Market Structure, Investment, and Technical Efficiencies in Mobile Telecommunications*

Jonathan T. Elliott,† Georges V. Houngbonon,‡
Marc Ivaldi,§ Paul T. Scott¶

June 4, 2024

Abstract

We develop a model of competition in prices and infrastructure among mobile network operators. Although consolidation increases market power, it can lead to more efficient data transmission due to economies of scale, which we derive from physical principles. After estimating our model with French consumer and infrastructure data, equilibrium simulations reveal that while prices decrease with the number of firms, so do download speeds. Our framework also allows us to quantify the impact of spectrum allocation. The marginal social value of spectrum exceeds firms' willingness to pay in our model as well as observed prices in spectrum auctions.

Keywords: Market structure, scale efficiency, antitrust policy, infrastructure, endogenous quality, queuing, mobile telecommunications.

JEL Classification: D22, L13, L40.

*This research has benefited from the support of Orange (through the TSE-Partnership research foundation) and the Agence Nationale de la Recherche grant ANR-17-EURE-0010. It has also benefited from the support of NYU IT High Performance Computing resources, services, and staff expertise. We thank Richard Chen, Stéphane Ciriani, Chris Conlon, Ying Fan, Fraida Fund, Mohamed Karray, Jakub Kastl, Michael Knox, Elena Krasnokutskaya, Paul LaFontaine, Marc Lebourges, Julienne Liang, Alessandro Lizzeri, Florin Maican, Chitra Marti, Thomas Marzetta, Shivendra Panwar, Nancy Rose, Greg Rosston, Maher Said, David Salant, Jan Philip Schain, David Sibley, Alan Sorenson, Patrick Sun, Dan Waldinger, Aleks Yankelevich, Ali Yurukoglu, and anonymous referees for insightful comments, discussions, and suggestions. The usual disclaimer applies. Elliott and Scott are lead authors. All estimation and simulation code (in Python) can be found at https://github.com/jonathantelliott/mobile-telecommunications.

†Johns Hopkins University, jonathan.elliott@jhu.edu

‡IFC-World Bank Group, ghoungbonon@ifc.org, gvivienh@gmail.com

§Toulouse School of Economics, marc.ivaldi@tse-fr.eu

¶New York University, ptscott@stern.nyu.edu

1 Introduction

In the mobile telecommunications industry, market structure is shaped by antitrust policy and the regulation of radio frequencies, or spectrum. Spectrum is necessary for the operation of a mobile network, and a firm's spectrum holdings (the set of frequencies it has the right to operate) impacts its quality of service, i.e., download speeds. Recently, mobile network operators in many countries have sought to merge and combine their spectrum holdings, with mixed responses from regulators.[1] In recent discussions regarding both antitrust policy and spectrum allocation, quality of service has emerged as a prominent concern.[2]

In this paper, we develop a structural model of the mobile telecommunications industry to capture the impact of changes in market structure (the number of network operators and the allocation of spectrum among them) on equilibrium outcomes such as prices, investment, download speeds, and welfare. We follow an interdisciplinary approach, relying on tools from empirical industrial organization to model market power, and using standard telecommunications engineering models to credibly quantify scale efficiencies. Our framework allows us to address the various impacts of regulating market structure, such as the trade-off between market power and economies of scale. Traditionally, the trade-off means that consolidation may lead to higher or lower prices (Williamson, 1968); in mobile telecommunications, we find that consolidation presents a trade-off between higher prices and improved quality of service. As our notion of market structure includes not only the number of firms but also their spectrum holdings, our framework also allows us to consider the impact of changes in the allocation of spectrum to and within the industry.

Our structural model comprises firms, consumers, and data transmission. Firms (mobile network operators) choose the prices of their mobile service plans and their level of investment in infrastructure, which consumers rely on for data consumption. Consumers choose a mobile phone plan as well as how much data to consume using that plan given its download speed. Our model of data transmission describes how download speeds emerge from firms' and consumers' decisions.

[1] Approved mergers include T-Mobile/Orange (UK, 2010), Hutchinson/VimpelCom (Italy, 2016), Sprint/T-Mobile (USA, 2020), and Teléfonica/Virgin (UK, 2020). Blocked mergers include AT&T/T-Mobile (USA, 2011), TeliaSonera/Telenor (Denmark, 2015), and Teléfonica/Hutchinson (UK, 2016). Anecdotally, network operators in some countries (e.g., France) have recently avoided proposing four-to-three mergers due to an expectation that they would be blocked by antitrust authorities.

[2] For instance, the Sprint/T-Mobile merger was allowed based on the finding "that quality benefits and dynamic competition serve as countervailing forces to the static analysis that substantially address its predicted harmful price effects" (Federal Communications Commission, 2019). Genakos, Valletti and Verboven (2018) study how concentration in mobile telecommunications is related to both prices and investment in infrastructure. Turning to spectrum allocation, the Federal Communications Commission's "National Broadband Plan" describes the potential consequences of insufficient spectrum allocation to mobile telecommunications: "higher prices, poor service quality, an inability for the U.S. to compete internationally, depressed demand and, ultimately, a drag on innovation" (Federal Communications Commission, 2010).

Download speeds, arguably the crucial measure of quality of service in this context, present two modeling challenges. First, due to congestion, download speeds depend on consumers' data consumption decisions as well as firms' investments. Second, even without considering congestion, there is no straightforward mapping from firms' investments to data transmission rates. This is because data transmission depends on, among other things, the spectrum operated and the distance over which data is transmitted. We model download speeds relying on engineering models of data transmission that capture how data is transmitted across space and how network load is handled (in particular, we follow Błaszczyszyn, Jovanovicy and Karray, 2014).[3]

These engineering relationships imply two types of economies of scale that have important economic implications, which we call *economies of density* and *economies of pooling*.

Economies of density result from path loss: as electromagnetic waves carrying data travel, they lose power. Therefore, a mobile network operator can serve a densely populated area more efficiently (meaning either a higher download speed at a given cost or the same download speed at a lower cost) than a sparsely populated area.[4] With symmetric firms, the population density served by each firm is inversely proportional to the number of firms. Consequently, for a given level of total investment in the industry, mobile data services are of higher quality when the number of firms is small.[5]

Economies of pooling relate to the allocation of mobile network resources. When many consumers request data at the same time, data requests enter a queue. Longer queues result in slower download speeds, and there are economies of scale in serving queues. For example, if two network operators were to combine both of their customer bases and spectrum, the combined firm could more efficiently allocate network capacity among customers. This would increase maximum potential download speeds, reduce congestion, and result in higher average download speeds. More generally, the allocation of resources serving a stochastic demand process features economies of scale (Mulligan, 1983; De Vany, 1976; Carlton, 1978).

We estimate a discrete-continuous model of demand for mobile plans and data consumption based on the French market in 2015. Our estimation relies on a consumer dataset that includes information about subscriptions and data consumption by plan and municipality (French communes) for a single mobile network operator, Orange Mobile. We also incorporate

[3] As we derive our production function and associated scale efficiencies transparently from physical principles, our study falls within the tradition of engineering production functions of Chenery (1949).

[4] For example, suppose that the number of base stations per person is held constant across different population densities, so that less population-dense areas have lower geographic base station density. Because signals in the sparsely populated areas will have to travel further on average, they will experience greater path loss, and they will have inferior service despite receiving the same level of investment per capita.

[5] Of course, the equilibrium level of investment (per firm and in total) may change with the number of firms. Our model allows for such changes endogenously, with firms strategically choosing investment in infrastructure.

measured download speeds from Ookla, detailed (publicly available) data on mobile network infrastructure from the radio frequency regulator (ANFR), and income distribution data from the French statistical office (INSEE). While we only observe consumers who subscribe to Orange Mobile, we also observe the prices and characteristics of all contracts available from the other network operators, and we prove that the estimation strategy of Berry, Levinsohn and Pakes (1995) extends to this setting.[6]

Using a model of the supply side derived from the engineering literature, we recover a small number of cost parameters from firms' first-order conditions. Intuitively, once we have estimated demand, we can quantify marginal revenue. We can then use the firms' first-order conditions and our understanding of marginal revenue to make inferences about firms' costs. Firms' pricing decisions provide information about their costs per user served. Furthermore, firms' investment decisions—specifically, the choice of how densely to build base stations—provide information about the costs of building base stations.

We use the estimated models of demand and supply to compute counterfactual equilibria under different market structures. Consolidation presents a trade-off for consumers: faster downloads with higher prices. Focusing on symmetric firms, we find that consumer surplus (currently the relevant barometer for antitrust policy) is maximized at eight firms. However, low-income consumers prefer more firms than high-income consumers since high income consumers have a higher willingness to pay for increased download speeds. Total surplus is maximized at four firms.

We also quantify the marginal social value of allocating more spectrum to mobile telecommunications, a crucial value in a regulator's decision of how to allocate spectrum among industries. We then compare this value to an individual firm's willingness to pay for a marginal unit of spectrum. We find that the marginal social value is about five times greater than an individual firm's willingness to pay.[7] This result highlights the importance of using a structural model to quantify the social value of spectrum. While spectrum auctions may reveal network operators' willingness to pay, willingness to pay may be a gross underestimate of spectrum's social value in mobile telecommunications.[8]

We simulate mergers between mobile network operators in France in the short-run, where infrastructure remains fixed and merging firms combine their network resources. We find that despite merger efficiencies, all mergers between two operators decrease consumer surplus.

[6]Our model predicts shares for all products from all providers in the market, but we only require that the model rationalize product-level market shares for Orange. For other firms, we impose firm-level demand shocks and require the model to rationalize firm-level market shares. Chu (2010) uses a similar approach.

[7]Rosston (2003) found social value to be more than ten times firm willingness to pay.

[8]Our model takes spectrum allocation as given. Thus, while our framework allows us to quantify the impact of spectrum allocation on outcomes, we abstract away from concerns about the spectrum allocation mechanism, in contrast to Milgrom and Segal (2020) and Doraszelski et al. (2019).

Related Literature Most theoretical studies on the relationship between competition and investment take total industry-wide investment as the outcome of interest (Arrow, 1962; Vives, 2008). However, mobile telecommunications networks feature important sources of economies of scale, introducing a potential wedge between industry-wide investment and performance. Even if total investment increases with the total number of firms, quality of service may decline as network resources are spread more thinly across firms. By augmenting a model of investment in infrastructure with an engineering-based model of data transmission, we can directly quantify these scale economies.

A few papers also study investment in mobile telecommunications infrastructure. Granja (2022) studies investment decisions under universal service regulation in Brazil. Lin, Tang and Xiao (2022) analyze 4G technology investment under a hypothetical merger, finding that the merger would reduce investment in this technology. Grajek and Röller (2012) argue that the empirical evidence suggests that access regulation (forcing incumbents to share their infrastructure with entrants) reduces incentives to invest in telecommunications infrastructure. Björkegren (2022) also models endogenous investment in infrastructure, finding that adding a competitor increases investment in rural areas. Björkegren's setting is a less-developed country where geographic coverage is the key product characteristic affected by network operators' investments; ours is a developed country where we take full geographic coverage for granted, and quality of service is the key non-price product characteristic.

There is a limited empirical literature studying imperfectly competitive markets in which firms optimally choose the quality of their products offered. In the seminal theory (Spence, 1975) and in well-studied empirical contexts such as newspapers (Fan, 2013) and cable television (Crawford and Shum, 2007; Chu, 2010; Crawford et al., 2018; Crawford, Shcherbakov and Shum, 2019), quality is a product characteristic that firms can directly control. However, in the context of mobile telecommunications, a challenge for accurately modeling quality of service is the simultaneous determination of download speeds and demand for data.

Consumer demand for a network operator's services depends on its quality of service, and due to congestion externatlities, its quality of service depends on consumer demand.[9] Most demand models for mobile services do not model the simultaneous determination of demand and quality of service (including Bourreau, Sun and Verboven (2021), Cullen, Schutz and Shcherbakov (2020), Fan and Yang (2020), Nevo, Turner and Williams (2016), Sinkinson (2020), Sun (2015)). El Azouzi, Altman and Wynter (2003) and Lhost, Pinto and Sibley (2015) use queuing theory to model the simultaneous determination of service quality and choice of service provider as we do; Malone, Nevo and Williams (2017) model congestion in

[9]Congestion externalities are negative network externalities. Related challenges arise in markets with positive network externalities; e.g., Lee (2013).

broadband services with a different framework. Our study builds on these by incorporating path loss (and therefore economies of density) and by estimating a product-level demand model using detailed consumption and quality data (therefore allowing us to tackle questions of market power). Meanwhile, in the engineering literature, Hua, Liu and Panwar (2012) examine the economies of density and pooling benefits from integrating network resources. However, they do not employ an economic equilibrium framework that endogenizes consumers' choices and firms' investments.

While our analysis assesses the impact of market structure on prices and quality of service, market structure in mobile telecommunications has broader potential impacts: on product proliferation and the types of contracts offered (Seim and Viard, 2011; Fan and Yang, 2020), on coordinated effects (Bourreau, Sun and Verboven, 2021), and on incentives to engage in vertical restrictions (Sinkinson, 2020).

Outline The remainder of this paper is organized as follows. Section 2 presents the data along with descriptive statistics and institutional details. Section 3 introduces the demand model and describes its identification and estimation. Section 4 presents the engineering-based industry model, explains the sources of economies of scale, and lays out the estimation of firms' cost parameters. Section 5 reports estimation results. Finally, section 6 presents several counterfactual analyses. Section 7 concludes.

2 Data and Background

2.1 Firms

We focus on the French telecommunications market in October 2015. During this period, the French mobile industry comprised four mobile network operators (MNOs): Orange Mobile (ORG), SFR-Numericable (SFR), Bouygues Telecom (BYG) and Free Mobile (FREE).

MNOs own and operate network infrastructure (with some network sharing, which we will describe in section 4.3). In contrast, mobile virtual network operators (MVNOs) sell plans to customers without owning their own network resources; instead, they rent access to MNOs' networks. Providing network access to MVNOs is mandatory and enforced by regulation, but the access charge is freely negotiated with the MNO. MVNOs accounted for 10.6% of all mobile contracts in late 2015 (ARCEP, 2016).

2.2 Products and Characteristics

We collect data on mobile phone plan terms (including monthly prices, data limits, and voice limits) from online quarterly catalogs of offers proposed by the four MNOs and the largest

MVNO, EI Telecom.

By 2015, wireless plans were largely differentiated based on data services, with more expensive plans having larger data allowances. Most plans featured unlimited voice allowances; only some low-end plans with zero or small data allowances had limited voice minutes. Furthermore, while data consumption continued to grow rapidly through 2015, voice and text message usage had stabilized.[10]

Table 1 describes our choice set, with monthly prices and data limits representing the main characteristics of interest. Monthly data limits are "soft," in the sense that customers can still use data services after exceeding the limit, but with significantly throttled download speeds.[11] We aggregate phone plans by data limit category (less than 500 MB, 500–2 999 MB, 3 000–6 999 MB, and more than 7 000 MB) and by whether they include unlimited voice services. The choice of four data limit groups is motivated by the observation that network operators typically advertise three or four tiers of plans. For each plan group and firm, we select a representative plan to include in our choice set.

Our representative plans do not feature bundled services like fixed broadband, fixed telephony, and television services. Moreover, plans have different contract lengths (no commitment, a 12-month commitment, or a 24-month commitment). Since most consumers subscribe to plans with a 24-month contract duration, our representative plans feature 24-month commitments. Within each group of a firm's plans, defined by data and voice limits, the representative plan we select is the one that is the least expensive among the 24-month commitment plans available to new subscribers (after adjusting the monthly price for a handset subsidy as described below). This plan always excludes home broadband and television services. Thus, our choice set of representative plans consists entirely of mobile-only plans.

The representative phone plans in our model's choice set reflect the characteristics of plans actually available in the market, with the exception of the monthly price. For plans that include a handset subsidy, we adjust the price by subtracting the value of the handset subsidy from the monthly price. Details on how we calculate this subsidy can be found in Appendix C.1.1.

In the customer database described below, we observe market shares for plans with only wireless services as well as plans with bundled services. Each actual plan in these data is then associated with a representative plan (grouped by data limit and whether voice minutes are limited), and our estimation method takes the market shares of the representative plans to be

[10] Source: Séries chronologiques annuelles (1998-2015) data released by ARCEP. Obtained from http://www.arcep.fr/fileadmin/reprise/observatoire/serie-chrono/series-chrono-annuelles-1998-2015p.xlsx September 23, 2022.

[11] The data allowances we measure are the phone plans' baseline allowances. We ignore add-on options.

the aggregate market share of all the actual products associated with them.[12] For instance, our empirical model features one high-data-limit plan for Orange. We treat the price of this plan as 38.74 €. This price corresponds to an observed price of 54.99 € for this plan minus 16.25 € for the value of the associated handset subsidy. We calculate the market share of this representative plan as the sum of market shares of all eleven high-data-limit contracts offered by Orange.

Table 1: The Choice Set

Operator	**Price (€)**	**Data Limit (MB)**	**Unlimited Voice**	**Plans Represented**	**Min Price (€)**	**Max Price (€)**	**Min Limit (MB)**	**Max Limit (MB)**
Orange	12.07	50	No	11	4.99	30.99	0	50
Orange	14.99	1000	No	4	14.99	14.99	1 000	1 000
Orange	22.91	1000	Yes	2	22.91	24.99	1 000	1 000
Orange	30.91	4000	Yes	5	19.99	48.99	3 000	5 000
Orange	38.74	8000	Yes	11	38.74	165.99	8 000	20 000
Bouygues	8.07	0	No	6	3.99	11.32	0	20
Bouygues	14.99	1000	No	3	14.99	14.99	1 000	1 000
Bouygues	20.91	3000	Yes	4	19.99	29.99	3 000	5 000
Bouygues	33.74	10000	Yes	4	32.70	72.70	10 000	20 000
Free Mobile	2.00	50	No	1	2.00	2.00	50	50
Free Mobile	19.99	3000	Yes	1	19.99	19.99	3 000	3 000
SFR	12.07	100	No	5	5.99	14.99	100	200
SFR	14.99	1000	No	3	14.99	19.99	1 000	1 000
SFR	22.91	1000	Yes	3	22.91	29.99	1 000	1 000
SFR	31.91	5000	Yes	5	19.99	43.99	3 000	5 000
SFR	37.74	10000	Yes	9	36.70	149.99	10 000	20 000
MVNO	7.99	0	No	13	7.99	18.99	0	200
MVNO	17.99	1000	No	5	9.99	17.99	500	1 000
MVNO	19.99	500	Yes	10	19.99	35.99	500	2 000
MVNO	42.99	5000	Yes	13	12.99	61.99	3 000	5 000
MVNO	64.99	10000	Yes	4	64.99	76.99	10 000	10 000

Note: Each row corresponds to an object in the choice set, i.e., a representative product. The minimum and maximum prices and data limits are over the set of plans represented by each representative product in the choice set. The representative plan is the least expensive within the group of plans *that has a commitment of 24 months*; therefore, because some plans have commitments of less than 24 months, it is possible for the minimum price to be lower than the representative plan's price.

We do not explicitly distinguish between pre- and postpaid phone plans. Most French consumers subscribe to postpaid plans, accounting for 83% of the market in late 2015 (ARCEP, 2016). While postpaid plans require consumers to pay for their consumption at the end of a monthly billing period, prepaid plans require customers to pay in advance. Prepaid plans generally involve low data limits and limited voice allowances.

[12]The Orange customer database includes consumers on plans that are no longer available. These plans, like available plans, are all mapped to a representative plan and consumers subscribing to these plans will contribute to the market share of the associated representative plan.

For MVNOs, our choice set includes one representative plan for each category; that is, we effectively assume that there is one representative MVNO firm. Representative MVNO contracts are selected from the largest MVNO (EI Telecom) in the same way we select representative contracts for MNOs.

2.3 Demand Data

Our main demand data source is a proprietary dataset based on the universe of mobile contracts for one operator, Orange Mobile, in October 2015. This dataset includes the number of subscribers to and the usage of mobile data services by plan and municipality. Note that we focus only on the residential market for mobile services, ignoring business customers. Residential customers represented 89% of mobile subscriptions in 2015 (ARCEP, 2016).

The customer dataset is complemented by data on the quality of mobile data services, as measured by download speeds. Due to congestion, delivered download speeds are not solely a function of infrastructure and geographic characteristics. Congestion arises because the available bandwidth is shared among users and, as a result, the greater the number of users, the lower the quality (as measured by download speed). At the same time, the number of users (and therefore the demand for data) on a network depends on quality. In our counterfactuals, we employ a model in which demand and quality of service are simultaneously determined, but for the purpose of estimation, we rely on a direct measure of download speeds as our measure of quality. Speedtest is a service offered by the firm Ookla that allows users to check their download and upload speeds. We use a proprietary dataset provided by Ookla on over one million speed tests in France from the second quarter of 2016 that include measured download speed, the time of the test, the location of the user, and the mobile network operator. Using these speed tests, we construct a measure of experienced download speeds for each mobile network operator in each municipality. Section C.2 in the data appendix explains the construction of this quality measure in detail.

Markets are defined as municipalities (French communes). We limit our analysis to relatively populous markets, specifically, those with a population greater than 10 000, resulting in a total of 589 markets.[13] We limit ourselves to populous markets because active network sharing (where network operators share the transmitting components of their infrastructure) is relatively common in rural areas but not practiced in urban areas, with the exception of Free Mobile's reliance on Orange's network for 2G and 3G (but not 4G) traffic, with caps on the speeds available to Free customers. Even in this case, Free must rely on its own 4G infrastructure to deliver competitive download speeds. Thus, for our sample, we are comfortable

[13] There are 592 municipalities with a population exceeding 10 000, but we exclude three municipalities because we lack a sufficient number of download speed tests to construct reasonably precise quality measures for them. This yields a total of 589 markets in our sample.

associating a firm's measured download speeds with that firm's own investments in infrastructure. Municipality-level market size is defined as the population age 12 and older, obtained from the French Bureau of Statistics, INSEE. While France has about 36,000 municipalities, the 589 in our sample contain 43.5% of France's population.

For network operators other than Orange, we only have market shares at the national level from GSMA Intelligence. Table 2 presents the market shares of each firm in October 2015.

Table 2: Aggregate Market Shares of Alternatives

Market Size (millions)	ORG	SFR	BYG	FREE	MVNO	Non-users
56.5	29.4%	13.4%	17.2%	21.5%	10.6%	8.0%

Note: Data reported by the regulator (ARCEP, 2016) provides the relative share of MVNOs and MNOs. Relative shares within MNOs are obtained from GSMA Intelligence. Shares are adjusted to allow for 8% outside option share, consistent with CREDOC (2015).

Our econometric approach makes use of the income distribution. We take income deciles (over households) for each municipality from the 2011 population census conducted by the French Bureau of Statistics (INSEE)

2.4 Infrastructure Data

Finally, we obtain detailed data on infrastructure from the national radio communications regulator (ANFR). These data describe the locations of all mobile base stations, along with the number of antennas and frequencies operated by each network operator.[14]

Ultimately, we want to quantify the typical cell for each municipality, characterized by the area served by base stations and the bandwidth operated. For bandwidth, we compute the total bandwidth operated across all base stations for each operator and municipality. For municipalities with an uneven population distribution, dividing municipality area by the number of base stations would be a misleading measure of the area of the cell size experienced by most users. The concentration of base stations within sparsely inhabited areas is typically low, for such areas have few users and low data demand. Instead, we consider a measure of the "adjusted area" of a commune. We compute the mean population density by averaging over persons rather than space (equivalently, we take the contraharmonic mean of population density across space).[15] The adjusted area is defined as the municipality's population divided by our mean population density.[16] If a municipality were to consist of a populated area with

[14] This database is publicly accessible at https://www.cartoradio.fr/.

[15] The data we use for this is the Gridded Population of the World, v4, available at https://sedac.ciesin.columbia.edu/data/collection/gpw-v4.

[16] For example, Fontainbleau is a relatively populous commune consisting of a town surrounded by a forest.

a uniform population density as well as an area with zero population density, then this way of measuring a municipality's population density captures the population density of the populated area, and the adjusted area would equal the populated area. We then measure our object of interest, the area served by a typical base station, as the adjusted area divided by the number of base stations.

In addition to infrastructure data from ANFR, we use traffic data from OSIRIS, which is an internal database provided by Orange. OSIRIS provides the total volume of data traffic per network cell over time. We use these volumes to calculate data demand rates, which we then use to calibrate parameters of the data transmission model.

2.5 Descriptive Statistics

Table 3 provides summary statistics for our variables of interest.

Table 3: Summary Statistics

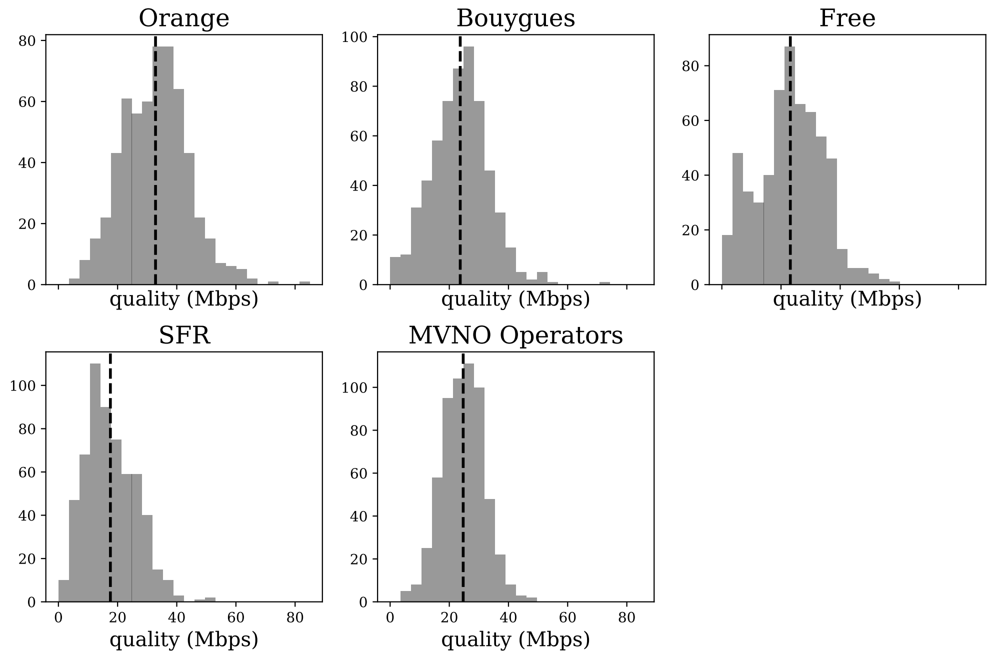
	Mean	**Std. Dev.**	**Min.**	**Max.**
Quality and market data				
Market average usage (MB)	1 128	204	579	1 829
Quality Orange (Mbps)	32.82	11.11	3.97	84.98
Quality Bouygues (Mbps)	23.70	9.65	0.60	72.97
Quality Free (Mbps)	23.15	11.03	1.56	56.74
Quality SFR (Mbps)	17.57	8.58	0.39	52.30
Quality MVNO (Mbps)	24.70	7.04	5.13	48.87
Median income (Euros)	13 035	3 177	5 152	31 320
Number of markets		589		
Tariff data				
Price	23.47	14.22	2.00	64.99
Price (Orange)	23.92	9.90	12.07	38.74
Price (Others)	23.33	15.32	2.00	64.99
Data limit	3 081	3 484	0	10 000
Number of phone plans		22		
Infrastructure data				
Bandwidth per firm (MHz)	70.69	30.42	0.00	140.20
Number of base stations	7.47	21.47	0	511
Effective cell radius (km)	1.44	0.93	0.26	7.64

Note: Customer, quality, market, and infrastructure data summary statistics are (unweighted) across markets. Tariff data summary statistics are across mobile phone plans.

Measured quality (download speeds) varies substantially both across and within markets. Across markets, the average standard deviation for an operator is 9.56 Mbps, and across

While the population density of the town is relatively high, the population density of the commune appears low if we divide by the commune's total area. Our measure of adjusted area for Fontainbleau is 69.6 square kilometers, while the raw municipality has an area of 172 square kilometers.

Figure 1: Histograms of Qualities by Operator



Note: This figure shows the average download speeds at the market level for each operator. Dashed vertical lines represent the average of the market-level averages by operator, unweighted by population. The scale of the x-axis is the same across all subplots, allowing for comparisons in the distribution of average download speeds across operators.

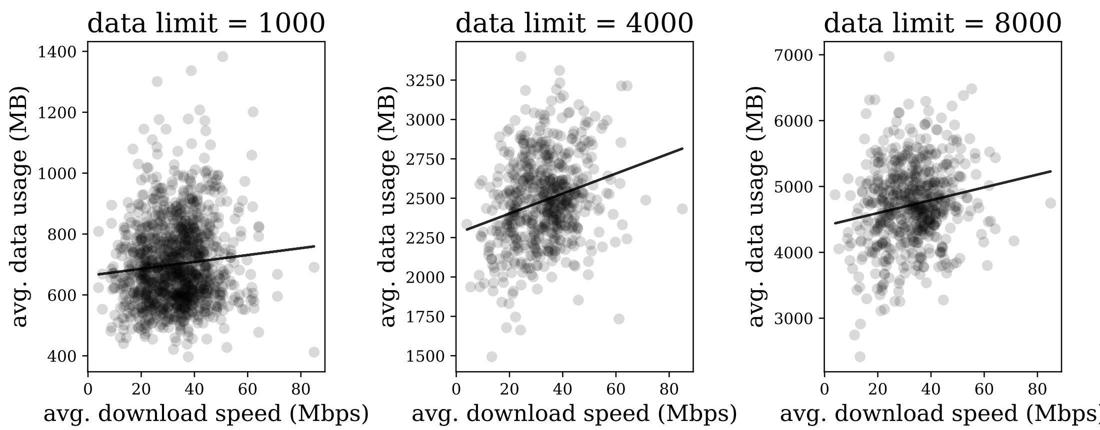
operators, the average standard deviation for a market is 7.92 Mbps. Figure 1 displays histograms of measured quality across markets for each mobile network operator.[17] The raw Ookla data include few measurements for MVNO operators, so we set MVNO speeds equal to the average of Orange, Bouygues, and SFR within each market. Anecdotally, MVNOs contracted with all three of these MNOs in 2015, but not with Free. Pooling the MVNO speed measurements that we do have across communes gives a roughly similar average download speed to this imputation: 22.7 Mbps in comparison to a mean of 24.7 Mbps in Table 3.

Data usage is positively correlated with measured quality. Figure 2 plots the relationship across markets between Orange download speeds and the observed average data usage for three different data limits.[18] Most consumers do not actually reach their data limit in a given

[17]There is a potential selection concern in these measures of download speeds. Because they come from voluntary speed tests, it may be the case that measurements tend to happen when consumers experience slow downloads. However, the levels of download speeds reported in Table 3 are consistent in the aggregate with the levels coming from other sources. We note that for Orange, Bouygues, and SFR, our average download speeds lie within the values reported by ARCEP for intermediate and urban density areas (the densities of areas in our sample). For Free, the 23 Mbps average download speed is actually higher than the 19 Mbps number reported by ARCEP. See https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/couverture-et-qualite-de-service-mobile-2g-3g-4g-5g/couverture-et-qualite-des-services-mobiles-juillet-2016.html (accessed November 7, 2022).

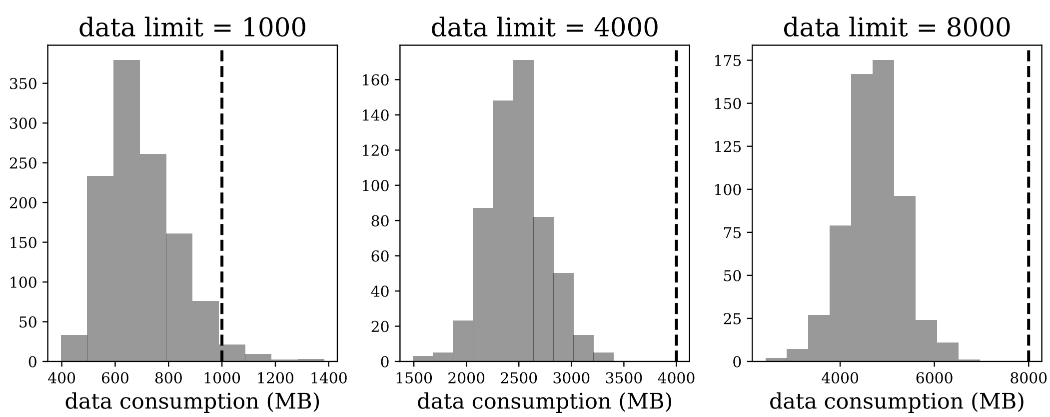
[18]The correlations for data limits 1 000 MB, 4 000 MB, and 8 000 MB are, respectively, 0.092, 0.254, 0.176.

Figure 2: Average Data Usage vs. Measured Quality across Markets



Note: This figure presents scatter plots that display for three different Orange phone plans the relationship between the average download speed and the average amount of data consumed per customer. Observations are at the market level. The line in each subplot is a line-of-best-fit for the observations.

Figure 3: Average Data Usage across Markets



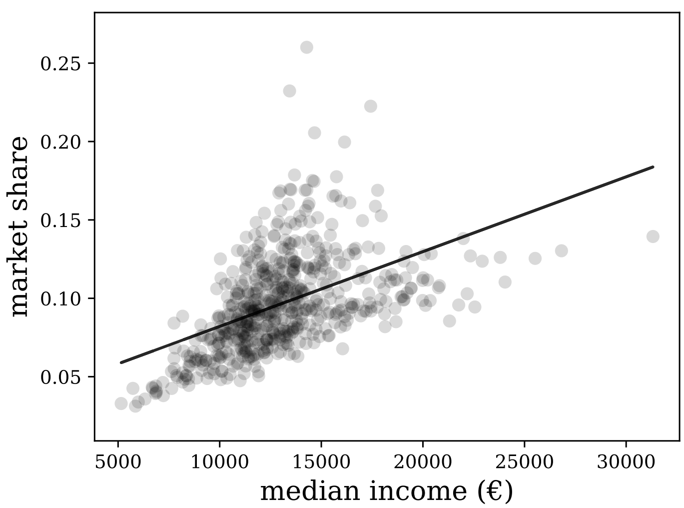
Note: Each subplot represents, for Orange phone plans with a particular data limit, a histogram of customers' average data consumption at the market level. Market-level average data consumption for each phone plan is obtained by, for each market, averaging data consumption across customers in that market subscribing to the phone plan. Dashed vertical lines represent the average of the market-level averages, unweighted by population.

month, as demonstrated in Figure 3, which plots the histograms of market-level average data consumption for three different data limits.[19]

When measuring data consumption, we must account for the fact that some consumers in our data subscribe to plans that have somewhat different data limits than that of the relevant representative plan. For example, a consumer might subscribe to an Orange plan with a 3 GB per month limit, but in our econometric model, we treat them as having subscribed to Orange's representative plan with a 4 GB limit. In our descriptive statistics and in model estimation, we average data consumption only over subscribers that subscribe to plans with the same data limit as the representative plan.

Markets with higher median incomes tend to have higher market shares for expensive phone plans. Figure 4 plots the relationship between the median income in each market and the joint market share of the two most expensive Orange phone plans, with prices of 30.91 € and 38.74 €. Median incomes are positively correlated with the market share of these expensive plans, with a correlation coefficient of 0.495.

Figure 4: Median Income vs. Expensive Contract Market Share



Note: The figure presents a scatter plot of the median income in a market against the market share of Orange's two most expensive phone plans (the 4 GB representative plan and the 8 GB representative plan). The line is a line-of-best-fit.

3 Demand Model

In this section, we describe a model of consumer choice capturing how consumers choose mobile phone plans and the amount of data to consume, taking prices and download speeds

[19]For the data limits 1 000 MB, 4 000 MB, and 8 000 MB, the fraction of the data limit that is consumed is, respectively, on average, 0.700, 0.621, and 0.590.

as given. We explain in section 4 how download speeds depend on consumer behavior through congestion, but since individual consumers are small, download speeds can be treated as exogenous for the purpose of an individual consumer's decision. In section 3.2, we discuss identification and estimation of the demand model's parameters.

3.1 Demand Model

We begin by introducing some notation that is used in this section as well as in section 4, which presents the model of supply. There exists a set of mobile phone plans, $\mathcal{J}$, indexed by j. Each plan j belongs to a particular firm, $f(j)$, and the set of plans provided by a firm is given by $\mathcal{J}_f$. Consumers belong to different geographic markets, indexed by m, which vary in demographics and geography (the latter matters for the efficiency of data transmission). To keep the analysis tractable, we do not model travel or any interdependence between markets. Table 9 in the Appendix provides a list of all parameters used in the model and their definitions.

Consumers make decisions about to which mobile phone plan (if any) to subscribe to and how much data to consume using that plan. Each mobile phone plan j in a market m is characterized by the download speed available in that market, $Q_{f(j),m}$;[20] the price of that phone plan, P_j; and a data consumption limit, $\bar{d}_j$. Note that download speeds are common within a market across plans offered by the same firm, as firms do not discriminate across plans in the download speeds they offer. Note also that prices and data limits do not depend on the market. In France, mobile phone plan prices and characteristics (except download speeds) are set nationally.

A consumer's indirect utility for a plan j depends on the utility that they derive from consuming x Gigabytes of data as well as the product characteristics. This indirect utility is given by

$$u_{jm}\left(x;\vartheta_i,\varepsilon_{ij},\theta_{pi}\right) = w_j\left(x, Q_{f(j),m}, \vartheta_i\right) + \theta_v v_j - \theta_{pi} P_j + \xi_{jm} + \varepsilon_{ij}, \tag{1}$$

where $w_j(\cdot)$ maps the plan j, data consumption x, and data quality $Q_{f(j),m}$ into the utility from consumption of mobile data services. This function depends on ϑ_i, a random variable that captures how much i values consuming data. Section 3.1.1 explains in detail how we model the value of data function $w_j(\cdot)$ and the role of ϑ_i. Other plan characteristics that enter the consumer's utility include the price, P_j; whether the plan has an unlimited voice allowance, captured by v_j (equal to 1 if plan j has an unlimited voice allowance, 0 otherwise); and ξ_{jm}, a product-market-specific demand shock. ε_{ij} is an idiosyncratic product-specific shock.

Before the realization of the shocks ϑ_i and $\boldsymbol{\varepsilon}_i$, there are two sources of *ex ante* consumer

[20]While consumers may be mobile, we assume that their choices depend on the network quality in their municipality of residence.

heterogeneity. The first is the price sensitivity parameter, θ_{pi}. The second is the distribution of the data value shock, ϑ_i. As we will explain in the following subsection, this data value shock is not only a random variable but also has a different *distribution* for different types of consumers.

We note that the relatively parsimonious utility function expressed in equation 1 leaves out product characteristics like geographic coverage, international roaming terms, and download speeds in markets other than the consumer's home market. Our specification is based on the idea that variation in prices and mobile data service quality were the most important aspects of product differentiation at the time. All operators offered close to complete coverage during our sample period, so there is very little variation in coverage. Meanwhile, identifying the importance of international roaming terms and/or coverage in neighboring municipalities would be challenging due to the high-dimensionality of these product characteristics. Additionally, anecdotal evidence suggests that these characteristics are not particularly important to most consumers.

3.1.1 Mobile Data Consumption

After subscribing to a particular plan j, a consumer chooses how much data to consume given the plan's data consumption limit and download speed, as well as the consumer's value of data consumption. They choose this level of consumption to maximize the utility from data consumption, $w_j(\cdot)$. To rationalize finite data consumption even when additional data consumption entails no monetary cost, our functional form of $w_j(\cdot)$ includes a term which corresponds to the disutility of download times. This disutility is proportional to the amount of data downloaded and is inversely proportional to the download speed. It can be thought of as the opportunity cost of time spent downloading. Consumers consume data until the marginal utility of extra data equals the marginal disutility of additional download time.

A consumer's utility of data consumption is given by the following functional form:

$$w_j(x, Q, \vartheta_i) = \vartheta_i \log(1 + x) - c_j(x, Q). \tag{2}$$

The first term captures the utility the consumer derives from consuming data. It exhibits decreasing marginal returns and depends on the parameter governing how much the consumer values data consumption, ϑ_i.

The second term in equation 2, $c_j(\cdot)$, reflects the cost of the time spent downloading. There is a discontinuity in download speeds when a consumer reaches their monthly data limit, $\bar{d}_j$, captured by the two cases in equation 3. Data consumed after reaching the data limit downloads at a throttled speed Q^L.[21] Thus, we use the following formula for the time cost of data downloads:

$$c_j(x,Q) = \begin{cases} \theta_c \frac{x}{Q} & \text{if } x \leq \bar{d}_j \\ \theta_c \left(\frac{\bar{d}_j}{Q} + \frac{x - \bar{d}_j}{Q^L} \right) & \text{if } x > \bar{d}_j, \end{cases} \tag{3}$$

where θ_c is a preference parameter capturing the disutility of waiting for downloads.

We let $x^*_{jm}(\cdot)$ denote the consumer's optimal data consumption:

$$x^*_{jm}(\vartheta_i) = \arg\max_{x \in \mathbb{R}_+} \left\{ w_j \left(x, Q_{f(j),m}, \vartheta_i \right) \right\}.$$

This discontinuity in download speeds creates a discontinuity in the marginal cost of data consumption. Consequently, the first order condition and the structure of the marginal cost of data consumption yield four possible cases that determine the optimal data consumption:[22]

$$x^*_{jm}(\vartheta_i) = \begin{cases} 0 & \text{if } \vartheta_i \leq \frac{\theta_c}{Q_{f(j),m}} \\ \frac{\vartheta_i}{\theta_c / Q_{f(j),m}} - 1 & \text{if } \frac{\theta_c}{Q_{f(j),m}} \leq \vartheta_i < \left(\frac{\theta_c}{Q_{f(j)}} \right) \left(\bar{d}_j + 1 \right) \\ \bar{d}_j & \text{if } \frac{\theta_c}{Q_{f(j),m}} \left(\bar{d}_j + 1 \right) \leq \vartheta_i < \frac{\theta_c}{Q^L} \left(\bar{d}_j + 1 \right) \\ \frac{\vartheta_i}{\theta_c / Q^L} - 1 & \text{if } \vartheta_i \geq \frac{\theta_c}{Q^L} \left(\bar{d}_j + 1 \right). \end{cases} \tag{4}$$

The first case captures consumer types that consume no data.[23] The second case captures consumer types that consume less than $\bar{d}_j$ even without throttling. The third case captures consumer types that consume more than $\bar{d}_j$ if download speeds were not throttled, but under throttling, the marginal cost of an additional unit of data is greater than the marginal benefit, so they consume exactly the data limit. The final case captures consumer types that consume more than $\bar{d}_j$ even under throttled download speeds.[24]

3.1.2 Mobile Phone Plan Decision

A consumer i chooses the mobile phone plan that maximizes their expected utility, where the expectation is with respect to the data value parameter, ϑ_i, assumed to be exponentially distributed:

$$\vartheta_i \sim \mathit{Exponential}\left(\theta_{di}\right).$$

[21] MNOs in France typically use a throttled speed of 128 Kbps (see section C.1.2 in the appendix for more information about throttled download speeds). We use this value for throttled speeds in our estimation of demand and cost parameters as well as in our counterfactuals.

[22] Here we assume that $Q^L < Q_{f(j),m}$, which holds in our data.

[23] We interpret such consumers as those who do not need their mobile plan (e.g., they went out of the country for the month). In the data, we observe a point mass of consumers who consume zero data—even among those with high data limit plans.

[24] Small data limit plans have hard data limits (i.e., there is no throttling). We therefore impose that all contracts with data limits less than 500 MB cannot consume greater than the associated data limit.

That consumers do not know their realization of ϑ_i prior to choosing a plan reflects that consumers may be unable to perfectly forecast their utility for data.[25] While consumers do not know their ϑ_i *ex ante*, they do know their θ_{di}.

The outside option, represented by $j = 0$, corresponds to not subscribing to any mobile phone plan. We normalize the outside option's indirect utility to ε_{i0}.

We assume a nested logit structure for these idiosyncratic shocks ε_{ij}:

$$\varepsilon_{ij} = \zeta_{ig(j)} + (1 - \sigma)\, \eta_{ij},$$

where η_{ij} is i.i.d. extreme value and ζ_{ig} is distributed such that ε_{ij} is extreme value. The value $\sigma \in [0, 1)$ is the nesting parameter. There are two nests: one containing only the outside option, and the other containing all mobile phone plans.

We adopt this nested logit structure to more flexibly model substitution to the outside option. Note that if $\sigma = 0$, the model is equivalent to a mixed logit model without nesting. As σ approaches 1, the model approaches the case where there is no outside option.

After observing their vector of idiosyncratic taste shocks $\boldsymbol{\varepsilon}_i$, but before knowing their data value shock ϑ_i, consumer i chooses the phone plan that maximizes their expected utility. Their choice of phone plan j^*_{im} is given by:

$$j^*_{im}\left(\boldsymbol{\varepsilon}_i; \boldsymbol{\theta}_i\right) = \arg \max_{j \in \mathcal{J} \cup \{0\}} \left\{ \mathbb{E}\left[u_{jm}\left(x^*_j\left(\vartheta_i\right); \vartheta_i, \varepsilon_{ij}, \theta_{pi}\right) \mid \theta_{di} \right] \right\}, \tag{5}$$

where the expectation is over ϑ_i, and θ_{di} is the parameter that controls ϑ_i's distribution as described above. We let $\boldsymbol{\theta}_i = (\theta_{pi}, \theta_{di})$ capture the heterogeneous preference parameters.

Integrating over idiosyncratic taste shocks, we obtain market shares for each mobile phone plan conditional on consumer type $\boldsymbol{\theta}_i$:

$$s_{ijm}\left(\boldsymbol{\theta}_i\right) = \int \mathbb{1}\left\{ j = j^*_{im}\left(\boldsymbol{\varepsilon}_i; \boldsymbol{\theta}_i\right) \right\} dF\left(\boldsymbol{\varepsilon}_i\right), \tag{6}$$

and integrating over consumer types we get market shares:

$$s_{jm} = \int s_{ijm}\left(\boldsymbol{\theta}_i\right) dF_m\left(\boldsymbol{\theta}_i\right). \tag{7}$$

These market shares, along with data consumption (given by equation 4), yield the average

[25] Another reason for including these exponential shocks is that they introduce a smoothness in data consumption that ultimately keeps our GMM objective function continuous.

data consumed in market m by consumers subscribed to phone plan j:

$$\bar{x}_{jm} = \int\int \frac{s_{ijm}(\boldsymbol{\theta}_i)}{s_{jm}} x^*_{jm}(\vartheta_i)\, dF(\vartheta_i|\boldsymbol{\theta}_i)\, dF_m(\boldsymbol{\theta}_i)\,. \tag{8}$$

3.2 Demand Estimation

In this section, we describe how we estimate the demand model's parameters using a modified version of Berry, Levinsohn and Pakes (1995).

We seek to estimate the distribution of consumer parameters, $(\theta_{pi}, \theta_{di}, \theta_c, \theta_v, \sigma)'$. Note that we have two heterogeneous parameters that we allow to vary by income. Specifically, we assume

$$\begin{pmatrix} \log(\theta_{pi}) \\ \log(\theta_{di}) \end{pmatrix} = \begin{pmatrix} \theta_{p0} \\ \theta_{d0} \end{pmatrix} + \begin{pmatrix} \theta_{pz} \\ \theta_{dz} \end{pmatrix} z_i, \tag{9}$$

where z_i is the consumer's income. We measure income in units of 10 000 €, data limits in GB, and quality in GBps.[26]

3.2.1 Unobserved Demand Component

As is standard in the demand estimation literature, the unobserved demand components $\boldsymbol{\xi}$ are computed to rationalize observed market shares. We observe the set of products (in our setting, phone plans) offered by all firms, but we only observe detailed market share data at the plan-market-level for Orange. For plans offered by other firms, we observe market shares at an aggregate firm-level. The standard BLP contraction mapping used to solve for $\boldsymbol{\xi}$ cannot recover the unobserved demand components with market shares at different levels of aggregation. We therefore use a modified technique (similar to Chu (2010)) that is able to handle market shares at different levels of aggregation.

Our modified estimation technique rationalizes plan-level market shares for Orange plans and only the firm-level aggregate market shares for all other firms. Formally, we assume

$$\forall j \in \mathcal{J}_{-ORG}, \forall m: \quad \xi_{jm} = \xi_{f(j)},$$

where $\mathcal{J}_{-ORG}$ is the set of non-Orange plans, and $f(j)$ is the firm that corresponds to plan j.[27]

[26] Note that the quality measures are in Giga*bytes* per second (GBps), not Giga*bits* per second (Gbps). This conversion is necessary so that the second term in Equation 3 has the interpretation of seconds spent downloading data.

[27] Since aggregate market shares are for all of France (and not just the urban communes in our sample), we would like to sum over all of France when computing the national shock $\xi_{f(j)}$ that rationalizes the national market shares for a firm other than Orange. We therefore construct a "Rest of France" municipality that aggregates the population and income distributions from all communes not included in our estimation sample. Download speeds in the Rest of France commune are computed as the average download speeds in all communes

Appendix B.1 describes a modified version of the BLP contraction mapping that is capable of solving for the unique vector $\boldsymbol{\xi}$ under the above assumption.

Given the lack of product-level market shares for the firms other than Orange, an alternative approach would be to aggregate products within each firm and treat each firm as having a single product. Doing so would rule out within-firm variation in observed characteristics as well as the unobserved ξ shocks. In contrast, our approach allows observable characteristics to vary while ruling out within-firm variation in ξ (for firms other than Orange). We have some ability to evaluate this assumption given that we recover product-level ξ shocks for Orange. Our estimates indicate that the variation in ξ's across firms is more substantial than the variation within firm across municipalities and products. The standard deviation of the estimated ξ_{jm} terms for Orange is 0.149, while the standard deviation of the firm-level $\xi_{f(j)}$ terms is 0.439 (including the mean value for Orange).

3.2.2 Identification

In this section, we formally present our identification approach and offer intuition for how the variation in the data maps to parameters of interest.

A limitation of our data is that prices are set nation-wide and do not vary by market. Moreover, prices showed very little intertemporal variation around our sample period.[28] Figure 5 illustrates the stability of prices over the two years prior to our sample period. Prices of Orange phone plans are in black, and the prices of other operator plans are in light gray. Given the lack of price variation, it is difficult to identify price elasticities during the period we study.

We therefore calibrate price elasticities based on an earlier study. Formally, we calculate the implied price elasticity of Orange products in market m based on a proportional increase in all Orange prices, defined as follows:

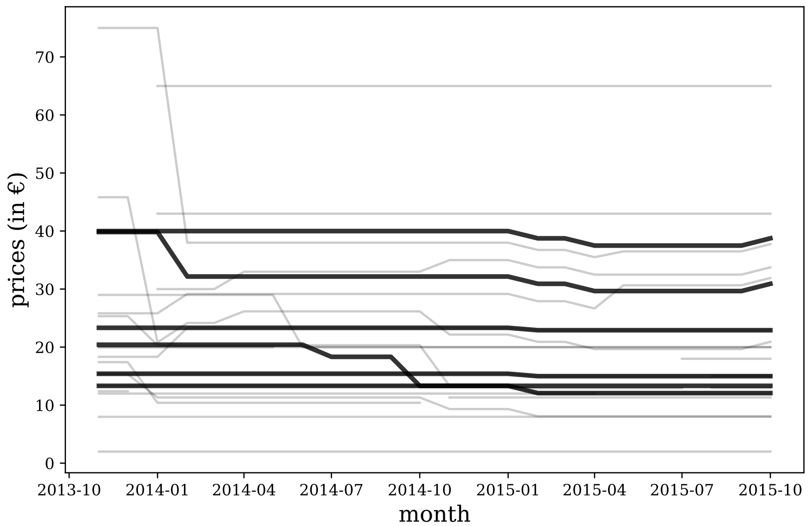
$$e_m^{ORG}(\theta) = \frac{\mathrm{d}\ln s_{ORG,m}(\mathbf{P};\theta)}{\mathrm{d}\ln P_{ORG}}, \tag{10}$$

where $s_{ORG,m}(\cdot)$ is the market share of phone plans offered by Orange and $\mathrm{d}\ln P_{ORG}$ represents a proportional change in the price of all Orange phone plans.

outside of the 589 in our sample. The "Rest of France" market plays a very limited role in the estimation; we include it only so that we can calculate aggregate market shares that are comparable to the observed national market shares.

[28] Note that Bourreau, Sun and Verboven (2021) consider a time period that includes the entry of Free Mobile in 2012. Following this entry, there were substantial price changes as the incumbent MNOs reacted to the new low-cost competitor. In contrast, during the two years leading up to our sample period, price variation was quite limited.

Figure 5: Prices of Orange Phone Plans over Two Years



Note: Black lines represent the prices of Orange phone plans over time. Light gray lines represent the prices of other mobile network operators' phone plans over time.

Our source for this calibration is Bourreau, Sun and Verboven (2021), hereafter BSV, who study the French market around the entry of Free Mobile, a few years earlier than our sample period. Free's entry was disruptive, resulting in considerable price and choice set variation, but prices settled down before the period covered by our data. BSV's estimates imply a value of -2.36 for $\mathbb{E}\left[e_m^{ORG}(\theta)\right]$.[29] We therefore require that

$$\mathbb{E}\left[e_m^{ORG}(\theta)\right] = -2.36 \tag{11}$$

as a moment in our estimation procedure, described below.

The main assumption behind this elasticity calibration is that Orange's own price elasticity was stable between 2013 and 2015, consistent with the lack of price variation during this period. Given this assumption, the calibration ensures that our demand model features the same degree of price sensitivity as BSV's model.

Importantly, we calibrate elasticities, which are functions of parameters, rather than the parameters themselves. Directly imputing parameters of the utility function from BSV's model would not achieve the same effect as imposing equation 11, as our model features several crucial departures from BSV, including incorporating download speeds, allowing for endogenous data consumption, and using data limits as a product characteristic. Because demand elasticities depend on all parameters of the utility function and not just the price

[29]We calculate the overall own-price elasticity for Orange $\mathbb{E}\left[e_m^{ORG}(\theta)\right]$ using elasticities in BSV's Table A.4, diversion ratios in Table A.3, and market shares in Table 3. See Appendix B.2 for the details of this calculation.

coefficient parameters, using some of BSV's parameters while changing other aspects of the demand model would result in demand elasticities that differ from BSV's model and therefore have no empirical basis.

Substitution to the outside option (controlled by the nesting parameter σ) is another feature of the model that our data provide little information on. Thus, we also calibrate our model to the level of diversion to the outside option implied by BSV's estimates. Specifically, BSV's estimates imply a diversion ratio of 0.036 for substitution to the outside option in response to a proportional increase in all of Orange's prices.[30] We impose this diversion ratio as another moment in our estimation procedure. That is, we define

$$DIV_m^{ORG,0}(\theta) = -\frac{\frac{\mathrm{d}s_{0,m}(\mathbf{P};\theta)}{\mathrm{d}\ln P_{ORG}}}{\frac{\mathrm{d}s_{ORG,m}(\mathbf{P};\theta)}{\mathrm{d}\ln P_{ORG}}},$$

where $\mathrm{d}\ln P_{ORG}$ again represents a proportional change in the price of all Orange phone plans, and require that

$$\mathbb{E}\left[DIV_m^{ORG,0}(\theta)\right] = 0.036. \tag{12}$$

While equations 11 and 12 pin down important derivatives of the demand system, there are other crucial aspects of the demand system that we must estimate using our data. In particular, we still need to estimate how consumers trade off prices and download speeds (and how consumers differ in such preferences).

Data utility parameters θ_{d0}, θ_{dz}, and θ_c are identified, in part, by matching predicted data consumption with observed data consumption. Specifically, matching observed and predicted data consumption effectively identifies the average data valuation parameter θ_{di} conditional on the cost of download time parameter θ_c. Equation 4 shows that the responsiveness of data consumption to download speeds, $\frac{\mathrm{d}x^*}{\mathrm{d}Q_{fm}}$, is a function of ϑ_i/θ_c. Then, as the distribution of ϑ_i is controlled by θ_{di}, the responsiveness of data consumption to download speeds informs the relative values of θ_{di} and θ_c.

However, download speeds may be endogenous to demand shocks due to congestion and the influence of local demand conditions on firms' investment decisions. Consequently, we instrument download speeds with (log) population densities, which influence experienced download speeds by changing the level of path loss. Another reason for using an instrument is attenuation bias. Our measures of download speeds are based on limited sample sizes (see Appendix C.2 for details), and therefore there is a degree of measurement error in the variable we use.

The heterogeneity in data valuations, θ_{di}, depends on the θ_{dz} parameter, which controls how

[30] Appendix B.2 explains how this number is calculated.

data valuations depend on income. We rely on the covariance between municipality-level median income and average data consumption to identify θ_{dz}.

While the imputed elasticity moment effectively identifies the average θ_{pi}, we need a moment to pin down the heterogeneity in price responsiveness. Variation in median incomes across markets helps identify how this parameter varies by income. We assume that the demand shocks $\boldsymbol{\xi}$ are uncorrelated with median incomes.

Finally, the unlimited voice characteristic is assumed to be uncorrelated with demand shocks in order to identify its coefficient, θ_v.

In summary, we have the following moments that we use to identify the distribution of preference parameters θ. Note that the moments are only evaluated for Orange plans since we only observe data consumption and plan-market shares for Orange. Because they are only evaluated for Orange plans, we demean the shocks in the moments below (i.e., we use $\xi_{jm}(\theta) - \theta_O$, where θ_O is a parameter that we estimate and has the interpretation of the average value of the Orange demand shocks).

Moments
$\mathbb{E}\left[-e_m^{ORG}(\theta) - 2.36\right] = 0$
$\mathbb{E}\left[DIV_m^{ORG,0}(\theta) - 0.036\right] = 0$
$\mathbb{E}\left[(\xi_{jm}(\theta) - \theta_O)\, inc_m^{med}\right] = 0$
$\mathbb{E}\left[\bar{x}_{jm}(\theta) - \bar{x}_{jm}\right] = 0$
$\mathbb{E}\left[(\bar{x}_{jm}(\theta) - \bar{x}_{jm})\, inc_m^{med}\right] = 0$
$\mathbb{E}\left[(\xi_{jm}(\theta) - \theta_O) \log(pop\_density_m)\right] = 0$
$\mathbb{E}\left[(\xi_{jm}(\theta) - \theta_O)\, \bar{d}_j\right] = 0$
$\mathbb{E}\left[(\xi_{jm}(\theta) - \theta_O)\, v_j\right] = 0$
$\mathbb{E}\left[\xi_{jm}(\theta) - \theta_O\right] = 0$

We use two-stage efficient GMM to estimate θ, searching for θ in an outer loop and solving for $\boldsymbol{\xi}(\theta)$ in an inner loop using the modified contraction mapping described in Appendix B.1.

One concern is that most mobile subscriptions involve 12- or 24-month contracts and cancellation is costly. In this context, it is natural to think that short-run substitution patterns may understate the long-run responsiveness to price and/or download speeds. We note that our estimation strategy does not rely on short-run changes in market shares and plan characteristics, which is the sort of estimation strategy that would be most vulnerable to this concern. Instead, we rely on cross-sectional variation in download speeds and market shares during a period in which prices were relatively stable; cross-sectional variation like this arguably approximates long-run responsiveness.

4 Industry Model

In this section, we present a model of data transmission and firm competition that endogenizes download speeds and prices. Section 4.1 presents the engineering model of data transmission. Section 4.2 then describes how download speeds and data consumption are simultaneously determined by the engineering model and demand model. Section 4.3 describes how firms choose the prices of mobile phone plans and the level of investment in infrastructure. Finally, section 4.4 explains how we estimate cost parameters.

4.1 Engineering Model

In this section, we lay out a formal model of how download speeds are determined by bandwidth allocations, investment in infrastructure, and the load imposed on a network by consumers. We rely on standard telecommunications engineering models and are particularly indebted to Błaszczyszyn, Jovanovicy and Karray (2014).

Section 4.1.1 explains the geometry of base stations and path loss, which is how signal strength declines with the distance of transmission. Next, in section 4.1.2, we consider the problem of serving a stochastic demand process and introduce a simple queuing theory model. Section 4.1.3 then discusses the economic implications of each component of the engineering model, explaining how the path loss of section 4.1.1 leads to economies of density, and how the queuing model of 4.1.2 leads to economies of pooling.

4.1.1 Base Stations and Path Loss

Channel capacity, $\bar{Q}_{fm}$, describes the rate at which network operator f can transmit data in municipality m, averaged across space. Channel capacity can also be understood as what delivered download speeds would be without congestion. In this section, we derive channel capacity from basic principles. Ultimately, channel capacity will be larger when firms operate more spectrum, and smaller when base stations serve larger areas because signals must travel farther (on average) as the area of a cell increases.

Network operators (not including virtual network operators) own and operate their own networks with no sharing of infrastructure. While *passive network sharing* (the sharing of the physical structure of base stations) is common, our cost function specification is in a sense robust to it, as we discuss below. During 2015, *active network sharing* (which occurs when equipment that transmits data is shared) occurred primarily in areas with low population density. Because we want to associate each firm's quality of service with its own investment decisions, we ultimately focus on the higher-density areas of France in our analysis. See Appendix C.3 for further discussion.

We assume that each municipality has homogeneous population density and that the full land area is divided into equally-sized hexagonal cells, so that each cell is identical for a given operator and municipality. We assume that each cell is served by a single base station transmitting an omni-directional signal. A crucial aspect of infrastructure is owned spectrum or bandwidth, B_{fm}, which measures the range of frequencies that firm f operates in municipality m.[31] Bandwidth is not a choice variable in our model, but it is an aspect of market structure that we vary in our counterfactual analysis.

The size of network operator f's cells in market m is characterized by R_{fm}, which is the cell radius (more precisely, a hexagonal cell's maximal radius, which is equal to its side length). We could also think of this choice variable as being the number of base stations in a given municipality, N_{fm}. The area served by each cell is $A_m/N_{fm} = \frac{3\sqrt{3}}{2}R_{fm}^2$, where A_m is municipality m's effective land area, and $3\sqrt{3}R_{fm}^2/2$ is the area of a regular hexagon with radius R_{fm}. We take for granted that firms will serve the municipality's full area.[32] Assuming full coverage is standard practice in recent engineering-based studies of mobile service provision in developed countries, reflecting the idea that quality, not coverage, is the relevant non-price characteristic that network operators now compete on in developed countries. We assume that the municipality's area can be divided into equally-sized hexagons, effectively ignoring municipality geometry and other spatially explicit details. Heterogeneity in municipality topography and other features that affect radio transmission can be captured in a municipality-level spectral efficiency parameter, introduced later in this section.

If a base station devotes its full bandwidth to serving a consumer at location ℓ, that consumer will receive download speed $B_{fm}q_{m\ell}$ in megabits per second (data transmission rates generally scale linearly with bandwidth used). We will soon introduce a precise function to describe how $q_{m\ell}$ depends on the consumer's location within the cell, but generally, $q_{m\ell}$ will be lower for consumers located further from the base station due to path loss, the phenomenon of signals losing power as they travel.

Suppose that data demand is uniformly distributed across space. Normalizing data demanded per unit area to unity (this is harmless as the demand rate per unit area would cancel out of equation 14 below), the total data demanded within a cell is equal to its area. We denote a cell's area by $A\left(R_{fm}\right) = \frac{3\sqrt{3}}{2}R_{fm}^2$.

To determine the amount of time consumers spend downloading data, we need to integrate

[31] French firms own the same spectrum in every municipality, but they do not operate their full holdings everywhere. The most important reason firms sometimes operate less than their full holdings appears to be that 4G was rolled out gradually—there are some municipalities where at least one of the firms does not yet operate any 4G spectrum. Our B_{fm} measures the frequencies that firm f actually uses in municipality m.

[32] When implementing the model empirically, we use an adjusted measure of land area because the raw land area may overstate the area that operators need to cover (at least with high quality download speeds) when large unpopulated areas are present. See section 2.4 above for details.

over $q_{m\ell}$. Specifically, for each unit of data downloaded, a consumer at location ℓ spends $(B_{fm}q_{m\ell})^{-1}$ seconds downloading. Integrating over consumers (or, equivalently, the cell's area, since consumers are uniformly distributed), the total time spent downloading is given by

$$\int_{\ell\in\mathcal{L}(R_{fm})} \frac{1}{B_{fm}q_{m\ell}} d\ell, \tag{13}$$

where $\mathcal{L}(R_{fm})$ is the set of locations composing a hexagon with radius R_{fm}.

To determine average download speeds, we divide total data downloaded by total time spent downloading. That is, we divide $A(R_{fm})$ by equation 13. This yields a harmonic mean of $q_{m\ell}$, multiplied by bandwidth:

$$\bar{Q}_{fm}(R_{fm}, B_{fm}) = \frac{B_{fm}A(R_{fm})}{\int_{\ell\in\mathcal{L}(R_{fm})} q_{m\ell}^{-1} d\ell}. \tag{14}$$

The above equation expresses *channel capacity*, capturing how feasible download speeds are influenced by the firm's choice of cell radius R_{fm} and its bandwidth B_{fm}. Importantly, channel capacity is not the same as *delivered* download speed. Below, we will model how delivered download speeds also depend on consumption (i.e., congestion) using queuing theory.

Next, we consider the individual download speed function $q_{m\ell}$, which gives download speed measured in bits per second (per Hertz of bandwidth):

$$q_{m\ell} = \gamma_m \log_2(1 + SINR_\ell), \tag{15}$$

where $SINR_\ell$ is the signal-to-noise-and-interference ratio, which we will explain below, and γ_m is a spectral efficiency parameter.

When the spectral efficiency parameter is set equal to unity ($\gamma_m = 1$), equation 15 represents the Shannon-Hartley Theorem (Shannon, 1948), which provides the theoretical upper bound on data transmission rates as a function of SINR. The Shannon-Hartley Theorem's bound is much higher than the data transmission rates typically achieved in practice. Actual rates of data transmission are affected by the encoding technology, topography, weather, and the presence of buildings and other physical barriers. Such factors vary by market, so we employ a market-specific spectral efficiency parameter γ_m. We calculate these spectral efficiency parameters to match our model's predicted delivered download speeds with observed download speeds. See Appendix A.2 for details.

This spectral efficiency parameter can absorb many aspects of the data transmission technology, and in particular, anything that affects the level of download speeds without affecting how they decline with distance. For instance, one might be concerned that our measure of spectrum includes all the frequencies owned by an operator, and therefore the frequencies used for downloads as well as uploads by mobile customers, but we are using this measure of bandwidth to model only download speeds. Operators could manage outgoing transmissions (downloads) and incoming transmissions (uploads) by using half of their spectrum for each (in practice, they use more sophisticated strategies). In this case, we could define B_{fm} as half of the operated bandwidth, but given that channel capacity scales linearly with bandwidth, setting $\gamma_m = .5$ would achieve the same effect. Avoiding interference between incoming and outgoing signals is just one of many factors that tends to make $\gamma_m < 1$. Ultimately, we find a mean value of γ_m equal to 0.165 across municipalities.

The signal-to-noise-and-interference ratio (SINR) is given by the ratio of signal power to the sum of noise and interference power:

$$SINR_\ell (R_{fm}) = \frac{S_\ell}{N + I_\ell (R_{fm})}, \tag{16}$$

where S_ℓ is signal power density (signal power per unit of bandwidth), N is noise power density, and $I_\ell (R_{fm})$ is interference density. Note that signal power S_ℓ depends on location due to path loss. Interference power $I_\ell (R_{fm})$ depends on both location and the cell's radius because cell size determines how far neighboring base stations are.

We assume that base stations transmit signals at the maximum power permitted by regulation. As the signal travels away from the base station, its power diminishes (path loss). We take this into account by using the Hata model of path loss (Hata, 1980), in which the signal power received by a consumer depends on their distance from the base station. In a vacuum, the signal power would be proportional to the squared inverse of the distance traveled. In telecommunications jargon, this is a *path loss exponent* of two. In the Hata model we use, the path loss exponent is 3.522, reflecting the fact that signal strength drops off more quickly as it travels along the Earth's surface than it would in a vacuum. See Appendix A.1.1 for the precise functional form of S_ℓ.

Noise power N is constant, and set equal to Johnson-Nyquist noise. Interference power I_ℓ is set equal to 30% of the signal power from the six adjacent cells. See Appendix A.1.2 for details regarding the units and formulas for the noise and interference variables.

4.1.2 Queuing and Congestion

Consumers' download requests do not arrive uniformly over time. This means that the channel capacity $\overline{Q}_{fm}$ derived above will not represent the actual delivered download speed in practice.

To derive a relationship between channel capacity and average delivered download speed, we follow Błaszczyszyn, Jovanovicy and Karray (2014) and Lhost, Pinto and Sibley (2015) and assume that download requests arrive according to a Poisson process and that download requests are served through a M/M/1 queue (a queuing system in which a single server serves jobs on a first-come, first-served basis). Then, the average download speed, Q_{fm}, will be

$$Q_{fm} = \overline{Q}_{fm} - Q^D_{fm}, \tag{17}$$

where Q^D_{fm} is the arrival rate of download requests among consumers served by the base station.[33] The data demand rate comes from the demand model and is provided explicitly in equation 19 below. Each of the terms in equation 17 should be understood as rates, i.e., as values measured in Megabits per second.

4.1.3 Economies of Scale

Our model allows for two sources of scale efficiencies: *economies of pooling* and *economies of density.*

Economies of Pooling It has long been recognized in the economics literature that "there are economies of scale in servicing a stochastic market" (Carlton, 1978).[34] In operations management, the same phenomenon has been referred to as the "Pooling Principle" (Cattani and Schmidt, 2005). Thus, we use "economies of pooling" to describe economies of scale that arise from consolidating bandwidth in the context of stochastic demand.

Mathematically, it is easy to see how economies of scale result from our queuing theory model. Equation 17 holds that the average delivered download speed corresponds to the difference between channel capacity and the download demand rate. Crucially, channel capacity (and potential download speeds) scale linearly with bandwidth. If two identical firms combine their bandwidth and customer bases (holding the data demand rate per customer fixed), then both

[33]To see this, suppose consumers submit data requests following a Poisson process with arrival rate λ. Each data request has size d. Then, our overall data demand rate is $Q^D = \lambda d$.

With channel capacity $\bar{Q}$, the requests that can be processed per second (service rate) is $\bar{Q}/d$. An M/M/1 queue with service rate $\bar{Q}/d$ and Poisson arrival rate λ will have a mean number of active users equal to $\frac{\lambda}{\bar{Q}/d-\lambda}$. (see Taylor, Karlin and Taylor (1998), pp. 548-549 for a derivation). This mean number of users includes times when the system is idle as well as when it is actively serving request.

Over time, the average rate of data transmission must be Q^D since throughput must equal demand. The average download speed experienced by users is equal to this throughput, Q^D, divided by the mean number of users, $\frac{\lambda}{\bar{Q}/d-\lambda}$:

$$\frac{Q^D\left(\bar{Q}/d-\lambda\right)}{\lambda} = Q^D\left(\frac{\bar{Q}}{d\lambda}-1\right) = Q^D\left(\frac{\bar{Q}}{Q^D}-1\right) = \bar{Q}-Q^D,$$

where the first equality distributes $1/\lambda$, the second substitutes $Q^D = \lambda d$, and the third simplifies.

[34]Robinson (1958) was perhaps the first to describe this phenomenon, under the heading of "Economies of Massed Reserves" (pp. 26-27). De Vany (1976) was an early application that used queuing theory to derive economies of scale. Mulligan (1983) shows formally how economies of scale result from queuing theory.

terms on the right-hand side of equation 17 would double. Consequently, delivered download speeds (the left-hand side) would also double.

One way in which combining queues improves efficiency is by avoiding inefficient waiting. For instance, in the context of grocery store customers queuing for checkout, having only one queue for two cashiers avoids the possibility that a customer could be queued for a busy cashier while the other cashier is idle.

Pooling has another advantage in the context of mobile telecommunications that relies on the role spectrum plays in data transmission. Returning to the example of grocery store checkout, combining queues may reduce idle registers and wait times, but it might not reduce the time needed to serve a customer after reaching the front of the queue—i.e., there may be no advantage to serving a customer with two cashiers and cash registers instead of one. However, in the context of data transmission, using more spectrum does allow for faster downloads (as if having two cashiers allowed customers to check out twice as fast).

While the queuing theory model we use is standard and relatively simple, it should be noted that it has strong implications regarding the extent of economies of pooling. Taken literally, our use of a single-server queue supposes that consumers are served on a first-come, first-served basis, with all of the base station's bandwidth being used by one customer at a time. Since channel capacity represents the average speed at which a base station can transmit data, when we increase channel capacity, we increase download speeds (at least for consumers being served rather than waiting in the queue).[35]

In principle, it is possible to attain pooling efficiencies without consolidating firms through active network sharing.[36] In practice, there was very little active network sharing during our study period, so we ignore this possibility for the purposes of our study. Active network sharing undermines quality differentiation, as well as incentives to invest in quality, potentially creating a free-riding problem. Thus, network sharing creates a trade-off between pooling efficiencies and investment incentives that represents an important topic for future work.

Economies of Density Due to path loss, captured by the $q_{m\ell}$ function, the closer users are to a base station, the more efficiently that station can serve them. Thus, if we increase the

[35] In practice, bandwidth is typically divided among multiple consumers, which might seem to call for a queuing theory model where bandwidth is divided into separate servers (or carriers, to use wireless telecommunications terminology). In such models, increases in operated bandwidth might not pass through to increases in bandwidth used to serve individual customers, and increases in channel capacity would not necessarily lead to proportional increases in download speeds. On the other hand, operators are able to assign large portions of bandwidth to individual customers, a practice known as carrier aggregation, and it would be inefficient if they never did so (e.g., when only one customer is downloading something within a cell at a given moment).

[36] Active network sharing, or radio access network (RAN) sharing, involves sharing the systems involved in data transmission.

density of users served by a firm while keeping constant the number of users per base station, users will, on average be closer to the base stations serving them, improving download speeds. Consequently, urban areas tend to be less costly to serve than rural areas (at a given quality level). Mapping this to our analysis, having fewer firms implies that individual firms face higher population densities. Consequently, at a given level of investment in terms of base stations per customer, a market with fewer firms will have less path loss and higher download speeds.

The economies of density can quantified by comparing the channel capacities arising from two network operators to the channel capacity arising from one network operator that has the same number of base stations as the two combined, but arranged into a hexagonal grid with smaller cell sizes. If each of the two operators have a cell radius R, then the combined operator would have a radius of $R/\sqrt{2}$. The combined firm enjoys higher channel capacity per cell due to decreased path loss, but the degree of improvement is very sensitive to the baseline cell radius. If the two-operator case has a radius of $R = 1$ km, then the single operator with the same number of base stations has a channel capacity that is just 0.1% larger (per unit of bandwidth operated). If $R = 5$ km, however, the combined operator would have a more substantial improvement of channel capacity, 19.4%. In our infrastructure data, the effective cell radii cover a range of values that includes both 1 km and 5 km (see Table 3), but they tend to be much closer to 1 km. This foreshadows one message from our counterfactual results: while economies of density can matter in principle, they have little impact for the typical cell sizes in our data. We revisit this discussion in Appendix D.3, where we simulate equilibria for different population densities.

Other studies of economies of density have focused on transportation costs (Holmes, 2011) or associated waiting times (Rosaia, 2023). The economies of density in our study relate to transmitting data with electromagnetic waves rather than transporting people or physical goods. One difference is that data transmission costs increase with distance not because of the time it takes, but because of the loss of signal quality. The economies of density that result are largely similar, however.

Other Scale Economies While our model incorporates the above sources of economies of scale, we note that there may be other sources of economies and diseconomies of scale. Perhaps the most important omitted source of positive economies of scale comes from backhaul, or the economies of scale in the physical telecommunications network that makes the traditional telecommunications market a natural monopoly. But we also may be ignoring important sources of diseconomies of scale, such as operational and managerial diseconomies of scale.

It should be noted that the engineering models we use necessarily imply scale efficiencies, and therefore there is no possibility that we would find that there are diseconomies of scale. In other words, this paper does not ask the question "are there scale economies in mobile telecommunications?" We take the scale efficiencies implied by engineering models for granted, and our aim is to quantify to what extent scale efficiencies are realized in equilibrium and how they trade-off with market power.

4.2 Simultaneity of Demand and Download Speeds

Our demand system describes how consumption depends on download speeds. The queuing theory model above describes how download speeds depend on consumption. We now consider how the engineering relationships described above come together with demand to simultaneously determine delivered download speeds, taking prices and infrastructure as given. Formally, we now consider the determination of download speeds, data consumption, and product market shares conditional on a vector of prices of mobile phone plans $\mathbf{P}$ and infrastructure variables $(\mathbf{R}_m, \mathbf{B}_m)$, where $\mathbf{R}_m$ and $\mathbf{B}_m$ are the stacked cell radii and bandwidths of the network operators.

In section 3.1, product market shares s_{jm} and average data consumption $\bar{x}_{jm}$ implicitly depend on prices and download speeds. Here, we make this dependence explicit, writing $s_{jm}(\mathbf{Q}_m, \mathbf{P}) = s_{jm}$ and $\bar{x}_{jm}(\mathbf{Q}_m, \mathbf{P}) = \bar{x}_{jm}$.

The total demand for downloads on network operator f's network over a month can be broken down into the product of three terms, which come from the demand component of our model:

$$X_{fm}(\mathbf{Q}_m, \mathbf{P}) = pop_m \sum_{j \in \mathcal{J}_f} s_{jm}(\mathbf{Q}_m, \mathbf{P}) \times \bar{x}_{jm}(\mathbf{Q}_m, \mathbf{P}), \tag{18}$$

where pop_m is the number of potential consumers in the market.[37]

The demand rate for downloads on network operator f's network is the total downloads serviced by operator f over a month, $X_{fm}(\cdot)$, distributed across time and across base stations. This rate is given by:

$$Q^D_{fm}(R_{fm}, \mathbf{Q}_m, \mathbf{P}) = \frac{X_{fm}(\mathbf{Q}_m, \mathbf{P})}{H \times N_{fm}(R_{fm})}, \tag{19}$$

where H is the number of seconds in a month and $N_{fm}(\cdot)$ is the number of base stations network operator f has in market m.[38]

[37] MVNOs use MNOs' infrastructure for their own plans. Therefore, in our empirical analysis, we incorporate the load that results from the plans offered by the MVNOs on the MNOs' networks. ORG, BYG, and SFR all allow MVNOs to use their infrastructure, and (lacking data on these relationships) we assume MVNO load is distributed equally among these three MNOs.

[38] In our empirical application and counterfactuals, we use $H = 31 \times 17 \times 3600$. That is, we try to capture download speeds during peak hours when most of the downloads occur, and we assume that days effectively consist of seventeen peak hours. In our traffic data (from OSIRIS), consumption rates display remarkably

Combining equations 14, 17, and 19, we have

$$\forall f = 1, \ldots, F: \qquad Q_{fm} = \bar{Q}_{fm}\left(R_{fm}, B_{fm}\right) - Q_{fm}^{D}\left(R_{fm}, \mathbf{Q}_m, \mathbf{P}\right). \tag{20}$$

Given prices and infrastructure variables, the vector of equilibrium download speeds $\mathbf{Q}_m^*$ is defined as the vector of values of Q_{fm} that solves equation 20. The download speed function $\mathbf{Q}_m^*\left(\mathbf{P}, \mathbf{R}_m, \mathbf{B}_m\right)$ describes equilibrium download speeds given prices and infrastructure variables.

4.3 Firm Competition

In this section, we present how firms choose prices and infrastructure to maximize profits. We can understand the endogenous determination of download speeds in the previous section as happening within each market m, with potentially different infrastructural variables in each market, $(\mathbf{R}_m, \mathbf{B}_m)$. However, prices are set nationally, so we will not introduce subscripts on the price vectors.

Firms set prices and infrastructure simultaneously in all markets in a static game. We consider the first-order conditions with respect to each competitive variable in turn.

4.3.1 Price Competition

Variable profits are given by

$$\left(\mathbf{P}_f - \mathbf{c}_f^u\right) \cdot \sum_m pop_m \mathbf{S}_{fm}^*\left(\mathbf{P}, \mathbf{R}_m, \mathbf{B}_m\right), \tag{21}$$

where $\mathbf{c}^u$ is the variable cost per customer, pop_m is the size of market m, and $\mathbf{S}_{mf}^*\left(\cdot\right)$ denotes the vector of product-level shares for phone plans offered by firm f in market m. This market share function is derived from the demand system and our download speed model (equation 20) as follows:

$$\mathbf{S}_{fm}^*\left(\mathbf{P}, \mathbf{R}_m, \mathbf{B}_m\right) = \mathbf{s}_{fm}\left(\mathbf{Q}_m^*\left(\mathbf{P}, \mathbf{R}_m, \mathbf{B}_m\right), \mathbf{P}\right),$$

where the $\mathbf{s}_{fm}\left(\cdot\right)$ corresponds to the stacked vector of firm f's phone plan-level market shares given by equation 7.

We assume that each firm chooses prices to maximize the variable profits expressed in equation 21. Note that equilibrium download speeds depend on price, so the first-order condition for

little variance during the waking hours of 8 a.m.–midnight, with between 4.7% and 6.3% of total consumption happening each hour. From 1 a.m.–7 a.m., data consumption is much lower, with each hour accounting for 0.6%-1.7% of daily consumption. Consumption is in-between these rates during the transition hours of midnight–1 a.m. and 7 a.m.–8 a.m. Thus, it is a reasonable approximation to assume that there are 17 identical peak hours during which all consumption occurs.

optimal price-setting must not only take into account the direct effect of lowering price on consumer demand, but also the indirect effect of endogenous download speeds. The indirect effect lowers price elasticities because as demand for firm f falls, its download speeds increase due to reduced network load, which has a positive effect on demand and thereby dampens the demand reduction. We discuss demand elasticities further in section 6.

4.3.2 Costs and Infrastructure Competition

Firms also decide on their infrastructural investments in each market, measured by R_{fm}. Infrastructure costs in market m are given by the following function:

$$C_{fm}\left(R_{fm}, B_{fm}\right) = c_{fm}^{s} \frac{A_m}{A\left(R_{fm}\right)} B_{fm}, \tag{22}$$

where A_m is the land area of market m, and c_{fm}^{s} captures costs per base station and unit of bandwidth (which may vary by network operator and by market), and $A\left(R\right) = 3\sqrt{3}R^2/2$ is the area of a hexagonal cell with radius R.

This cost function reflects the idea that the main costs associated with a base station are the electricity costs, the cost of installing antennas, and other costs that are proportional to the amount of bandwidth being operated. An advantage of this cost function is that, if we suppose that all firms operate at the same base station locations, then redistributing bandwidth among firms and/or changing the number of firms does not change the total costs incurred within the industry. Thus, this cost function shuts down a potential source of economies of scale associated with the duplication of fixed costs.[39]

This cost function also rules out any gains from passive network sharing. Because costs are proportional to bandwidth, firms would not change their total costs by combining their network resources at a given location. While our analysis does not explicitly incorporate passive network sharing, this does *not* lead us to overstate the case for consolidation. That is, one might worry that some of the predicted counterfactual efficiency gains from consolidation will be overstated because those efficiency gains can be realized among firms without consolidating. Because this source of cost savings does not exist in our baseline model, this is not a concern when interpreting our main counterfactuals.

That said, it is natural to think that there are some fixed costs associated with operating a base station, such as rents or setup costs, which are independent of the bandwidth being operated. To address this concern, we conduct robustness exercises with an alternative cost function that treats all infrastructure costs as fixed costs per base station (that is, we drop the

[39]See Peha (2017) for an analysis of economies of scale in mobile services coming from fixed costs per base station (without the economies of density and economies of pooling we consider).

B_{fm} term from equation 22). Appendix D includes results for this alternative cost function.

We can define market-level profits as follows:

$$\Pi_{fm}\left(\mathbf{P},\mathbf{R}_m,\mathbf{B}_m\right) = pop_m\left(\mathbf{P}_f - \mathbf{c}_f^u\right)\cdot \mathbf{S}_{fm}^*\left(\mathbf{P},\mathbf{R}_m,\mathbf{B}_m\right). \tag{23}$$

Finally, we can define the national profit function (including infrastructure costs) for each firm f:

$$\Pi_f\left(\mathbf{P},\mathbf{R},\mathbf{B}\right) = \sum_m \Pi_{mf}\left(\mathbf{P},\mathbf{R}_m,\mathbf{B}_m\right) - \sum_m C_{fm}\left(R_{fm},B_{fm}\right). \tag{24}$$

where $\mathbf{R}$ and $\mathbf{B}$ stack the infrastructure variables across markets.

Equation 24 defines the profit function for each firm, summing across all 589 markets. We assume that each firm unilaterally and simultaneously chooses a (national) price vector $\mathbf{P}_f$ and a vector of cell radii (a cell radius for each municipality) $\mathbf{R}_f$ to maximize their profits, taking other firms' price and infrastructure choices as given, yielding equilibrium prices $\mathbf{P}^*$ and radii $\mathbf{R}^*$.

4.4 Cost Estimation

Using our demand estimates and the industry model presented above, we infer firms' costs based on the assumption that firms set prices and invest in infrastructure to maximize profits.

We model operators as playing a static investment game. In reality, operators upgrade their networks over time. Furthermore, the demand for mobile data services has grown rapidly, accompanied by rapid technological innovations. Abstracting away from these dynamic features, we aim to only capture the long-run trade-offs presented by investments in mobile telecommunications infrastructure.

We argued in section 3.2.2 that our cross-sectional approach to demand can capture the long-run trade-offs that consumers face. We estimate firms' cost parameters by making the marginal cost of improving infrastructure match the marginal revenue from improved infrastructure implied by the demand model. Thus, our cost model should be understood as a long-run cost model, in the sense that it is the cost function that rationalizes the investments we see if they had been made in a one-time static game that permanently determines infrastructure.

There are two types of cost parameters to be estimated: c_j^u, the cost per user of phone plan j, and c_{fm}^s, the cost per base station per unit of bandwidth in market m for network operator f.

4.4.1 Costs per User

From equation 21, the first-order condition from the price setting game is

$$\sum_m pop_m \mathbf{S}^*_{mf}\left(\mathbf{P}, \mathbf{R}_m, \mathbf{B}_m\right) + \left(\sum_m pop_m J_f \mathbf{S}^*_{mf}\left(\mathbf{P}, \mathbf{R}_m, \mathbf{B}_m\right)\right)\left(\mathbf{P}_f - \mathbf{c}^u_f\right) = 0, \tag{25}$$

where J_f represents the Jacobian operator with respect to $\mathbf{P}_f$.

Therefore, an estimate of per-user marginal costs is given by

$$\hat{\mathbf{c}}^u_f = \mathbf{P}_f + \left(\sum_m pop_m J_f \mathbf{S}^*_{mf}\left(\mathbf{P}, \mathbf{R}_m, \mathbf{B}_m\right)\right)^{-1} \sum_m pop_m \mathbf{S}^*_{mf}\left(\mathbf{P}, \mathbf{R}_m, \mathbf{B}_m\right). \tag{26}$$

4.4.2 Infrastructure Costs

Given the demand estimates and the model of how the infrastructure variables $(\mathbf{R}, \mathbf{B})$ map to delivered quality, we can simulate how equilibrium revenues change as the infrastructure is changed. Intuitively, we can measure the marginal revenue of infrastructure, and this allows us to infer its marginal cost.

Formally, we use numerical differentiation to approximate the marginal operating income with respect to cell radius for each market based on a 0.01 km change in cell radius:

$$MR^R_{fm}\left(\mathbf{R}_m, \mathbf{B}_m\right) = \frac{\Pi_{fm}\left(\mathbf{P}, \left(R_{fm} + 0.01, \mathbf{R}_{-f,m}\right), \mathbf{B}_m\right) - \Pi_{fm}\left(\mathbf{P}, \left(R_{fm} - 0.01, \mathbf{R}_{-f,m}\right), \mathbf{B}_m\right)}{0.02}. \tag{27}$$

Note that these profit functions are defined in terms of the equilibrium download speeds that result from the infrastructural investment and prices. Thus, the above expressions for marginal operating income should be understood as implicitly taking into account how quality changes as infrastructural investment changes. Furthermore, note that profits Π_{fm} include per-user costs; hence our use of "operating income" rather than "revenue."

Next, assuming that infrastructural investments are chosen to maximize profits, we can use the marginal operating income above to recover the remaining cost function parameters. Specifically, the marginal cost of increasing R_{fm} is obtained by differentiating the cost function in equation 22. For each firm and municipality, our estimated cost parameter c^s_{fm} sets this marginal cost equal to the marginal operating income in equation 27.

Table 4: Demand Parameter Estimates

Estimates	$\hat{\theta}_{p0}$	$\hat{\theta}_{pz}$	$\hat{\theta}_v$	$\hat{\theta}_O$	$\hat{\theta}_{d0}$	$\hat{\theta}_{dz}$	$\hat{\theta}_c$	$\hat{\sigma}$
	−1.859	−0.727	0.460	2.375	0.597	0.335	1.405e−4	0.683
	(0.687)	(0.221)	(0.180)	(0.272)	(0.380)	(0.039)	(4.108e−5)	(0.111)
Willingness to pay for				10th %ile	30th %ile	50th %ile	70th %ile	90th %ile
1 GB plan → 4 GB plan				2.73 €	3.16 €	3.52 €	3.95 €	4.77 €
unlimited voice				3.81 €	5.53 €	7.44 €	10.43 €	20.19 €
10 Mbps → 20 Mbps				1.90 €	2.39 €	2.84 €	3.41 €	4.68 €

Note: Rather than estimating θ_c and σ directly, we estimate $\log(\theta_c)$ and $\log\left(\frac{\sigma}{1-\sigma}\right)$ since $\theta_c > 0$ and $0 < \sigma < 1$, respectively. We derive the associated standard errors using the Delta Method. The estimates for θ_c are in scientific notation. Measures of willingness to pay implied by the demand parameter estimates are provided in the lower section. Willingness to pay for upgrading from a 1 GB phone plan to a 4 GB one is calculated holding download speeds fixed at the median speed observed in the data. Willingness to pay for upgrading from a download speed of 10 Mbps to 20 Mbps is calculated for a 10 GB phone plan.

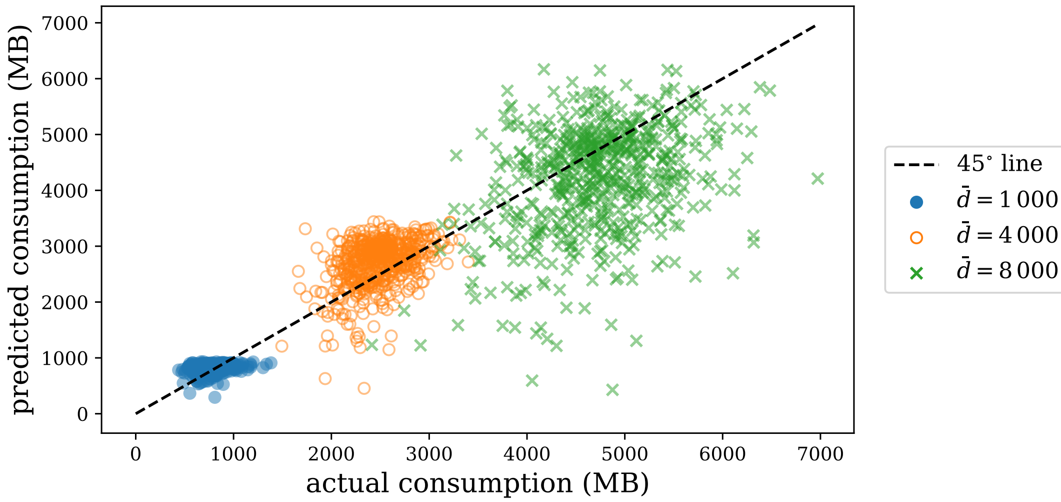
5 Results

5.1 Demand Estimates

Demand parameter estimates are presented in Table 4. Price sensitivity is decreasing in income. Meanwhile, the data utility parameter (θ_{di}) is increasing in income, which implies an inverse relationship between income and the value of data consumption (since the mean of ϑ_i is $1/\theta_{di}$), suggesting a higher opportunity cost of time spent downloading for higher income individuals. To help interpret the parameter estimates, Table 4 converts these estimates into willingness to pay for certain phone plan characteristics across income percentiles. Our estimates suggest that higher income consumers are willing to pay considerably more for phone plans with larger data limits, unlimited voice minutes, and faster download speeds than lower income consumers. These high-income consumers are willing to pay more despite an inverse relationship between income and the value of data consumption since they are less price sensitive.

Figure 6 examines how well our model predicts actual data consumption by plotting predicted and actual average data consumption across markets for three Orange phone plans with different data limits. Our estimated model accurately predicts the average level for each data limit category. This result is non-trivial, for the fraction of the data limit used varies across categories, and our model does not include parameters that allow us to directly fit this usage level by category. Instead, to generate different consumption levels by plan, we must have heterogeneity in preferences that maps to heterogeneity in bliss points, with consumers self-selecting into plans based on these bliss points and their price responsiveness.

Figure 6: Predicted vs. Actual Average Data Consumption



Note: This figure presents a scatter plot at the market-level of actual average data consumption versus the consumption predicted by our demand parameter estimates for three Orange phone plans. The dashed line is a 45-degree line.

Table 5: Cost Estimates

Per-user costs		$\bar{d} < 1\,000$	$1\,000 \leq \bar{d} < 5\,000$	$\bar{d} \geq 5\,000$
$\hat{c}^u_j$		(in €)	(in €)	(in €)
		4.95	10.33	20.53
		(0.65)	(0.66)	(2.02)
Per-base station costs	Orange	SFR	Free	Bouygues
$\hat{C}_f$	(in €)	(in €)	(in €)	(in €)
	182 197	140 556	142 792	201 733
	(60 698)	(40 035)	(46 587)	(67 896)

Note: Per-user cost values represent the average estimated cost per user, calculated across all products within the data limit range specified in the corresponding column. Values in parentheses are the average standard errors. Per-base station costs are in long-run per base station terms (rather than monthly per base station-units of bandwidth terms, as introduced in the text). To create long-term per base station costs, we first use the monthly per base station-units of bandwidth costs that we recover from our estimates and multiply them by 75. This corresponds to the estimated monthly cost of a base station operating 75 MHz of bandwidth, which corresponds to the average amount of bandwidth per firm across markets. To recover the cost of long-lived base stations from our estimates based on monthly profits, we assume a monthly discount rate of 0.5%. The above results are therefore 201 times the per-base station costs we recover. Values in parentheses are standard deviations of the distribution of estimated costs across markets (not standard errors of the estimates).

5.2 Cost Estimates

Table 5 reports per-user and per-base station cost estimates. These estimates are recovered by inverting prices and radii, as described in section 4.4. Estimated per-user costs increase considerably with the size of the data limit: small data limit plans (those with data limits less than 1 000 MB) have an average per-user cost of 4.95 €, medium-sized data limit plans (between 1 000 and 5 000 MB) have an average of 10.33 €, and large data limit plans (over 5 000 MB) have an average of 20.53 €.

Estimates of the average sunk cost of a 75 MHz base station are between 140 000 € and about 200 000 €, depending on the firm. As described in the table's notes, we converted from monthly costs to one-time sunk costs by assuming a 0.5% monthly discount rate. Using this conversion of our estimates, we can compare these costs to industry estimates. Our estimates fit squarely within the range of estimates of the costs of constructing large base stations, generally between 50 000 € and 250 000 € (Nikolikj and Janevski, 2014; Analysys Mason, 2015; Smail and Weijia, 2017). Note that per-base station costs do vary across markets. For instance, the estimated standard deviation in the cost per base station across markets for Orange is 61 000 €, reflecting differences in land acquisition costs, labor costs, and other factors.

6 Counterfactual Simulations

Our framework can address questions of market structure, where market structure is defined by a vector that describes the number of firms and how much spectrum is allocated to each. In section 6.1, we consider the trade-off between market power and scale economies, investigate the optimal number of firms, and consider how to allocate spectrum among firms. Then, in section 6.2, we investigate the marginal value of spectrum allocated to mobile telecommunications and find that the marginal contribution to consumer surplus far exceeds firms' willingness to pay. In section 6.3 we simulate mergers between MNOs in France, focusing on the short-run where infrastructure is held fixed.

In all of our counterfactual simulations, we take market structure as exogenous. This may seem at odds with the fact that spectrum allocations are endogenously determined through auctions. However, the recent literature has expressed increasing concerns about whether spectrum auctions lead to efficient post-auction outcomes, especially given that achieving efficiency in a spectrum auction is not as simple as achieving an efficient allocation of resources among auction participants; the regulator also cares about outcomes for consumers, not just the bidding firms.[40]

Economists have considered how to design auctions that take into account post-auction market structure (Cramton et al., 2011; Rey and Salant, 2017). Our framework can complement this

[40] For instance, Jehiel et al. (2003) argue that features of multi-unit auction design that lead to more efficient allocation among auction participants can exacerbate post-auction market structure concerns. Eső, Nocke and White (2010) point out that increasing capacity, when capacity is efficiently allocated from the perspective of firms, can actually lead to a reduction in consumer welfare. Ershov and Salant (2022) present empirical evidence that some spectrum auctions have adverse impacts on market structure.

paradigm; the outcome of a spectrum auction is a market structure, and our framework provides an understanding of how market structure maps to equilibrium outcomes of concern to regulators, such as prices, download speeds, and welfare.

6.1 Market Power, Scale Efficiencies, and Bandwidth Allocations

In this section, we solve for equilibria using a representative commune in which each firm offers two mobile phone plans: one with a moderately low data limit of 1 GB and one with a very high data limit (in 2015) of 10 GB, which is the largest of the representative contracts. Focusing on a representative commune allows us to avoid solving for equilibrium investment levels in all markets, which would be computationally impractical because, with national prices, optimal prices and investment levels are interdependent across markets. The representative commune that we construct has a population density equal to the population-weighted mean population density in France (2 792 people / km^2).[41] Furthermore, this representative commune has an income distribution matching the overall income distribution in our sample, available bandwidth equal to the population-weighted mean of the sum of frequencies operated in each market, a spectral efficiency parameter equal to the population-weighted mean, and base station cost parameters equal to the mean across markets and operators, weighted by population.[42] Both phone plans have an unlimited voice allowance, demand shocks equal to the average of those estimated for the Orange phone plans ($\hat{\theta}_O$), and per-user costs equal to the average of the estimated per-user costs for similar phone plans (those with $\bar{d}_j < 5$ GB for the low data limit plan and those with $\bar{d}_j \geq 5$ GB for the high data limit one).[43]

One might worry about whether the focus on a representative commune yields results that hold for France when considered as a whole. In particular, does the representative commune, with its moderate population density, yield the same optimal number of firms that we would find for France, which comprises a mixture of high and low population-density areas? In Appendix D.3, we find that the optimal number of firms is basically invariant to population density. The optimal number of firms for the representative commune is also optimal for high- and low-density areas, and, consequently, for France as a whole.

[41] Equivalently, this is the mean population density integrating over people, or the contraharmonic population density integrating over space. We focus on the population density experienced by people because the raw population densities (population density divided by land area) of countries like the US or France are much lower than what most residents experience, and mobile network operators typically invest less intensively in sparsely-populated areas. Raw population densities are therefore not representative.

[42] For these results we use the cost specification presented in section 4.3.2. Appendix D.1 presents our counterfactual results under an alternative specification in which base station costs do not scale with bandwidth. While some counterfactual exercises look very similar, this specification does favor more consolidation; however, we believe that these gains are substantially overstated, as discussed in the appendix section.

[43] Equilibrium multiplicity is a potential concern. To address this issue, we employ a wide range of starting values when searching numerically for an equilibrium. While our algorithm does not always converge, when it does converge, it always converges to the same equilibrium. This is the case even for the asymmetric equilibria of section 6.1.2.

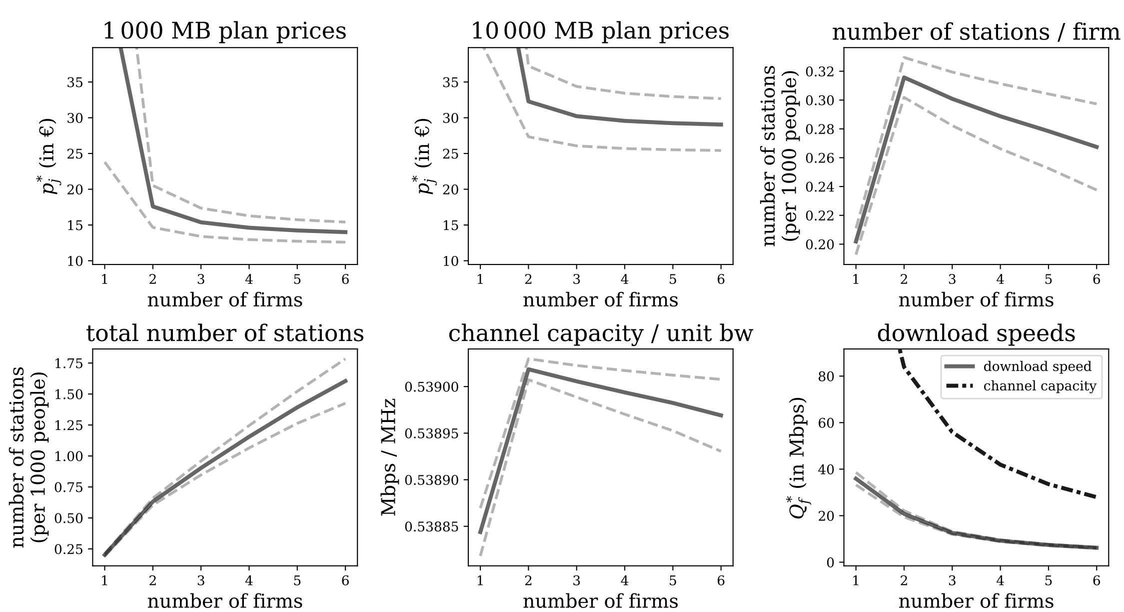
6.1.1 Optimal Number of Firms

In this section, we explore the trade-off between market power and economies of scale by considering the optimal number of firms in a static equilibrium. Fewer firms grants each firm more market power but also results in a higher density of consumers (lowering average path loss) and more pooling of consumers (improving the allocation of network resources). Given the gradual nature of network deployment in the industry, this exercise cannot hope to capture the short-run impacts of a potential merger; instead, we aim to capture the long-run trade-offs associated with consolidation. In section 6.3 we explore the short-run impacts of mergers between MNOs in France.

The optimal number of firms depends on how equilibrium prices, investment, and download speeds vary based on the number of firms. Figure 7 displays these endogenous variables for symmetric equilibria with between one and six firms. Total bandwidth available to the industry is divided equally among the firms, which optimally set prices and investment levels. That is, each firm owns and operates spectrum $B_{fm} = B_0/n$, where B_0 is the total bandwidth available to the industry, and n is the number of firms.

In this exercise we focus on symmetric equilibria involving firms with identical spectrum endowments. When evaluating the optimal number of firms, one reason for our focus on symmetric spectrum holdings is that asymmetric spectrum allocations can be inefficient (Peha, 2017). In section 6.1.2, we consider asymmetric bandwidth allocations.

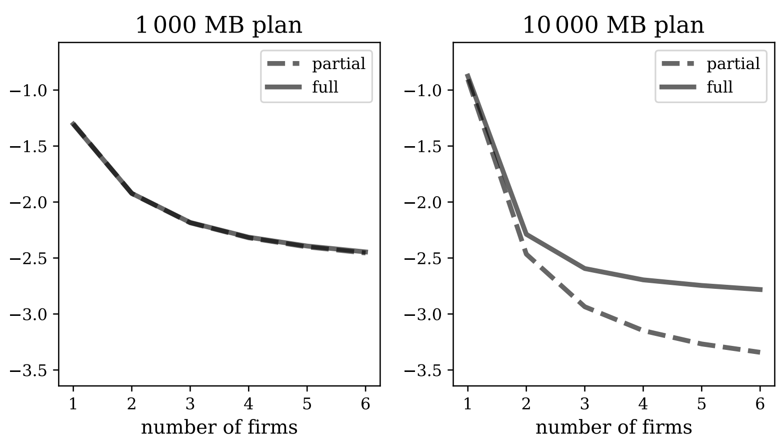
Figure 7: Counterfactual Prices and Qualities



Note: Channel capacity is per base station. Download speeds are the average speed of transmission received by a user, including wait times. Dashed lines represent 95% confidence intervals.

Equilibrium prices are declining in the number of firms but remain well above per-user marginal costs, which are 8.18 € and 20.53 € for the low and high data limit plans, respectively. Prices determine to which plan consumers subscribe and therefore the amount of data consumed. As a firm lowers its price, it attracts more customers, causing the load on its network to increase, lowering download speeds. Lower download speeds dampen the appeal of the lowered price. The relevant elasticity for the purpose of setting optimal prices, therefore, involves a full derivative that takes into account this indirect effect of changing prices on download speeds. Figure 8 displays how this indirect effect from download speeds influences optimal price-setting behavior by displaying two elasticities: *partial price elasticities* and *full price elasticities*. Partial price elasticities are the price elasticities that hold the quality of service fixed, evaluated at equilibrium prices. Full price elasticities, on the other hand, allow the quality of service to adjust with the price. Note that the full price elasticities decline less with the number of firms than the partial elasticities. This divergence stems from the intensification of the indirect quality effect as the number of firms grows. When there are many firms, a firm's own capacity is small relative to the number of consumers that it can potentially attract from other firms. This causes the quality of service to degrade more sharply for a given price reduction.

Figure 8: Full and Partial Price Elasticities



Note: This figure shows price elasticities for the two types of phone plans. Partial elasticities are derivatives in which download speeds are held fixed. Full elasticities take into account how download speeds change endogenously as prices change. Price elasticities are evaluated at the equilibrium prices and quantities.

Our cost specification holds that the cost per firm per base station, $c^s_{fm}B_{fm}$, is proportional to a firm's bandwidth allocation. As we assume that a firm's bandwidth allocation is inversely proportional to the number of firms, the cost per base station in these counterfactuals is $c^s_{fm}B_0/n$. Then, in a symmetric equilibrium in which each firm builds N_{fm} base stations, the industry-wide expenditure on infrastructure is $N_{fm}c^s_{fm}B_0$. Therefore, the number of base stations per firm—the third panel of Figure 7—is proportional to total industry expenditure on infrastructure.

Investment patterns display a non-monotonic relationship in the number of firms. Moving from monopoly to duopoly, the number of base stations for each firm increases (alternatively, the cell radius decreases). Increasing the number of firms beyond two, however, decreases investment per firm: for each increase in the number of firms, each firm builds fewer base stations (increasing the cell radius).

Despite this non-monotonicity in investment, download speeds are always decreasing in the number of firms. Demonstrating economies of scale, download speeds are higher under a monopoly than a duopoly, even though a monopolist deploys fewer base stations.

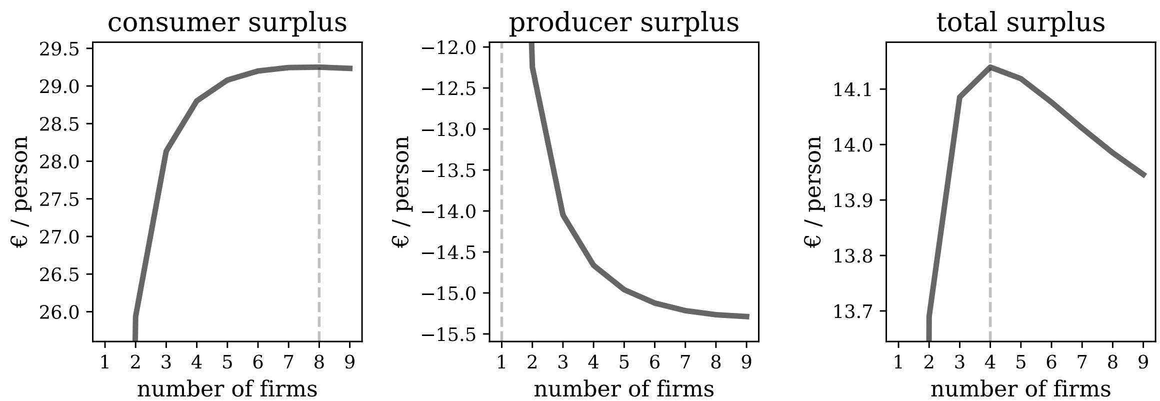
Closer inspection reveals that these economies of scale are driven largely by economies of pooling, rather than economies of density. Path loss will reduce channel capacity per unit of bandwidth, and when firms invest in more base stations, those base stations will serve closer customers, reducing path loss. As expected, channel capacity per unit of bandwidth follows the same shape as the number of base stations per firm, but note the scale of the graph for channel capacity per unit of bandwidth; the differences are trivial. In other words, firms are not seeing significant gains in data transmission by avoiding path loss.

In contrast, economies of pooling have a large impact. We see that channel capacity is roughly inversely proportional to the number of firms, which is driven by channel capacity's proportionality to bandwidth operated (see equation 14 and the fact that total available bandwidth is being spread across the firms, i.e., $B_{fm} = B_0/n$).

With both prices and quality declining in the number of firms, the optimal number depends on the trade-off between price and quality. Figure 9 considers welfare compared to the monopoly case as the number of firms varies. We find that the optimal number of firms is four in terms of total surplus, and eight in terms of consumer surplus, which is currently the relevant barometer for antitrust policy. We present here consumer surplus calculated without including the logit error terms (ε_{ij}). This means that our consumer surplus results reflect differences in prices and download speeds rather than being mechanically driven by the number of firms.

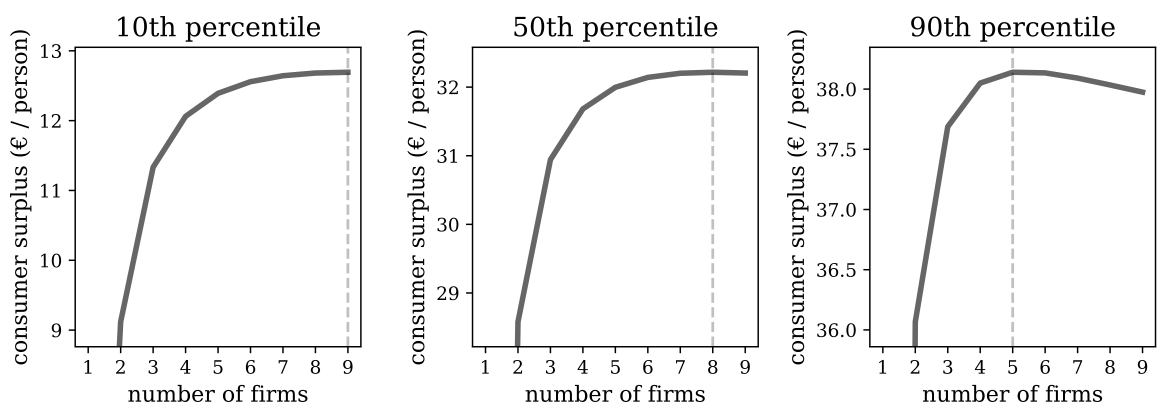
However, as figure 10 illustrates, consumers do not agree on the optimal number of firms. We plot welfare for various income deciles against the number of firms for our preferred specification. While consumer surplus is increasing in the number of firms for most consumers (up to eight or nine firms), the optimal number of firms for high-income consumers is five.

Figure 9: Counterfactual Welfare



Note: This figure displays measures of welfare as a function of the number of firms. Welfare is measured in euros per capita relative to the monopoly case, so for each plot the welfare value at 1 firm is 0. Dashed vertical lines indicate the number of firms that maximizes that measure of welfare.

Figure 10: Counterfactual Welfare by Income Level



Note: Welfare is measured in euros per capita relative to the case of monopoly. Dashed vertical lines indicate the number of firms that maximizes that measure of welfare.

6.1.2 Asymmetric Spectrum Allocations

How should spectrum be allocated within the industry, and what is the impact of asymmetries in this allocation? The previous exercise of determining the optimal number of firms considered the case in which firms are symmetric. In practice, MNOs tend to differ in terms of their plan offerings, costs, and spectrum allocations. We consider in this subsection the impact that an asymmetric bandwidth allocation has on equilibrium outcomes and welfare. In section 6.3 we conduct a merger simulation in which we allow for asymmetries in these other dimensions based on the asymmetries observed in France.

Here, we consider a three-firm equilibrium with the same setup as in section 6.1.1 but with asymmetric bandwidth allocations. Specifically, rather than allocating one-third of the total bandwidth to each of the three firms as in the previous exercise, we allocate half of the bandwidth to one firm and one-quarter each to the other two firms. This allocation is equivalent to the one that would result from two firms in the symmetric four-firm case merging their spectrum holdings.

Table 6: Three-firm Equilibria under Different Spectrum Allocations

	firm's bandwidth	1 000 MB plan price (in €)	10 000 MB plan price (in €)	download speed (in Mbps)
symmetric allocation				
equal allocation firm	$\frac{1}{3}B$	15.377 (1.010)	30.222 (2.120)	12.589 (0.317)
asymmetric allocation				
large allocation firm	$\frac{1}{2}B$	15.498 (1.059)	30.313 (2.149)	15.473 (0.491)
small allocation firm	$\frac{1}{4}B$	15.319 (0.984)	30.187 (2.103)	10.949 (0.237)

	Δ CS (in €/person)	Δ PS (in €/person)	Δ TS (in €/person)
symmetric allocation	−0.668 (0.154)	0.614 (0.118)	−0.054 (0.044)
asymmetric allocation	−0.701 (0.161)	0.625 (0.122)	−0.076 (0.047)
difference	0.032 (0.007)	−0.010 (0.003)	0.022 (0.004)

Note: Rows correspond to firms' phone plan characteristics under either a symmetric spectrum allocation or an asymmetric one (for which there are two types of firms—one with $\frac{1}{2}$ of the total bandwidth and two with $\frac{1}{4}$ of the total bandwidth). The second set of results presents the surplus relative to the four-firm symmetric case. The final row presents the difference between the symmetric and asymmetric values.

In line with Peha (2017), our results suggest that asymmetric spectrum allocations are inefficient. Table 6 presents equilibrium prices, download speeds, and welfare under the symmetric and asymmetric allocations. Unsurprisingly, for the firm with more spectrum, download speeds in the asymmetric allocation case are faster than in the symmetric case, and the reverse is true for the firms with low spectrum holdings. Reflecting these differences in download speeds, compared to the symmetric case, phone plan prices are higher for the firm with more bandwidth and lower for those with less. While heterogeneity in plan characteristics may be beneficial with consumer heterogeneity (e.g., more price-sensitive consumers can benefit from lower prices at the cost of slower download speeds), we find that both consumer and total surplus are lower under the asymmetric allocation than the symmetric one (producers, meanwhile, benefit in the aggregate from the asymmetric allocation).

6.2 Value of Spectrum to the Industry

Regulators such as the FCC in the US and ARCEP and ANFR in France are tasked with bandwidth allocation. This involves determining which industries (and firms) are allowed to operate which frequencies of electromagnetic spectrum and for what purposes. It is therefore crucial for such agencies to understand how allocating bandwidth to mobile telecommunications affects social welfare. For a regulator to allocate the optimal amount of spectrum to mobile telecommunications, they must understand both the marginal social value of spectrum in mobile telecommunications and its opportunity cost (the marginal value for other purposes). Our model provides quantification of the value of spectrum in this industry; quantifying its opportunity cost calls for a model of other industries and is beyond the scope of this paper.

In this section, we quantify how allocating more bandwidth to the telecommunications industry affects firm profits, consumer welfare, and total surplus. When a firm receives a larger bandwidth allocation—holding prices and infrastructure fixed—its download speeds increase and it gains market share. The derivative

$$\frac{\mathrm{d}\Pi_f \left(\mathbf{R}^* \left(B_f, \mathbf{B}_{-f}\right), \left(B_f, \mathbf{B}_{-f}\right)\right)}{\mathrm{d}B_f} \tag{28}$$

captures how a firm's profit changes when just that firm receives a larger bandwidth allocation, taking into account how equilibrium investment and prices respond. This value captures an individual firm's willingness to pay for more bandwidth at the margin. The derivative

$$\frac{\mathrm{d}\Pi_f \left(\mathbf{R}^* \left(B_f, B_{f'}, \mathbf{B}_{-f,f'}\right), \left(B_f, B_{f'}, \mathbf{B}_{-f,f'}\right)\right)}{\mathrm{d}B_{f'}} \tag{29}$$

captures the impact of this increase in allocated bandwidth on *other* firms' profits.[44]

In a simple spectrum auction, the firms' bids will be related to the difference between these two expressions. A firm's bid reflects its own gain in profits from the increased bandwidth should it win the auction relative to losing the auction and the spectrum being allocated to another firm.

A regulator's decision of whether to allocate spectrum to mobile telecommunications should be based not on the firms' bids, however, but on the marginal social value of allocating the bandwidth to the industry (compared to the marginal social value of allocating it to other industries and purposes). This marginal social value is captured by the following two derivatives. The first derivative,

$$\frac{\mathrm{d}\Pi_f \left(\mathbf{R}^* \left(B\mathbf{1}\right), B\mathbf{1}\right)}{\mathrm{d}B}, \tag{30}$$

captures how the equilibrium profit of an individual firm changes when *all* firms are allocated

[44] We are assuming, as in section 6.1.1, that firms are symmetric (prior to changing bandwidth allocations), so the identity of f' does not matter so long as $f' \neq f$.

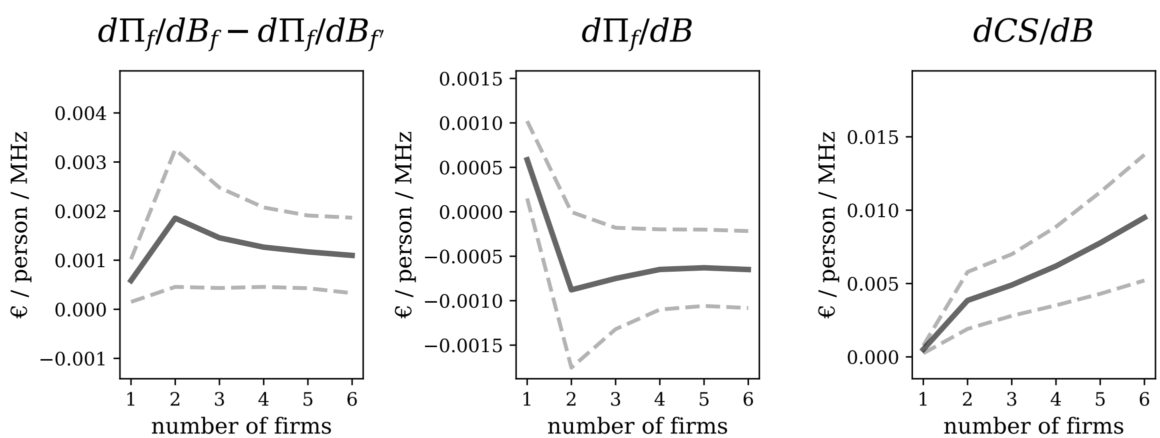
more bandwidth. The second derivative,

$$\frac{\mathrm{d}CS\left(\mathbf{R}^{*}\left(B\mathbf{1}\right), B\mathbf{1}\right)}{\mathrm{d}B}, \tag{31}$$

captures how consumer surplus changes as all firms are allocated more bandwidth.

Our framework and estimates allow us to calculate each of these values. This allows us to not only calculate the marginal social value of allocating more bandwidth to the industry, but also to compare that value to the difference between expressions 28 and 29. Since spectrum auctions provide a signal of the marginal willingness to pay for spectrum (rather than it being allocated to a rival), this comparison sheds light on the similarity between that willingness to pay abd the social value relevant to a regulator allocating spectrum across industries and uses.

Figure 11: Bandwidth Derivatives



Note: Derivatives are evaluated at the symmetric equilibrium values. Values are rates per month. The derivative of own profits with respect to another firm's bandwidth ($d\Pi_f/dB_{f'}$) is undefined in the monopoly case. In the first subplot, therefore, what is reported in the case of only one firm is simply the derivative of own profits with respect to own bandwidth ($d\Pi_f/dB_f$). Dashed lines represent 95% confidence intervals.

We compute the marginal value of spectrum based on symmetric equilibria as in section 6.1.1 As Figure 11 shows, with four firms, the firm's willingness to pay for additional bandwidth (the left panel) is about five times less than what a unit of bandwidth allocated to the industry would add to consumer surplus (the right panel). This reflects the importance of using a structural model such as ours to quantify the social value of bandwidth. While auctions may allow us to observe signals of operators' willingness to pay for spectrum, such measures may be far lower than the social value of spectrum.[45]

[45] Of course, a regulator seeking to maximize total surplus would also need to consider the middle panel, but these values are small relative to the right one since firms compete away the surplus from additional bandwidth, so the point that the value of additional bandwidth is many times larger than that captured by spectrum auctions still stands.

In late 2015, the same time as our analysis, France auctioned off 60 MHz of spectrum (20-year licenses) in the 700 MHz band that was previously used by television broadcasters. The auction raised 2.8 billion euros, which, dividing by France's population at the time (66.55 million), results in 0.70 €/person/MHz. At four symmetric firms, we find that a firm's profits are increased by 0.00127 €/person/MHz/month. At a 0.5% monthly discount rate, this yields a willingness to pay for spectrum of 0.25 €/person/MHz. While this suggests that our value for a firm's willingness to pay is on the low end, it may be that firms' bids reflect an expectation that spectrum's value will grow over time. Once again using a 0.5% monthly discount rate again, the marginal consumer surplus from an additional unit of spectrum in a four-firm equilibrium is 1.24 €/person/MHz. Therefore, whether we consider firms' willingness to pay as implied by the model, or as implied by the bids in the 2015 auction, the marginal social value of spectrum exceeds what firms are willing to pay.

6.3 Short-Run Merger Analysis

Our comparative statics with respect to the number of firms in section 6.1 should be interpreted with caution when extrapolating to merger analysis. Because those counterfactuals involve static equilibria, they certainly cannot capture the short-run impacts, for infrastructure cannot be rearranged instantaneously and costlessly in response to a change in market structure. Moreover, since firms have asymmetric bandwidth allocations and offered plans, mergers between these firms will also lead to asymmetries across these variables.

In this section, we consider the impact of mergers between two MNOs in the short-run. We allow ownership structures and the prices they imply to change; however, we keep product sets and infrastructure fixed. Presumably, prices and the response from consumers can change more quickly than investments in base stations or plan offerings. We model a merger between two MNOs f and f' as the merged MNO, $\tilde{f}$, offering plans $\mathcal{J}_{\tilde{f}} = \mathcal{J}_f \cup \mathcal{J}_{f'}$, (with each product maintaining the same ξ_{jm} and c_u).[46]

MNOs frequently co-locate base stations. Following a merger base station locations are fixed in the short-run. If the two merging firms had base stations in the same location, the merged firm's base station would be no closer to the customers it serves (resulting in no economies of density). We therefore use as the number of base stations of the merged firm, $N_{\tilde{f}m}$, the number of *separately* located base stations of the two firms. If we observe in the data that both MNO f and f' have a base station in the same location, we only count that location once, so $\max\{N_{fm}, N_{f'm}\} \leq N_{\tilde{f}m} \leq N_{fm} + N_{f'm}$.

[46] These mergers result in highly asymmetric firms. Over time, operators will re-optimize their networks, plan offerings, and bandwidth holdings. Assessing whether such adjustments maintain or reduce asymmetries requires a dynamic model, and potentially a model that endogenizes bandwidth holdings, which we leave for future work.

The merged firm has access to the bandwidth holdings of the two pre-merger firms, resulting in economies of pooling. We use for the merged firm's bandwidth an upper bound on its effective bandwidth: the sum of the two merged MNOs' allocations ($B_{\tilde{f}m} = B_{fm} + B_{f'm}$). In the short-run, not every base station may be able to utilize the merged firms' full bandwidth allocation. While a base station created by merging two stations that were at the same location may be able to use the full merged bandwidth allocation, a base station at a location previously operated by only one of the original MNOs would need additional investment to operate the combined bandwidth. This means that $B_{\tilde{f}m}$ likely overstates the bandwidth that can be used by the merged MNO in the short-run, leading to faster predicted download speeds, which increase consumer surplus. This analysis will therefore tend to *overstate* post-merger efficiencies; however, as a preview of our findings, we find that all mergers between MNOs imply a short-run *decrease* in consumer surplus.

The synergies implied by a merger depend on the two pre-merger MNOs' bandwidth allocations and the degree of co-location of their base stations in the market. These variables can vary from market-to-market, so rather than using a single representative market (as in the exercises in section 6.1), we use multiple markets to capture this heterogeneity across France. Using all markets in France would be computationally infeasible,[47] so we use several representative markets corresponding to population categories and weight those markets in firms' profits by the fraction of France's population that those categories represent.[48]

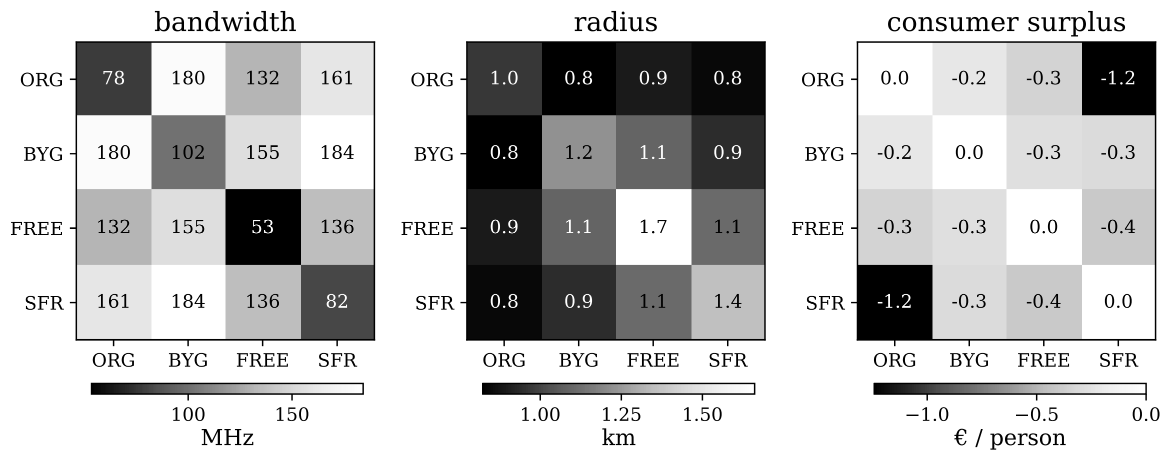
Despite efficiencies from merging infrastructure, we find that all possible mergers between two MNOs would decrease consumer surplus. Figure 12 presents the implied bandwidth and radius of the merged MNOs as well as the impact on consumer surplus for each merger. All mergers decrease consumer surplus, with reductions ranging in size from a minimum of 0.22 €/person/month (Orange and Bouygues) to a maximum of 1.24 €/person/month (Orange and SFR).[49]

[47]Solving for the firms' equilibrium prices requires solving for the equilibrium download speeds (equation 20) for every candidate vector of prices.

[48]Specifically, we use three population categories: $<$35 000, 35 000–100 000, and $\geq$100 000, which respectively correspond to 32.03%, 29.56%, and 38.41% of the population of France in our sample markets (and are used as the weights the markets receive in the firms' profit functions). Within each category, we choose a market based on which one has the most similar estimated base station costs ($\hat{c}^s_{fm}$) to the population-weighted average within the category (based on the sum of the squared differences). The municipalities we use are therefore Muret, Épinay-sur-Seine, and Amiens.

[49]For comparison, in section 6.1.2 we found that consumer surplus is 0.67 €/person/month lower in the three-firm asymmetric bandwidth allocation equilibrium than in the four-firm symmetric allocation equilibrium. This allocation would result from a merger in spectrum allocations between two firms in the four-firm symmetric case, similar to the merger exercise in this section (though the exercises differ in a few other dimensions, including that the exercise in 6.1.2 captures long-run equilibrium investment).

Figure 12: Short-Run Counterfactual Consumer Surplus under Mergers



Note: For each subplot, the four original MNOs are displayed along the x- and y-axes. A point corresponds to the merger of the MNOs along the x- and y-axes, and the diagonal corresponds to no merger. The bandwidths and radii of the merged and non-merged MNOs are provided in the first and second subplots, respectively. Values along the diagonal are population-weighted averages of the bandwidth or radius of the MNO. Off-diagonal values are the population-weighted averages of the bandwidth or radius of the merged MNO (while for non-merged firms in that merger scenario, the values remain the same as the ones for those MNOs along the diagonal). The third subplot depicts the change in consumer surplus (relative to the no-merger case) for the different merger scenarios.

7 Conclusion

The regulation of the mobile telecommunications industry, encomparssing antitrust policy and spectrum allocation, calls for an understanding of scale efficiencies as well as market power. Our approach has been an interdisciplinary one, drawing from tools in empirical industrial organization to understand market power, and from wireless engineering to understand scale efficiencies. Our framework allows us to quantify the relationship between market structure—defined as the number of firms and the spectrum allocation among them—and equilibrium outcomes such as prices, download speeds, and welfare.

References

Analysys Mason. 2015. "MCT review 2015-2018: Mobile network cost modelling."

Arrow, Kenneth Joseph. 1962. "Economic Welfare and the Allocation of Resources for Inventions." *The Rate of Direction of Inventive Activity: Economic and Social Factors.*

Asker, John, and Michael L Katz. 2022. "The Sprint/T-Mobile Merger." In *Antitrust Economics at a Time of Upheaval.* , ed. John Kwoka, Tommaso Valletti and Lawrence White. Competition Policy International.

Autorité de Régulation des Communications Électroniques (ARCEP). 2011*a*. "Décision no 2011-0597 de l'Autorité de régulation des communications électroniques et des postes."

Autorité de Régulation des Communications Électroniques (ARCEP). 2011*b*. "Décision no 2011-0599 de l'Autorité de régulation des communications électroniques et des postes."

Autorité de Régulation des Communications Électroniques (ARCEP). 2016. "Observatoire des marchés des communications électroniques 4 février 2016."

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–90.

Björkegren, Daniel. 2022. "Competition in network industries: Evidence from the Rwandan mobile phone network." *RAND Journal of Economics*, 53(1): 200–225.

Błaszczyszyn, Bartlomiej, Miodrag Jovanovicy, and Mohamed Kadhem Karray. 2014. "How user throughput depends on the traffic demand in large cellular networks." 611–619, IEEE.

Bourreau, Marc, Yutec Sun, and Frank Verboven. 2021. "Market entry, fighting brands, and tacit collusion: Evidence from the French mobile telecommunications market." *American Economic Review*, 111(11): 3459–99.

Carlton, Dennis W. 1978. "Market behavior with demand uncertainty and price inflexibility." *The American Economic Review*, 68(4): 571–587.

Cattani, Kyle, and Glen M Schmidt. 2005. "The pooling principle." *INFORMS Transactions on Education*, 5(2): 17–24.

Centre De Recherche Pour L'étude Et L'observation Des Conditions De Vie (CREDOC). 2015. "Baromeètre du numérique (Édition 2015)."

Chenery, Hollis B. 1949. "Engineering Production Functions." *The Quarterly Journal of Economics*, 63(4): 507–531.

Chu, Chenghuan Sean. 2010. "The effect of satellite entry on cable television prices and product quality." *The RAND Journal of Economics*, 41(4): 730–764.

Conlon, Christopher, and Jeff Gortmaker. 2020. "Best practices for differentiated products demand estimation with pyblp." *The RAND Journal of Economics*, 51(4): 1108–1161.

Cramton, Peter, Evan Kwerel, Gregory Rosston, and Andrzej Skrzypacz. 2011. "Using spectrum auctions to enhance competition in wireless services." *The Journal of Law and Economics*, 54(S4): S167–S188.

Crawford, Gregory S, and Matthew Shum. 2007. "Monopoly quality degradation and regulation in cable television." *The Journal of Law and Economics*, 50(1): 181–219.

Crawford, Gregory S, Oleksandr Shcherbakov, and Matthew Shum. 2019. "Quality overprovision in cable television markets." *American Economic Review*, 109(3): 956–95.

Crawford, Gregory S, Robin S Lee, Michael D Whinston, and Ali Yurukoglu. 2018. "The welfare effects of vertical integration in multichannel television markets." *Econometrica*, 86(3): 891–954.

Cullen, Joseph, Nicolas Schutz, and Oleksandr Shcherbakov. 2020. "The Welfare Effects of Early Termination Fees in the US Wireless Industry."

De Vany, Arthur. 1976. "Uncertainty, waiting time, and capacity utilization: A stochastic theory of product quality." *Journal of Political Economy*, 84(3): 523–541.

Doraszelski, Ulrich, Katja Seim, Michael Sinkinson, and Peichun Wang. 2019. "Ownership concentration and strategic supply reduction." National Bureau of Economic Research.

El Azouzi, Rachid, Eitan Altman, and Laura Wynter. 2003. "Telecommunications network equilibrium with price and quality-of-service characteristics." In *Teletraffic science and engineering.* Vol. 5, 369–378. Elsevier.

Ershov, Daniel, and David J Salant. 2022. "Auctions and Mobile Market Competition: Evidence from European 4G Auctions." *Available at SSRN 4168166.*

Eső, Péter, Volker Nocke, and Lucy White. 2010. "Competition for scarce resources." *The RAND Journal of Economics*, 41(3): 524–548.

Fan, Ying. 2013. "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market." *American Economic Review*, 103(5): 1598–1628.

Fan, Ying, and Chenyu Yang. 2020. "Competition, product proliferation, and welfare: A study of the US smartphone market." *American Economic Journal: Microeconomics*, 12(2): 99–134.

Federal Communications Commission. 2010. "Connecting America: The National Broadband Plan."

Federal Communications Commission. 2019. "Memorandum opinion and order, declaratory ruling, and order of proposed modification." *WT Docket 18-197.*

Fund, Fraida, Shahram Shahsavari, Shivendra S Panwar, Elza Erkip, and Sundeep Rangan. 2017. "Resource sharing among mmWave cellular service providers in a vertically differentiated duopoly." 1–7, IEEE.

Genakos, Christos, Tommaso Valletti, and Frank Verboven. 2018. "Evaluating market consolidation in mobile communications." *Economic Policy*, 33(93): 45–100.

Grajek, Michał, and Lars-Hendrik Röller. 2012. "Regulation and investment in network industries: Evidence from European telecoms." *The Journal of Law and Economics*, 55(1): 189–216.

Granja, João. 2022. "Regulation and Service Provision in Dynamic Oligopoly: Evidence from Mobile Telecommunications."

Grigolon, Laura, and Frank Verboven. 2014. "Nested Logit or Random Coefficients Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation." *Review of Economics and Statistics*, 96(5): 916–935.

Hata, Masaharu. 1980. "Empirical Formula for Propagation Loss in Land Mobile Radio Services." *IEEE Transactions on Vehicular Technology*, 29(3): 317–325.

Holmes, Thomas J. 2011. "The Diffusion of Wal-Mart and Economies of Density." *Econometrica*, 79(1): 253–302.

Hua, Sha, Pei Liu, and Shivendra S Panwar. 2012. "The urge to merge: When cellular service providers pool capacity." 5020–5025, IEEE.

Jehiel, Philippe, Benny Moldovanu, Marco Ottaviani, and Carol Propper. 2003. "An economic perspective on auctions." *Economic Policy*, 18(36): 269–308.

Kim, Haesik. 2015. "Coding and Modulation Techniques for High Spectral Efficiency Transmission in 5G and Satcom." 2746–2750, IEEE.

Lee, Robin S. 2013. "Vertical integration and exclusivity in platform and two-sided markets." *American Economic Review*, 103(7): 2960–3000.

Lhost, Jonathan, Brijesh Pinto, and David Sibley. 2015. "Effects of spectrum holdings on equilibrium in the wireless industry." *Review of Network Economics*, 14(2): 111–155.

Lin, Zhongjian, Xun Tang, and Mo Xiao. 2022. "Endogeneity in Games with Incomplete Information: U.S. Cellphone Service Deployment."

Malone, Jacob, Aviv Nevo, and Jonathan W Williams. 2017. "The tragedy of the last mile: Economic solutions to congestion in broadband networks." *Available at SSRN 2849869*.

Milgrom, Paul, and Ilya Segal. 2020. "Clock auctions and radio spectrum reallocation." *Journal of Political Economy*, 128(1): 1–31.

Mulligan, James G. 1983. "The economies of massed reserves." 73(4): 725–734.

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

Nikolikj, Vladimir, and Toni Janevski. 2014. "A cost modeling of high-capacity LTE-advanced and IEEE 802.11ac based heterogeneous networks, deployed in the 700 MHz, 2.6 GHz and 5 GHz Bands." *Procedia Computer Science*, 40: 49–56.

Peha, Jon M. 2017. "Cellular economies of scale and why disparities in spectrum holdings are detrimental." *Telecommunications Policy*, 41(9): 792–801.

Rey, Patrick, and David Salant. 2017. "Allocating essential inputs."

Robinson, E. A. G. 1958. *The Structure of Competitive Industry.* Cambridge University Press.

Rosaia, Nicola. 2023. "Competing platforms and transport equilibrium: Evidence from New York City."

Rosston, Gregory L. 2003. "The long and winding road: the FCC paves the path with good intentions." *Telecommunications Policy*, 27(7): 501–515.

Seim, Katja, and V Brian Viard. 2011. "The effect of market structure on cellular technology adoption and pricing." *American economic journal: Microeconomics*, 3(2): 221–51.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell Systems Technical Journal*, 27: 379–423,623–656.

Sinkinson, Michael. 2020. "Pricing and entry incentives with exclusive contracts: Evidence from smartphones."

Smail, Ghoul, and Jia Weijia. 2017. "Techno-economic analysis and prediction for the deployment of 5G mobile network." 9–16, IEEE.

Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *Bell Journal of Economics*, 6(2): 417–429.

Sun, Patrick. 2015. "Quality competition in mobile telecommunications: Evidence from Connecticut."

Taylor, H.M., S. Karlin, and H.E. Taylor. 1998. *An Introduction to Stochastic Modeling.* Elsevier Science.

Varadhan, Ravi, and Christophe Roland. 2008. "Simple and globally convergent methods for accelerating the convergence of any EM algorithm." *Scandinavian Journal of Statistics*, 35(2): 335–353.

Vives, Xavier. 2008. "Innovation and competitive pressure." *The Journal of Industrial Economics*, 56(3): 419–469.

Williamson, Oliver E. 1968. "Economies as an Antitrust Defense: The Welfare Tradeoffs." *The American Economic Review*, 58(1): 18–36.

A Technical Appendix (for online publication)

A.1 Data Transmission Details

A.1.1 Signal Power

Equation 33 in section 4.1 provides the formula we use for signal power. It is based on the Hata model of path loss (Hata, 1980). We use the Hata model for urban environments since we focus our analysis on urbanized areas. This model provides us with the following formula for path loss:

$$L\left(r\right)=68.75+27.72\log_{10}\left(f\right)-13.82\log_{10}\left(h\right)+\left(44.9-6.55\log_{10}\left(h\right)\right)\log_{10}\left(r\right), \tag{32}$$

where $L\left(r\right)$ is in decibels, r is the distance from the antenna (in km), h is the height of the base station antenna (in m), and f is the frequency (in MHz).[50]

The specific values in our path loss equation can be derived as follows. We assume a base station height of 30 m and a signal frequency of 1900 MHz, which is approximately the median operated frequency in France in 2015. These values yield

$$L\left(r\right)=139.2232+35.2249\log_{10}\left(r\right).$$

The signal power in dBm at a distance r from the antenna is

$$A-L(r),$$

where A is the transmitted power. We assume a signal power of 61 dBm (or 1259 W) per 5 MHz of bandwidth at the base station, which corresponds to the regulated limit on effective isotropic radiated power for the 2600 band (ARCEP, 2011a); similar limits apply for lower frequencies (ARCEP, 2011b).

Converting the units to watts, this yields the following formula for signal power, in W per 5 MHz of bandwidth:

$$S_{\ell}=S\left(r_{\ell}\right)=\exp\left(-24.92\right)r_{\ell}^{-3.522}, \tag{33}$$

[50] In the Sprint/T-Mobile merger, heterogeneity in the merging parties' spectrum holdings played an important role in the claimed efficiency gains. T-Mobile had substantial holdings of low-frequency spectrum, and Sprint owned only high-frequency spectrum (Asker and Katz, 2022). Notice that frequency f enters positively into equation 32, meaning that the signal power of high-frequency spectrum will be lower, and its signal power level will approach the levels of noise and interference power at shorter distances. It would be a straightforward extension to our model to capture this heterogeneity in spectrum holdings using equation 32 and integrating over the appropriate set of frequencies for each firm. Such a model could capture how a firm holding only low-frequency spectrum would experience higher costs of service, especially in areas of low population density.

where r_ℓ is location ℓ's distance from the base station. These values yield a path loss exponent of 3.522. Most engineering studies use a path loss exponent between 3.5 and 4.[51] In contrast, signal strength in a vacuum would have a path loss exponent of 2, but signals decay more quickly on the Earth's surface.

A.1.2 Noise and Interference

Noise power N is set equal to Johnson-Nyquist noise, -107.01 dBm per 5 MHz of bandwidth. Note that our expression for signal strength in equation 33 yields approximately $\exp(-24.92) = 1.5\mathrm{e}{-11}\mathrm{W}$ signal power density at a distance of 1 km from the base station. The noise power density is $10^{(-107.01/10)}/1000 \approx 2\mathrm{e}{-14}\mathrm{W}$. Thus, signal power is orders of magnitude larger than noise power at such distances from a base station.

Interference power is set equal to 30% of the signal power from the six adjacent cells. The 30% number follows Błaszczyszyn, Jovanovicy and Karray (2014) and reflects the fact that adjacent cells are not always be in use, and modern systems use directional signals to limit interference. To illustrate its magnitude, note that the edge of a 1 km cell, at the midpoint between the serving base station and an adjacent identical cell's base station, we would have $1.5\mathrm{e}{-11}\mathrm{W}$ of signal power (per 5 MHz) from the cell being used, and $0.3 \cdot 1.5\mathrm{e}{-11}\mathrm{W}$ of interference power from the immediately adjacent cell's base station. Ignoring interference from other neighboring cells, this would lead to a SINR ratio of approximately 0.3^{-1}, the ratio of signal to interference. At these signal levels, interference dominates the denominator, and noise power plays little role.

Ultimately, the way we calculate interference power (at each point within a cell) is to sum interference from the neighboring six cells, pictured in Figure 13. For a given point in the center cell, we compute the distances between that point and the centroids of the adjacent cells, which is the location of the antennas corresponding to each cell.

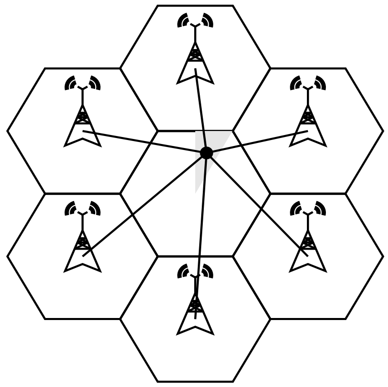
Let $\mathcal{L}$ be the locations of the centroids of the six adjacent hexagons to a hexagon centered at the origin of a Euclidean plane, when all seven hexagons are regular with a (maximum) radius of unity. That is,

$$\mathcal{L} \equiv \left\{ \left(0, \sqrt{3}\right), \left(\frac{3}{2}, \frac{\sqrt{3}}{2}\right), \left(\frac{3}{2}, -\frac{\sqrt{3}}{2}\right), \left(0, -\sqrt{3}\right), \left(-\frac{3}{2}, -\frac{\sqrt{3}}{2}\right), \left(-\frac{3}{2}, \frac{\sqrt{3}}{2}\right) \right\}.$$

These points correspond to the locations of the adjacent base stations pictured in Figure 13. Let $d(\ell, \ell')$ represent the Euclidean distance between two points ℓ and ℓ'. Ultimately,

[51] For instance, Błaszczyszyn, Jovanovicy and Karray (2014) assume a path loss exponent of 3.8.

Figure 13: A Hexagonal Cell and its Six Adjacent Cells



Note: The figure depicts the distance between an individual at a random location in the center cell and the base stations that correspond to the six adjacent cells. In determining the channel capacity of the cell, we integrate over the entire area of the center cell, taking into account this interference at each point.

interference power is calculated as follows:

$$I_\ell\left(R_{fm}\right) = 0.3 \sum_{\ell' \in \mathcal{L}} S\left(R_{fm} d\left(\ell, \ell'\right)\right), \tag{34}$$

where the signal power function $S(\cdot)$ is defined in equation 33 above. In other words, interference power is 30% of the summed signal powers from the adjacent six hexagons.

To calculate channel capacity (equation 14), we need to integrate signal power S_ℓ and interference power $I_\ell\left(R_{fm}\right)$ over the locations ℓ within a hexagonal cell. To perform this integration, it suffices to focus on one of the twelve right triangles that compose the hexagon and then multiply by twelve (each of the twelve triangles has the same distribution of interference). Specifically, we integrate over the shaded triangle in Figure 13:

$$\bar{Q}_{fm}\left(R_{fm}, B_{fm}\right) = \gamma_m B_{fm} A\left(R_{fm}\right) \left[12 \int_0^{\frac{\sqrt{3}}{2} R_{fm}} \int_0^{\frac{y}{\sqrt{3}}} \frac{1}{\log_2\left(1 + \frac{S_{(x,y)}}{N + I_{(x,y)}\left(R_{fm}\right)}\right)} \mathrm{d}x \mathrm{d}y\right]^{-1}. \tag{35}$$

A.2 Spectral Efficiency Calibration

We calibrate the spectral efficiency parameter γ_m using delivered download speed data for each municipality. This is done by solving for the value of γ_m that makes equation 20 hold for Orange (we do not have usage data for other operators). In this calibration, the average experienced download speed Q_{fm} is the average download speed in Mbps in the delivered download speed data obtained from Ookla. Q^D comes from the OSIRIS infrastructure usage data. For each market, we determine Q^D by calculating the amount of data requested of

Orange per second between noon and 1 pm and dividing by the number of Orange base stations in that market. Solving for the γ that makes equation 20 hold yields a spectral efficiency parameter, $\hat{\gamma}_m$, for each market.[52] Across municipalities, the mean value of $\hat{\gamma}_m$ is 0.165, and its standard deviation is 0.048.

We recover γ_m focusing on Orange (for which we have data on infrastructure usage to construct Q^D) and assume that this commune-level spectral efficiency parameter applies to every firm. This is reasonable given that the firms have access to the same technologies. One potential reason for differences across firms would be heterogeneous spectrum holdings. In the Hata model of path loss, higher frequency spectrum is associated with greater loss (see equation 32). Thus, a firm with spectrum holdings primarily in high-frequency bands would experience greater loss than a firm with considerable holdings of low-frequency spectrum. Reassuringly, the spectrum holdings of Orange, SFR, and Bouygues were very similar in 2015.

Free owned little low-frequency spectrum in 2015. This is a potential cause for concern when estimating Free's costs using our structural model with the value of γ_m recovered using Orange. Free could have a different spectral efficiency parameter. Furthermore, another reason to worry about using equation 27 to recover Free's infrastructure costs is that Free benefited from active network sharing with Orange, meaning its delivered quality did not only depend on its own infrastructure investments, but also on Orange's (see section C.3). Despite these concerns, our estimates of Free's infrastructure costs are not drastically different from those of the other firms (see Table 5). Thus, dropping Free's cost estimates before running the counterfactuals described below (which involve symmetric firms with costs parameters that are averages across our estimates) makes little difference to the results.

B Demand Estimation Details (for online publication)

B.1 Contraction Mapping

Here we consider an alternative version of the Berry, Levinsohn and Pakes (1995) (BLP) contraction mapping in which we observe market shares at the product-market level for Orange products but only aggregate firm-level market shares for the other products. We first show in section B.1.2 that if we observe market shares at the firm-market level, the problem can

[52] Different frequencies are used for 3G and 4G technology. We account for differences between generations by calculating the channel capacity for each technology separately (i.e., $\bar{Q}(R, B^{3G})$ and $\bar{Q}(R, B^{4G})$). We then adjust the 3G channel capacity to its 4G-equivalent by using the ratio of 3G-to-4G maximum link spectral efficiencies (respectively, 2.5 and 4.08 (Kim, 2015)). Therefore, in determining $\hat{\gamma}_m$ in each market, we calculate channel capacity as

$$\bar{Q}_{fm}\left(R_{fm}, B^{3G}_{fm}, B^{4G}_{fm}\right) = \frac{2.5}{4.08}\bar{Q}_{fm}\left(R_{fm}, B^{3G}_{fm}\right) + \bar{Q}_{fm}\left(R_{fm}, B^{4G}_{fm}\right).$$

be rewritten in such a way that the BLP contraction mapping proof holds. In section B.1.3 we extend this result to the nested logit setting. Finally, in section B.1.4 we show that if we observe some firm market shares only at the aggregate level (as is our case), the problem can still be rewritten to fit into the BLP contraction mapping proof setup.

B.1.1 Standard BLP Contraction Mapping Setup

We will start with the standard BLP setting in order to introduce notation. In this setting, there are products $j \in \mathcal{J} = \{1, \ldots, J\}$, and we observe market shares ς_{jm} for each product. We can express an individual's utility for a product as $u_{ijm} = \delta_{jm} + \mu_{ijm} + \varepsilon_{ijm}$, which yields the type-specific market shares

$$s_{ijm} = \frac{\exp\left(\delta_{jm} + \mu_{ijm}\right)}{\sum_{j'} \exp\left(\delta_{j'm} + \mu_{ij'm}\right)}.$$

Aggregate market shares are given by

$$s_{jm}\left(\delta\right) = \int \frac{\exp\left(\delta_{jm} + \mu_{ijm}\right)}{\sum_{j'} \exp\left(\delta_{j'm} + \mu_{ij'm}\right)} dF\left(\mu_m\right).$$

The existence of the contraction mapping implies that there is a unique vector δ such that $s_m\left(\delta\right) = \varsigma_m$ for any observed vector of shares ς_m.

B.1.2 Grouped Products Extension

Our setting is one in which market shares are observed only for certain groupings of products. That is, let $\mathcal{J}$ be partitioned into subsets $\mathcal{J}_f$ with $f \in \mathcal{F} = \{1, 2, \ldots F\}$. For each f, we observe only the market share ς_{ft} for all the products within $\mathcal{J}_f$. The subsets $\mathcal{J}_f$ may include individual products (i.e., in our application each Orange product would have its own $\mathcal{J}_f$ set) or several products (i.e., each non-Orange firm has one $\mathcal{J}_f$ group that includes all that firm's products).

Providing a parametric form, let $\delta_{jm} = \theta_1 x_{jm} + \xi_{jm}$, where θ_1 would capture what is often referred to as "linear parameters," i.e., parameters that can typically be estimated outside of the contraction mapping because they only shift the mean utility component δ_{jm} that the contraction mapping aims to recover. In this extension, the θ_1 parameters must be included in the contraction mapping.

We cannot recover δ_{jm} (or ξ_{jm}) separately for different $j \in \mathcal{J}_f$. We assume $\xi_{jm} = \xi_{fm}$ for all $j \in \mathcal{J}_f$ for each f.

Let $\bar{x}_{fm}$ be the mean value of x_{fm} for those products within $\mathcal{J}_f$. Then, we have $\delta_{jm} = \theta_1 \bar{x}_{fm} + \theta_1 x^d_{jm} + \xi_{fm}$, where $x^d_{jm} := x_{jm} - \bar{x}_{fm}$. We define $\widetilde{\delta}_{fm} = \theta_1 \bar{x}_{fm} + \xi_{fm}$, and $\widetilde{\mu}_{ijm} = \theta_1 x^d_{jm} + \mu_{ijm}$.

This very nearly allows us to re-define the model in terms where we could apply the original BLP proof strategy to establish the contraction mapping. The only problem is that $\widetilde{\mu}_{ijm}$ is defined over j, where we would need it to be defined over f in order to apply the same proof strategy. Let's consider the aggregation over j to f:

$$s_{ifm}\left(\widetilde{\delta}\right)=\sum_{j\in\mathcal{J}_f}\frac{\exp\left(\widetilde{\delta}_{fm}+\widetilde{\mu}_{ijm}\right)}{\sum_{j'\in\mathcal{J}}\exp\left(\widetilde{\delta}_{f(j')m}+\widetilde{\mu}_{ij'm}\right)},$$

where $f\left(j'\right)$ refers to the f associated with product j'.

Defining $\widetilde{\mu}_{ifm}=\log\left(\sum_{j\in\mathcal{J}_f}\exp\left(\widetilde{\mu}_{ijm}\right)\right)$, it follows that

$$\sum_{j\in\mathcal{J}_f}\exp\left(\widetilde{\delta}_{fm}+\widetilde{\mu}_{ijm}\right)=\exp\left(\widetilde{\delta}_{fm}+\widetilde{\mu}_{ifm}\right),$$

and therefore

$$s_{ifm}\left(\widetilde{\delta}\right)=\sum_{j\in\mathcal{J}_f}\frac{\exp\left(\widetilde{\delta}_{fm}+\widetilde{\mu}_{ifm}\right)}{\sum_{f'}\exp\left(\widetilde{\delta}_{f'm}+\widetilde{\mu}_{if'm}\right)}.$$

We can then aggregate up to market-level shares s_{fm} by integrating over the $\widetilde{\mu}_{ifm}$, and we have rewritten our extended setting in a way that allows us to apply the BLP proof strategy.

B.1.3 Grouped Products Extension with Nested Logit

In the more general random coefficients nested logit (RCNL) model introduced by Grigolon and Verboven (2014) (henceforth, GV), we can construct analogous formulas that will allow us to recover group-specific mean demands $\widetilde{\delta}$.

In the RCNL model, type-specific market shares are as follows:

$$s_{ijm}=\frac{\exp\left(\frac{\delta_{jm}+\mu_{ijm}}{1-\sigma}\right)}{\exp\left(\frac{I_{ig(j)}}{1-\sigma}\right)}\frac{\exp\left(I_{ig(j)}\right)}{\exp\left(I_i\right)},$$

where $\sigma\in[0,1)$ is the nesting parameter, $g\left(j\right)$ return the nest to which j belongs,[53] and

$$\begin{aligned}I_{ig}&=(1-\sigma)\log\left(\sum_{j\in\mathcal{J}_g}\exp\left(\frac{\delta_{jm}+\mu_{ijm}}{1-\sigma}\right)\right),\\ I_i&=\log\left(1+\sum_{g\in\mathcal{G}}\exp\left(I_{ig}\right)\right).\end{aligned}$$

[53] We will assume that products produced by the same firm belong to the same group. Formally, for each f, $g\left(j\right)=g_f$ for all $j\in\mathcal{J}_f$.

In this extension, we redefine $\widetilde{\delta}_{fm}$ and $\widetilde{\mu}_{ifm}$ to incorporate σ. Let $\widetilde{\delta}_{fm} = \frac{\theta_1 \bar{x}_{fm} + \xi_{fm}}{1-\sigma}$, $\widetilde{\mu}_{ijm} = \frac{\theta_1 x^d_{jm} + \mu_{ijm}}{1-\sigma}$, and $\widetilde{\mu}_{ifm} = \log\left(\sum_{j \in \mathcal{J}_f} \exp\left(\widetilde{\mu}_{ijm}\right)\right)$. Then

$$s_{ifm} = \frac{\exp\left(\widetilde{\delta}_{fm} + \widetilde{\mu}_{ifm}\right)}{\exp\left(\frac{I_{ig(f)}}{1-\sigma}\right)} \frac{\exp\left(I_{ig(f)}\right)}{\exp\left(I_i\right)}$$

where $I_{ig} = (1-\sigma)\log\left(\sum_{f \in \mathcal{F}_g} \exp\left(\widetilde{\delta}_{fm} + \widetilde{\mu}_{ifm}\right)\right)$ and $\mathcal{F}_g = \{f \in \mathcal{F} : g(f) = g\}$.

GV note that, substituting in our notation,

$$f\left(\widetilde{\delta}\right) = \widetilde{\delta} + \log(\varsigma) - \log\left(s\left(\widetilde{\delta}\right)\right)$$

is a contraction mapping if

$$1 - \frac{1}{s_f} \frac{\partial s_f}{\partial \widetilde{\delta}_f} \geq 0.$$

Unlike in GV, this holds in our case. Explicitly,

$$\frac{\partial s_f}{\partial \widetilde{\delta}_f} = \left(1 - \frac{\sigma}{1-\sigma} s_{f|g} - s_f\right) s_f,$$

and so

$$1 - \frac{1}{s_f} \frac{\partial s_f}{\partial \widetilde{\delta}_f} = \frac{\sigma}{1-\sigma} s_{f|g} + s_f \geq 0 \quad \Leftrightarrow \quad \sigma s_{f|g} + (1-\sigma) s_f \geq 0.$$

This condition holds for all $\sigma \in [0, 1)$.

B.1.4 Market Aggregation Extension

In our setting we observe market shares only at the aggregate level for some firms. We assume in this extension $\xi_{jm} = \xi_{f(j)}$ for all j, m and recover ξ_f for each f. We will proceed in this section using the non-nested setting introduced in section B.1.2, but the results hold using the analogues to the RCNL expressions introduced in section B.1.3.

Analogous to the previous setup, let $\bar{x}_f$ be the mean value of x_{jm} across products $j \in \mathcal{J}_f$ and markets m, $\bar{x}_f = \frac{1}{MJ_f} \sum_m \sum_{j \in \mathcal{J}_f} x_{jm}$. Then, $\delta_{jm} = \theta_1 \bar{x}_{f(j)} + \theta_1 x^d_{jm} + \xi_{f(j)}$. where we now define $x^d_{jm} := x_{jm} - \bar{x}_{f(j)}$. Analogously defining $\widetilde{\delta}_f = \theta_1 \bar{x}_f + \xi_f$, $\widetilde{\mu}_{ijm} = \theta_1 x^d_{jm} + \mu_{ijm}$, and $\widetilde{\mu}_{ifm} := \log\left(\sum_{j \in \mathcal{J}_f} \exp\left(\widetilde{\mu}_{ijm}\right)\right)$, then

$$\bar{s}_{if}(\widetilde{\delta}) = \sum_m w(m) \frac{\exp\left(\widetilde{\delta}_f + \widetilde{\mu}_{ifm}\right)}{\sum_{f'} \exp\left(\widetilde{\delta}'_f + \widetilde{\mu}_{if'm}\right)}.$$

We can aggregate up to aggregate firm shares $\bar{s}_f$ by integrating over $\tilde{\mu}_{ifm}$:

$$\bar{s}_f = \int \sum_m w(m) \frac{\exp\left(\widetilde{\delta}_f + \widetilde{\mu}_{ifm}\right)}{\sum_{f'} \exp\left(\widetilde{\delta}_{f'} + \widetilde{\mu}_{if'm}\right)} dF(\widetilde{\mu}_{ifm}) = \int \frac{\exp\left(\widetilde{\delta}_f + \widetilde{\mu}_{ifm}\right)}{\sum_{f'} \exp\left(\widetilde{\delta}_{f'} + \widetilde{\mu}_{if'm}\right)} dG(\widetilde{\mu}_{ifm}).$$

The final expression makes clear that the BLP contraction mapping proof strategy still holds in this aggregate setting.

Since we observe product-level market shares for every market for Orange products, we allow ξ_{jm} to differ by product and market for all $j \in \mathcal{J}_{ORG}$.

When coding the contraction mapping, we follow Conlon and Gortmaker (2020) in implementing the SQUAREM algorithm (Varadhan and Roland, 2008).

B.2 Calibration

In this section, we explain how we derived the values used from calibration.

First, we have the own price elasticity for Orange, defined in equation 10, which is an elasticity with respect to a proportional change in the prices of all of Orange's products at once. An equivalent way of writing this elasticity is

$$\begin{aligned} \frac{\mathrm{d}\ln s_{ORG}}{\mathrm{d}\ln P_{ORG}} &= \textstyle\sum_{j\in\mathcal{J}_{ORG}} \frac{\partial s_{ORG}}{\partial P_j}\frac{P_j}{s_{ORG}} \\ &= \textstyle\sum_{j\in\mathcal{J}_{ORG}} \frac{P_j}{s_{ORG}} \sum_{j'\in\mathcal{J}_{ORG}} \frac{\partial s_{j'}}{\partial P_j} \\ &= \textstyle\sum_{j\in\mathcal{J}_{ORG}} \frac{\partial s_j}{\partial P_j}\frac{P_j}{s_j}\frac{s_j}{s_{ORG}} \sum_{j'\in\mathcal{J}_{ORG}} \frac{\partial s_{j'}}{\partial P_j} \Big/ \frac{\partial s_j}{\partial P_j} \\ &= \textstyle\sum_{j\in\mathcal{J}_{ORG}} e_j \frac{s_j}{s_{ORG}} \sum_{j'\in\mathcal{J}_{ORG}} -DIV^{j,j'}, \end{aligned} \tag{36}$$

where $\mathrm{d}\ln P_{ORG}$ represents a proportional change in the prices of Orange's plans, e_j denotes product j's own price elasticity and $DIV^{j,j'}$ is the diversion ratio from product j to j',

$$DIV^{j,j'} = -\frac{\frac{\partial s_{j'}}{\partial P_j}}{\frac{\partial s_j}{\partial P_j}}.$$

The third line follows by dividing and multiplying by $\frac{\partial s_j}{\partial P_j}/s_j$.

In Bourreau, Sun and Verboven (2021), we take elasticities (e_j) from Table A.4, diversion ratios ($DIV^{j,j'}$) from Table A.3, and quantities (shares) from Table 3. Plugging in all these numbers, we find an elasticity of -2.36.

Next, we have Orange's diversion ratio to the outside option, defined as

$$DIV^{ORG,0} = -\frac{\frac{\mathrm{d}s_0}{\mathrm{d}\ln P_{ORG}}}{\frac{\mathrm{d}s_{ORG}}{\mathrm{d}\ln P_{ORG}}}, \tag{37}$$

where $\mathrm{d}\ln P_{ORG}$ represents a proportional change in all of Orange's prices. By the chain rule,

$$\frac{\mathrm{d}s_0}{\mathrm{d}\ln P_{ORG}} = \sum_{j\in\mathcal{J}_{ORG}} \frac{\partial s_0}{\partial P_j} P_j.$$

We then have

$$\frac{\mathrm{d}s_0}{\mathrm{d}\ln P_{ORG}} = \sum_{j\in\mathcal{J}_{ORG}} DIV^{j,0}\frac{\partial s_j}{\partial P_j} P_j = \sum_{j\in\mathcal{J}_{ORG}} DIV^{j,0} e_j s_j, \tag{38}$$

where $DIV^{j,0}$ is the diversion ratio from j to the outside option.

Turning to the denominator of equation 37, we can write

$$\frac{\mathrm{d}s_{ORG}}{\mathrm{d}\ln P_{ORG}} = e_{ORG} s_{ORG}. \tag{39}$$

Finally, we substitute equations 38 and 39 into equation 37:

$$DIV^{ORG,0} = \frac{\sum_{j\in\mathcal{J}_{ORG}} DIV^{j,0} e_j s_j}{e_{ORG} s_{ORG}}. \tag{40}$$

Taking elasticities, diversion ratios, and shares from the same sources in BSV as above, this yields a diversion ratio of 0.036.

Being precise, the moment we use for estimation averages over market-level diversion ratios, while the diversion ratio calculated here is based on objects that are already sample averages. That is, there is a difference in whether we average before or after taking the ratio expressed in equation 40. However, this difference appears to have a trivial impact on the value of the diversion ratio. In our sample at our parameter estimates, the difference in the diversion ratio if averaging is done before taking the ratio versus if done after is 0.00087.

C Data Appendix (for online publication)

This appendix provides additional description of our main datasets and variables. Section C.1 presents the characteristics of mobile tariffs and the tariff dataset. Section C.2 describes the measurement of the quality of mobile data. Section C.3 discusses network sharing in France in 2015.

C.1 Product Data

C.1.1 Product Characteristics

We collect data on mobile phone plans released between November 2013 and October 2015, along with their characteristics, from operators' quarterly catalogs. It includes postpaid plans from the four MNOs and the largest MVNO (EI Telecom) as well as their prepaid plans.[54] Promotional plans, typically released during summer and Christmas, are not included in the dataset.

Plan characteristics include tariff, voice and data limits, handset subsidy, length of commitment, and whether or not plans were bundled with fixed services. As described in section 2.2, we choose representative mobile-only plans for each firm and adjust monthly prices based on contract duration and handset subsidies.

We take over 100 contracts from catalogs, and from them we construct 21 representative products in our model's choice set. We define categories of plans according to their level of data limits: less than 500 MB, 500–3 000 MB, 3 000–7 000 MB and more than 7 000 MB. These thresholds are chosen following discussions with industry experts and the statistical distribution of chosen plans. The second data limit category—that is, contracts with 500–3 000 MB—we have further split according to their voice allowances: unlimited or not, making a total of five categories of phone plans. Low data limit plans typically do not have unlimited voice, and high data limit contracts typically come with unlimited voice allowance, so we do not split these categories by the voice limit. We exclude plans bundled with fixed broadband or television.

We choose the least expensive plan in each category as the category's representative plan. Some customers keep old plans that are no longer available, so we fill these missing data by using the most similar representative plan. While some plans with handset subsidies have corresponding standalone versions, some do not. We adjust the prices of these latter plans using data on the price of handsets and the upfront payment required by Orange. We collect these data for both iPhone and Samsung, the two most popular handsets. We then distribute the handset cost over 24 months and update the monthly plan price by subtracting off the monthly cost of the handset. In addition, we assume that Orange's handset subsidies apply to other operators' subsidized contracts because we do not observed their upfront costs.

[54] ORG's contracts include not only those that are sold through its main brand, but also others sold under alternative brands such as SOSH, BNP Paribas Mobile, FNAC Mobile, Click Mobile, Carrefour Mobile, etc.

C.1.2 Soft Data Limits

For plans with data limits, the download speed is reduced for usage above allowance if no add-on is purchased. The maximal download speed under throttling is typically 128 Kbps. With this download speed, it would take over half-an-hour to download a 30 MB file, compared to 2 minutes under a theoretical non-throttled speed of 2 Mbps in a 3G network, and 24 seconds given a moderate 4G download speed of 10 Mbps. Basically, only emails and light web pages can be opened under throttling. As presented in Table 7 below, this download speed is not always specified by operators in their contracts. When it is, it may depend on the location of the usage (local or abroad). The actual download speed experienced by customers is a function of the number of simultaneous users, its location and handset. In our demand model, however, we assume that any data consumption over the data limit yields a speed of exactly 128 Kbps.

Table 7: Maximal Download Speed under Throttling (Kbps)

Operator	National	Roaming
ORG	128*	ns
SFR	ns	ns
BYG	128	32
FREE	ns	ns

*:except video streaming.

ns ≡ not specified.

Source: operators' contracts

C.2 Quality Data

Quality measures are constructed using download speed test results provided by Ookla. Test results come from users who use Ookla's free Internet speed test, called "Speedtest," using a web browser or within an app. Using speed tests in France in the second quarter of 2016 yields 1 056 285 individual speed tests. Each speed test records the download speed, mobile network operator, and the user's location. We aggregate speed tests by averaging measured download speeds over tests for a given operator and geographic market, yielding an operator-market quality measure. An operator-market quality measure is, on average, an average of 284 test results. Note that our estimates rely on an instrument for these quality measures (see section 3.2.2), alleviating concerns about attenuation bias.

C.3 Network Sharing

Network sharing occurs when a network operator shares a part or the whole of its network resources with a retail competitor. These resources can be passive network elements, such as antenna supports, masts, or active network elements, such as frequency bandwidths. Passive network sharing affects coverage differentiation but not necessarily quality differentiation. It typically consists of operators sharing the same tower and potentially the cost of electricity. In general, it is any agreement between MNOs that do not involve the sharing of available frequency bandwidths.

In contrast, under active network sharing (Radio Access Network-Sharing), operators cannot differentiate in terms of quality, defined as the frequency bandwidth available per customer. Typically, it consists of the sharing of frequency bands and the network elements involved in data transmission. Roaming agreements, whereby an operator's customers rely on the network of a host operator to communicate, is the highest level of active network sharing. It does not offer any possibility for quality or coverage differentiation.

Table 8 below presents the network sharing agreements reached between 2012 and 2015. These agreements apply to two types of areas according to their population density. "White Areas" or "Zones Blanches" correspond to areas where population density is so low that network deployment by several operators is not profitable. These areas, which are typically rural, are designated by the regulator and represent roughly 1% of the population and 10% of the national surface. Only ORG, SFR and BYG have invested in these areas.

The most widespread network technologies in the White Areas are 2G, EDGE and GPRS. [55] However, 3G technology has been recently deployed. As of the end of December 2015, half of ORG and BYG's networks in these areas were covered by 3G, compared to 35% for SFR. In general, only one operator invests in a given White Area, and 64% of antennas in these areas are involved in a roaming agreement. Rival operators roam over the network of the only operator that invests in the area. As a result, there is no quality differentiation. For the remaining 36% of antennas, operators share passive network elements.

At the national level, FREE's customers can roam over ORG's 2G and 3G networks as long as there is no FREE antenna nearby. As a result, FREE cannot differentiate from ORG on 2G and 3G technologies, except when a FREE antenna is nearby its customer. In addition, FREE does not have access to networks in Zones Blanches where BYG or SFR is the leader. MVNOs have roaming agreements with their hosts and therefore cannot differentiate in terms of quality or coverage.

Our model focuses on high-density areas to avoid the need to explicitly model network shar-

[55] EDGE and GPRS are suitable for low-speed mobile data services.

ing. During our period of study, the only active network sharing in such areas would have involved FREE's customers receiving data from 2G and 3G infrastructure owned and operated by ORG. Meanwhile, ORG and FREE each owned and operated their own distinct 4G network infrastructure, At the margin, 4G investments were how firms were differentiating and competing in download speeds in 2015.

Table 8: Network Sharing Agreements 2012-2015

		FREE		ORG		SFR		BYG
Zone Blanche	Roaming: 64% of 2G & 3G antenna				↔			
	Passive sharing: 36% of antenna				↔			
Low Density	2G and 3G RAN-Sharing	✗		✗			↔	
	4G Roaming	✗		✗			→	
High Density		✗		✗		✗		✗
National	Passive sharing				↔			
	2G and 3G Roaming		→			✗		✗

Source: Summary from discussions with ORG's experts.

Note: ↔: two-way (reciprocal) sharing, $A \to B$ one-way sharing hosted by operator B.

D Supplementary Results (for online publication)

D.1 Alternative Cost Specification

In this section we consider the robustness of our counterfactual results to the specification of infrastructure costs. In the main text, we use a specification in which base station costs are proportional to bandwidth. In this section, we consider the impact on our results of an alternative specification in which we assume that all of the infrastructure costs are fixed per base station. That is, we replace the infrastructure cost function specification (equation 22) with

$$C_{fm}\left(R_{fm}, B_{fm}\right) = c^s_{fm} \frac{A_m}{A\left(R_{fm}\right)}.$$

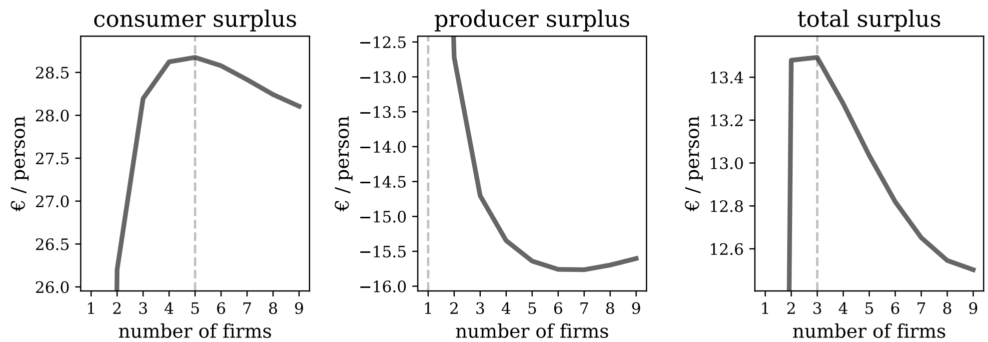
We use this specification of infrastructure costs to recover $\hat{c}^s_{fm}$ for each firm f and market m, and these parameters now have the interpretation of costs per base station (rather than costs per base station per unit of bandwidth).[56]

At four firms, base station costs are the same for both cost specifications – this is just the average cost per base station recovered from the data. At fewer firms, base station costs are cheaper in this specification where costs do not scale with bandwidth; at more firms, base stations are more expensive in this specification.

[56] Each $\hat{c}^s_{fm}$ is simply $B_{fm}\tilde{c}^s_{fm}$, where $\tilde{c}^s_{fm}$ is the infrastructure cost parameter recovered using the specification used in the main text.

Figure 14 plots different measures of welfare as we change the number of symmetric firms, analogous to figure 9 in section 6.1.

Figure 14: Counterfactual Welfare



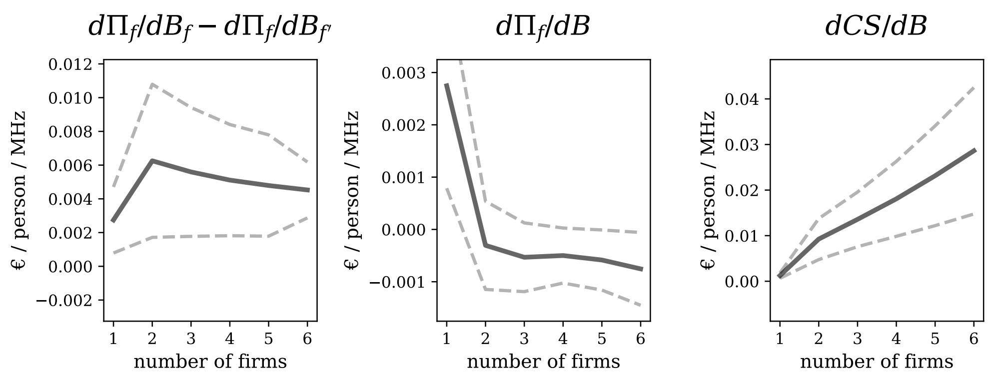
Note: Welfare is measured in euros per capita relative to monopoly.

Relative to the case in which costs are proportional to bandwidth, this cost specification implies fewer firms maximize both consumer and total surplus. However, this specification probably overstates the extent of scale efficiencies.

With this cost specification, there is the introduction of another source of economies of scale from the duplication of fixed costs. Thus, in this case, if we hold the number of base stations per firm fixed, more firms means more base stations, which means higher costs. In contrast, when base station costs are proportional to bandwidth, if we hold the number of base stations per firm fixed, then total base station costs do not change as we change the number of firms, given that the total bandwidth in the industry is fixed. While base stations certainly involve some fixed costs, firms can and do avoid duplicative fixed costs by engaging in passive network sharing, in which firms share some base station infrastructure (such as the land or the tower). Thus, we view as unrealistic these counterfactuals in which base stations involve fixed costs that are unavoidably replicated as we increase the number of firms. In contrast, our main cost specification is consistent with an equilibrium in which firms co-locate their base stations and share fixed costs.

The results for this alternative specification point to the importance of scale efficiencies that can be attained without integration. One might wonder in particular whether the gains from economies of pooling, which play an important role in our main results, can be attained without consolidation. We note that such gains would require firms from sharing their active network infrastructure (also known as Radio Access Network (RAN) sharing). Such network sharing is rare, while the sharing of passive infrastructure is common. This may be because firms do not find it profitable to share their active network infrastructure as in Fund et al. (2017). Active network sharing undermines firm's incentives to differentiate on quality. An open question is whether there is a possible regulatory framework that would allow firms to attain efficiencies from pooling network infrastructure without undermining incentives to invest.

Figure 15: Bandwidth Derivatives



Note: Derivatives are evaluated at the symmetric equilibrium values. The derivative of own profits with respect to another firm's bandwidth ($d\Pi_f/dB_{f'}$) is undefined in the monopoly case. In the first subplot, therefore, what is reported in the case of only one firm is simply the derivative of own profits with respect to own bandwidth ($d\Pi_f/dB_f$). Dashed lines represent 95% confidence intervals.

We also assessed the marginal value of spectrum with this alternative cost specification. Results were similar: marginal surplus exceed firms' willingness to pay by a factor of four instead of the factor of five we saw with the main specification in section 6.2.

D.2 Equilibrium without Path Loss

Here we show that in symmetric equilibria the optimal number of base stations per consumer is constant with respect to population density when there is no path loss or interference.

Let N_{fm} represent the number of base stations operated by operator f in municipality m. The number of consumers within each cell is given by $\frac{D_m A_m}{N_{mf}}$, where D_m is the population density and A_m is the municipality's area. We now rewrite equation 19 as

$$Q_{fm} = \overline{Q}_{fm} - \frac{D_m A_m}{N_{mf}} q^D \left(\boldsymbol{P}_{fm}, \boldsymbol{Q}_{fm}, \boldsymbol{P}_{-fm}, \boldsymbol{Q}_{-fm} \right), \tag{41}$$

where $q^D \left(\boldsymbol{P}_{fm}, \boldsymbol{Q}_{fm}, \boldsymbol{P}_{-fm}, \boldsymbol{Q}_{-fm} \right)$ represents equilibrium data consumption per capita. Note that channel capacity per base station $\overline{Q}_{fm}$ is exogenous without path loss and interference. Bandwidth is endowed, so there are no choice variables to influence channel capacity. The firm's only infrastructure choice here is effectively how many consumers they want to serve with each base station.

Consider firm f's variable profit function, equation 21, now written in per-consumer terms and as a function of quality:

$$\Pi_{fm}^{V}\left(\boldsymbol{P}_f, \boldsymbol{Q}_{fm}\right) \equiv \left(\boldsymbol{P}_f - \boldsymbol{c}_f^u\right) \cdot \boldsymbol{s}_f\left(\boldsymbol{P}_{fm}, \boldsymbol{Q}_{fm}, \boldsymbol{P}_{-fm}, \boldsymbol{Q}_{-fm}\right).$$

Let $\lambda_{fm} = \frac{D_m}{N_{fm}}$, and note that λ_{fm} can represent the firm's infrastructure choice variable. Rewrite variable profits as

$$\Pi_{fm}^{V}\left(\boldsymbol{P}_f, \lambda_{fm}\right) \equiv \left(\boldsymbol{P}_f - \boldsymbol{c}_f^u\right) \cdot \boldsymbol{s}_f\left(\boldsymbol{P}_{fm}, \lambda_{fm}, \boldsymbol{P}_{-fm}, \boldsymbol{\lambda}_{-fm}\right),$$

noting that the share function can be expressed as a function of λ_{fm} since delivered download speeds are determined by the congestion equation 41, and here $\lambda_{fm} = \frac{D_m}{N_{fm}}$ defines the congestion equation above.

Given the cost function expressed in equation 22, infrastructure costs are $c_{fm}^s B_{fm} N_{fm}$, and costs per capita can be expressed as

$$c_{fm}^s B_{fm} \frac{N_{fm}}{D_m A_m} = c_{fm}^s B_{fm} \lambda_{fm}^{-1} A_m^{-1}.$$

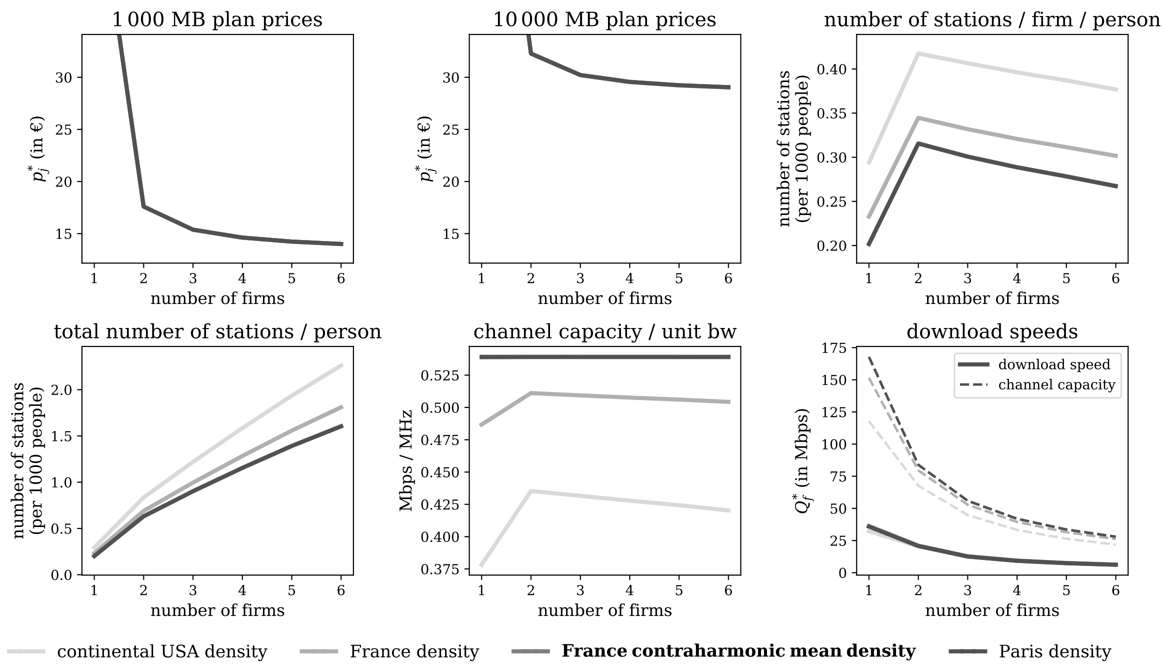
Both variable profits and infrastructure costs depend on population density D_m and the number of base stations N_{fm} only through their ratio $\lambda_{fm} = \frac{D_m}{N_{fm}}$. Therefore, the firm's optimum and the equilibrium level of investment entail a value for λ, or a number of base stations per consumer. Therefore, when we do comparative statics with respect to population density, the equilibrium number of base stations will be proportional to population density.

D.3 Impact of Population Density

Our main counterfactual simulations consider a market with moderate population density. This density of 2 792 persons / km^2 roughly corresponds to a high-density suburb. A natural question is whether the population density affects the trade-off between market power and scale efficiencies, perhaps changing the optimal number of firms. We first note that, without path loss, the equilibrium comparative statics with respect to population density would be very straightforward.

As shown above, without path loss, channel capacity is fixed by the bandwidth owned and operated by the firm. The cell radius will not affect channel capacity. The decision of cell radius amounts to a decision of how many customers to serve with each base station, with the firm effectively choosing the optimal level of congestion. The population density will not affect this choice when we think about it in terms of the optimal number of consumers per base station (or the optimal level of congestion). As population density increases, the optimal number of consumers per station remains constant, implying base station area will be inversely proportional to population density. Equilibrium outcomes like prices and delivered download speeds remain the same. See section D.2 for a more formal account.

Figure 16: Counterfactual Prices and Qualities by Density



Note: Each line in a subplot corresponds to a different population density, with the darker the line, the higher the density. Channel capacity is per base station. Download speeds are the average speed of transmission received by a user, including wait times.

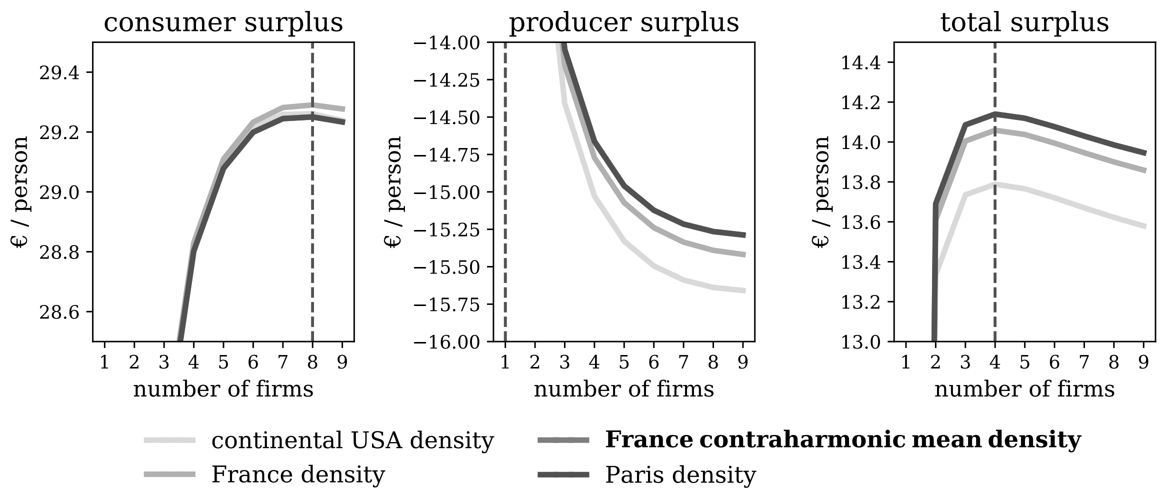
In addition to France's population-weighted mean population density (2 792 people/km^2), we consider three alternative population densities: the raw population densities of the continental USA (43.1) and France (123.9)—note that these are both quite low densities as both countries involve large unpopulated areas—and the population density of Paris (20 588).

Figure 16 illustrates how equilibrium outcomes for these different population densities. Certain outcomes are indeed affected by population density. Naturally, path loss is more severe when serving a less dense market, demonstrated by lower channel capacities per unit of bandwidth in Figure 16 (despite higher levels of investment per person).[57]

Otherwise, the comparative statics with respect to population density are very similar to what we would expect without path loss. In other words, we do not see substantial economies of density. Figure 18 depicts channel capacity as a function of the cell's radius. For radii in

[57] For each of these densities, we use the Hata model of path loss presented in Appendix A.1.1. This Hata model is for small cities. We have also simulated these counterfactual densities with rural and suburban Hata models of path loss for the associated densities, which exhibit less path loss as a function of distance. Results look similar but correspond more closely to the case of no path loss (in which the density does not matter).

Figure 17: Counterfactual Welfare by Density

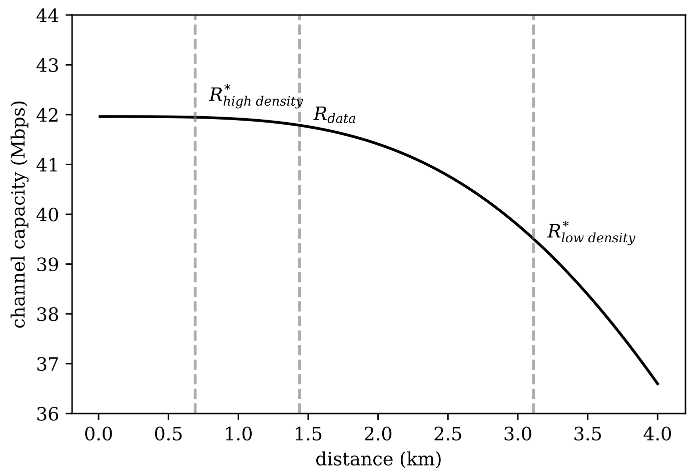


Note: Depicted are measures of welfare as a function of number of firms. Each line in a subplot corresponds to a different population density, with the darker the line, the higher the density. Welfare is measured in euros per capita relative to monopoly, so for each plot the value at 1 firm is 0. Dashed vertical lines denote the number of firms that maximizes that measure of welfare.

the range of the equilibrium radii in our counterfactuals, this function is quite flat, which is consistent with economies of density not being substantial at these population densities, although they may be at extremely low densities.

The optimal number of firms (for consumer or total surplus), depicted in Figure 17, is quite robust to the population density. Equilibrium outcomes like prices and delivered download speeds are extremely similar for different population densities. A takeaway is that, given the equilibrium cell sizes we observe, economies of density only appear to be a significant concern in very sparely populated areas.

Figure 18: Channel Capacity as Function of Radius



Note: R_{data} corresponds to the average radius of a cell in our data. $R^*_{\text{low density}}$ and $R^*_{\text{high density}}$ correspond to the equilibrium radius chosen in the four-firm equilibrium when the market has, respectively, a density of France and a density equal to the population-weighted mean population density of France. Bandwidth is set equal to the same total bandwidth as in the rest of our counterfactuals divided by four (for four firms), and spectral efficiency is also set to the same value as in the rest of our counterfactuals.

Table 9: Notation

Symbol	Description
f	indexes firms
i	indexes consumers
j	indexes mobile phone plans
$\mathcal{J}$	set of mobile phone plans
ℓ	indexes a location
$\mathcal{L}(R)$	set of locations within hexagon of radius R
m	indexes markets (municipalities)
γ_m	data transmission efficiency in market m
ε_{ij}	idiosyncratic, consumer-plan-level demand shock
θ	demand parameters
θ_{pi}	price coefficient
θ_{p0}	parameter controlling the mean of the price coefficient
θ_{pz}	parameter controlling the heterogeneity in the price coefficient
θ_v	coefficient on dummy for unlimited voice
θ_O	average Orange demand shock
θ_c	opportunity cost of time spent downloading data coefficient
θ_{di}	parameter of exponential distribution that defines distribution from which a consumer's utility of data consumption is drawn
θ_{d0}	parameter controlling the mean of θ_{di}
θ_{dz}	parameter controlling the heterogeneity in θ_{di}
ϑ_i	random shock to consumer's utility of data consumption, distributed exponentially with parameter θ_{di}
$\boldsymbol{\theta}_i$	vector containing θ_{pi} and θ_{di}
ξ_{jm}	market-level demand shock
σ	nesting parameter
B_{fm}	bandwidth (in Megahertz)
c_j^u	cost per user
c_{fm}^s	cost per base station and unit of bandwidth
$\overline{d}_j$	data consumption limit of phone plan j
D_m	population density
F	used for CDFs
H	number of hours in a month
$I_\ell(R_{fm})$	interference power at location ℓ when cell radius is R_{fm}
N_{fm}	number of base stations for firm f in market m
$q_{m\ell}$	data transmission speed at location ℓ in municipality m (in Mbits/second)
$\overline{Q}_{fm}$	channel capacity (in Mbits/second)
Q_{fm}	download speed (in Mbits/second) of firm f in market m
Q^L	throttled download speed (in Mbits/second)
Q_{fm}^D	demand requests (in Mbits/second)
P_j	price of phone plan j
R_{fm}	radius of area served by one base station (in km)
S_ℓ	signal power at location ℓ
s_{jm}	market share
$\mathbf{s}$	vector of market shares
u	utility of a phone plan
v_j	dummy variable for whether plan j has an unlimited voice allowance
w	utility from data consumption over course of month
x	monthly data consumption
z_i	consumer i's income