in many cases only one or two carriers cover regions with low population density, it suggests that the costs are right around the NPV for larger carriers. Consequently, the merger of smaller carriers can potentially incentivize them to invest in rural markets.

The overall direction of the total investment agrees with the intuition that carriers with worse initial coverages have to invest more to get on par with the leading firm at the time, Verizon. Having the best network, Verizon did not have a lot of pressure to invest in

Figure 10: Total investments implied by the model.



Notes: Investments shown in $bn and grouped by carrier and initial technology.

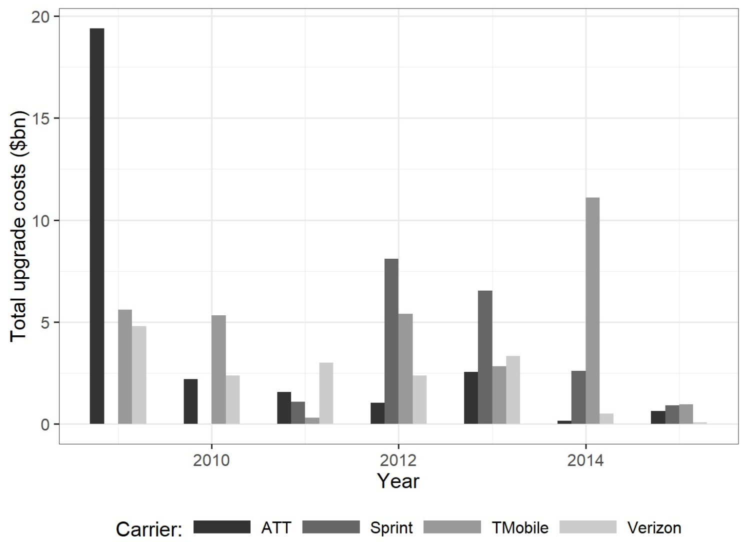
early periods, while AT&T and T-Mobile had to invest substantially early on to remain competitive. AT&T and Verizon are the largest investors in 3G-to-4G upgrade, while Sprint and T-Mobile are the leaders in none-to-4G deployments. The total deployment investment over time by all four firms is slightly above $90bn, which is in the bulk part of total CAPEX over the same period of around $190bn that includes other capital related costs besides coverage improvement.

Another important breakdown is the carrier-timing grouping of the investment costs. Figure 11 provides this breakdown. Verizon smooths out investments over multiple years, while AT&T's strategy of closing the gap involved substantial investment in 2009. Like Verizon, T-Mobile invested relatively uniformly over the period with a bump in 2014, a year following the acquisition of MetroPCS, being an additional anecdotal evidence of suboptimal market consolidation in the lower quality part of the market. Sprint generally struggled with investing over the whole period with most investment being done in 2012-2013, a few years later than the competitors. This struggle cost the company almost half of their initial market share.

Finally, I demonstrate how σ can be interpreted. In simple words, σ trades-off the immediate and the option value of investment decision. As long as investment has not been done, the firm can hope for a better deal in the future, foregoing some demand in the meantime. This being said, the shape of $\mathbb{E}[u|p]$, the expected cost share being paid for investment decision happening with probability p, depends on σ. Higher σ would mean cheaper investments for $p < 1$.[11] Figure 18 in the Appendix B shows this function. Interestingly, the function is convex over most part of the $[0, 1]$ region. The increase in costs is especially steep when going from probability 0.8 to probability 1. This 0.2 increase in deployment probability almost doubles the average costs, leading to largest

[11] At $p = 1$, $\mathbb{E}[u|p] = 1$ by construction.

Figure 11: Total investments implied by the model.



Notes: Investments shown in $bn and grouped by carrier and year.

markets being disproportionately costly. For lower part of distribution, that is below 0.5, the relationship is approximately linear.

8 Counterfactual simulations

8.1 Solution method

As the model has three different levels of aggregation and an 11-dimensional state space for dynamic investment decision, the use of exact methods can be problematic. Rather than using exact methods, I apply a combination of neural nets based reinforcement learning and evolutionary game theory to find an approximate equilibrium of the game. I now describe the steps needed to get counterfactual results.

As a starting point for the algorithm, I have a flexible parametrization for both strategies and value functions. I use neural nets to allow for sufficient flexibility. As the problem is non-stationary, I will have a collection of $\{\sigma(d_{jzt}, \lambda; \theta^{\sigma}_{jt})\}_{jt}$ and $\{V(d_{jzt}, \lambda; \theta^{V}_{jt})\}_{jt}$, i.e., the time-carrier specific strategies and value functions. Value is defined as the difference between NPV of all profits and costs of deployments followed by optimal investment policy when investment has not been done before t and NPV of all profits when it has. As I show in Appendix D, this is a correct renormalization of the value function for this particular problem, and a function defined in this manner is always non-positive. The initialization of the parameters θ^{σ}_{jt} and θ^{V}_{jt} is done by assigning small random values, which is a commonly used practice in machine learning. Another initialization I have to make is the initialization of strategies and values for some sufficiently large $\bar{t}$, so that the backward induction to update values and strategies can be applied. The assumption is that for some $\bar{t}-1$ the strategy will be to invest with probability 1. This uniquely defines value functions from $\bar{t}$ onwards. The assumption is innocuous as for $\bar{t}$ sufficiently large the deployment costs are essentially 0, meaning that investment is a strongly dominant strategy.

The algorithm sequentially updates parameters $\widehat{\theta}^{\sigma}_{jt,k}$ and $\widehat{\theta}^{V}_{jt,k}$. I now describe details for some fixed j and t while updating from $k-1$ to k and explain how to iterate over j and t later. Given $\{\sigma(d_{jzt}, \lambda_c; \theta^{\sigma}_{jt})\}_{jt}$, one can simulate the transition to the next period state, λ'_c, for any initial λ_c and distribution of $\{d_{jzt}\}_{z\in c,jt}$. Mean-field approximation for the county level state variable transition ensures that $T(\lambda'_c; \lambda_c)$ does not depend on action in any particular zip code, so it is completely exogenous for a particular decision on investment in z. I approximate the integral over this distribution by averaging over random sample of $\{\lambda'_{c,r}\}_{r=1}^{R}$ drawn according to $T(\lambda'_c; \lambda_c)$. If $d_{jzt} = 4G$, the value is 0 and a firm faces no choice, so the only case to consider is $d_{jzt} \neq 4G$. If $d_{jzt} \neq 4G$, the firm can either switch to 4G forever starting next period undertaking costs of $c_0^l (c_\gamma)^{t-2009} pop_c^{c^l_{pop}} TRI_z^{c^l_{TRI}} u_{jzt}$, or do nothing in the current period, facing no costs at current period but carrying d_{jzt} into next period, resulting in $V(d_{jzt}, \lambda'; \theta^V_{j,t+1})$ value. Denote the best response policy corresponding to current guess on strategies and value functions as $BR_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$, and the corresponding optimal value as $\Psi_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$. The optimality condition for deployment in period t then implies:

$$c_0^l (c_\gamma)^{t-2009} pop_c^{c^l_{pop}} TRI_z^{c^l_{TRI}} F_u^{-1}(BR_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})) = -\frac{\delta}{R} \sum_{r=1}^{R} V(d_{jzt}, \lambda'_{c,r}; \theta^V_{j,t+1}) \quad (30)$$

and:

$$\begin{aligned} \Psi_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k}) = -(p_{jt} - MC_{jt}) \left[\frac{\partial s_{jct}}{\partial q_{jct}} \Delta q_{jzt,t} \right] - \\ BR_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k}) \times \mathbb{E}\left[u | u < F_u^{-1}(BR_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})) \right] + \\ \left(1 - BR_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})\right) \times \delta \mathbb{E}[V(d_{jzt+1}, \lambda'; \theta^V_{j,t+1})], \end{aligned} \quad (31)$$

where $\Delta q_{jzt,t}$ is given by Equation 25. The first equation allows to obtain $BR_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$ in a closed form, while the second one allows to calculate for $\Psi_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$ by plugging $BR_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$ into it.

Having defined $BR_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$ and $\Psi_{jt}(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$, all necessary ingredients to update $\widehat{\theta}^{\sigma}_{t,k}$ and $\widehat{\theta}^{V}_{t,k}$ are now given. The pseudo-code for this updating algorithm is described below:

Algorithm 1: Update of $\widehat{\theta}^{\sigma}_{jt,k}$ and $\widehat{\theta}^{V}_{jt,k}$

Draw county c:
- `function` $BR(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$:
  - **Input** : Current technology d_{jzt} and county state λ_c
  - **Output:** Probability of investment
- `function` $\Psi(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$:
  - **Input** : Current technology d_{jzt} and county state λ_c
  - **Output:** Updated value corresponding to optimal probability of investment
- **for** z **in** $Zips_c$ **do**
  - $B_j = BR(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$
  - $P_j = \Psi(d_{jzt}, \lambda_c; \widehat{\theta}^{\sigma}_{t,k}, \widehat{\theta}^{V}_{t,k})$
  - $L_1(\theta^{\sigma}_{jt}) = (B_j - \sigma(d_{jzt}, \lambda_c; \theta^{\sigma}_{jt}))^2$
  - $L_2(\theta^{\sigma}_{jt}) = (P_j - V(d_{jzt}, \lambda_c; \theta^{V}_{jt}))^2$
  - $\widehat{\theta}^{\sigma}_{jt,k+1} := \widehat{\theta}^{\sigma}_{jt,k} + \epsilon_{\sigma} \nabla_{\theta^{\sigma}_{jt}} L_1(\widehat{\theta}^{\sigma}_{jt,k})$
  - $\widehat{\theta}^{V}_{jt,k+1} := \widehat{\theta}^{V}_{jt,k} + \epsilon_{V} \nabla_{\theta^{V}_{jt}} L_2(\widehat{\theta}^{V}_{jt,k})$
- **end**

In the pseudo-code above ϵ_σ and ϵ_V stand for learning rates for corresponding learning procedures. Before moving forward, I'd like to point out a couple details. First, the functions σ and V are, in general, *inconsistent* in the proposed algorithm. This is very different from other algorithms used for solving dynamic games. Though the updates for actions and values are done towards the values obtained by reinforced learning, it is only done in an incremental step. This is to avoid radical strategies changes in response to competitors also learning the game. Slow updating mitigates potential flip-flopping issues at the cost of two sets of functions not being fully consistent. Second, though the steps of updating strategies and values look similar, they are different in nature. Strategies update is essentially evolutionary dynamics, while value update is a straightforward value iteration step.

Once the updating algorithm for θ's is given, I can explain how the complete algorithm for finding approximate equilibrium in investment game works. Algorithm 1 will be a part of this larger algorithm.

Algorithm 2: Finding approximate equilibrium of the investment game

Initiate $\bar{t}$, $\widehat{\theta}^{\sigma}_{jt,0}$ and $\widehat{\theta}^{V}_{jt,0}$
while $k \leq K$ **do**
    $t := \bar{t}$
    while $t \geq 0$ **do**
        for j **in** ListOfCarriers **do**
            Algorithm 1 for $\widehat{\theta}^{\sigma}_{jt,k}$ and $\widehat{\theta}^{V}_{jt,k}$
        end
        $t := t - 1$
    end
end

The final piece for solving the full model is to find prices that are consistent with the investment game equilibrium and vice versa. This is done by iterating between Algorithm 2 and solving for prices satisfying Equation 6.

As the model may lead to an equilibrium that is slightly different from observed realization of the data for the factual simulations, I compare factual equilibrium to counterfactual equilibrium rather than comparing data to counterfactual equilibrium. The changes proposed by the model are then interpreted as estimates for actual effects one would have observed in the data if a merger happened. I conduct a comparison of actual data with factual equilibrium in Appendix C. As a main take away of the analysis, the benefits from merger part of my simulations can be viewed as lower bounds on actual benefits from merger.

8.2 Counterfactual 1: T-Mobile acquires Sprint in 2009

The first counterfactual analyzes economic consequences of a recently proposed merger between T-Mobile and Sprint had it happened in 2009. I evaluate the merger on several dimensions: (i) change in prices and profits, (ii) change in deployments and investments, (iii) change in surplus and distribution of this surplus.

I start with the effects the merger would have on deployments. Figure 19 in Appendix B demonstrates that in equilibrium the investment incentives of the non-merging parties are barely affected, leading to almost identical roll-out plans. In contrast, the merger would have substantial effects on non-merging parties' incentives. The merged company has a roll-out strategy that is almost identical to the strategy of AT&T. This effect intensifies competition in the segment that cares about quality, benefiting consumers with either high tastes for quality or those that do not have a viable alternative. On the other hand, for customers who do not care about quality or fortunately live in the areas that have a good coverage in the absence of merger, the same merger leads to a loss of a cheap option.

Now I move to the effect on average price paid by consumers which is shown in Figure 20 in the Appendix B. In early periods, the prices are higher under the approved merger since the concentration in the industry increases immediately, while it takes

Table 4: Surplus changes under T-Mobile/Sprint merger.

	No merger ($bn)	Merger ($bn)	Change ($bn)	Change (%)
Consumer surplus	2,579.4	2,594.5	15.1	0.6
Profits	1,007.1	1,012.8	5.7	0.6
Profits (non-merging)	874.4	870.4	-4.0	-0.5
Investment	80.0	76.5	-3.5	-4.4
Average investment	20.0	25.5	5.5	27.5
Total surplus	3,506.5	3.530.8	24.3	0.7

time for the firms to fully implement their deployments. Between years 2011 and 2016, higher competition in the high quality segment of the market results in *lower* average paid prices. After 2016 the prices under both scenarios are virtually the same.

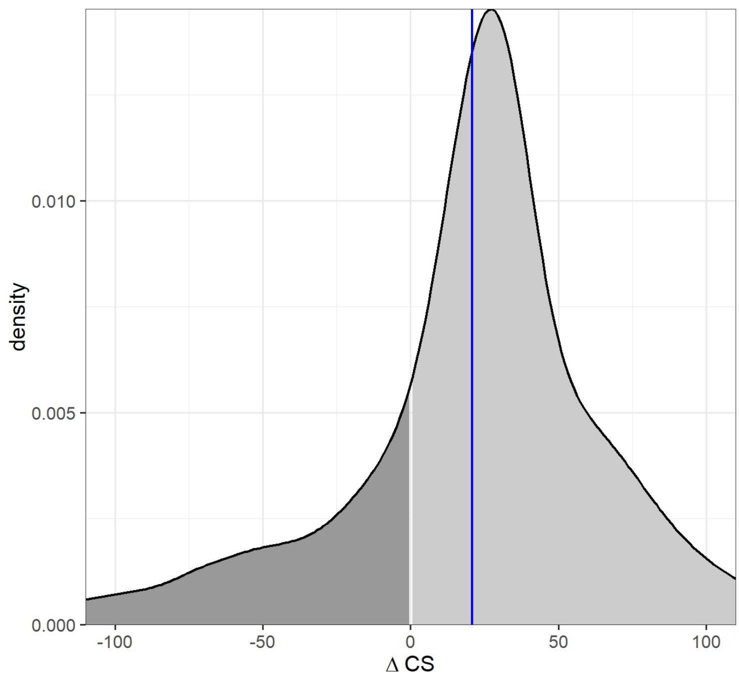
Mobile penetration mimics the behavior of the prices and can be seen in Figure 21 in the Appendix B. In early years, penetration goes down as the quality has not yet seen any substantial changes while the prices have increased due to higher industry concentration. However, as time passes, both, prices and quality start to move in the same direction, leading to higher mobile penetration. In the very late periods, as the average price and deployments are getting close in the presence and in the absence of the merger, the gap in penetration rates also closes.

Table 4 summarizes the results of the counterfactual simulation and provides a breakdown of welfare changes for different market participants. All surpluses go up by 0.6-0.7% which results in $15.1bn, $5.7bn, and $24.3bn for consumer surplus, profits, and total surplus respectively. Increased competitive pressure in the high quality segment causes profits of the non-merging firms to decrease by $4bn, or 0.5%. The merger also results in substantial elimination of duplicative investment, as the total investment decreases by 4.4%, or $3.5bn. The quality of investment, on the other hand improves significantly. As the merged firm invests on par with two others, the average per firm investment goes up by more than 25%.

Finally, I examine the distribution of consumer surplus over different areas. As both demand and investment decisions are highly heterogeneous, there is a high dispersion in the distribution of change in consumer surplus, which is shown in Figure 12. Though the average and modal change in consumer surplus is around $20 in the life-time monetary metric, the range of values varies from -$200 to $200. Those differences do not occur at random, but are rather driven by underlying structural parameters, compositions of demand, and county characteristics. It is thus important to understand systematic differences between counties to evaluate potential winners and losers under the merger scenario.

To understand what affects the size of change in consumer welfare, I run a linear regression with change in consumer surplus on the left hand side and the following set

Figure 12: Distribution of changes in consumer surplus over different geographic areas.



Notes: The measure of consumer surplus is the total net present value in U.S. dollars for an average consumer in the county. The difference in consumer surplus between counterfactual and factual equilibria is plotted along the x-axis. The blue line is the average (over counties) change in per-person consumer surplus.

Table 5: Estimation results for distribution of change in consumer surplus by county.

Coefficient	**Value**	**s.e.**
log of population	-0.94	1.10
TRI	0.55^{***}	0.20
Initial AT&T quality	6.98^{**}	2.80
Initial Verizon quality	10.94^{***}	2.14
Initial quality of Sprint+T-Mobile	-27.95^{***}	3.40
median income in the county	-7.48^{***}	1.84
median age in the county	-5.39^{***}	1.13
# of obs.	2,481	
R^2	0.06	

p-values: *=0.1, **=0.05, ***=0.01.

of controls: log of population, TRI, initial deployment of AT&T, initial deployment of Verizon, average initial deployment of merging parties, median income in the county, and median age in the county. The output for this regression is in Table 5.

Population is not that important for the change in consumer surplus, however, initial deployment of the parties is extremely important. As this deployment is itself endogenous, population is negatively correlated with change in consumer surplus, making rural areas major winners from the merger. Indeed, rural and/or rugged areas typically face lower initial presence of one or both of the merging firms. This lower presence leads to substantial gains received under merger scenario.

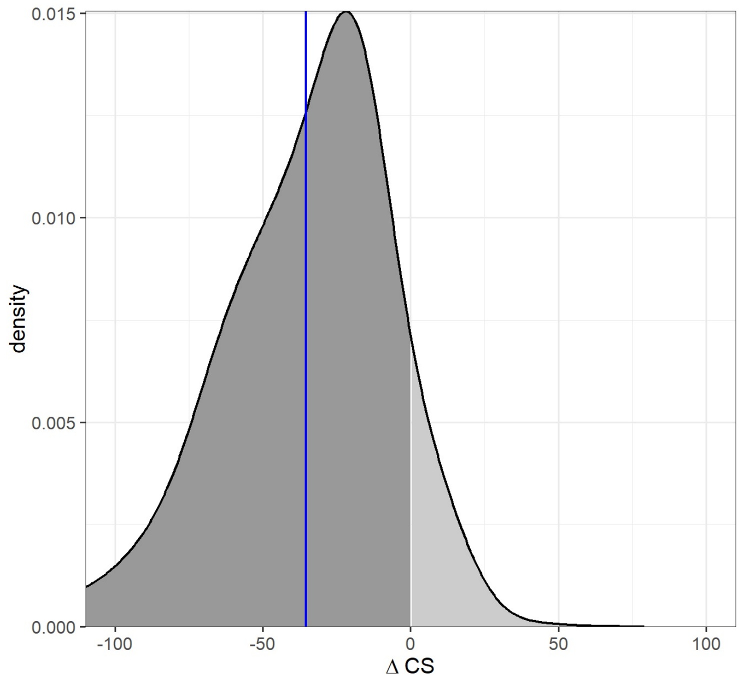
Summarizing the results, they align with the rationale provided by FCC and DoJ in their merger revision – the positive impact of closing the quality gap would result in tougher competition and welfare increase for large share of the consumers. In addition, the push back by some groups could be explained by substantial heterogeneity in the effects within the proposed framework.

8.3 Counterfactual 2: AT&T acquires T-Mobile in 2009.

Having analyzed T-Mobile/Sprint merger, I now analyze merger that was not approved by the FCC – the AT&T/T-Mobile merger. Like before, I run a counterfactual simulation for the situation if the merger had happened in 2009. Unlike in T-Mobile/Sprint case, the vast majority of consumers have negative change in consumer surplus that is followed by the merger as shown in Figure 13.

A similar decomposition of change in consumer surpluses demonstrates that most of consumers are hurt on both directions, quality and price. Figure 22 shows that the

Figure 13: Distribution of changes in consumer surplus over different geographic areas.



Notes: The measure of consumer surplus is the total net present value in U.S. dollars for an average consumer in the county. The difference in consumer surplus between counterfactual and factual equilibria is plotted along the x-axis. The blue line is the average (over counties) change in per-person consumer surplus.

merger would have little impact on any of the market participants. The increase in market concentration without substantial quality bump leads to higher average price paid, as shown in Figure 23. The effects lead to some decline in mobile penetration as shown in Figure 24. All figures can be found in Appendix B.

The decrease in consumer surplus occurs through multiple channels. First, acquisition of T-Mobile by AT&T removes one of the competitors from the market. This leads to upward pressure on the equilibrium prices in both high and low quality segments which hurts every consumer that has previously chosen any but the outside good. Second, there is no increase in investments. Since the initial incentives to invest for AT&T were high without the merger, and the share of T-Mobile is considerably smaller, the merger creates almost no extra incentives for a resulting firm compared to the benchmark. Finally, some subscribers, that are pushed away from the market because they lose their only viable option. These consumers have high price sensitivity and/or low utility from wireless service. However, they may still choose T-Mobile as a cheap option if the coverage in a local area is good enough. The same consumers may find Sprint inappropriate option because of either relatively poor local quality or a lower taste shock for Sprint. Losing a cheap carrier with appropriate quality pushes those customers to an outside option. Combination of these effects have unambiguously negative implications for the consumer welfare according to the model.

9 Conclusion

This paper evaluated the effects of several recently-proposed mergers in the U.S. wireless telecommunications industry. By constructing a model that allows for endogenous investment and pricing decisions, I can account for two competing effects in mergers: static price competition and long term returns on investments. The main finding of the paper is that the merger counterfactual outcomes align with the FCC's decisions on two of the largest wireless telecommunications mergers proposed in the last decade — AT&T/T-Mobile and T-Mobile/Sprint. The AT&T/T-Mobile merger would have harmed consumers by around $30bn, while the T-Mobile/Sprint merger would benefit them by approximately $15bn.

The applicability of the model proposed in the paper, as well as the techniques used in the model's estimation and solution, are not limited to either the U.S. telecommunications market or wireless telecommunications more generally. The modeling framework can be used in other industries where endogenous quality is a first-order concern. Estimation techniques that flexibly account for consumer heterogeneity can also find its' place in various economic applications.

There are a few potential opportunities for further research. First, the paper simplifies the pricing game to the choice of a single variable, ARPU. As contracts in the industry are non-linear, it may be worthwhile to explore the effects the mergers would have had on different types of contracts. Second, the paper leaves out the wholesale market. Though this market is not as relevant for the given time period, it has grown in recent years. The dynamic effects on the size and structure of this market, as well as the implications for consumers, may be obtained with more recent and detailed data. Finally, the analysis of switching costs and their evolution over time may be important for the wireless industry. However, one would need panel data rather than a repeated sample to achieve this goal.

References

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728, 2005.

Aguiar, Luis and Joel Waldfogel. Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music. *Journal of Political Economy*, 126(2):492–524, 2018.

Arrow, Kenneth Joseph. Economic welfare and the allocation of resources for invention. In *Readings in industrial economics*, pages 219–236. Springer, 1972.

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

Berry, Steven, James Levinsohn, and Ariel Pakes. Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of political Economy*, 112(1):68–105, 2004.

Berry, Steven T. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262, 1994.

Björkegren, Daniel. Competition in network industries: Evidence from the rwandan mobile phone network. *Available at SSRN 3049364*, 2018.

Compiani, Giovanni. Nonparametric demand estimation in differentiated products markets. *Available at SSRN 3134152*, 2018.

Crawford, Gregory S and Ali Yurukoglu. The welfare effects of bundling in multichannel television markets. *American Economic Review*, 102(2):643–85, 2012.

Doudchenko, Nick and Evgeni Drynkin. Micro-level BLP: A Machine Learning Approach. *working paper*, 2019.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. How well targeted are soda taxes? 2017.

FCC, . Mobile wireless competition report (20th annual), 2017.

Gillen, Benjamin J, Matthew Shum, and Hyungsik Roger Moon. Demand estimation with high-dimensional product characteristics. In *Bayesian Model Comparison*, pages 301–323. Emerald Group Publishing Limited, 2014.

Gillen, Benjamin J, Sergio Montero, Hyungsik Roger Moon, and Matthew Shum. Blp-lasso for aggregate discrete choice models of elections with rich demographic covariates. *USC-INET Research Paper*, (15-27), 2015.

Goettler, Ronald L and Brett R Gordon. Does amd spur intel to innovate more? *Journal of Political Economy*, 119(6):1141–1200, 2011.

Griffith, Rachel, Martin O'Connell, and Kate Smith. Tax design in the alcohol market. *Journal of Public Economics*, 172:20–35, 2019.

Krusell, Per and Anthony A. Jr. Smith. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896, 1998.

Nunn, Nathan and Diego Puga. Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36, 2012.

Petrin, Amil and Boyoung Seo. Identification and estimation of discrete choice demand models when observed and unobserved characteristics are correlated. *working paper*, 2016.

Schumpeter, Joseph A. Capitalism, socialism, and democracy. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*, 1942.

Shapiro, Carl. *Competition and Innovation: Did Arrow Hit the Bull's Eye?*, pages 361–404. University of Chicago Press, April 2011.

Weintraub, Gabriel Y, Lanier Benkard, and Benjamin Van Roy. Oblivious equilibrium: A mean field approximation for large-scale dynamic games. In *Advances in neural information processing systems*, pages 1489–1496, 2006.

Wollmann, Thomas G. Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, 108(6):1364–1406, 2018.

A Details on demand estimation technique

A.1 Model

Consider the following discrete choice setting. There are M separate markets populated by N_m individuals for $m = 1, \dots, M$. Each individual $i = 1, \dots, N_m$ is characterized by a set of observable covariates $Z_{im} \in \mathbb{R}^p$ and her product choice $d_{im} \in \{0, 1, \dots, J\}$, where J is the number of products available in each market and $d_{im} = 0$ corresponds to the outside good. Product $j = 1, \dots, J$ is characterized by an observable k-dimensional variable X_{jm} and an unobservable $\xi_{jm} \in \mathbb{R}$. We also assume the existence of a set of instimental variables $W_{jm} \in \mathbb{R}^l$ that are uncorrelated with ξ_{jm}. Utility u_{ijm} derived by individual i from buying good $j = 1, \dots, J$ in market m is given by

$$u_{ijm} \quad = \quad \beta(Z_{im})^T X_{jm} - \alpha P_{jm} + \xi_{jm} + \varepsilon_{ijm},$$

where P_{jm} is the price of product j in market m and ε_{ijm} is an idiosyncratic error distibuted according to a Generalized Extreme Value Type-I (Gumbel) distribution. The utility of the outside good is $u_{i0m} = \varepsilon_{i0m}$.

We assume that each person chooses the good that provides the highest level of utility in which case the probability that individual i chooses good $j = 1, \dots, J$ in market m is:

$$s_{jm}(Z_{im}) \quad = \quad \frac{\exp \beta(Z_{im})^T X_{jm} - \alpha P_{jm} + \xi_{jm}}{1 + \sum_{j'} \exp \beta(Z_{im})^T X_{j'm} - \alpha P_{j'm} + \xi_{j'm}}.$$

Averaging across individuals we obtain the market shares,

$$s_{jm} \quad = \quad \frac{1}{N_m} \sum_{i=1}^{N_m} \frac{\exp \beta(Z_{im})^T X_j - \alpha P_{jm} + \xi_{jm}}{1 + \sum_{j'} \exp \beta(Z_{im})^T X_{j'} - \alpha P_{j'm} + \xi_{j'm}}.$$

A.2 Estimation

The method we use in this paper is based on the following idea. If in each market the market shares for a given $Z = Z_0$ were observed, the problem would reduce to a simple discrete choice estimation. Indeed,

$$s_{jm}(Z_0) \quad = \quad \frac{\exp \beta(Z_0)^T X_{jm} - \alpha P_{jm} + \xi_{jm}}{1 + \sum_{j'} \exp \beta(Z_0)^T X_{j'm} - \alpha P_{j'm} + \xi_{j'm}}.$$

We can let $\beta = \beta(Z_0)$ and obtain the estimates $(\widehat{\alpha}, \widehat{\beta})$ using steps 2–3 from the algorithm described above. As $s_{jm}(Z_0)$ are unobserved, we attempt to estimate them using the individual level data, (Z_{im}, d_{im}).

We use the following algorithm:

1. Use a prediction method of choice to fit $s(z, j, m) = \mathbb{P}\Big(d_{im} = j | Z_{im} = z\Big)$.

2. Pick Z_0 within the support of Z.

3. Predict $\widehat{s}(Z_0, j, m)$ for every $j = 1, \ldots, J$ and $m = 1, \ldots, M$.

4. Initialize the parameters to be estimated: $(\alpha, \beta) = (\alpha_0, \beta_0)$.

5. Iterate (α, β) until convergence. At step n:

   (a) Solve for ξ_{jm}, $j = 1, \ldots, J$, $m = 1, \ldots, M$ by inverting the predicted market shares—find $\widehat{\xi}_{jm}$ such that the implied market shares $s_{jm}(\alpha_n, \beta_n, \widehat{\xi}_{jm})$ coincide with the predicted market shares $\widehat{s}_{jm}$.

   (b) Compute the cost function $L(\alpha_n, \beta_n, \widehat{\xi}_{jm})$.

   (c) Update the parameters $(\alpha, \beta) = (\alpha_{n+1}, \beta_{n+1})$.

A.3 Adding aggregate moment

With an aggregate moment the full estimation problem can be written as:

$$\mathbb{E}_W[m(\theta(Z); s(Z), W)] = 0, \quad \forall Z \tag{32}$$

$$\mathbb{E}_{(W,Z)}[g(\gamma(Z); s(Z), W)] = 0 \tag{33}$$

where $m(\cdot)$ are the individual moments, and $g(\cdot)$ is the aggregate moment. We assume that $\gamma \subset \theta$, that is $\theta = (\gamma, \gamma')$ with $\gamma(Z)$ being of much smaller dimension than $\theta(Z)$. In demand estimation setup, elasticities only depend on parameters through α and s, i.e., $\alpha(Z)$ is one-dimensional no matter how large is $\theta(Z)$. To handle the problem, we first fix some finite base set $\mathcal{Z} = \{Z_1, \ldots, Z_K\}$ and some interpolation method providing us $f(Z)$ for any Z once we have values $f_1, \ldots, f_K$ for $\mathcal{Z}$. We can then approximate the problem above as:

$$\mathbb{E}_W[m(\theta(Z); s(Z), W)] = 0, \quad \forall Z \in \mathcal{Z} \tag{34}$$

$$\mathbb{E}_{(W,Z)}[g(\gamma(Z); s(Z), W)] = 0 \tag{35}$$

Next, we assume that the loss function has the form:

$$\mathcal{L}(\theta, \gamma) = \frac{1}{K} \sum_{k=1}^{K} \bar{m}_k' R_k \bar{m}_k + \underbrace{h}_{\text{relative weight}} \bar{g}' R_0 \bar{g} \tag{36}$$

where $h > 0$ is the constant allowing to scale up and down the importance of the aggregate moment. The key observation that allows us to greatly simplify the optimization is that the gradient is only hard to compute for $\alpha(Z)$ responsible for only a small fraction of the parameters. Differentiating w.r.t. γ' we obtain:

$$\frac{\partial \mathcal{L}}{\partial \gamma_k'} = \frac{2}{K} \frac{\partial \bar{m}_k'}{\partial \gamma_k'} R_k \bar{m}_k \tag{37}$$

which is the same exact form as the one for the problem without aggregate moment. Note that it only depends on single k and small number of parameters, θ_k, as $\bar{m}_k$ only depends on θ_k. The last piece is to differentiate the loss function w.r.t. γ. This can be done numerically, as the number of parameters in γ is small. After obtaining $\frac{\partial \mathcal{L}}{\partial \theta}$, one can use gradient descent method (with any machine learning stabilization technique, such as ADAM, RMSprop, etc) to find optimal values of parameters.

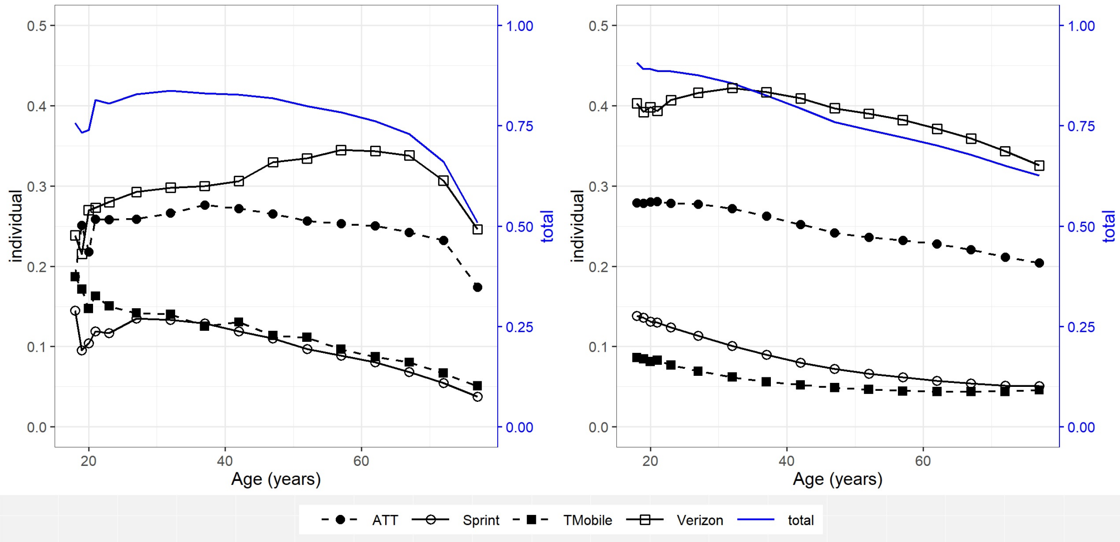
B Extra figures and tables

Table 6: Reduced form regressions results.

	OLS				IV			
					First stage		Second stage	
	$\log(s_{jmt}) - \log(s_{0mt})$		$\log(s_{jmt}) - \log(s_{0mt})$		$4G$		$\log(s_{jmt}) - \log(s_{0mt})$	
Coefficient	**Value**	**s.e.**	**Value**	**s.e.**	**Value**	**s.e.**	**Value**	**s.e.**
2G	1.05^{***}	0.07	-0.55^{***}	0.06	-0.79^{***}	0.01	0.26^{*}	0.14
3G	1.72^{***}	0.05	0.25^{***}	0.04	-0.73^{***}	0.01	1.01^{***}	0.13
4G	2.03^{***}	0.05	–		–		1.03^{***}	0.17
TRI	–		-0.015^{***}	0.002	-0.014^{***}	0.000	–	
ARPU	-0.02^{***}	0.005	-0.008	0.005	0.006^{***}	0.001	-0.014^{***}	0.005
log(ads)	0.09	0.07	0.01	0.07	-0.04^{***}	0.01	0.05	0.07
FE_j	Yes		Yes		Yes		Yes	
FE_t	Yes		Yes		Yes		Yes	
county controls	Yes		Yes		Yes		Yes	
# of obs.	22,851		22,851		22,851		22,851	
R^2	0.43		0.38		0.79		0.42	

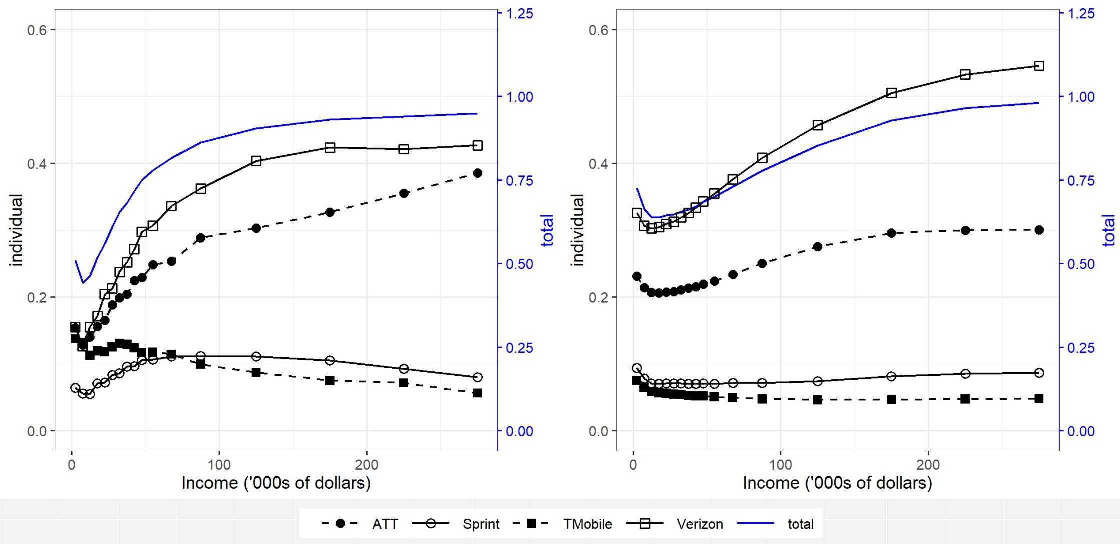
Notes: The results of reduced form estimation. A data point is a county-carrier-year combination. County controls are mean age and income in the county. R^2 is adjusted to the number of degrees of freedom. **p-values**: *=0.1, **=0.05, ***=0.01.

Figure 14: Comparison of observed and predicted shares by age.



Notes: Based on GfK MRI survey data. Left panel plots actual data, and right panel plots the data predicted by the model. Left y-axis and black lines show shares of individual carriers. Right y-axis and blue line show total penetration rate (sum of four shares). Age is the age of the head of household.

Figure 15: Comparison of observed and predicted shares by income.



Notes: Based on GfK MRI survey data. Left panel plots actual data, and right panel plots the data predicted by the model. Left y-axis and black lines show shares of individual carriers. Right y-axis and blue line show total penetration rate (sum of four shares). Income stands for income of a household.

Figure 16: Estimated function $\alpha(Z)$ and corresponding t-statistics.

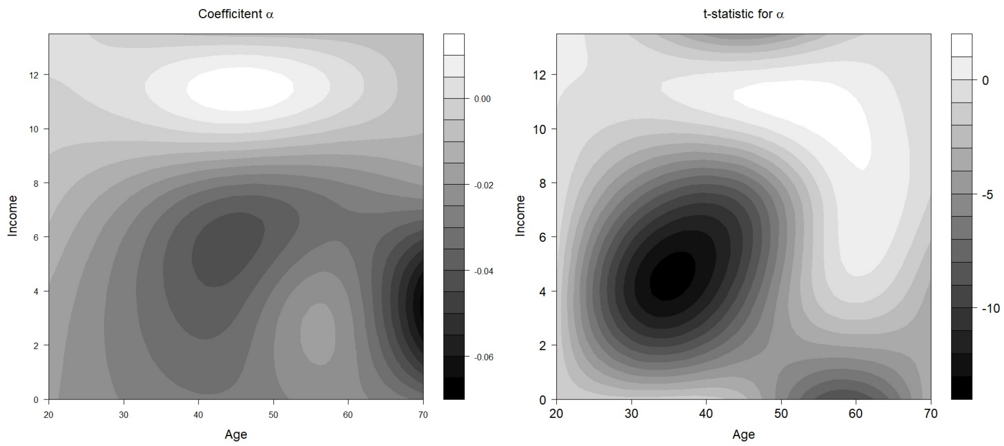
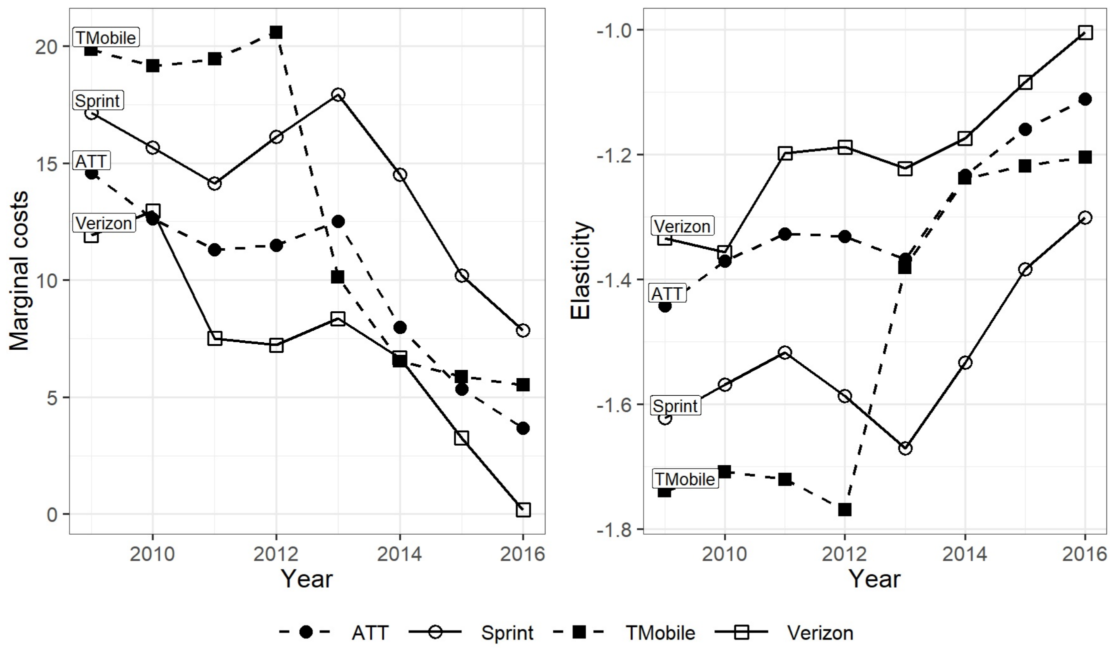
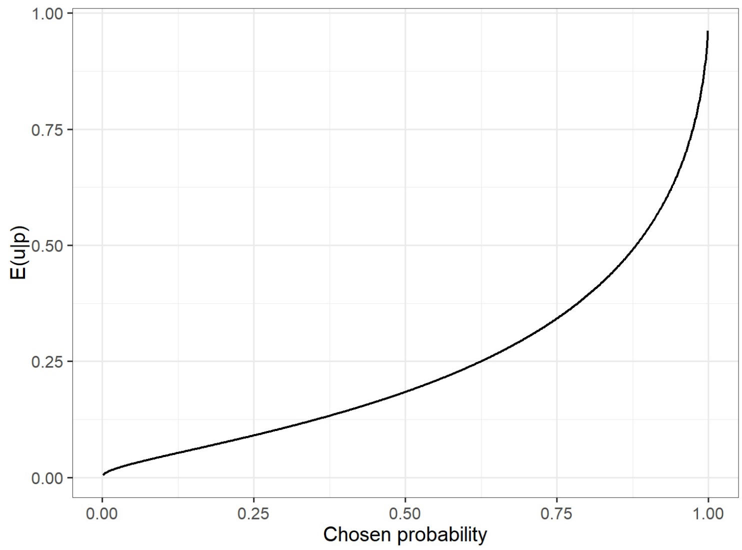


Figure 17: Dynamics of estimated marginal costs and own price elasticities.



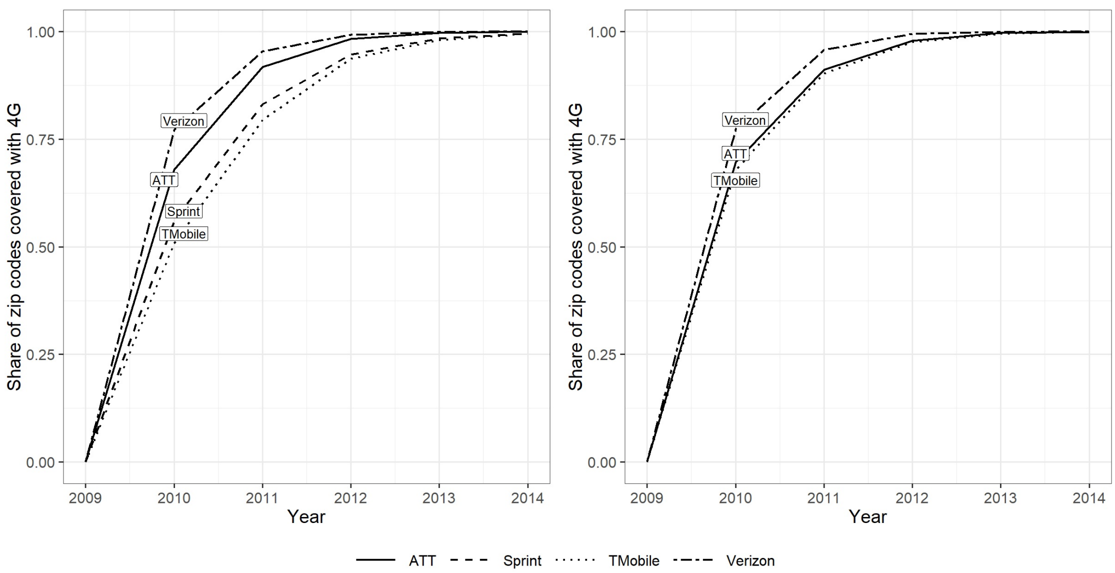
Notes: Imputed values of elasticities (left panel) and marginal costs (right panel) by carrier and year. Marginal costs are reported on $ per month basis.

Figure 18: $\mathbb{E}[u|p]$ as a function of p.



Notes: $E(u|p)$ denotes the expectation of cost shock under optimal investment if the chosen probability level is p.

Figure 19: Factual and counterfactual equilibrium 4G deployment levels.



Notes: Factual dynamics is shown in the left panel and counterfactual dynamics is shown on the right panel. y-axis is the share of zip codes covered by 4G technology for a given carrier in a particular year.

Figure 20: Factual and counterfactual equilibrium ARPU for the whole industry.

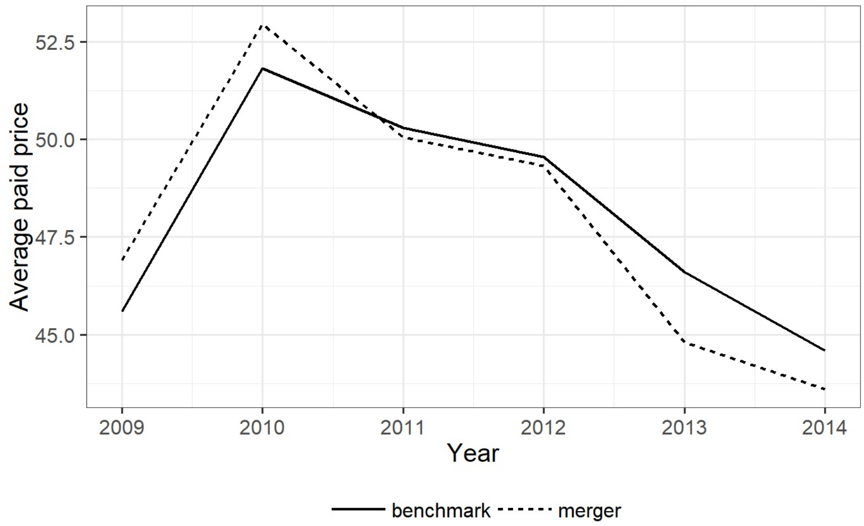


Figure 21: Factual and counterfactual mobile penetration.

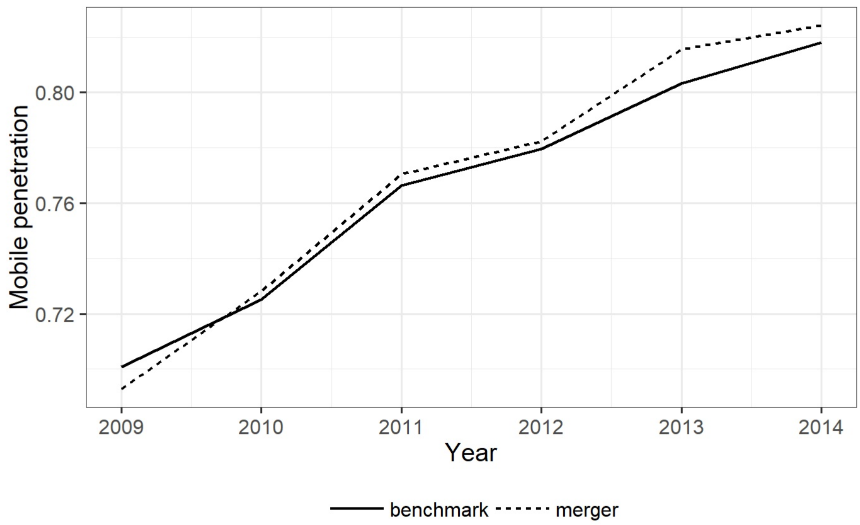
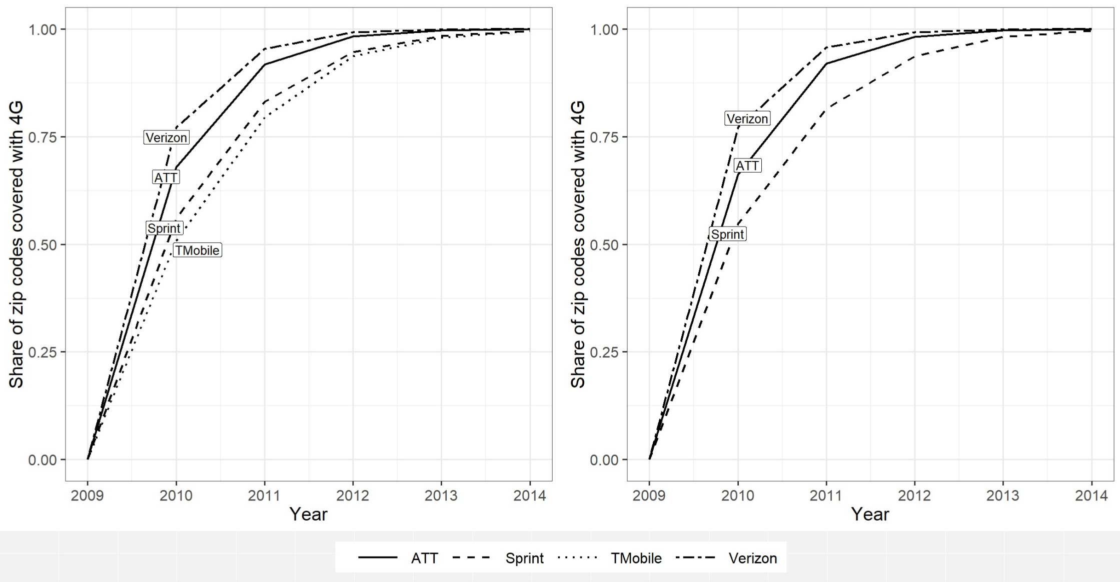


Figure 22: Factual and counterfactual equilibrium 4G deployment levels.



Notes: Factual dynamics is shown in the left panel and counterfactual dynamics is shown on the right panel. y-axis is the share of zip codes covered by 4G technology for a given carrier in a particular year.

Figure 23: Factual and counterfactual equilibrium ARPU for the whole industry.

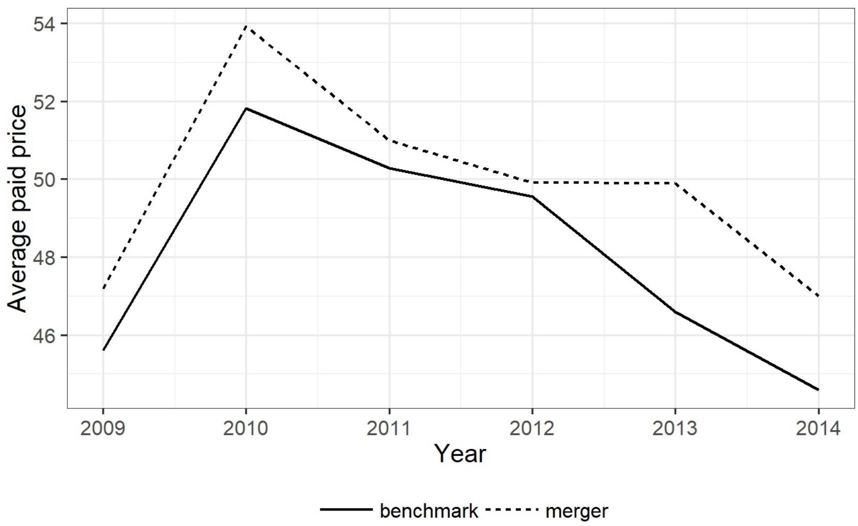
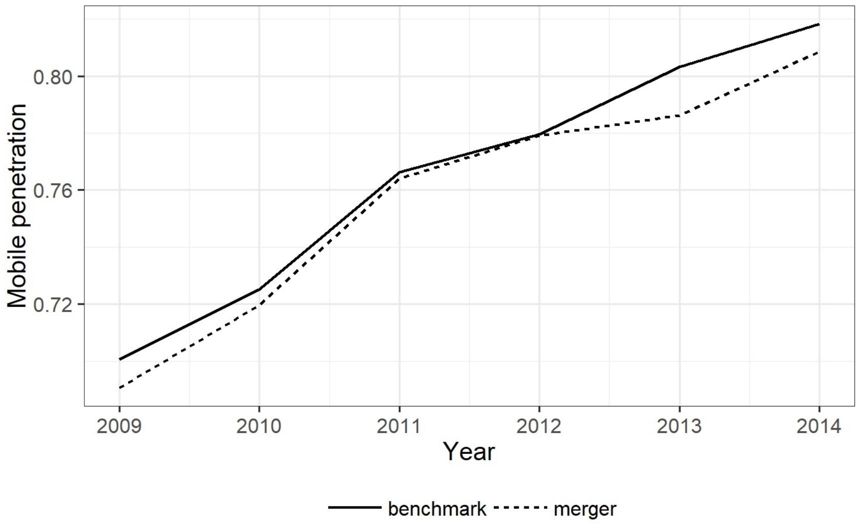


Figure 24: Factual and counterfactual mobile penetration.



C Comparison of actual data and factual equilibrium

We concentrate on three major measures of comparison between the model equilibrium and actual data: (i) deployment, (ii) penetration, and (iii) prices. We start with deployments as those define qualities, and all other equilibrium objects in the model. A comparison between model factual and actual deployments can be seen in Figure 25. Note that the actual data is different in a couple respects. First, Sprint has delayed the upgrade of its' network to 4G technology until 2011. This can be clearly seen in the data, yet the equilibrium we calculate does not account for this fact. Slower deployment creates less competitive pressure on other market participants, especially the one that has highest cross elasticity with Sprint, T-Mobile, leading to an overall slower 4G roll out. As a result, our counterfactual exercise can be viewed as the lower bound on potential gains. Second, in the model all carriers reach full 4G deployment by 2014, while in the actual data only AT&T and Verizon do so. This can be explained by the model being too lenient in explaining how expensive it is to cover some areas in the extreme of TRI and population distribution. Extremely rural and/or rugged areas thus may experience even stronger effects than our counterfactual analysis would suggest, resulting in our findings being even a more conservative lower bound.

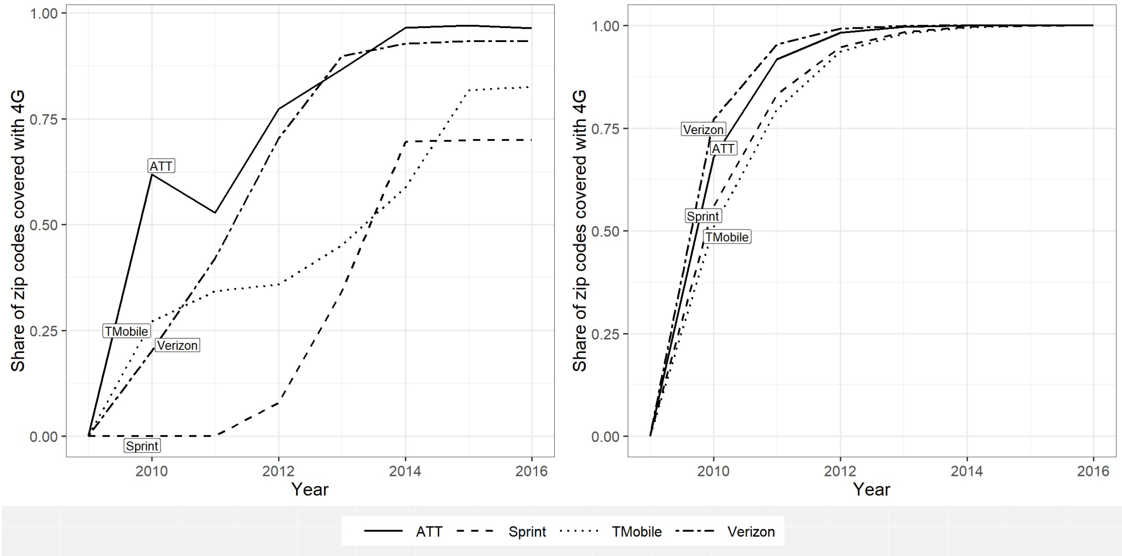


Figure 25: Actual (left panel) and factual equilibrium (right panel) 4G deployment levels. y-axis is the share of zip codes covered by 4G technology for a given carrier in a particular year.

Those differences between actual and equilibrium deployment decisions result in some differences in other market performance variables. Figure 26 shows model and actual average paid price dynamics. The overall shapes of the curves agree in the sense that the prices are slightly growing between years 2009 and 2012, while they start to decline after that period. As quality of service is better in the model equilibrium scenario, the prices are slightly higher in early periods. Starting 2013, however, the model predicted prices are slightly lower as the effective degree of competition becomes stronger than in actual data.

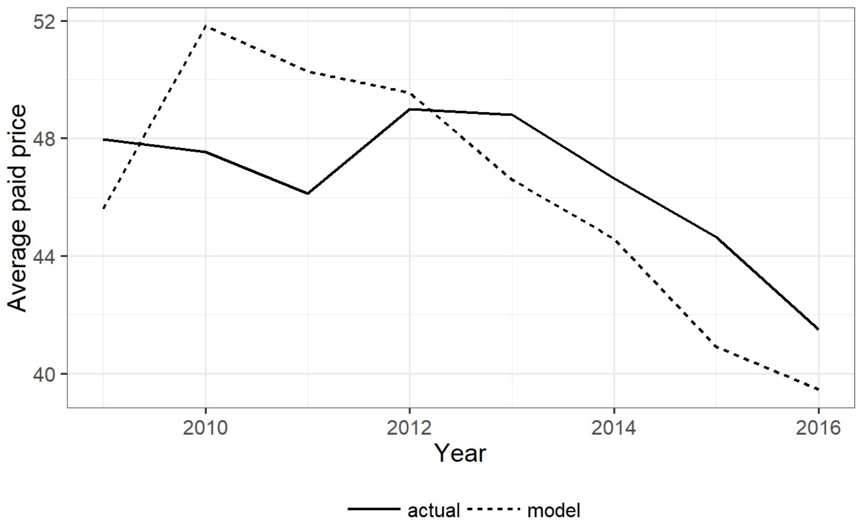


Figure 26: Actual and factual equilibrium ARPU for the whole industry.

The counterplay between prices and deployments also affect mobile penetration. Figure 27 demonstrates this dynamics. As qualities are higher and prices lower in the model equilibrium than in the actual data, mobile penetration rate is inevitably higher. The major divergence happens after 2012 and accounts for the fact that model is too optimistic on quality provided by Sprint and T-Mobile under equilibrium scenario.

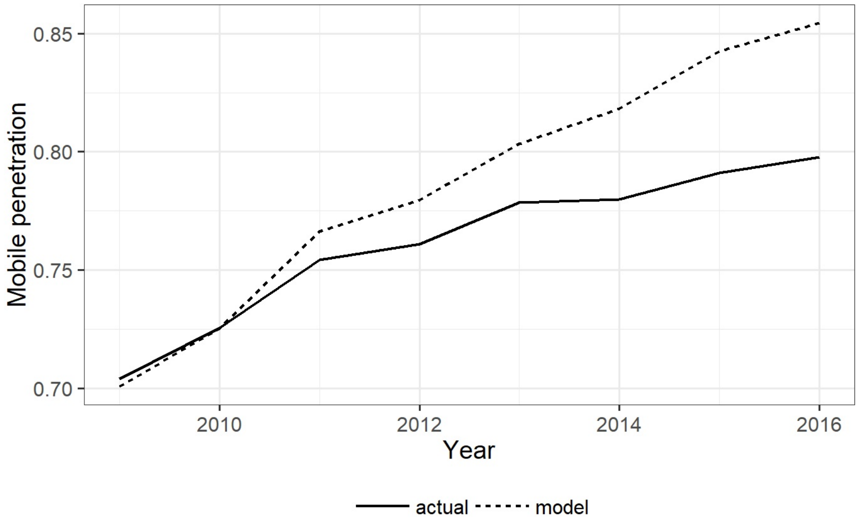


Figure 27: Actual and factual equilibrium choosing one of the four available options rather than an outside option.

Summing up data comparison between actual data and the one generated as equilibrium of factual model, we can conclude that the main difference comes from the fact that the model is more generous on *pre-merging investment incentives* between the weaker firms, Sprint and T-Mobile. As a result, the findings of our counterfactual analysis are likely to bee conservative, demonstrating the *lower bound* on the consequences of a potential merger.

D Extra proofs

D.1 Correctness of renormalization of the value function

From the definition of V^1 and V^0:

$$V^1_{jt}(\lambda) = \pi_{jt}(\lambda) + \delta \int V^1_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' \tag{38}$$

$$V^0_{jt}(\lambda) = \max_p \left\{ -pc_{jzt}\mathbb{E}[u|p] + \delta \left(p \int V^1_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' + (1-p)\int V^0_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' \right) \right\} \tag{39}$$

The second equation can be rearranged as:

$$V^0_{jt}(\lambda) = \max_p \left\{ -pc_{jzt}\mathbb{E}[u|p] - \delta p \int \Delta V_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' \right\} + \delta \int V^0_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' \tag{40}$$

where $\Delta V = V^0 - V^1$. Substituting equation for V^1 from this expression:

$$\Delta V_{jt}(\lambda) = \max_p \left\{ -pc_{jzt}\mathbb{E}[u|p] - \delta p \int \Delta V_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' \right\}$$
$$- \pi_{jt}(\lambda) + \delta \int \Delta V_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' =$$
$$\max_p \left\{ -\pi_{jt}(\lambda) - pc_{jzt}\mathbb{E}[u|p] + \delta(1-p) \int \Delta V_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' \right\} \quad (41)$$

Since the last expression depends only on ΔV's, that proves the correctness. To prove non-positivity of ΔV, assume that for some $t+1$ non-positivity has been proven. Then the following chain of inequalities holds:

$$\Delta V_{jt}(\lambda) \leq \sup_p \{-\pi_{jt}(\lambda) - pc_{jzt}\mathbb{E}[u|p]\} + \sup_{\lambda,p} \left\{ \delta(1-p) \int \Delta V_{jt+1}(\lambda')T(\lambda';\lambda)d\lambda' \right\} \leq$$
$$\leq -\pi_{jt}(\lambda) < 0 \quad (42)$$

Finally, for large enough $\bar{t}$ and $\forall t \geq \bar{t}$, $c_{jt} \approx 0$ and $p^*_{jt} \approx 1$ as was discussed in the paper. Hence, $\Delta V_{jt}(\lambda)$ is uniformly within ϵ of $-\pi_{jt}(\lambda) < 0$. ■

D.2 Proof of Equation 25.

For the cast when investment at t' has been done, the future quality that a zip code adds to the consumer utility is:

$$q^1_{jzt,t'} = \frac{1}{I_c} \sum_{i=1}^{I_c} \frac{1}{|zip_c|} \beta^{4G}(Z_i) \quad (43)$$

If no investment has been done at t', then the quality at $t'+1$ remains the same as it was initially with probability one. Moreover, if it is also true that no investment has been done by t, the quality also stays the same, and equals to:

$$q^0_{jzt,t'} = \frac{1}{I_c} \sum_{i=1}^{I_c} \frac{1}{|zip_c|} \sum_{l \in \{2G,3G,4G\}} \beta^l(Z_i)\mathbb{I}[d_{jz,t'} = l] = \frac{1}{I_c} \sum_{i=1}^{I_c} \frac{1}{|zip_c|} \sum_{l \in \{2G,3G,4G\}} \beta^l(Z_i)\mathbb{I}[d_{jz,t} = l] \quad (44)$$

The difference in quality in any future period t happens if and only if no investment has been done at every period from $t'+1$ till $t-1$. The probability of such event:

$$\mathbb{P}[\text{no investment by } t|\text{no investment by } t'+1] = \prod_{t''=t'+1}^{t-1} (1 - \sigma_{jt''}(\lambda_{zt''})) \quad (45)$$

Combining those equations, the average difference between q^1_{jzt} and q^0_{jzt} is:

$$\Delta q_{jzt} = \frac{1}{I_c} \sum_{i=1}^{I_c} \frac{1}{|zip_c|} \Big(\beta^{4G}(Z_i) - \sum_{l \in \{2G,3G,4G\}} \beta^l(Z_i) \mathbb{I}[d_{jz,t} = l] \Big) \prod_{t''=t'+1}^{t-1} (1 - \sigma_{jt''}(\lambda_{zt''})) \quad (46)$$

■